RESEARCH ARTICLE

IN-SILICO CHARACTERIZATION AND HOMOLOGY MODELING OF PEPCK ENZYME OF *MEDICAGO TRUNCATULA*

Nitin Kumar Verma[1]*, Atal Bihari Bajpai[2], Vibha[3] and Saurab Kumar Yadav[4]

[1.] Department of Biotechnology and Bioinformatics, Uttaranchal College of Science and Technology, Dehradun
[2.] Department of Botany, D.B.S. (P.G.) College, Dehradun
[3.] Genetics and Tree Propagation Division, Forest Research Institute, Dehradun
[4.] Department of Biotechnology, OPJS University, Rajasthan

Received-01.12.2016, Revised-10.03.2017

Abstract: Phospho*enol*pyruvate carboxykinase (PEPCK) is an enzyme in the lyase family. PEPCK is an ATP-dependent that is involved in the metabolic pathway of gluconeogenesis. It converts oxaloacetate into phospho*enol*pyruvate and carbon dioxide. In this study, the results of structural and physiochemical study of *Medicago truncatula* PEPCK has explored. The conceptual three-dimensional structure investigated while there was no structural information available in any other database. Computational analysis performed on *Medicago truncatula* PEPCK and developed a three-dimensional structure of PEPCK enzyme using comparative modeling approach. The modeled enzyme includes N-terminal and C-Terminal domains with a mixed α/β topology. The energy of constructing models was minimized and the quality of the models was evaluated by VERRIFY_3D and PROCHECK. Ramachandran plot analysis showed the confirmation of 100 % amino acid residues was within the most favored regions. Multiple sequence alignment of the PEPCK protein sequence of different plant sources revealed the conserved region and constructed a phylogenetic tree. The stability of model checked through Gromacs 4.5. The final three-dimensional structure submitted in the protein model database (PMDB). This study may play keystone role in in-vivo and in-vitro studies.

Keyword Phospho*enol*pyruvate carboxykinase, phylogenetic tree, Gromacs, MD simulation, Homology Modeling

INTRODUCTION

Phospho*enol*pyruvate carboxykinase (EC: 4.1.1.49) is an inducible enzyme, which is found in plant, animal and microorganism cells (Walker *et al*., 2002). PEPCK is a cytosolic enzyme that catalyzes the reversible reaction in plants (Malone *et al*., 2007):

$$OAA + ATP \leftrightarrow PEP + ADP + CO_2$$

PEPCK occurs in phloem companion cells, roots, the flesh of fruits, stomata guard cells, simple and glandular trichomes, latex-producing ducts, developing seeds, germinating seeds and in the leaves of many C4 and CAM (Crassulacean acid metabolism) plants (Borland *et al.*, 1998; Kim and Smith, 1994; Leegood and Walker, 2003; Walker and Chen, 2002). The function of PEPCK has clearly established in some of the plant cell type. In germinating seeds, it catalyzed an essential reaction in the conversion of lipids and some amino acids to sugar by gluconeogenesis (Leegood and Ap Rees, 1978; Rylott *et al.*, 2003). In the leaves of some C_4 and CAM plants, it functions as a decarboxylase in photosynthetic CO_2-concerning mechanism (Burnell and Hatch, 1988; Dittrich et al., 1973). This evidence showed that PEPCK plays a vital role in the metabolism of nitrogenous compound in some other cells and tissues (Chen *et al.*, 2004; Delgado-Alvarado et al., 2007; Leegood and Walker, 2003). The mechanism and regulation of C_4 acid decarboxylation in phospho*enol*pyruvate (PEP) carboxykinase-type C_4 plants examined in isolated bundle sheath cell strands. The ATP is required to derive this reaction and derived from mitochondrial respiration, due almost exclusively to malate oxidation via a NAD-specific malic enzyme. The contribution of photosynthesis was negligible to the ATP consumed by this cytosolic reaction (Carnal *et al.*, 1993).

Protein executes most of the work of the living cell. With the innovation of sequence technology, it is now significantly easier to obtain the uncharacterized function of protein in the organism. Still, there are the protein sequences with their functions yet to be discovered or experimentally confirmed (Miller and Attwood, 2003). *Insilico* analysis help in determining the protein functions, which can be divide into three broad categories: sequence, expression and interaction based methods. Sequence based methods rely on the ability to construct the alignment among the protein sequences (Pellegrini, 2001). The sequence analysis approach was used for the prediction of protein domain family and their identification was based on the amino acid sequence similarity. However, structural analysis and comparative study permit confirming the function of the protein and all these predictions are based on the sequence alignment (Pearson, 2013).

