

## Insights into the structure–function relationship of disease resistance protein HCTR in maize (*Zea mays* L.): A computational structural biology approach



Budheswar Dehury<sup>a,b</sup>, Mousumi Sahu<sup>a</sup>, Mahesh Chandra Patra<sup>c</sup>, Kishore Sarma<sup>a</sup>, Jagajjit Sahu<sup>a</sup>, Priyabrata Sen<sup>a</sup>, Mahendra Kumar Modi<sup>a</sup>, Manabendra Dutta Choudhury<sup>b</sup>, Madhumita Barooah<sup>a,\*</sup>

<sup>a</sup> Agri-Bioinformatics Promotion Programme, Department of Agricultural Biotechnology, Assam Agricultural University, Jorhat 785013, Assam, India

<sup>b</sup> Department of Life Science and Bioinformatics, Assam University, Silchar 788011, Assam, India

<sup>c</sup> BIF-Centre, Department of Bioinformatics, Orissa University of Agriculture and Technology, Bhubaneswar 751003, Odisha, India

### ARTICLE INFO

#### Article history:

Received 30 May 2013

Received in revised form 6 August 2013

Accepted 8 August 2013

Available online 18 August 2013

#### Keywords:

HC-toxin reductase

Dihydroflavonol reductase

Disease resistant

Rossmann fold

Cofactor

### ABSTRACT

The disease resistance gene *Hm1* of maize encodes a NADPH-dependent reductase enzyme, HC-toxin reductase (HCTR) that detoxifies the HC toxin secreted by the race specific fungus *Cochliobolus carbonum* race 1. HCTR enzyme shares 29.6% sequence identity with dihydroflavonol reductase (DFR) of grape, a key enzyme involved in flavonoid biosynthesis. Here we report the comparative modelling, molecular dynamics simulation and docking studies to explain the structure–function relationship and the mode of cofactor (NADPH) binding in HCTR enzyme at the molecular level. The nucleotide binding domain of modelled HCTR adopts a classic Rossmann fold and possesses a consensus glycine rich GxGxxG motif. Molecular simulation studies suggested that HCTR model retained stability throughout the simulation in aqueous solution. HCTR model showed considerable structural identities with the cofactor binding site of DFR, but significant difference in the catalytic site might be the reason of functional divergence between these families of proteins. Similarly electrostatic surface potential analysis of both HCTR and DFR revealed profound variations in the charge distribution over the substrate binding site, which can be correlated with the sequence variability and may suggest distinct substrate-binding patterns and differences in the catalytic mechanism. Docking results indicated Phe19, Gly21, Arg40, Thr90, Gly208, Arg218, Glu221 and Thr222 are important residues for cofactor (NADPH) binding through strong hydrogen bonding and electrostatic interactions. Alanine scanning and analysis of docking energies of mutant proteins suggested that Phe19, and Arg40 are two critical residues for the cofactor binding. The result from the present study is expected to pave the way for exploration of similar genes in other economically important crop varieties.

© 2013 Elsevier Inc. All rights reserved.

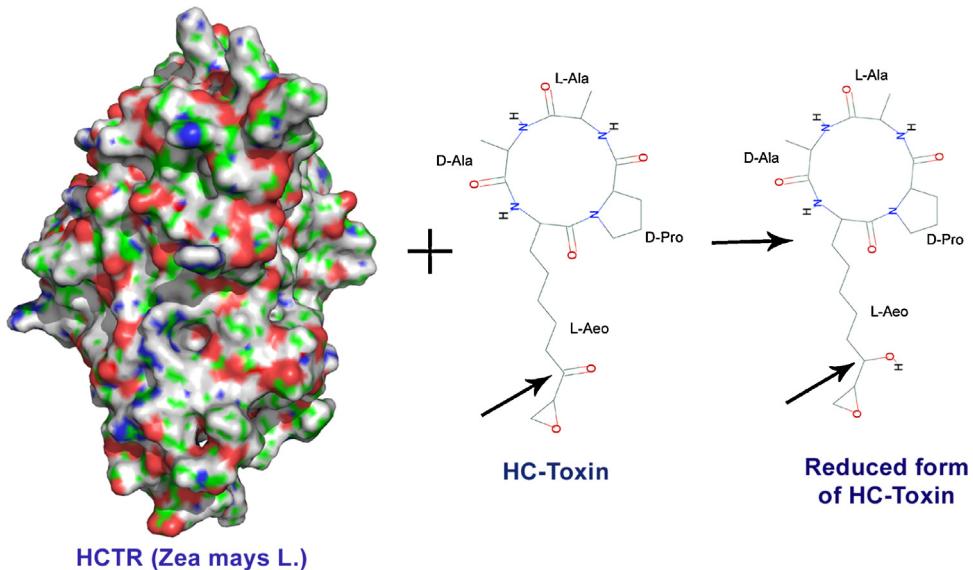
### 1. Introduction

Plants are continuously assailed with multitude of pathogenic invaders and, therefore, must enlist an equally diverse set of defense response mechanisms to protect themselves. Knowledge of the plant–pathogen interaction mechanisms is of critical importance to create disease-resistant crops. Pathogenicity of most fungal pathogens rely on the production of toxins to infect their host plants but the mechanism of toxin mediated disease causing and host defence response to these toxins remain elusive. Nature has endowed the crop plants with an inherent mechanism to defend themselves from the invasion of pathogens. This feature restricts

the invasion and proliferation of potent pathogens which is termed as resistance. The complex network of plant innate immune system comprises of three important steps viz., pathogen detection, signal transduction, and defense response initiation [1–3]. The resistance genes encoded by plants recognize the pathogen-produced effectors molecules and induce a defence signal. The transmission and amplification of defense signal require kinase genes in the plant cell that initiates cascade of host transcription machinery i.e., defence response genes including the pathogenesis-related (PR) gene which can confer local or systemic resistance to the plant [2,4]. The proteins encoded by a number of disease resistance *R* genes conferring resistance to a diverse spectrum of pathogens have been isolated and cloned from a wide range of plant species [5].

The disease resistance *R* genes plays a crucial role in activating the defense mechanism of the plants through specific recognition of certain effector molecules produced by corresponding avirulence

\* Corresponding author. Tel.: +91 0376 2340095; fax: +91 0376 2340001/2340101.  
E-mail address: [m17barooah@yahoo.co.in](mailto:m17barooah@yahoo.co.in) (M. Barooah).



**Fig. 1.** The *Hm1* gene of maize encodes HC-toxin reductase (HCTR), which detoxifies HC-toxin secreted by the fungus *Cochliobolus carbonum* race 1 (CCR1). The chemical reaction scheme shows the structure of HC-toxin and its reduced form. The reduction of the 8-carbonyl group of Aeo to an alcohol in the HC-toxin has been marked with arrow (Aeo stands for 2-amino-9,10-epoxy-8-oxodecanoic acid).

(*avr*) genes from pathogens [6,7]. They are being widely used in the breeding of disease-resistant cereal crops including maize cultivars. In comparison to other cereal crop species, very few *R* genes such as *Hm1*, *Hm2*, *Rcg1*, *Rp1-D*, and *Rxo1* have been cloned and reported in maize [8–10].

Out of the several devastating diseases caused by various pathogens in maize, the leaf blight disease caused by the fungus *Cochliobolus carbonum* race 1 (CCR1) is considered as one of the most destructive disease that affects the net yield of the crop. The asexual form of this ascomycetes fungus is known as *Helminthosporium carbonum* (HC) (synonym *Bipolaris zeicola*). Unlike the other plant pathogens, CCR1 affects every part of the host, causing blight of the leaves, rot of the roots and the stalk at any stage of development [11].

*Hm1* was the first disease resistance gene cloned in maize. The gene confers resistance against leaf spot causing fungus CCR1. The *Hm1* mediated disease resistance mechanism evolved naturally in maize by encoding a reductase enzyme (*i.e.*, HC toxin reductase: HCTR) which detoxifies fungal HC-toxin and in turn provides resistance against the fungus [12]. Although it was the first disease resistance (DR) gene to be cloned, it retains its novelty by disarming the pathogen directly instead of participating in the plant recognition and response system as most DR genes do. The gene encodes a NADPH (reduced form of nicotinamide adenine dinucleotide phosphate)-dependent reductase (HCTR) enzyme that inactivates the key virulence factor, HC-toxin (an epoxide containing cyclic tetrapeptide). The cyclic tetrapeptide of HC-toxin structure comprised of cyclo (D-Pro-L-Ala-D-Ala-L-Aeo), where Aeo represents 2-amino-9,10-epoxy-8-oxodecanoic acid [13]. The HC-toxin is considered as one the microbial secondary metabolite whose ecological function in nature has been studied in detail. The molecular reaction of CCR1 and HC-toxin is controlled by the *Hm1* and *Hm2* loci in maize and these loci encode HC-toxin reductase, which detoxifies HC-toxin by reducing the 8-carbonyl group of Aeo. The chemical reaction scheme containing the chemical structures of HC-toxin and its reduced form has been shown in Fig. 1. Furthermore, this HC-toxin act as an inhibitor of histone deacetylases (HDACs) in diverse organisms, including plants, insects, and mammals, but the inhibition of HDACs during infection by CCR1 fungus that lead to disease is not understood yet.

In plants NADPH plays a major role in cellular redox homeostasis and is an indispensable electron donor in numerous enzymatic reactions, biosynthetic pathways, and detoxification processes [14]. This enzyme is a reduced form of NADPH-dependent reductase which shows significant homology to several NADPH-dependent enzymes like dihydroflavonol 4-reductase (DFR), vestitone reductase and anthocyanidin reductase involved in the biosynthesis of flavonoids and anthocyanins in plants. *Hm1* mediated disease resistance is an imitable mechanism in which the corresponding *avr* component from the fungal pathogen CCR1 is not involved in toxin degradation and thus unique in nature from the other reported *R* genes till date. *Hm1* provides complete protection against CCR1 in every part of the maize plant, but certain lines of maize have been reported to contain a second defence gene named *Hm2* that confers effective resistance in mature maize cultivars [15,16]. The *Hm2* gene is structurally truncated homologue of *Hm1* [17].

CCR1 exhibits a high degree of host specificity and can cause disease only in maize. *Hm1* gene of maize is an ortholog of HCTR like sequences from several grass species including sorghum, rice, barley, wheat, rye, oats, millet, fescue, bluegrass, and bamboo species. In barley, rice, and sorghum, these homologs are syntenic to the maize *Hm1* gene, indicating that they are truly orthologs and derived from a common ancestor.

Several crystal structure of reductase enzymes have been reported in Protein Data Bank (PDB) which are involved in biosynthesis of secondary metabolites, such as dihydroflavonol reductase (DFR)(PDB ID: 2C29) [18] in grape (*Vitis vinifera* L.), vestitone reductase (PDB ID: 2P4H) in Alfalfa (*Medicago sativa* L.) [19], and apo anthocyanidin reductase (PDB ID: 2RH8) in grape [20]. It has been observed that these enzymes share a considerable sequence identity to HCTR of *Zea mays* L. As such three-dimensional structure of HCTR protein is expected to provide an insight into the possible mechanism of action in maize.

Notwithstanding the knowledge generated in the last two decades to understand the structural and functional behaviour of disease resistance proteins of various crops, several questions remain to be answered about what molecular mechanism governs the novel function of *R* genes belonging enzymatic class in maize. The disease resistant dominant *Hm1* gene, which is mapped in the long arm of the chromosome 1 in maize provides broad

spectrum of complete resistance against the fungal pathogen *C. carbonum* race1 (CCR1). Although the NADPH-dependent enzyme produced by *Hm1* acts on the fungal HC-toxin to detoxify it, the biochemical characterization of this enzyme remains elusive. In the present study, based on the available crystal structures, a homology model of HCTR was built and subsequently refined through molecular dynamics simulation. The cofactor NADPH (reduced form of nicotinamide adenine dinucleotide phosphate) was docked into equilibrated model in order to investigate the binding site of HCTR, enzyme-cofactor interactions, the amino acid residues involved in binding, and to facilitate the understanding of the cofactor specificity and mode of catalysis in enzymatic class of resistance gene in maize.