In this study, the theoretical 3D model of PEPCK enzyme of *Medicago truncatula* was constructed in order to elucidate the structural properties. The various computational tools used to validate the model structure, and its stability was checked by Molecular dynamics (MD) simulation.

*Corresponding Author

MATERIAL AND METHOD

Sequence Retrieval and Template selection

The protein sequence of Phospho*enol*pyruvate carboxykinase enzyme of *Medicago truncatula* was retrieved from NCBI (National Center for Biotechnology Information) (Wheeler et al., 2003) database for modeling, which have 658 amino acid residues and their accession number is XP_003612927.1. Comparative modeling usually starts by searching the PDB (Protein Data Bank) (Berman et al., 2000) of the known protein structure using the target sequence as a query. This search is generally done by comparing the target sequence with the sequence of each structure in the database. The protein sequence of Phospho*enol*pyruvate carboxykinase (PEPCK) of "*Medicago truncatula*" used as a target sequence. For identification of suitable template, BLASTp (Johnson et al., 2008) was performed against PDB, and the template was selected for generating the putative 3D structures.

Secondary structure analysis

Secondary structure of PEPCK protein comprises mainly of coils, sheets and helix. In the present study secondary structure of PEPCK protein was predicted by using PDBsum (De Beer et al., 2014) and SOPMA (Geourjon and Deléage, 1995). The predicted secondary structure helped us to improve the alignment between target and template protein.

Comparative Modeling

The method of comparative modeling requires for the identification of homologous sequences with known structure. The first step of comparative modeling is scanning and selection of template protein structure using the target sequence as query (Berman et al., 2000). BLASTp is generally used for template selection procedure (Johnson et al., 2008). BLAST results output potential template for modeling is 1II2, and the high degree of primary sequence identity was 49.40% between Phospho*enol*pyruvate carboxykinase of "*Medicago truncatula*" (target) and *Trypanosoma cruzi* (template) indicates the crystallographic structure that was good model to be used as template for PEPCK enzyme. The alignment of the *Mt*PEPCK target and *Trypanosoma cruzi* (PDBID: 1II2) is shown in Fig. 1. Fifty models were constructed through MODELLER 9v10 (John and Sali, 2003). MODELLER implements comparative protein modeling by satisfaction of spatial restraints.

Analysis of the model

The several models generated through Modeller for the same target and the best model selected for further analysis. We evaluated the model with the lowest value of the Modeller objective function and PROCHECK statistics (Laskowski et al., 1993). The overall stereo-chemical quality of the models for PEPCK enzyme checked through Ramachandran's map calculation it was computed through the Structure Analysis and Verification Server (PROCHECK and VERIFY-3D) program that is available on Naional Institute of Health (NIH) server (http://services.mbi.ucla.edu/SAVES/) (Eisenberg et al., 1997).

Visualization and structural analysis

Visualization and Structural analysis of the final protein model was carried out through PyMol software (DeLano, 2002). 3dSS (Sumathi et al., 2006) and CHIMERA (Pettersen et al., 2004) were used for the structural alignment and RMSD calculation with model and template protein structure.

Active site analysis

After building of protein model, the possible deep cleft active site of PEPCK was explored through CASTp program (Binkowski et al., 2003) (http://sts.bioengr.uic.edu/castp).

Physiochemical characterization:

For physiochemical characterization theoretical pI (isoelectric point), molecular weight, -R and +R (total number of positive and negative residue), EI (extinction coefficient), II (instability index), AI (aliphatic index), and GRAVY (grand average hydropathy) were computed using the Expasy's ProtParam server for set of proteins (http://us.expasy.org./tools/protparam.html), and the results are shown in Table 2.

Multiple sequence alignment and Phylogenetic analysis

The amino acid sequence of M. *truncatula* PEPCK and their all identified homologs were subjected to multiple sequence alignment for recognition of conserved residues and pattern using CLUSTRAL-W program implemented in MEGA-5 program (Tamura et al., 2011). Phylogenetic analysis is a method to find the evolutionary relationship between the species using their protein sequences. The NJ method implemented in MEGA-5 program was used to construct the phylogenetic tree.

Molecular Dynamics Simulation of PEPCK

MD simulation was performed through GROMACS (GROningen MAchine for Chemical Simulation) (Van Der Spoel et al., 2005) package, with Gromacs 96.1 (43A1) force field. The accurate force field is necessary for reproducing the conformational and dynamic behavior of condensed-phase system. The Gromos 96.1 force field has well parameterized for protein. For solvate the model, it has 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 applied in all directions. The system was neutralized by adding six Na+ counter ions since the overall protein charge was negative. To get rid of the high-energy interactions and steric clashes of the system, the energy of the system minimized using the steepest descent method until a tolerance of 1000 KJ/mol. All the bond lengths were constrined with the LINear Constrains Solver (LINCS) (Hess et al., 1997) method, whereas the geometry of water molecules was constrained with SETTLE algorithm

(Miyamoto and Kollman, 1992). The energy minimization system treated for 200 ps at constant temperature (300 K), pressure (1 atm) and without any position restraints. All the analysis of MD simulation was performed using XMGRACE software (Hartmann, 2009). The best model for further analysis selected at 1000 ps MD simulation and the structure has low RMSD value.

Submission of the modeled protein in protein model database (PMDB)

The generated model of *Medicago truncatula* PEPCK protein submitted successfully in Protein model database without any stereochemical errors. Submittion id of the model is PMID: PM00779121.

RESULT

Physiochemical characterization

Phospho*enol*pyruvate carboxykinase protein (72.475 Kda) of *Medicago truncatula* was extracted from Genbank. Glycine 8.20 and Tryptophan 1.22% (fig. 1) was identified the most and the least number of amino acids in the sequence. shown

The primary sequence analysis PEPCK enzyme is illustrated in Table 2 since, the isoelectric point (pI), solubility is minimized and mobility of an isoelectro focusing system is zero due to calculated pI will be useful. Isoelectric point (pI) is the pH at which the surface of the protein covered by the charge, but the net charge of the protein is zero. At pI, protein is stable and dense. In processing of buffer system, purification by isoelectric focusing method the computed isoelectric point (pI) will be valuable. While Expasy's Protparam computes the extinction coefficient of 276, 278, 279, 280 and 282 nm wavelengths, 280 nm has elected since protein absorbed light strongly. Extinction coefficient of PEPCK enzyme of *Medicago truncatula* at 280 nm was 77280 M^{-1} cm^{-1}. The computed extinction coefficient can help in the quantitative study of protein-protein and protein-ligand interaction in solution. The instability index provides and determines of the stability of protein in a test-tube. There are definite dipeptides and the occurrence of which is particularly divergent in the unstable protein compared with those in the stable once. This method assigned a weight value of instability, which is feasible to compute an instability index (II). A protein whose instability index is slighter than 40 is estimated as stable and a value above 40 estimates that the protein may be unstable. The instability index of PEPCK enzyme of *Medicago truncatula* was found 35.20, which indicates that the protein is stable. The aliphatic index (AI) is elucidated as the relative volume of a protein occupied by aliphatic side chains (A, V, I and L) that is estimated as a positive factor for increase of the thermal stability of globular proteins. Aliphatic index (AI) for the PEPCK enzyme was 75.91. A very high aliphatic index of the protein sequence indicates that the protein may be stable for a vast temperature range. The minimal thermal stability of protein was indicative for a more flexible structure then compared to other protein. The Grand Average hydropathy (GRAVY) value for a peptide or protein is calculated and the sum of hydropathy values of all the amino acids, divided by the number of residues in the sequence. A GRAVY index of PEPCK enzyme was -0.395. This low value shows the probability of better interaction with water.

Sequence analysis

The accession number of retrieved sequences along with the organism's name is listed in Table 1. The multiple sequence alignment was performed for nine plant species (Table 1) using ClustralX with Bioedit. The tool is used with default parameters, and results showed the presence of some conserved regions among all the sequences from different sources and only in some regions have a point different that can environmental effect such as mutation.

Besides, the sequence similarity among these enzymes is shown in Table 3 and the conserved protein portion or residues with their respective positions in *Medicago truncatula* PEPCK (fig. 3) and other organism's sequences from residues 110 to 145, 180 to 270 and from 272 to 660. The phylogenetic analysis was used through MEGA 6 software, and the phylogenetic tree constructed using NJ plot (fig. 4). The results revealed that the PEPCK protein *L.alabamica* of Brassicaceae family was closely related to *A.thaliana*, *P.trichocarpa* and *M.truncatula* while *M.notabilis*, *N.tabacum and S.lycopersicum* are closely related, and these two clusters were distantly related to *Z.mays* and *C.sativus* (fig. 4).