## 2. Materials and methods

### 2.1. Sequence retrieval and domain analysis

The reviewed primary amino acid sequence of HCTR protein (GenBank accession no: NP\_001105920) conferring resistance against leaf spot disease was retrieved from GenBank database of NCBI. Putative conserved domain of HCTR was computed using Pfam (<http://pfam.janelia.org/>) [21], SMART (<http://smart.embl-heidelberg.de/>) [22] and CDD (<http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>) [23]. InterProScan tool (<http://www.ebi.ac.uk/Tools/pfa/iprscan/>) [24] was used to predict the protein family, super family and the domain arrangement within the protein. SignalP4.1 server (<http://www.cbs.dtu.dk/services/SignalP/>) [25] was used to predict the signal peptides of the protein.

### 2.2. Primary structure analysis

To have a broad chemistry of HCTR, primary structure analysis was done using ProtParam tool (<http://expasy.org/cgi-bin/protparam>) [26] of Expasy Proteomic Server. Computational analysis on various physico-chemical properties of the protein such as molecular weight, isoelectric point, instability index, aliphatic index and grand average hydropathy (GRAVY) were performed. The secondary structure of HCTR was predicted using CONCORD server (<http://helios.princeton.edu/CONCORD/>) [27] which uses consensus-based method for protein secondary structure prediction integrating several popular tools, such as PSIPRED, DSC, GOR IV, Predator, Prof, PROFphd, and SSpro.

### 2.3. Comparative modelling of HCTR protein

Comparative modelling is considered as one of the most accurate computational methods to generate reliable theoretical 3-D model of protein from its primary amino acid sequence. This method is capable of yielding models for wide spectrum of application and is being routinely used in various biological applications. HCTR has considerable amount of sequence similarity and identity with secondary metabolites synthesizing enzymes of grape and alfalfa. Till date, no crystal structure of the same has been reported in any literature or in PDB. Hence, the comparative modelling of HCTR was carried out to generate its tertiary structure.

Suitable templates of HCTR for model building were identified by performing DELTA-BLAST (domain enhanced lookup time accelerated BLAST) against PDB. DELTA-BLAST (<http://blast.ncbi.nlm.nih.gov>) [28] searches a database of pre-constructed position specific score matrices (PSSMs) before searching a protein-sequence database to yield better homology detection. DELTA-BLAST was preferred against normal BLASTP because the retrieval accuracy and sensitivity towards protein

analysis is more in case of DELTA-BLAST than normal BLAST. To ensure the correctness in the template identification process, apart from DELTABLAST, HCTR was subjected to various meta-servers like 3D Jury [29], Pcons.net [30], GeneSilico [31] and Geno3D [32] to find reliable templates with conserved domains. In addition, the protein threading approach implemented by I-TASSER [33] and protein fold recognition server Phyre Version 2.0 (Protein Homology/analogY Recognition Engine) [34] were also used to determine the best templates in terms of fold recognition. Considering the suitable templates obtained from DELTA-BLAST and various meta-servers search, we carried out 3-D model building of the HCTR by multi-template approach utilizing crystal structures of the dihydroflavonol reductase (DFR) (PDB ID: 2C29), vestitone reductase (PDB ID: 2P4H) and apo anthocyanidin reductase (PDB ID: 2RH8) as the most appropriate template structures. Multiple sequence alignment of the target and template protein sequences was carried out by ClustalW program (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>). Modeller tool generated 20 different models and were ranked according to their discrete optimized protein energy (DOPE) score (a statistical potential to assess the quality of the models) and the model with lowest DOPE score was selected as the best model. The best model was subjected for loop modelling and refinement in the homology modelling protocol of Discovery Studio version 3.5 (DS3.5; Accelrys Inc. San Diego, CA, USA). Moreover, to ensure the correctness and sensitivity of the HCTR model generated by Modeller, we cross-checked it with the best generated 3-D models of LOMETS [35], Phyre and I-TASSER server.

### 2.4. Energy minimization

The refined model of HCTR was subjected to energy minimization by DS3.5 with the minimization protocol. The minimization protocol employs the steepest descent and conjugate gradient methods of minimization algorithms with a generalized Born implicit solvent model [36]. Parameters of a distance-dependent dielectric constant = 1, non-bonded radius of 14 Å, CHARMM force field [37], spherical electrostatic cut-off, and the steepest descent algorithm were used to remove close van der Waals contacts for a maximum steps of 5000 with 0.1 minimizing RMS gradient. The potential energy, van der Waals energy and electrostatic energy of HCTR model was computed using the calculate energy protocol in DS3.5.

### 2.5. Quality assessment and validation

The quality, internal consistency and reliability of the energy minimized HCTR model was evaluated by a number of computational tools. PROCEHCK [38] was used to check the stereo-chemical quality of the model, which quantifies the residues in the available zones of Ramachandran plot. A Ramachandran plot provides the position of the torsion angles phi ( $\phi$ ) and psi ( $\psi$ ) between  $\text{Ca-C}$  and  $\text{N-Ca}$  atoms of the residues contained in a peptide. ERRAT tool [39], which finds the overall quality factor of the protein, was used to check the statistics of non-bonded interactions between different atom types. VERIFY\_3D program [40] determines the compatibility of the atomic model (3D) with its own amino acid sequence (1D) where a high Verify\_3D profile score indicates the high quality of a protein model. All the above analyses were carried out using Structural Analysis and Verification Server (SAVES) (<http://nihserver.mbi.ucla.edu/SAVES/>). Protein Structure Analysis (ProSA) tool (<https://prosa.services.came.sbg.ac.at/prosa.php>) [41] was employed for the refinement and validation of the modelled structure which checks the native protein folding energy of the model by comparing the energy of the model with the potential mean force derived from large set of known protein structures.

The energy was transformed to a score called Z-score. Residues with negative Z-score indicate reasonable side chain interactions. Finally the root mean square deviation (RMSD) between backbone chain atoms of the model and the respective templates was calculated by structural superimpositions using iPBA web server ([http://www.dsmb.inserm.fr/dsmb\\_tools/ipba/](http://www.dsmb.inserm.fr/dsmb_tools/ipba/)) [42]. The pair-wise 3-D structural alignment and RMSD of the equivalent C $\alpha$  atoms, of the model and template was performed using MATRAS (MArkovian TRAnsition of Structure evolution) web server ([http://strcomp.protein.osaka-u.ac.jp/matrass/matras\\_pair.html](http://strcomp.protein.osaka-u.ac.jp/matrass/matras_pair.html)) [43]. To ensure the correctness in the arrangement of secondary structure elements in the model with respect to its most closest homologue, a brief comparison was made using STRIDE (STRuctural IDEntification) [44] and DSSP (Dictionary of Secondary Structure of Proteins) [45] where both predicts the secondary structure elements from the 3-D coordinates of the protein.

## 2.6. Refinement of HCTR model using molecular dynamics simulation

The refinement in the side chain orientations and to gain a better relaxation as well as best possible arrangement of the atoms in the theoretical model of HCTR, molecular dynamics (MD) simulation was performed. The MD simulation was conducted using GROMOS96 43A1 force field [46] and the flexible SPC water model (to create the aqueous environment) in GROMACS version 4.5.4 package (GROningen MACHine for Chemical Simulations) [47]. To solvate the model, it was placed in a cubic box maintaining a distance of 1.5 nm between the box edges and the protein surface. Particles mesh Ewald (PME) electrostatic and periodic boundary conditions were applied in all directions. The system was neutralized by adding seven Na<sup>+</sup> counter ions. To get rid of the high energy interactions and steric clashes of the system, it was subjected to a steepest descent energy minimization until a tolerance of 1000 kJ/mol reached. All the bond lengths were constrained with the LINear Constraints Solver (LINCS) [48] method whereas the geometry of water molecules was constrained with SETTLE algorithm [49]. The energy minimized system was treated for 100 ps equilibration run. The pre-equilibrated system was consequently subjected to 10 ns production MD simulation, with a time-step of 200 ps at constant temperature (300 K), pressure (1 atm) and without any position restraints. Snapshots of the trajectory were taken on every 1 ps and all the analyses of the MD simulation were performed using XMGRACE software. From the 10 ns MD simulation, average structure with low RMSD value was selected as the best model for further analysis. The final selected structure was subjected for model validation process through number of tools in SAVES server and MolProbity web server (<http://molprobity.biochem.duke.edu/>) [50].

## 2.7. Docking of NADPH to the active site of HCTR using CDOCKER in DS3.5

Molecular docking is a computational method for predicting the binding accuracy of ligand (poses) at specific orientation into the active site of receptor molecule. The protein-ligand complex provides structural insights into the interactions and complementarities between the ligand and the active site of the receptor. In this study, CDOCKER module of DS 3.5 was used to find the binding specificity of the cofactor NADPH into the active site of HCTR model. CDOCKER module implements a CHARMM-based molecular dynamics method to dock ligands within the active site of enzymes and receptors. First, high-temperature MD is used to generate a set of random ligand conformations and then the produced conformations are translated into the binding site. The generated

poses are further created using random rigid-body rotations followed by simulated annealing. Finally the ligand poses are refined by minimization [51].

For docking study, initially the simulated HCTR protein was prepared using the “prepare protein” protocol in Discovery Studio 3.5 by removing water molecules and adding hydrogen ions based on the CHARMM force field. Similarly cofactor NADPH (ligand) was prepared by adding charges, hydrogen and applying force field in Discovery Studio environment. Energy was also minimized with CHARMM force field before performing docking. The random conformations search of NADPH was conducted utilizing a high temperature simulated annealing dynamics scheme. Ligands were heated to 700 K in 2000 steps, followed by annealing to 300 K in 5000 steps. Ten random conformations were generated and a final minimization was introduced to each docking poses.

After preparation of both the protein and cofactor, the active site of the protein was identified. DS3.5 binding site module contains two different algorithms viz., eraser and flood-filling to define protein binding site residues in protein. The first method identifies the binding site based on the shape of the receptor and the second method does it by acquiring previous knowledge of the volume occupied by the known ligand pose already in an active site. We employed the second method for identification of binding site in HCTR. Furthermore the active site of the HCTR was delineated based on the ligand-binding pocket of the template protein assuming that the mode of cofactor binding of the template structure and the modelled structure are similar. Both the analysis predicted the same binding site cavities with intended residues. Finally to ensure the correctness in the identification of binding site, MetaPocket (<http://projects.biotech.tu-dresden.de/metapocket/>) and GalaxySite (<http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=SITE>) web servers were used to detect the binding site residues of both target and template protein separately and the result revealed a consensus binding site along with intended residues for both the proteins. The cofactor NADPH was docked in to the active site of the modelled HCTR to elucidate its structural and functional relevance in terms of cofactor binding. During the docking process top 10 different ligand conformations for NADPH was prepared based on the top orientation of the molecules in the active site of the protein. The determination of cofactor binding affinity was calculated by choosing the consensus scoring scheme based on CDOCKER ENERGY, CDOCKER Interaction Energy, LigScore1\_Dreiding, LigScore2\_Dreiding, PLP1, PLP2, Jain, PMF and PMF4 to estimate the ligand-binding energy implemented in the protein-ligand interaction module of DS3.5. Finally from the resulting docking arrangements for both the HCTR model, the one with expected orientation of the carbonyl group close to NADPH and the active site in each case was used for further energy refinement to capture slight induced fit modifications of the protein active site. The binding energy calculation and 2-D interaction maps were generated to show proposed HCTR-NADPH.interactions.

## 2.8. Alanine scanning

Alanine substitution mutations were carried out by replacing the side chain ligands of the important NADPH binding residues using DS 3.5 software. The grid dimension was 70 × 70 × 70. Ligand preparation and docking parameters were same as that of the docking of HC-toxin described above. The mutant protein structures were used as receptor to dock the NADPH using Autodock 4.2 software [52]. The binding energy of the wild type HCTR-NADPH.complex was used as reference to judge the binding affinity of the mutant complexes.

**Table 1**

Templates selected for comparative model building of HCTR from DELTA-BLAST search against PDB.

Templates (PDB ID with their chain)	Source	Total score	Query coverage (%)	E value	% of identity	Resolution (Å)
2C29-D	<i>Vitis vinifera</i>	172	94	1e-49	31	1.81
2P4H-X	<i>Medicago sativa</i>	169	94	9e-49	32	1.40
2RH8-A	<i>Vitis vinifera</i>	154	91	5e-43	30	2.22

## 2.9. Docking of HC-toxin with HCTR–NADPH.complex

To identify the HC-toxin and substrate binding site, we docked the chemical structure of HC-toxin into the already docked HCTR–NADPH.complex using Autodock. Two dimensional coordinates of the ligand (in SDF format) were obtained from PubChem Compound database through the unique chemical structure identifiers CIDs: 107864. The initial geometries of the molecules were converted into PDB format, thoroughly cleaned, and minimized using DS 3.5 using CHARMM all atom force field. The partial charges of the ligand molecule were calculated by Automated Topology Builder tool [53] that uses PM3 method for quantum calculation of molecular electronic structure. Structure preparation and binding site definition was carried out using Autodock Tools 1.5.2. Kollman charges were assigned for the protein atoms and Gasteiger parameters were set for partial charges of ligand atoms. Due to lack of X-ray crystal structure, the catalytic site of the protein is still ambiguous. Therefore, we assumed that the substrate molecule must bind the protein alongside the cofactor in physiological condition. The binding site grid was centred on the already docked NADPH with a grid dimension of  $80 \times 80 \times 80$  grid points with  $0.375 \text{ \AA}$  of grid spacing that almost covered the whole protein. One hundred iterations of genetic algorithm run were carried out and empirical energy functions were used for docking calculation with an initial population size of 300. The maximum number of energy evaluations was  $2.5 \times 10^7$ . In the present study, we performed rigid docking where all amino acid residues were kept rigid, but the ligand was flexible. The docked poses were clustered at  $2.0 \text{ \AA}$  RMSD cut off and ranking of the ligand poses was done on the basis of binding free energy scores.

## 3. Results and discussion

### 3.1. Sequence analysis of HCTR

The domain prediction of HCTR by SMART and Pfam revealed four putative domains that are overlapping with each other viz., NmrA (11–119), Epimerase (11–278), 3Beta.HSD (12–207) and NAD binding 4 domains (13–262). Conserved domain (CD) search of HCTR revealed that it belongs to super family cl09931 comprising of Rossmann-fold NADPH binding proteins which shares a Rossmann-fold NAD(P)H/NAD(P)(+) binding (NADB) domain. This NADB domain is reported in numerous dehydrogenases enzymes involved in several metabolic pathways and also in many other redox enzymes. Overall, the proteins in this family contain a second domain in addition to the NADB domain responsible for binding to the substrate and catalysis of enzymatic reaction. Analysis with SignalP tool did not predict any signal peptide in HCTR protein.

The primary sequence analysis of HCTR (357 amino acid long) showed that this protein is acidic in nature (isoelectric point = 5.69) and has molecular mass of 38,521 Dalton (38.5 kDa). The aliphatic index of HCTR is very high (97.56) indicating the stability of the protein over a wide range of temperature. It is well known that a protein whose instability index is smaller than 40 is predicted as stable, whereas a value above 40 indicates that the protein may be unstable [54]. The HCTR is unstable in nature as its instability index was 51.42 (>40). The Grand Average hydropathicity (GRAVY) value for a peptide or protein is calculated as the sum of hydropathy

values of all the amino acids, divided by the total number of residues in the sequence. The GRAVY index of HCTR is very low ( $-0.090$ ) indicating the possibility of its better interaction with water. The CONCORD secondary structure prediction server showed that HCTR is dominated by random coils (55.47%) followed by helices (33.33%) and strands (11.20%) among the various secondary structure element.

### 3.2. Comparative modelling and energy minimization of HCTR model

Comparative modelling of proteins is often considered as a method of choice when a clear relationship of homology between the target and template is found. This approach provides reasonable result based on the assumption that the tertiary structure of the two proteins will be similar if their sequences are related. Absence of the experimental determined 3-D structure of HCTR prompted us to construct its theoretical 3-D models. Templates were retrieved by performing DELTA-BLAST search against PDB. Because of its low sequence identity in the PDB database from DELTA-BLAST search, HCTR was also scanned in meta-servers like Pcons.net, 3D Jury, Geno3D and Gensilico to find reliable templates. Considering the suitable templates obtained from the meta-servers search, we carried out 3D model building of the HCTR by advance modelling techniques utilizing crystal structures of the dihydroflavonol reductase (DFR) (PDB ID: 2C29), vestitone reductase (PDB ID: 2P4H) and apo anthocyanidin reductase (PDB ID: 2RH8) as the most appropriate template structures (as shown in Table 1). Cross-checking of this model with the best models obtained from I-TASSER, PHYRE (Protein Homology/analogy Recognition Engine) Version 2.0 and LOMETS showed better secondary structure conservation in our proposed HCTR model. Moreover, comparative modelling multi-template approach is recommended when multiple templates are available to avoid biasing the models towards one protein or one set of side chain conformations as it improve the quality of models [55]. Thus modelling by constraining the multi-template alignment in HCTR suggested a reliable model for structural analysis.

Based on the target-template alignment, Modeller9v11 was used to build the three-dimensional models of HCTR. It is evident from the target-template alignment (Fig. 2) that NADPH binding consensus motif GXGXXG (Gly16-X-Gly18-X-X-Gly21) of HCTR is highly conserved with the templates with few exceptions. For instances Gly16 of HCTR is substituted by Ala18 in the template 2C29 (D-chain) and Gly21 of HCTR is substituted by Ala21 in 2RH8 (A-chain). The best model (model with the lowset DOPE score) generated in the modeller was subjected to energy minimization in DS3.5. The total energy of the model before and after refinement was calculated by CHARMM force field in DS3.5.

### 3.3. Model quality assessment

The detailed residue-by-residue stereo-chemical quality of the modelled protein structure was evaluated by the Ramachandran plot (Fig. 3) in Procheck tool. The reliability of the backbone torsion angles  $\phi$  and  $\psi$  distribution of the target protein and template was examined and evaluated for comparison in Procheck. In HCTR model, 90.4% residues  $\phi$  and  $\psi$  angles are in the core region of the

2C29_D	----MGSQSETVCVTGASGFIGSWLVMRLLERGYTVRATVRD-PTNVKKVKHLLDLPKAE	55
2RH8_A	MATQHPIGKKTACVVGGTGTVASLLVKKLQQKGYAVNTTWRD-PDNQKKVSHLLELQELG	59
2P4H_X	-----KGRVCVTGGTGFLGSWIISLLENGYSVNTRRADPERKRDVSFLTNLPGAS	52
Hm1	-MAEKESENNGVRVCVTGGAGFIGSWLVRKLLEKGYTVHATLRNTGDEAKAGLLRRRLVPGAA	59
	.*.*.*:.*: : * : * : * : * : . : :	
2C29_D	THLTILWKADILADEGSFDEAIKGCTGVFHVATPMDFESK-DPENEVIKPTIEGMLGIMKSC	114
2RH8_A	-DLKIFRDLTDLSFEAPIAGCDFVFHVATPVHFASE-DPENDMIKPAIQGVVNVMKAC	117
2P4H_X	EKLHFFNADLSNPDSFAAAIEGCVGIFHTASPIDFAVS-EPEEIVTKRTVDGALGILKAC	111
Hm1	ERLRLFQADLFDAATFAPAAIAGCQFVFLVATPFGLDSAGSQYKSTAEEAVVDAVHAILRQC	119
	* : : * : * : * : * : * : * : * : : : : : :	*
2C29_D	AAAATVRRRLVFTSSAGTVNIQEHQLP-----VYDESCWSDMEFCRAKKMTAWM-YF	164
2RH8_A	TRAKSVKRVVILTSSAAAVTINQLDGTL-----VVDEKNWTDIEFLTSAKPPTWG-YP	169
2P4H_X	VNSKTVKRFIYTSSGSASFNGKD-----VLDES DWSDVDLLRSVKPFGWN-YA	161
Hm1	EESRTVKRVVIHTASVAAAASPLLEEVPAASGVGYRDFIDEESCWTSLNVDYPLRSAHFDKYI	179
	::*: : * : * : * : .. : . * : * : * : : * : :	*
2C29_D	VSKTLAEQAAWKYA--KENNIDFITIIPITLVVGPFISSMPPSLITALSPITGNEAHYSI	222
2RH8_A	ASKTLAEKAQWKF--EENNIDLITVIPTLMAGSSLTSDVPSIGLAMSLITGNEFLING	227
2P4H_X	VSKTLAEKAVLFG--EQNGIDVVTLILPFIIVGRFVCPKLPDSIEKALVVLGKKEQIGV	219
Hm1	LSKLQSEQELLSYNNGESPAFEVVVTPLGLVAGDTVLGRAPETVESAVAPVSRSEPYFGL	239
	** : * : .. : .. : : * : .. : * : * : * : : ..	.
2C29_D	IR-----QGQFVHLDDLCNAHIYLHEN-PKAEGRYICSSHDCIIILDLAKMLREKYPEY	274
2RH8_A	MKGMQMLSGSVSIAHVVEDVCRAHIFVVAEK-ESASGRYICCAANTSVPPELAFLSKRYPQY	286
2P4H_X	TR-----FHMVHVDDVARAHIYLLEN-SVPGGRYNCSPIFIVPIEEMSQLLSAKYPEY	270
Hm1	LRILQQLLGSLPLVHVDDVCDALVFCMERRPSVAGRFLCAAAAYPTIHDVVAHYASKFPHL	299
	: .. : * : * : * : .. : * .. : * : * : : : : : :	**.
2C29_D	NIP-TEFKGVDENLKSVCFCSSKKLTDLGFEPKYSLEDMFTGAVDTCRAKGLLPPSHEKPV	333
2RH8_A	KVP-TDFGDFPPKSK-LIISSEKLVKEGFSFKYGIIEEEYDESVEYFKAKGLLQN-----	338
2P4H_X	QILTVDLKEIKGARLPDLNTKKLVDAFGDFKYTIEDMFDDAIQCCKEKGYL-----	322
Hm1	DIL---KETTEAVATVRPARDRLGELGFKYKGMEELDSSVACARLGSLDGSKLGLQ	355
	.. : .. : * : .. : * : .. : * : .. : * : .. : * : ..	**.
2C29_D	DGKT 337	
2RH8_A	----	
2P4H_X	----	
Hm1	KG-- 357	

**Fig. 2.** Sequence alignment of dihydroflavonol reductase (DFR) from grape (*Vitis vinifera*) (PDBID: 2C29-D chain), apo anthocyanidin reductase from grape (*Vitis vinifera*) (PDBID: 2RH8-A chain), vestitone reductase from alfalfa (*Medicago sativa*) (PDBID: 2P4H-X chain) and HCTR from maize (*Zea mays L.*) aligned in ClustalW. The identical (asterisk) or similar (dot or colon) amino acids are also marked.

**Table 2**

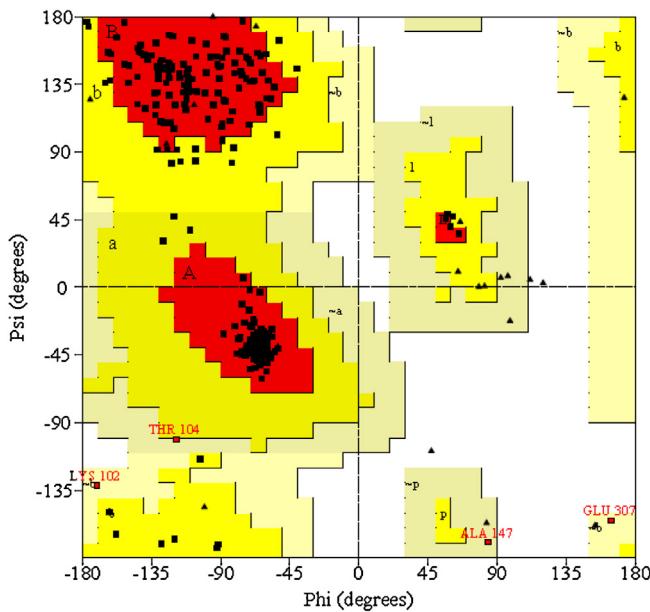
Comparison of Ramachandran plot statistics of HCTR model with closest structural homologue 2C29.