The secondary structure analysis showing 39.06% amino acid residues involved in random coil formation it is higher and 7.90% amino acid residues involved in β-turn formation. Others 30.55% residues involved in α-helix formation and 22.40% residues involved in extended strand formation (Table 4). According to topology analysis, 27 β-strands and 19 α-helixes were found (fig. 6) and fig. 4 showing the α-helix, β-strands and coils position in the structure.

Comparative Modeling

Comparative modeling of protein provides a significant hypothesis of homology between the target and template. This approach provides reasonable results based on the assumption that the tertiary structure of the two proteins will be similar if their sequences are related. Absence of the experimentally determined three-dimensional structure of Phospho*enol*pyruvate carboxykinase protein of *Medicago truncatula* is in experimentally proved database (Protein Data Bank). Comparative modeling method was utilized to construct its theoretical three-dimensional structure. BLAST (Basic Local Alignment Search Tool) scanning results revealed more identical with carystallographic structure *Trypanosoma cruzi* (PDBID: 1II2 with 2.0

Å resolution) while the template was determined on the basis of higher sequence identity. Sequence identity is 49.4% with 330 conserved residues and 66.5% sequence similarity. Comparative modeling predicts the three dimensional structure of the hypothetical model of a given protein sequence (target) based primarily on this alignment to the template. The resulting 3-D structure of Phospho*enol*pyruvate carboxykinase was sorted according to the score calculated from discrete optimization protein (DOPE) scoring function. The final model was selected for further study (fig. 7), which has the lowest root mean square deviation (RMSD) and relative to the trace of the crystal structure.

Model quality assessment

The detailed residue-by-residue stereo-chemical quality of the modeled protein structure was evaluated by the Ramachandran plot (fig. 8) using procheck tool. The reliability of the backbone torsion angle Φ and Ψ distribution of the protein and the template was evaluated by the Ramachandran plot in Procheck tool. The received Ramachandran plot (Phi-Phi) pairs had 84.0% residues in the most favored regions, 13.20% core residues in additional allowed regions, 0.90% residues generously allowed regions and 1.30% residues in disallowed regions Table 5. This value indicates a good quality model, whenever the crystal structure of *Trypanosoma cruzi* (PDB ID: 1II2) shows 89.70% residues in most favored regions. In order to characterize the model, structural motif and mechanically important loops were assigned to build a final 3D model of PEPCK protein. The packing quality of each residue of the model was assessed by Verify_3D program where the compatibility of the model residue with their environment is assessed by score function. Residue with a 3D-1D score of >0.2 should be considered reliable. As shown in Table 5 the score of the refined model is 68.24% of the residues have an averaged 3D-1D score. ProSA revealed a Z-score of -5.85 (fig. 9) for modeled protein.

RMSD (root mean square distance) between the equivalent Cα atoms pair (target and template) was measured to check the degree of structural similarity. The best-modeled structure for fitting into the template (crystal structure) was examined, the prepare model and its closest relative were superimposed based on Cα and backbone atom pairs. Pairwise 3D alignment search of the template protein with the modeled structure through 3dSS and CHIMERA showed the massive RMSD of 0.259 Å (fig. 10) on their backbone atom. These server's results conclude that PEPCK protein and its homologues share strong structural conservation and similarity in the structural folding. It signifies that the generated model is reasonably good for further studies. The final model was scanned through Castp with 1.4 Å probe radius, it was showed 12415 volume of the cavity with 4242.8 area. The amino acid residue involved in the formation of cavity shown in (fig. 11).

Molecular dynamics Simulation of PEPCK enzyme

MD simulation using Gromacs 4.5 confirmed the high stability of the selected model. Energy minimization was performed using the steepest descent methods. The RMSD and RMSF were found to be optimum as per the Gromacs manual and are given graphically. The steepest descent energy minimization for the solvated modeled protein revealed the maximum force reached the threshold of 1000 $kJ^{-1}nm^{-1}$ in 400 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. Compared to the starting coordinates, the RMSD of the backbone atom increased in the first 500 ps and then reached a plateau in the subsequently simulation time (fig. 12). The relative flexibility of the model was also characterized by plotting the root mean-square flexibility (RMSF) related to the average structure obtained for the MD simulation trajectories. Three flexible regions have predicted for the modeled protein structure of PEPCK considering the RMSF value and represented in (fig.12). One flexible region has been located in the N-terminal end, and the RMSF value of the region is 0.92. The second flexible region is nearby middle region of the protein, and this region processes several continuous peaks. The RMSF value of the peaks in this region is 0.62 and 0.59. The C-terminal end has a single flexible region with a 0.45 RMSF value. It can be summarized from the RMSF analysis the middle region of the modeled protein is more flexible in comparison of N-terminal and C-terminal.