Ramachandran plot statistics	Modelled structure of HCTR		Template (2C29-D)	
Residues in most favoured regions	Residues	Percentage (%)	Residues	Percentage (%)
Residues in additionally allowed regions	282	90.4	262	91.0
Residues in generously allowed regions	26	8.3	24	8.3
Residues in disallowed regions	4	1.3	2	0.7
Number of non-Glycine and non-Proline	0	0	0	0
Number of end residues (excluding Gly and Pro)	312	100.0	288	100.0
Number of Glycine residues	1		2	
Number of Proline residues	28		18	
Overall G factor	16		16	
	0.19			0.14

**Table 3**

Comparison of model validation scores from different server between modelled HCTR and its closest structural homologue 2C29 (D-chain).

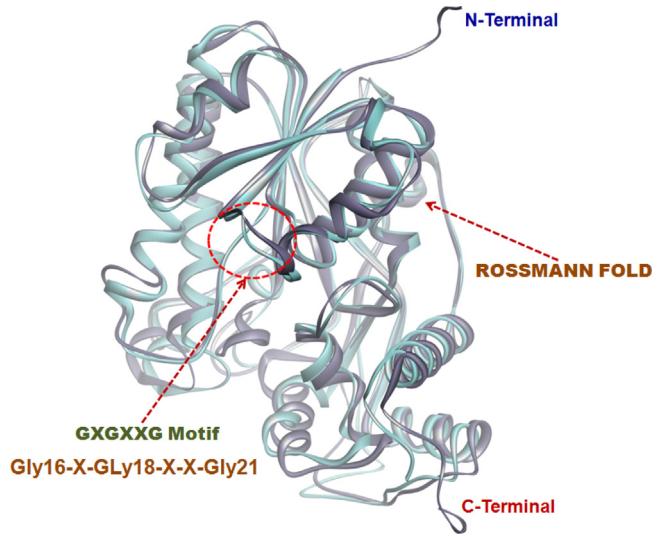
Target/template	Verify_3D	Errat	Prove (Z score)	ProSA	MolProbity clash	C $\alpha$ RMSD (Å)	Backbone RMSD (Å)
HCTR ( <i>Zea mays</i> )	95.81%	71.387	0.757	-8.41	2.10	1.38	1.22
2C29 ( <i>Vitis vinifera</i> )	99.69%	97.428	0.287	-10.93	2.27	-	-



**Fig. 3.** Ramachandran plot of the modelled HCTR. The plot was calculated by PROCHECK program.

Ramachandran plot, whereas the template has 91.0% residues in the most allowed region (Table 2). A good homology model should have more than 90% residues in the favourable region. Our model protein has 98.7% of the residues in the most allowed and allowed regions with a relatively low percentage of residues having general torsion angles. This result confirms the accuracy of the generated HCTR model. It is also observed that none of the active site residues are present in the disallowed region. All the bond distances and angles of modelled protein lie within the allowed range of the standard dictionary values indicating that HCTR model is reasonably good in geometry and stereochemistry. The  $\Phi$  and  $\psi$  distribution of Ramachandran plot of non-glycine, non-proline residues in the model and template (2C29-D chain) are summarized in Table 2. The packing quality of each residue of the model was assessed by Verify\_3D program where the compatibility of the model residues with their environment is assessed by a score function. Residues with a score over 0.2 should be considered reliable. As shown in Table 3; the score of the refined model maximally was above 0.2 which corresponds to acceptable side chain environment. ProSA revealed a Z-score of  $-8.41$  for modelled HCTR, where the template has Z-score of  $-10.93$  reflecting the overall quality of the model. The MolProbity server predicted that no residues with poor rotamer bonds (goal <1%), none of the residues had bad angles (goal <0.1%), and none of the C $\beta$  deviations were  $>0.25\text{ \AA}$  (goal 0%). The results of HCTR from MolProbity server when compared with the template reveals almost the same statistics which further strengthened the reliability of the proposed model.

RMSD (root-mean-square distance) between equivalent C $\alpha$  atom pairs (target and template) was measured to check the degree of structural similarity. To investigate how well the modelled structure fits to the crystal structure of the template, the prepared model and its closest relative was superimposed based on C $\alpha$  and backbone atom pairs. A pair-wise 3-D alignment search of the template protein with the modelled structure in iPBA web server showed an overall identity of mere 29% for 311 aligned residues with a very low RMSD of  $1.22\text{ \AA}$  on their backbone atoms (Fig. 4). Whereas, pair-wise 3-D alignment of the equivalent C $\alpha$  atom pairs of target and template showed a RMSD of  $1.38\text{ \AA}$ . Both the results of iPBA and MATRAS web server are in assumption that HCTR and its structural homologues share strong structural conservation and similarity in



**Fig. 4.** Structural superimposition of modelled HCTR (grey) and crystal structure DFR (PDB Id: 2C29) (aqua turquoise teal). The Rossmann fold and the GxGxxG motif have been marked. The structural superposition was performed using iPBA web server and image was generated in DS3.5.

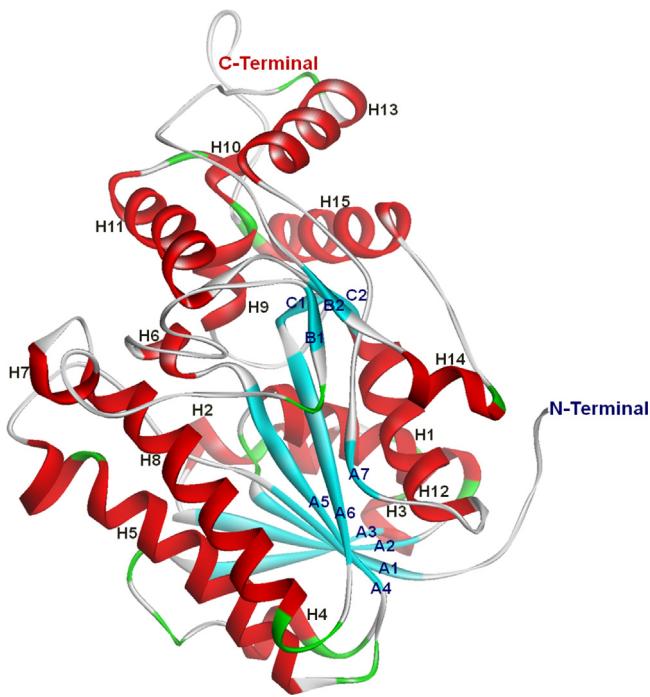
the structural folding. It also signifies that the generated model is reasonably good for further studies. From the above analysis, it is evident that the geometric quality of the backbone conformation, the residue contact, the residue interaction, and the energy profile of the structure are well within the limits which confirms the reliability of the modelled structure.

### 3.4. Detailed structural study of HCTR protein

To understand the arrangement of secondary structural elements in 3-D space, detailed proteomic structural characterization of HCTR was performed. HCTR is composed of 15  $\alpha$ -helices (40.1% of amino acid, 143 residues), 3 beta-sheets (A, B and C) of 11  $\beta$  strands (12.6% of amino acid, 45 residues) and other secondary structure elements (47.3%, 169 residues). The HCTR monomer is composed of two domains: the long N-terminal domain which is the dinucleotide binding domain that adopts a classic Rossmann fold [56], and C-terminal substrate binding domain. The active site residues lie within the deep cleft of the two discrete domains.

The large N-terminal domain that adopts a Rossmann fold consists of seven  $\beta$ -strands forming a large parallel  $\beta$ -sheet flanked by seven  $\alpha$ -helices. Unlike the template protein, presence of one  $\beta$ -strand (A7) and one  $\alpha$  helix (H9) within the Rossmann fold disrupts the overall symmetry of the two halves of  $\beta$ - $\alpha$ - $\beta$ - $\alpha$ - $\beta$  fold in HCTR. The N-terminal domain is stabilized by four beta-alpha-beta units. In contrast to N-terminal domain, the C-terminal domain is very small comprising of six  $\alpha$ -helices (H10, H11, H12, H13, H14 and H15), and four  $\beta$ -strands (B1, B2 and C1, C2) which forms two small beta-sheets (B and C) respectively (Fig. 5). Moreover, the HCTR protein bears 21 helix-helix interactions, 26  $\beta$ -turns and 2  $\gamma$  turns. The beta turns which join the adjacent helix and strands are dominated by hydrogen bonds, are involved in stabilization of the respective fold and maintaining the overall stability of the protein. The retention of most number of beta-alpha-beta folds is one of the characteristics of NADPH and NADH dependent reductase enzymes, which was also reported in the crystal structure of template protein.

Among the retrieved structural hits, DFR of grape (PDB ID: 2C29) was found to be the best hit based on maximum total score of 172 (with 31% identity) was considered as the closest structural homologue of HCTR as predicted by DELTA-BLAST (as shown in Table 1). Selectivity of the closest structural homologue from



**Fig. 5.** Homology model of HCTR from maize. Solid ribbon representation of the HCTR model coloured by its secondary structure elements. The secondary structure elements such as  $\alpha$  helices,  $\beta$  strands and the N and C termini are labelled. The image was prepared with Discovery Studio.

DELTA-BLAST search was also cross checked using HHpred, PHYRE, SWISS-MODEL and RaptorX. So as to understand the active site architecture of modelled HCTR, the pairwise structural superposition of HCTR with its closest structural homologue, DFR of grape was extensively studied. As compared to DFR, the active site of modelled HCTR also consists of the two pockets: NADPH binding pocket and the substrate binding pocket (a usual characteristics seen in almost all the members of this family of proteins). The region of binding of the NADPH (GxGxxG motif) and the groove for substrate binding was well conserved, with all the side chain residues involved in some interaction present in both enzymes in the same positions. As compared to the template a small change was noticed in HCTR model, close to the shift of the chain around the substrate binding site (data not shown) which might contribute towards different substrate specificities of this class of enzymes.

Finally, an attempt was made to correlate the assignment of secondary structure elements in the modelled HCTR predicted by DSSP and STRIDE (both uses the 3-D structure for prediction of their secondary structure elements) with CONCORD tool (uses a consensus method and predicts the secondary structure of protein from its primary amino acid sequence). Both the sequence-based and the 3-D model based secondary structure approaches perfectly correlate with each other which signifies that the model predicted by comparative modelling as accurate. In addition, CONCORD search also revealed that among the secondary structure elements turns/coils dominated over helices followed by sheets in the modelled HCTR.

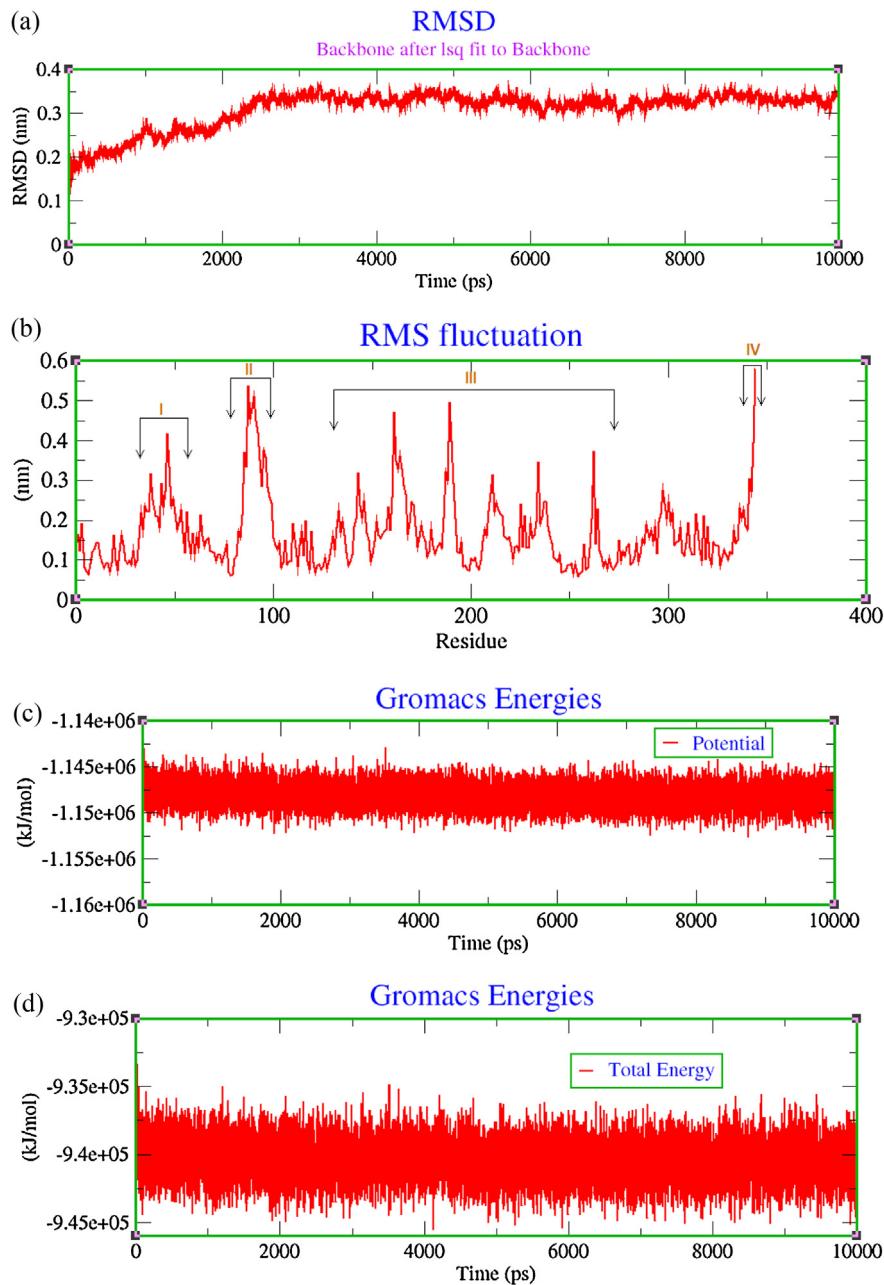
### 3.5. Comparison of electrostatic surface potential of modelled HCTR with the template

The Adaptive Poisson-Boltzmann Solver (APBS) package [57] was used to generate the electrostatic surface potentials for modelled HCTR and the template (DFR of grape). APBS package offers a state-of-the-art suite for performing Poisson–Boltzmann

electrostatic calculations on biomolecules. The charge distribution and patches that differentiate the HCTR and DFR was examined in detail (data not shown). Fascinatingly, it was observed that the overall topologies of both the NADPH dependent enzymes were similar with minute variations in the substrate binding region as depicted in the isocontour representation data. A specific electrostatic potential distribution pattern of the NADPH binding domain was observed for both the proteins. The charge distribution patterns (isocontour ranging from  $-5 \text{ kT}$  to  $+5 \text{ kT}$ ) of both the proteins could correlate with their different activity properties. In addition, the distribution of these charges denotes differences in the mechanism of action and/or interaction with other proteins and intracellular localization in both HCTR of maize and DFR of grape. The surfaces of the substrate binding pocket contained the most profound differences in charge distributions. In contrast to substrate binding pocket of HCTR, the NADPH binding pocket possess positively charged surface.

### 3.6. Molecular dynamic simulation

The refinement of the modelled protein was performed through MD simulation to get an optimized and stable structure suitable for docking with cofactor. The stability and dynamic properties of HCTR model were observed by MD simulation of 10 ns duration. The position-restrained dynamics for 2000 ps generated a structure with proper side chain conformations without any deviation of native backbone atoms. The steepest descent energy minimization for the solvated modelled protein revealed the maximum force reached the threshold of  $1000 \text{ kJ mol}^{-1} \text{ nm}^{-1}$  in 1143 steps. The RMSD of the protein backbone atoms are plotted as a function of time to check the stability of the system throughout the simulation. The RMSD reached the overall equilibrium at 3230 ps with a RMSD of 0.36 after which a plateau graph can be seen in Fig. 6a. During the last 1 ns it was observed that the RMSD of the system tends to be converged, indicating the system is stable and well equilibrated. The relative flexibility of the model was also characterized by plotting the root mean-square fluctuation (RMSF) relative to the average structure obtained from the MD simulation trajectories. Four flexible regions have been predicted for the modelled protein structure of HCTR considering the RMSF value and represented in Fig. 6b. Two flexible regions have been located in N-terminal end involving the residues Glu30-Leu54 and Pro77-Lys102. The highest RMSF values of the above regions are of 0.41 (Glu45) and 0.53 (Vala88). Some of the key residues of the N-terminal end fell in flexible regions (Arg40, Lys47 and Thr90). Both the residues (Arg40 and Lys47) possess basic polar and positively charged side chain which again signifies their affinity towards ligand or protein interaction. The third flexible region is the largest middle region of the protein involving the residues (Ala132-Ser271). This region possesses several continuous peaks. The RMSF values of the peaks in this region are 0.31 (Leu141), 0.47 (Cys160), 0.50 (Val188), 0.31 (Val210), 0.35 (Ser233) and 0.37 (Ala261). This longest flexible region possesses three active residues (Val188, Val210 and Thr222) whereas Val188 and Val210 formed two corresponding peaks in the Fig. 6b, indicating their potential binding affinity. Several catalytic residues such as Phe152, Lys156, Tyr163 and Lys167 also fell in the above flexible region. The C-terminal end has a single flexible region with a dominant RMSF value of 0.5 (Leu344). It can be summarized from RMSF analysis that the middle portion of the modelled HCTR enzyme is more flexible in comparison to the N-termini and C-termini. The MD trajectories of 10 ns showed the compactness of structure through radius of gyration of an average value of  $\sim 2.0 \text{ nm}$  from nm to 1.5 the end of the simulation. During simulation, variation in potential energy (Fig. 6c) and total energy was also calculated to evaluate the stability of the model (Fig. 6d).



**Fig. 6.** The RMSD, RMSF, potential energy and total energy graph of the modelled HCTR during MD simulation. (a) RMSD of backbone C $\alpha$  atoms of the maize HCTR modelled structure. (b) RMSF analysis of amino acid residues of the maize HCTR model structure. (c) Potential energy (kJ/mol) during 10 ns trajectory. (d) Total energy (kJ/mol) during 10 ns trajectory. All the images were generated using XMGRACE software.

### 3.7. Molecular docking of NADPH

The residue conservation of the binding site and structural comparisons of NADPH-dependent HCTR with known NADPH-dependent forms are crucial for predicting the cofactor specificity and the enzymatic mechanism. Molecular docking was performed using CDOCKER module in DS, so as to understand the recognition specificity and mode of cofactor binding to HCTR. The structural surfaces of the modelled protein provides insight into the shape of the HCTR catalytic cleft and enable us to study the important structural features that dictate cofactor specificity (the NADPH binding pocket). So as to gain insight into the most probable binding conformation of the cofactor NADPH to the stable protein structure, superimposition of HCTR and its closest structural homologue DFR was performed.

Structural superimposition showed that the key residues of the active site i.e., Gly18 (in the template Gly15), Phe19/Phe16, Ile20/Ile17, Arg40/Arg37, Lys47/Lys44, Asp68/Asp64, Leu69/Leu65, Val88/Val84, Thr90/Thr86 and Val210/Val193 (Fig. 7) are fully conserved and follow a common folding pattern. Despite of high structural similarity within the active site, few of the active site residues of the template protein i.e., Ser14, Tyr163, Lys167 and Ser205 do not superpose with the modelled structure (Fig. 7). The active site residue Ser14 and Ser205 of the template protein are replaced by Ala17 and Thr222 respectively in the modelled structure. Ser is a small, polar amino acid, though it is fairly neutral with regard to mutations that always prefer to be substituted with other polar or small amino acid in particular threonine which differs only in existence of a methyl group in place of a hydrogen group found in serine. The substitution of Serine residues by Alanine and Threonine

```

      : E1      H1      E2      H2      H3      E3
SecA : EEEEEESTTSHHHHHHHHHHHHHHHT EEEEEE S HHHHHHHHHH TTTHHHHHEEEEEE
      9 : VRVCVTGGAGFIGSWSLVRKLLEKGTYVHATIRNTGDEAKAGLLRLLVPGAAERLRLFQAD: 68
          *****  *****  ***  ***  *  *  *  *  *  *  *  *  *  *
      6 : ETVCVTGAASGFIWSWLVMRLLERGYTVRATVDPNVVKVHLDDL-PKAETHLTLWKA: 64
          *****  *****  ***  ***  *  *  *  *  *  *  *  *
SecB : EEEETTTTSHHHHHHHHHHHHHTT EEEEEE TT HHHHHHHHTS-TTHHHHHHEEEEEE
      : E1      H1      E2      H2      - H3      E3

      :      H4      E4      H5      E5
SecA : SS TTSSHHHHHTT SEEEE      SS SSIIIIITHHHHHHHHHHHHHHHHH SS EE
      69 : LFDAATFAPAIAGCQFVFIWATPFGLDSAGSQYKSTAEAVVDAVHAILRQCEESRTVKRV: 128
          *  *  *  *  *  *  *  *  *
      65 : LADEGSFDEAIKGCTGVFHATPMDFESKDPE-NEVIKPTIEGMLGIMKSCAAKTVRRL: 123
SecB : TTSTTTTHHHHTT SEEEE      SS SSHH-HHTHHHHHHHHHHHHHHHHHS EE
      :      H4      E4      H5-     H6      E5

      :      H6      E6      H7      H8
SecA : EEEEEESSHHHHS S SS SSS SEE TT      SSSSHHHH TTHHHHHHHHHHHH
      129 : IHTASVAAASPLLEEVVPASGVGYRDFIDESCWTSLNVDYPLRSAHFDKYILSKIQSEQE: 188
          *  *  *  *  *  *  *  *
      124 : VFTSSAG-T-----VNIQEHLQLPVYDESCWSDME---FCRAKKMTAWMYFVSKTLAE: 171
SecB : EEE GG-G-----TS SSS SEE TT      HH---HHHHH TTHHHHHHHHHHHH
      :      - ----- E6      H7---     H8

      :      E7      E8      H9      H10     H11
SecA : HHHHHHHHTTT EEEEEE EEEES SSS HHHHHHHHTHHHHH HHHHHHHHHHHHHHS
      189 : LLSYNNGESPAFEVVTPLGLIVAGDTVLGRAPETVESAVAPSRSEPYFGLLRILQQLG: 248
          *  *  *  *  *  *  *  *
      172 : QAAWKYAKENNIDFITIIPITIVVGPFIMSSMPESLITALSPI--TG-NEAH-YSIIR--: 224
SecB : HHHHHHHHHHTT EEEEEE EEEES S SS HHHHHHHHTHHH-HT- GGG-HHHHT---
      :      E7      E8      H9      H10--   - H11   ---
      :      E9 H12      E10      H13
SecA : S EEEHHHHHHHHHHHS S EE B SB HHHHHHHHHHHH TTS S SS
      249 : SLPLVHVDDVCDALVFCMERRPSVAGRFLCAAYPTIHDVVAHYASKFPHLDILKETTEA: 308
          *  *  *  *  *  *  *  *  *  *  *  *
      225 : QGQFVHLDLCAHIYLFENP-KAEGRYICSSHDCIILDLAKMREKYPEYNIPTEFKGV: 283
SecB : EEEEEEHHHHHHHHHHHHH T-T EEEEEE EEEEEEHHHHHHHHHHHS TTS S TT
      : E9      H12      - E10      E11 H13

      :---      E11 H14      H15
SecA :--- EE HHHHHHT HHHHHHHHHHHHHHTSS SS
      :--- VATVRPARDRLGELGFKYKYGMEILDSSVACAARLGSLDGSK: 351
          *  *  *  *  *  *  *  *  *
      284 : DENLKSVCFSSKKLTDLGFEFKYSLEDMFTGAVDTCRAKGLLPPSH: 329
SecB : TT EE HHHHHHT HHHHHHHHHHHHHHTSS S
      : E12 H14      H15