CONCLUSION

In this study, both sequence and structural activities of phospho*enol*pyruvate kinase enzyme through comparative genomics and molecular modeling approaches were examined. The aggregation of Bioinformatics tools was focused not only on sequence analysis but also structural information, guided us to suggest the uncharacterized information of phospho*enol*pyruvate kinase (PEPCK) protein of *Medicago truncatula*. The functional information of PEPCK protein suggested the involvement in breaking of various chemical bonds. This type of elimination reaction generates a double bond but not hydrolytic or oxidative reaction. There are two distinct types of feature found in that protein (i) PEPCK-ATP active domain with residue 135 to 603. It catalyses the first committed (rate-limiting) step in the diversion of tricarboxylic acid cycle intermediates toward gluconeogenesis. It catalyzes the reversible decarboxylation and phosphorylation of oxaloacetate to yield phospho*enol*pyruvate and carbondiooxide, using a nucleotide molecule (ATP) for the phosphoryl transfer, and has a strict

requirement for the divalent metal ion for activity. PEPCK consist of N terminal and C terminal domain, with the active site and metal ion located in the cleft between them. Both domains have α/β topology that is partly similar to one another. (ii) AAA (ATPases Associated with diverse cellular Activity) domain, they share common conserved module residues 350 to 377 amino acids. This domain belongs to functionally diverse protein family of ring shaped P-loop NTPases, which exert their activity through the energy dependent remodeling or translocation of macromolecule. AAA protein couple chemical energy provided by ATP hydrolysis to conformational changes, which are transduction into mechanical force exerted on a molecular substrate. The secondary structure analysis through SOPMA showing random coils percentage is very high 39.06%, β-turns having lowest percentage 7,90% and α helix having 30.55% whereas Extended strands having 22.49%. Secondly, a 3D model of PEPCK protein using the comparative modeling approach was constructed. The physiochemical characterization of protein by different parameter like as isoelectric point, molecular weight, total number of positive and negative residues, extinction coefficient, instability index, aliphatic index and grand average hydropathy (GRAVY) was determined. The PEPCK protein model had a stable confirmation in response to the atomic flexibility. Computational analysis used to distinguish its structural impulsive feature and validated by the SAVES. The structure-structure alignment between the best-constructed model and template (crystallographic structure) was performed. The calculated RMSD value is 0.259Å and submitted this structure into Protein Model Database (PMDB) and PMDB-ID is PM0080219. In last, we will conclude that this study will be helpful for the further experimental analysis of PEPCK.

Competing Interests

The authors declare that they have no competing interests.

```
mtPEPCK         140 NLSPAELYEQAIKYEKGSFITSNGAMATLSGAKTGRSPRDKRVV-KDKVT    188
                    ||...||.:.|:|.||.|.:|:.||:|.:|.|||||||.|||:| .|.|.
1II2_A            7 NLLSPELVQWALKIEKDSRLTARGALAVMSYAKTGRSPLDKRIVDTDDVR     56

mtPEPCK         189 ENELWWGKGSPNIEMDEETFMVNRERAVDYLNSLDKVFVNDQFLNWDPEN    238
                    ||..|   |..|:::.||:|...|:.|.::|::.:.:||.|.|...|...
1II2_A           57 ENVDW---GKVNMKLSEESFARVRKIAKEFLDTREHLFVVDCFAGHDERY    103

mtPEPCK         239 RIKVRIVSARAYHSLFMHNMCIRPSPEELENFGTPDFTIYNAGKFPCNRF    288
                    |:|||:.:.|.||:|||.:|.|.|:||||..||.||:.|||||:...:..
1II2_A          104 RLKVRVFTTRPYHALFMRDMLIVPTPEELATFGEPDYVIYNAGECKADPS    153

mtPEPCK         289 THYMTSSTSIDLNLARREMVILGTQYAGEMKKGLFSVMHYLMPKRQILSL    338
                    ...:||:|.:.||...||.|||||:||||||||:.:||..|||:...|.:
1II2_A          154 IPGLTSTTCVALNFKTREQVILGTEYAGEMKKGILTVMFELMPQMNHLCM    203

mtPEPCK         339 HSGSNMGKGGDVALFFGLSGTGKTTLSTDHNRYLIGDDEHCWSENGVSNI    388
                    |:.:|:||.|||.:|||||||||||||.|.:|.|||||||.|::.||.||
1II2_A          204 HASANVGKQGDVTVFFGLSGTGKTTLSADPHRNLIGDDEHVWIDRGVFNI    253

mtPEPCK         389 EGGCYAKCIDLSREKEPDIWNAIRFGTVLENVVFDEHTREVDYSDKSVTE    438
                    |||||||.|.|:.:.|.||::|:|||.|.||.|.|:.|.|:|:.|:|:.:
1II2_A          254 EGGCYAKAIGLNPKTEKDIYDAVRFGAVAENCVLDKRTGEIDFYDESICK    303

mtPEPCK         439 NTRAAYPIEYIPNAKLPCVGPHPKNVILLACDAFGVLPPVSKLTLAQTMY    488
                    |||.|||:.:|..|....:..||||||.|..|||||:|||::||.||.|:
1II2_A          304 NTRVAYPLSHIEGALSKAIAGHPKNVIFLTNDAFGVMPPVARLTSAQAMF    353

mtPEPCK         489 HFISGYTALVAGTE-DGIKEPQATFSACFGAAFIMLHPTKYAAMLAEKME    537
                    .|:.||||.|.|.| .|.:..:..||:|||..|::.|.|.|...||||:
1II2_A          354 WFVMGYTANVPGVEAGGTRTARPIFSSCFGGPFLVRHATFYGEQLAEKMQ    403

mtPEPCK         538 SHGATGWLVNTGWSGGSYGTG-NRIKLSYTRKIIDAIHNGSLLGAEYKKS    586
                    .|.:..||:|||::||....| .|:.|..||.||||||:|:|...||::.
1II2_A          404 KHNSRVWLLNTGYAGGRADRGAKRMPLRVTRAIIDAIHDGTLDRTEYEEY    453

mtPEPCK         587 EIFGLQIPTEVEGVPSEILDPINAWSDKNAYNATLLKLGGLFKKNF    632
                    ..:||..|..|..||..:|:|..||.|...:|.|..:|..:|:::|
1II2_A          454 PGWGLHIPKYVAKVPEHLLNPRKAWKDVRQFNETSKELVAMFQESF    499
```

Fig. 1. Sequence Alignment of *Medicago truncatula* (XP_003612927.1) with *Trypanosoma cruzi* (PDB ID: 1II2). The alignment was performed using EMBOSS.

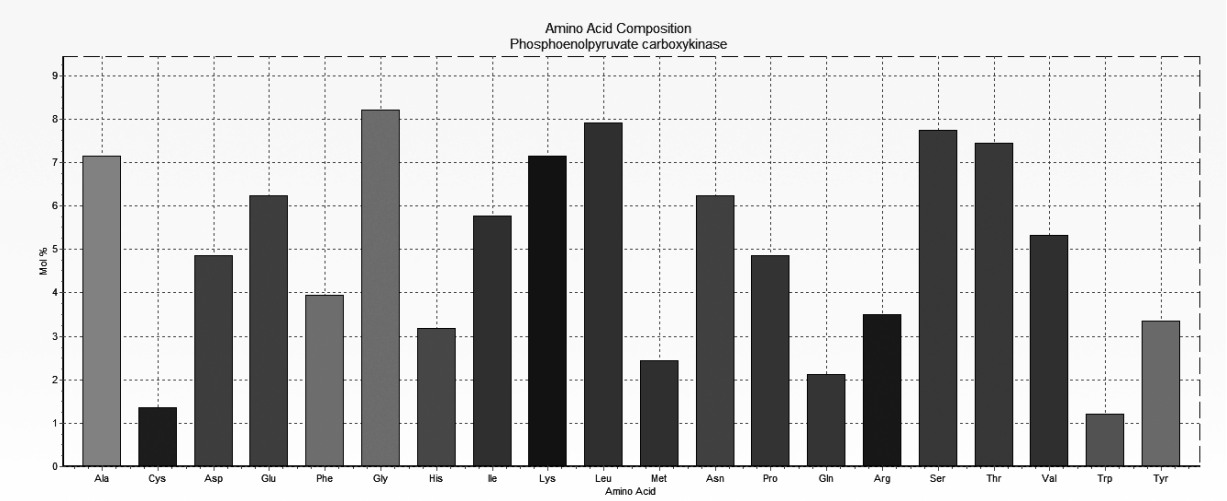

Fig. 2. Amino acid composition in PEPCK enzyme of *Medicago truncatula*

Amino Acid		Site
SLC	TLC	
G	Gly	135, 141, 204, 211, 218, 222, 243, 245, 319, 329, 359, 364, 369, 389, 393, 396, 403, 406, 408, 422, 432, 438, 439, 462, 506, 521, 541, 548, 552, 565, 588, 591, 597, 600, 601, 604, 624, 638, 647, 674, 703
A	Ala	125, 128, 192, 198, 212, 214, 219, 263, 295, 297, 328, 363, 399, 442, 458, 490, 500, 516, 519, 544, 547, 558, 562, 566, 567, 577, 581, 589, 620, 629, 701, 702
L	Leu	67, 70, 127, 130, 180, 186, 194, 216, 240, 267, 270, 280, 301, 315, 350, 370, 377, 384, 386, 400, 404, 412, 420, 447, 465, 514, 515, 523, 529, 531, 545, 580, 593, 611, 626, 627, 653, 669, 672, 675, 695, 700
M	Met	251, 303, 306, 340, 355, 366, 374, 378, 392, 535, 570, 579, 584
F	Phe	106, 206, 256, 302, 318, 323, 331, 371, 401, 402, 461, 470, 520, 538, 560, 564, 568, 637, 676
W	Trp	241, 242, 282, 428, 456, 592, 598, 659
K	Lys	73, 137,203, 220, 228, 244, 289, 367, 368, 394, 409, 443, 451, 501, 528, 554, 583, 616, 633, 662
Q	Gln	71, 119, 197, 278, 361, 533
E	Glu	133, 193, 196, 202, 250, 253, 261, 314, 316, 354, 365, 425, 437, 450, 501, 528, 554, 583, 616, 633, 662