```

**Fig. 7.** Comparison of NADPH-binding amino acids in modelled HCTR protein and the crystal structure of DFR (2C29) from grape. In MATRAS, the pair-wise 3-D structural superposition of amino acids from HCTR and DFR, the matching NADPH-interacting amino acids in HCTR and 2C29 were marked with green square boxes and mismatch were marked in red square boxes respectively. “SecA” and “SecB” represents their secondary structures of the modelled protein HCTR and template 2C29, respectively. These secondary structures are determined by the DSSP program. The DSSP codes for secondary structures are H=alpha helix, B=residue in isolated beta-bridge, E=extended strand, participates in beta ladder, G=3-helix (3/10 helix), I=5 helix (pi helix), T=hydrogen bonded turn, S=bend. (For interpretation of the references to color in this text, the reader is referred to the web version of the article.)

in the active site of modelled protein is a preferred substitution for Serine and may not affect the overall functionality of HCTR but may play an important role in substrate recognition or specificity. The other two substitutions in the active site of the template protein Tyr163 and Lys167 by Leu180 and Gln184 in the modelled protein do not affect the overall size and hydrophobicity of HCTR active

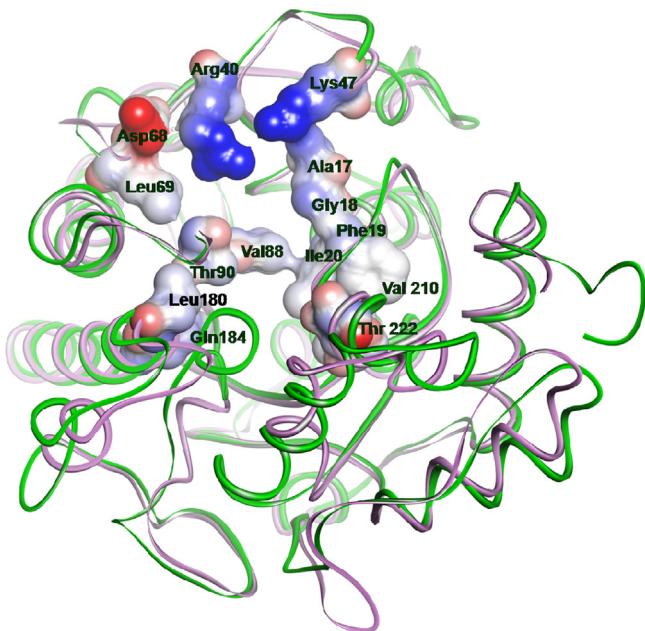
site. The structural superposition of HCTR (along with the active site residues) and the template is shown in Fig. 8.

The catalytic cleft of modelled HCTR was analyzed in detail from the pair-wise structural superposition with DFR of grape. It is evident from the structural superposition that a significant difference was observed in HCTR model as compared to the template. The

**Table 4**

Comparison of docking features of modelled HCTR and template protein with cofactor NADPH using CDOCKER in DS3.5.

Target/template	Ligand name	Binding energy (kcal/mol)	Ligand energy (kcal/mol)	Protein energy (kcal/mol)	Complex energy (kcal/mol)	Entropic energy (kcal/mol)
Target (HCTR)	NADPH	-698.6306	-235.24485	-9458.5	-10,122.35885	23.5822
Template (2C29)	NADPH	-752.54872	-50.12602	-13,090.0	-13,892.99115	23.8691



**Fig. 8.** The binding pocket for the co-factor NADPH has been marked in the structural superposition of HCTR (green colour) and its template 2C29 (mauve colour). The residues forming the binding pocket have been labelled and marked in green. (For interpretation of the references to color in this text, the reader is referred to the web version of the article.)

active site residue, Ser128 in the template structure corresponds to Ser133 in the modelled protein whereas the other four key catalytic residues (*i.e.*, Phe152, Lys156, Tyr163 and Lys167) are not conserved (Fig. 7) reflecting a different mode of catalytic mechanism in both the proteins.

Docking of NADPH to HCTR was performed by employing the CDOCKER, a grid-based molecular docking method which employs CHARMM force field and assigns the partial charges of the atoms with those found in Merck Molecular Force Field (MMFF) [58]. Initially, the template protein was used as receptor and the bound cofactor NADPH was used as ligand for docking. Both the ligand and template protein were prepared and docked into the active site of the protein. The docking result was analyzed to validate the CODCKER module and to verify whether the selected parameters were able to produce the suitable orientation of the ligand for active binding to its corresponding receptor. The result of both docked conformation (CODCKER) and the bound conformation (*i.e.*, binding conformation from the crystal structure) showed similar orientation with intended interactions. However, minute deviation was observed in the docked conformation in comparison to the bound conformation having a RMSD of 1.12 Å. The consensus scoring system *i.e.*, CDOCKER ENERGY, CDOCKER.Interaction Energy, Ligscore1.Dreiding, LigScore2.Dreiding, PLP1, PLP2, Jain, PMF and PMF4 used for scoring various poses of the cofactor (NADPH) with the template protein is shown in Supplementary Table 1. The pose with highest CDOCKER energy was selected as the best pose and analyzed. It was observed that NADPH firmly binds

to active site residues of the template protein through hydrogen bonds and hydrophobic interactions (Supplementary Fig. S1 and S2). The pose with the highest CDOCKER energy was selected to represent the studied HCTR ligand inside the binding cavity (Fig. 9a), and employed MMFF to minimize the final structure. The consensus scoring system for the ligand (NADPH) poses for HCTR are shown in Supplementary Table 2. The cofactor NADPH binding site is formed by a cleft spanning the C-terminal face of the Rossmann fold domain of HCTR. While studying the binding pocket of HCTR it was noticed that the NAD ring was more protected and deeper in the binding pocket of HCTR (Fig. 9b). The NADPH is bound in an extended conformation, through a strong network of hydrogen bonds and electrostatic interactions, as well as hydrophobic contacts. In HCTR, these interactions involve residues including the glycine-rich motif where Phe19, Gly21, Arg40, Thr90, Arg218, Gly208, Glu221 and Thr222 to be involved in the interaction with NADPH molecule through hydrogen bonds (Fig. 9b). The H-bond distances were in the range of 1.7–2.6 Å. The residue Arg40 alone made four H-bonds and three with Arg218 while the rest active site residues forms only one H-bond with cofactor. Furthermore, Gly15, Al17, Gly18, Leu39, Arg40, Val88, Ala89, Pro91, Phe92, Ala136, Leucine140, Val210 and Pro220 were also involved in hydrophobic interaction with NADPH (Fig. 10). The central diphosphate group of the NADPH binds the main-chain atoms of the glycine rich motif through direct H-bonds which are highly conserved in Rossmann dinucleotide-binding domains. Similarly the monophosphoester group of the adenine mononucleotide phosphate moiety is tightly bound with electrostatic interactions to charged residues. Moreover, theoretical calculation showed that free energy of binding score (negative CDOCKER energy) of NADPH to HCTR is −1043.27 kcal/mol confirms the affinity of HCTR model towards NADPH (Supplementary Table 2).

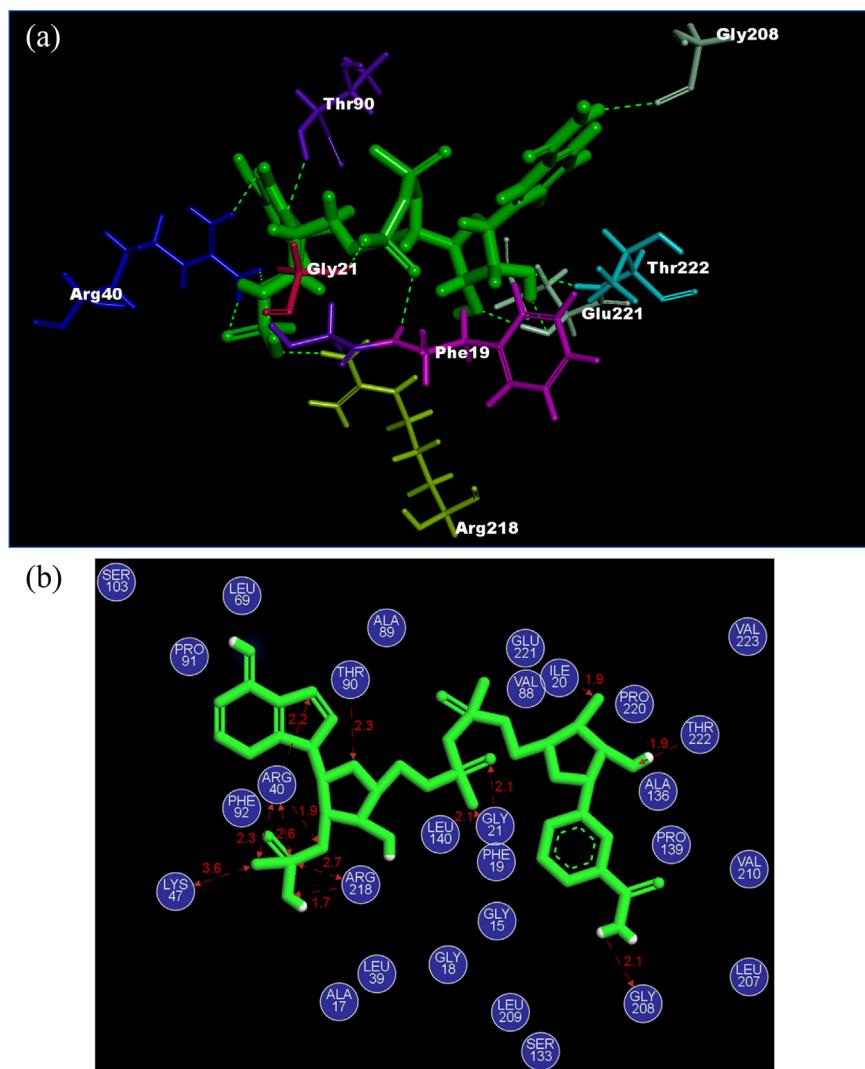
The total interaction energy of NADPH with HCTR reported was −10,122.50 kcal/mol arising from the hydrogen bonding, in which the van der Waals and electrostatic energies are −11,996 kcal/mol and −614.901 kcal/mol, respectively (Table 4). The binding energy of NADPH to HCTR was reported to be −698.6306 kcal/mol (Table 4). It is evident from the Figs. 10 and 11; NADPH is stabilized by hydrogen bonding and hydrophobic interactions in the centre of the active site. The active site predicted by different active site prediction tools is well supported by docking experiment which confirms the direct involvement of predicted residues in cofactor binding through hydrogen bonding, electrostatic and hydrophobic interaction. The results from the HCTR–NADPH.interaction analysis are exposed for the comparison of hydrophobic and hydrogen bond interacting residues with that of template (DFR) and NADPH (Table 5). It can be concluded from both the analysis that DFR of grape and HCTR are NADPH dependent enzymes and follow the same mode of binding and possess high affinity towards cofactor.

In order to validate the importance of each of the predicted key residues for cofactor binding, we performed alanine scanning mutations and the NADPH molecule was docked with the Ala-variants of HCTR protein (Supplementary Table 3). From the Supplementary Table 3, it can be seen that the F19≥A variant of HCTR showed least binding affinity. Furthermore, a similar result was obtained for R40≥A Ala-variant, suggesting that Phe19 and

**Table 5**

Docking features (hydrogen bond counts, hydrogen bond forming residues and hydrophobic interacting residues of modelled HCTR and template with NADPH).

Properties	No. of hydrogen bonds	Interacting residues forming hydrogen bonds	Hydrophobic interaction forming residues
Modelled HCTR	13	Phe19, Gly21, Arg40, Thr90, Arg218, Gly208, Glu221, Thr222	Gly15, Al17, Gly18, Leu39, Arg40, Val88, Ala89, Pro91, Phe92, Ala136, Leucine140, Val210 and Pro220
Template 2C29	14	Gly12, Ser14, Gly15, Phe16, Ile17, Arg37, Lys44, Asp64, Thr86, Tyr163, Thr191, Thr208	Val36, Leu65, Ala85, Pro87, Met88, Phe90, Pro190, Leu192, Val193, Pro204, Ser205



**Fig. 9.** The docked complex of HCTR–NADPH. In HCTR, amino acids in the NAD Binding Rossmann fold pocket that interacted with the cofactor NADPH were shown in 3-D graphical representation. (a) The amino acids forming hydrogen bond with NADPH has been labelled where green dashed lines indicate hydrogen bonds and (b) residues involved in hydrogen bonding with the cofactor NADPH are represented in Disks. The hydrogen bond is marked with red dotted arrow and the atomic distance ( $\text{\AA}$ ) between the interacting residues with NADPH has been marked. (For interpretation of the references to color in this text, the reader is referred to the web version of the article.)

Arg40 are two important residues for the cofactor interaction, and the substitution with alanine showed detrimental effect on the binding affinity. However, all other residues may be presumed to be essential for the stability and interaction of the cofactor. Thus, our result suggests that both hydrophobic and positively charged residues are essential for cofactor binding and enzyme catalysis. Conclusively, we suggested that further experimental site specific mutagenesis is required to validate the exact role of these residues.

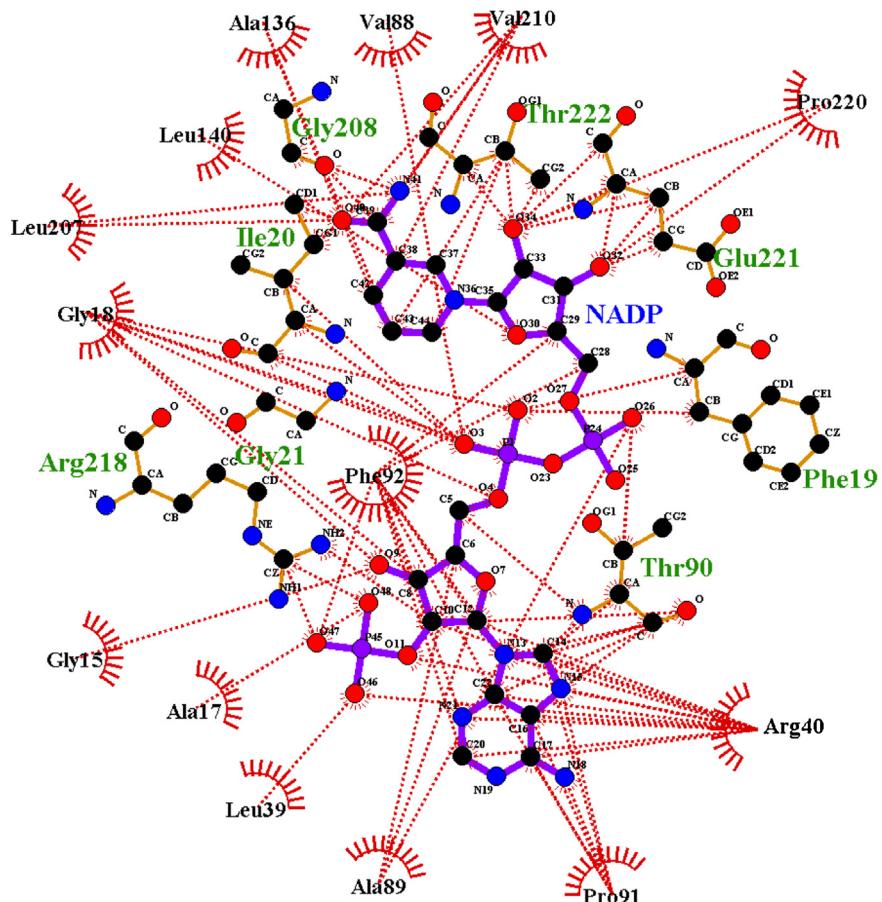
### 3.8. Docking of HC-toxin to HCTR–NADPH complex

To better identify the substrate binding residues and to have an idea about the catalytic site within HCTR protein, molecular docking of HC-toxin was performed into an already docked HCTR–NADPH complex. A total of 10 docking conformations were generated, of which the one with strongest binding affinity of  $-5.07 \text{ Kcal/mol}$  (Table 6) was chosen for further analysis. A closer visual

**Table 6**

The docking result of HC-toxin against the already docked complex of HCTR–NADPH using autodock.

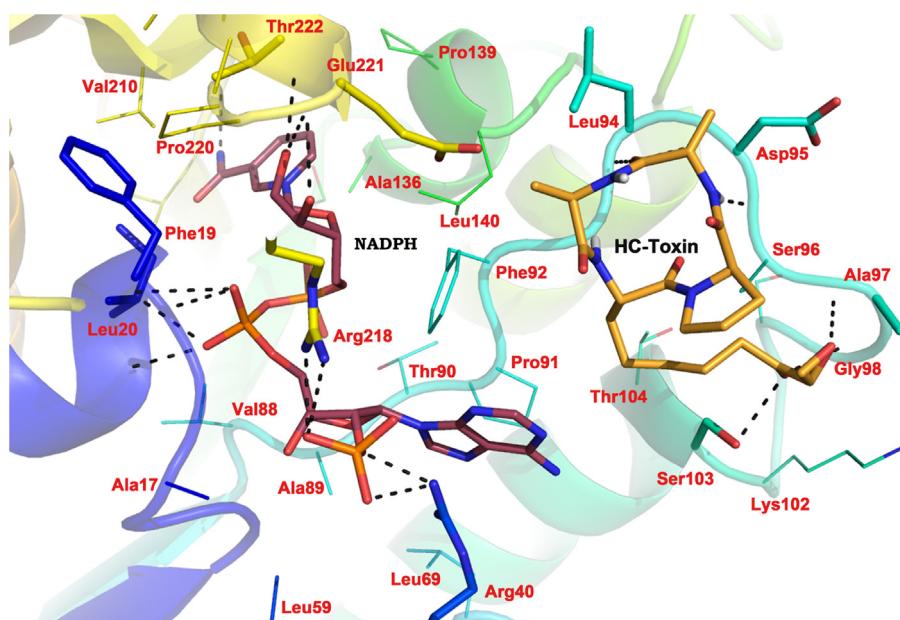
Rank	Binding energy (kcal/mol)	kI	Intermolecular energy	Internal energy	Torsional energy	Unbound extended energy	Cluster RMS	Ref. RMS
1	-5.07	193.05 $\mu\text{M}$	-7.16	-1.21	2.09	-1.21	0.0	92.35
2	-3.99	1.18 mM	-6.08	-1.42	2.09	-1.42	1.1	92.0
3	-4.79	306.19 $\mu\text{M}$	-6.88	-0.42	2.09	-0.42	0.0	93.96
4	-4.38	614.71 $\mu\text{M}$	-6.47	-0.57	2.09	-0.57	0.0	98.95
5	-4.14	930.97 $\mu\text{M}$	-6.22	-1.69	2.09	-1.69	0.0	99.33
6	-3.69	1.97 mM	-5.78	-1.75	2.09	-1.75	0.65	99.5
7	-4.02	1.13 mM	-6.11	-0.9	2.09	-0.9	0.0	100.1
8	-3.96	1.26 mM	-6.05	-2.0	2.09	-2.0	0.0	100.64
9	-3.36	3.47 mM	-5.44	-2.3	2.09	-2.3	1.11	99.95
10	-3.52	2.62 mM	-5.61	-1.53	2.09	-1.53	0.0	99.26



**Fig. 10.** Schematic diagram depicting interactions between cofactor NADPH and its amino acid binding partners with HCTR, identified with the program LIGPLOT. The hydrophobic interactions are marked in red dotted lines. (For interpretation of the references to color in this text, the reader is referred to the web version of the article.)

investigation revealed that the one of the HC-toxin conformations with lowest binding energy (highest binding affinity) prefers to bind at a position which is close to the docked NADPH structure. HC-toxin was monitored to form four strong hydrogen bonds with

average intermolecular distances of 2.5 Å. From Fig. 11 it can be seen that the HC-toxin is enclosed by a binding pocket lined by mostly hydrophobic residues, however, one negatively charged (Asp95) and one positively charged residue (Lys102) were also observed.



**Fig. 11.** Diagram showing HC-toxin docked into the already docked complex of HCTR–NADPH.structure. Residues involved in HC-toxin and NADPH binding are marked in red and the black dashed lines indicate hydrogen bonds. (For interpretation of the references to color in this text, the reader is referred to the web version of the article.)

## 4. Conclusion

The fungal diseases are the potential threats for agricultural crops as they cause severe loss in the crop yield across the globe. Introducing resistant genes to different crop varieties has provided effective results in combating the diseases caused by various fungal pathogens. The disease resistant dominant *Hm1* gene which is mapped in the long arm of the chromosome 1 in maize provides broad spectrum of complete resistance against the fungal pathogen *C. carbonum* race1 (CCR1). Although the NADPH-dependent enzyme produced by *Hm1* acts on the fungal HC-toxin to detoxify it, the molecular proteomic characterization of this protein remains elusive. Based on the available crystal structures, the 3-D structure of the HCTR was modelled and structural refinement was performed by energy minimization, and MD simulation. Although HCTR shares a mere sequence identity of 29.6% with its closest homolog DFR of grape, the structural superposition between equivalent C $\alpha$  and backbone atom pairs of HCTR and DFR showed that the key secondary structural elements i.e., beta-alpha-beta folds, the key components of NADPH and NADP dependent enzymes are strongly conserved and follow a common folding pattern with a very low RMSD of 1.22 Å (backbone atoms) and 1.38 Å (C $\alpha$  atoms) respectively. The HCTR model had a stable conformation in response to the atomic flexibility and interaction, when subjected to MD simulation at 10 ns in aqueous solution. Molecular docking experiment confirmed DFR and HCTR are NADPH dependent enzymes and follow the same mode of binding having high affinity towards cofactor NADPH. From the present study we can conjecture that Phe19, Gly21, Arg40, Thr90, Arg218, Gly208, Glu221 and Thr222 as the important determinant residues in cofactor binding, as they have strong hydrogen bonding interactions and electrostatic interactions with NADPH. The other conserved residues within the active site of HCTR are involved in hydrophobic interaction with the cofactor molecule, that aid in the overall stability of the docked complex. Although the modelled HCTR structure showed considerable structural identities with the cofactor binding site of the template, there was a significant difference in the catalytic site and substrate binding site represents the uniqueness of this enzyme which may have lead to their functional difference. Molecular docking of HC-toxin to the docked complex of HCTR-NADPH revealed that HC-Toxin is enclosed by a binding pocket lined by mostly hydrophobic residues and prefers to bind at a position which is close to the docked NADPH structure. Alanine scanning followed by docking study revealed that Phe19 and Arg40 are two crucial residues for strong affinity between cofactor and the enzyme. However, future studies involving site-directed mutagenesis along with wet-lab experiments may be carried out to confirm these predictions. In future, potential of such R-genes in diseases crop management can be achieved by the exploration of similar genes in other economically important crop varieties for developing novel disease resistant crops.

## Conflict of interest

No conflict of interests exists.

## Acknowledgements

This work was financially supported by the grant of Bioinformatics Initiative Division, Department of Electronics and Informatics Technology (DEITY), Ministry of Communications and Information Technology, Government of India. We also gratefully acknowledge the financial support by Biotechnology Information System Network (BTISNET), Department of Biotechnology, Government of India.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.jmgm.2013.08.011>.