S	Ser	55, 69, 83, 111, 122, 124, 126, 129, 176, 178, 190, 205, 209, 217, 246, 269, 294, 300, 342, 343, 345, 385, 388, 405, 413, 434, 527, 540, 561, 599, 650
P	Pro	85, 88, 136, 143, 191, 225, 247, 284, 310, 321, 332, 379, 453, 493, 498, 503, 509, 524, 525, 556, 573, 642, 649, 655, 704
V	Val	139, 231, 264, 273, 290, 356, 398, 433, 464, 468, 469, 477, 484, 505, 512, 522, 526, 546, 594
I	Ile	123, 199, 207, 249, 288, 292, 308, 325, 357, 383, 421, 436, 455, 459, 494, 497, 513, 539, 569, 609, 618, 621, 652, 699
C	Cys	307, 333, 427, 440, 444, 504, 517, 563
Y	Tyr	195, 201, 266, 298, 326, 339, 362, 376, 441, 479, 492, 536, 542, 576, 603, 613, 631
H	His	68, 184, 299, 304, 338, 375, 387, 426, 473, 508, 537, 572, 622
R	Arg	115, 132, 229, 260, 262, 291, 296, 309, 335, 352, 381, 475, 489, 608, 615
N	Asn	13, 188, 248, 259, 281, 286, 305, 334, 349, 391, 417, 435, 457, 467, 487, 511, 595, 679
D	Asp	53, 142, 177, 227, 233, 252, 265, 283, 322, 347, 397, 415, 423, 424, 446, 478, 518, 619
T	Thr	39, 63, 87, 131, 183, 208, 215, 221, 236, 255, 320, 324, 337, 341, 344, 360, 407, 410, 411, 414, 463, 474, 485, 488, 534, 543, 549, 559, 590, 596, 614

Fig. 3. Identified conserved residues with single letter code and their respective positions

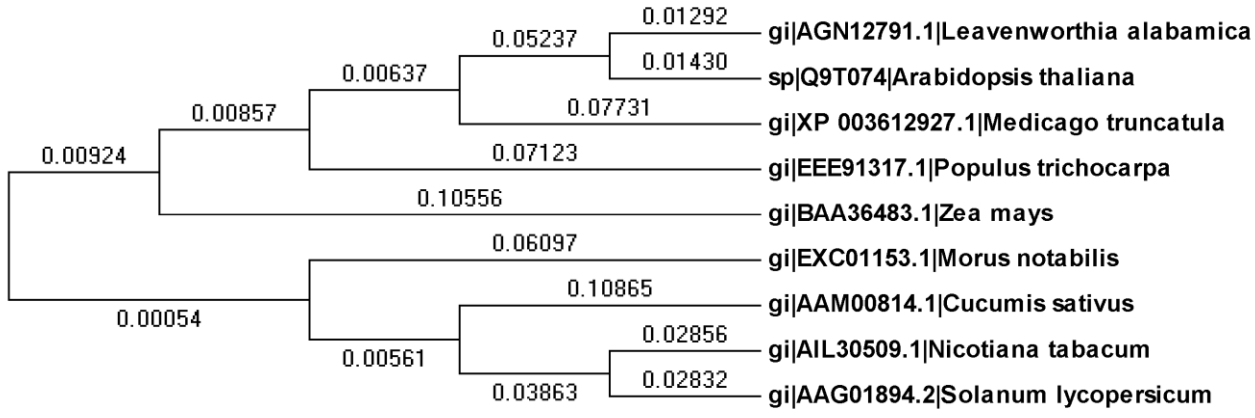

Fig. 4. Phylogenetic tree constructed through MEGA 6 in between different organism.