## References

- [1] Z. Nimchuk, T. Eulgem, B.F. Holt, J.L. Dangl, Recognition and response in the plant immune system, *Annu. Rev. Genet.* 37 (2003) 579–609.
- [2] P. Tiffin, D.A. Moeller, Molecular evolution of plant immune system genes, *Trends Genet.* 22 (2006) 662–670.
- [3] S.T. Chisholm, G. Coaker, B. Day, B.J. Staskawicz, Host-microbe interactions: shaping the evolution of the plant immune response, *Cell* 124 (2006) 803–814.
- [4] J.G. Bishop, A.M. Dean, T. Mitchell-Olds, Rapid evolution in plant chitinases: molecular targets of selection in plant-pathogen coevolution, *Proc. Natl. Acad. Sci. USA* 97 (2000) 5322–5327.
- [5] G.B. Martin, A.J. Bogdanove, G. Sessa, Understanding the functions of plant disease resistance proteins, *Annu. Rev. Plant Biol.* 54 (2003) 23–61.
- [6] J.L. Dangl, J.D.G. Jones, Plant pathogens and integrated defence responses to infection, *Nature* 411 (2001) 826–833.
- [7] J.D.G. Jones, J.L. Dangl, The plant immune system, *Nature* 444 (2006) 323–329.
- [8] L.Q. Zhang, A.S. Peek, D. Dunams, B.S. Gaut, Population genetics of duplicated disease-defense genes, *hm1* and *hm2*, in maize (*Zea mays* ssp. *mays* L.) and its wild ancestor (*Zea mays* ssp. *parviglumis*), *Genetics* 162 (2002) 851–860.
- [9] R.J. Wisser, P.J. Balint-Kurti, R.J. Nelson, The genetic architecture of disease resistance in maize: a synthesis of published studies, *Phytopathology* 96 (2006) 120–129.
- [10] P.J. Balint-Kurti, G.S. Johal, Maize disease resistance, in: J.L. Bennetzen, S.C. Hake (Eds.), *Handbook of Maize: Its Biology*, Springer, New York, 2009, pp. 229–250.
- [11] A. Ullstrup, Two physiologic races of *Helminthosporium maydis* in the corn belt, *Phytopathology* 31 (1941) 508–521.
- [12] J.D. Walton, HC-toxin, *Phytochemistry* 67 (2006) 1406–1413.
- [13] G.S. Johal, S.P. Briggs, Reductase activity encoded by the *HM1* disease resistance gene in maize, *Science* 258 (1992) 985–987.
- [14] G. Noctor, Metabolic signalling in defence and stress: the central roles of soluble redox couples, *Plant Cell Environ.* 29 (2006) 409–425.
- [15] D.S. Multani, R.B. Meeley, A.H. Paterson, J. Gray, S.P. Briggs, G.S. Johal, Plant-pathogen microevolution: molecular basis for the origin of a fungal disease in maize, *Proc. Natl. Acad. Sci. USA* 95 (1998) 1686–1691.
- [16] O.E. Nelson, A.J. Ullstrup, Resistance to leaf spot in maize, *J. Hered.* 55 (1964) 195–200.
- [17] S. Chintamanani, D.S. Multani, H. Ruess, G.S. Johal, Distinct mechanisms govern the dosage-dependent and developmentally regulated resistance conferred by the maize *Hm2* gene, *Mol. Plant-Microbe Interact.* 21 (2008) 79–86.
- [18] P. Petit, T. Granier, B.L. d'Estaintot, C. Manigand, K. Bathany, J.M. Schmitter, V. Lauvergeat, S. Hamdi, B. Gallois, Crystal structure of grape dihydroflavonol 4-reductase, a key enzyme in flavonoid biosynthesis, *J. Mol. Biol.* 368 (2007) 1345–1357.
- [19] H. Shao, R.A. Dixon, X. Wang, Crystal structure of vestitone reductase from alfalfa (*Medicago sativa* L.), *J. Mol. Biol.* 369 (2007) 265–276.
- [20] M. Gargouri, C. Manigand, C. Maugé, T. Granier, B.L. d'Estaintot, O. Cala, I. Pianet, K. Bathany, J. Chaudière, B. Gallois, Structure and epimerase activity of anthocyanidin reductase from *Vitis vinifera*, *Acta Crystallogr. D Biol. Crystallogr.* 65 (2009) 989–1000.
- [21] M. Punta, P.C. Coggill, R.Y. Eberhardt, J. Mistry, J. Tate, C. Boursnell, N. Pang, K. Forslund, G. Ceric, J. Clements, A. Heger, L. Holm, E.L. Sonnhammer, S.R. Eddy, A. Bateman, R.D. Finn, The Pfam protein families database, *Nucleic Acids Res.* 40 (2012) D290–D301.
- [22] I. Letunic, T. Doerks, P. Bork, SMART 7: recent updates to the protein domain annotation resource, *Nucleic Acids Res.* 40 (2012) D302–D305.
- [23] A. Marchler-Bauer, S. Lu, J.B. Anderson, F. Chitsaz, M.K. Derbyshire, C. DeWeese-Scott, J.H. Fong, L.Y. Geer, R.C. Geer, N.R. Gonzales, M. Gwadz, D.I. Hurwitz, J.D. Jackson, Z. Ke, C.J. Lanczycki, F. Lu, G.H. Marchler, M. Mullokandov, M.V. Omelchenko, C.L. Robertson, J.S. Song, N. Thanki, R.A. Yamashita, D. Zhang, N. Zhang, C. Zheng, S.H. Bryant, CDD: a conserved domain database for the functional annotation of proteins, *Nucleic Acids Res.* 39 (2011) D225–D229.
- [24] E.M. Zdobnov, R. Apweiler, InterProScan – an integration platform for the signature-recognition methods in InterPro, *Bioinformatics* 17 (2001) 847–848.
- [25] T.N. Petersen, S. Brunak, G. von Heijne, H. Nielsen, SignalP 4.0: discriminating signal peptides from transmembrane regions, *Nat. Methods* 8 (2011) 785–786.
- [26] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M.R. Wilkins, R.D. Appel, A. Bairoch, in: J.M. Walker (Ed.), *Protein identification and analysis tools on the Expasy server*, The Proteomics Protocols Handbook Humana Press, 2005, pp. 571–607.
- [27] Y. Wei, J. Thompson, C.A. Floudas, CONCORD: a consensus method for protein secondary structure prediction via mixed integer linear optimization, *Proc. R. Soc. A* (2011) (doi: rspa.2011.0514v1-rspa20110514).
- [28] G.M. Boratyn, A.A. Schäffer, R. Agarwala, S.F. Altschul, D.J. Lipman, T.L. Madden, Domain enhanced lookup time accelerated BLAST, *Biol. Direct* 7 (2012) 12.
- [29] K. Ginalski, A. Elofsson, D. Fischer, L. Rychlewski, 3D-Jury: a simple approach to improve protein structure predictions, *Bioinformatics* 19 (2003) 1015–1018.

- [30] J. Lundstrom, L. Rychlewski, J. Bujnicki, A. Elofsson, Pcons: a neural-network-based consensus predictor that improves fold recognition, *Protein Sci.* 10 (2001) 2354–2362.
- [31] M.A. Kurowski, J.M. Bujnicki, GeneSilico protein structure prediction meta-server, *Nucleic Acids Res.* 31 (2003) 3305–3307.
- [32] C. Combet, M. Jamion, G. Deléage, C. Geourjon, Geno3D: automatic comparative molecular modelling of protein, *Bioinformatics* 18 (2002) 213–214.
- [33] A. Roy, A. Kucukural, Y. Zhang, I-TASSER: a unified platform for automated protein structure and function prediction, *Nat. Protoc.* 5 (2010) 725–738.
- [34] L.A. Kelley, M.J.E. Sternberg, Protein structure prediction on the web: a case study using the Phyre server, *Nat. Protoc.* 4 (2009) 363–371.
- [35] S. Wu, Y. Zhang, LOMETS: a local meta-threading-server for protein structure prediction, *Nucleic Acids Res.* 35 (2007) 3375–3382.
- [36] B.N. Dominy, C.L. Brooks, Development of generalized Born model parameterization for proteins and nucleic acids, *J. Phys. Chem.* 103 (1999) 3765–3773.
- [37] B.R. Brooks, R.E. Brucolieri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus, CHARMM: a program for macromolecular energy, minimization, and dynamics calculations, *J. Comput. Chem.* 4 (1983) 187–217.
- [38] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK – a program to check the stereochemical quality of protein structures, *J. Appl. Cryst.* 26 (1993) 283–291.
- [39] C. Colovos, T.O. Yeates, Verification of protein structures: patterns of non-bonded atomic interactions, *Protein Sci.* 2 (1983) 1511–1519.
- [40] D. Eisenberg, R. Luthy, J.U. Bowie, VERIFY3D: assessment of protein models with three-dimensional profiles, *Methods Enzymol.* 277 (1997) 396–404.
- [41] M. Wiederstein, M.J. Sippl, ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins, *Nucleic Acids Res.* 35 (2007) W407–W410.
- [42] J.C. Gelly, A.P. Joseph, N. Srinivasan, A.G. de Brevern, iPBA: a tool for protein structure comparison using sequence alignment strategies, *Nucleic Acids Res.* 39 (2011) W18–W23.
- [43] T. Kawabata, MATRAS: a program for protein 3D structure comparison, *Nucleic Acids Res.* 31 (2003) 3367–3369.
- [44] D. Frishman, P. Argos, Knowledge-based protein secondary structure assignment, *Proteins* 23 (1995) 566–579.
- [45] W. Kabsch, C. Sander, Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, *Biopolymers* 22 (1983) 2577–2637.
- [46] R.P.S. Walter, H.H. Philippe, G.T. Ilario, E.M. Alan, R.B. Salomon, F. Jens, E.T. Andrew, H. Thomas, K. Peter, F.G. Wilfred, The GROMOS biomolecular simulation program package, *J. Phys. Chem.* 103 (1999) 3596–3607.
- [47] D.V.D. Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, H.J. Berendsen, GROMACS: fast, flexible, and free, *J. Comput. Chem.* 26 (2005) 1701–1718.
- [48] B. Hess, H. Bekker, H.J.C. Berendsen, J.G.E.M. Fraaije, LINCS: a linear constraint solver for molecular simulations, *J. Comput. Chem.* 18 (1997) 1463–1472.
- [49] S. Miyamoto, P.A. Kollman, Settle: an analytical version of the SHAKE and RATTLE algorithm for rigid water models, *J. Comp. Chem.* 13 (1992) 952–962.
- [50] V.B. Chen, W.B. Arendall III, J.J. Headd, D.A. Keedy, R.M. Immormino, G.J. Kapral, L.W. Murray, J.S. Richardson, D.C. Richardson, MolProbity: all-atom structure validation for macromolecular crystallography, *Acta Crystallogr. D Biol. Crystallogr.* 66 (2010) 12–21.
- [51] G.S. Wu, D.H. Robertson, C.L. Brooks, V. Michal, Detailed analysis of grid-based molecular docking: a case study of CDOCKER-A CHARMM-based MD docking algorithm, *J. Comput. Chem.* 24 (2003) 1549–1562.
- [52] R. Huey, G.M. Morris, A.J. Olson, D.S. Goodsell, A semi-empirical free energy force field with charge-based desolvation, *J. Comput. Chem.* 28 (2007) 1145–1152.
- [53] A.K. Malde, L. Zuo, M. Breeze, M. Stroet, D. Pogar, P.C. Nair, C. Oostenbrink, A.E. Mark, An Automated force field Topology Builder (ATB) and repository: version 1.0, *J. Chem. Theory Comput.* 7 (2011) 4026–4037.
- [54] K. Guruprasad, B.V. Reddy, M.W. Pandit, Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting *in vivo* stability of a protein from its primary sequence, *Prot. Eng.* 4 (1990) 155–164.
- [55] P. Larsson, B. Wallner, E. Lindahl, A. Elofsson, Using multiple templates to improve quality of homology models in automated homology modelling, *Protein Sci.* 17 (2008) 990–1002.
- [56] S. Rao, M. Rossmann, Comparison of super-secondary structures in proteins, *J. Mol. Biol.* 76 (1973) 241–256.
- [57] N.A. Baker, D. Sept, S. Joseph, M.J. Holst, J.A. McCammon, Electrostatics of nanosystems: application to microtubules and the ribosome, *Proc. Natl. Acad. Sci. USA* 98 (2001) 10037–10041.
- [58] H.A. Thomas, Merck molecular force field I basis, form, scope, parameterization, and performance of MMFF94, *J. Comp. Chem.* 17 (1996) 490–519.