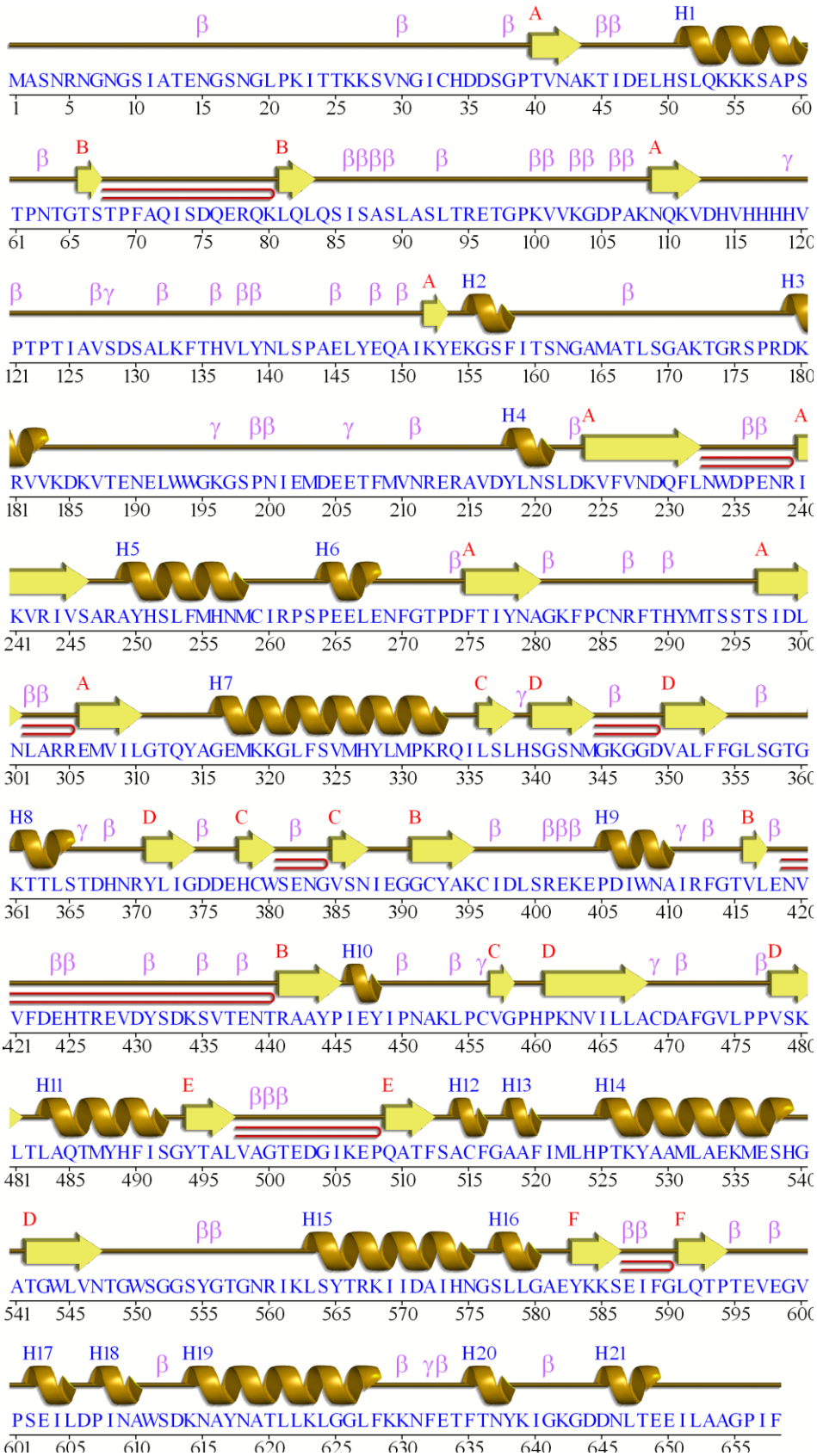

Fig. 5. The graphical representation of PEPCK protein with its secondary structure element

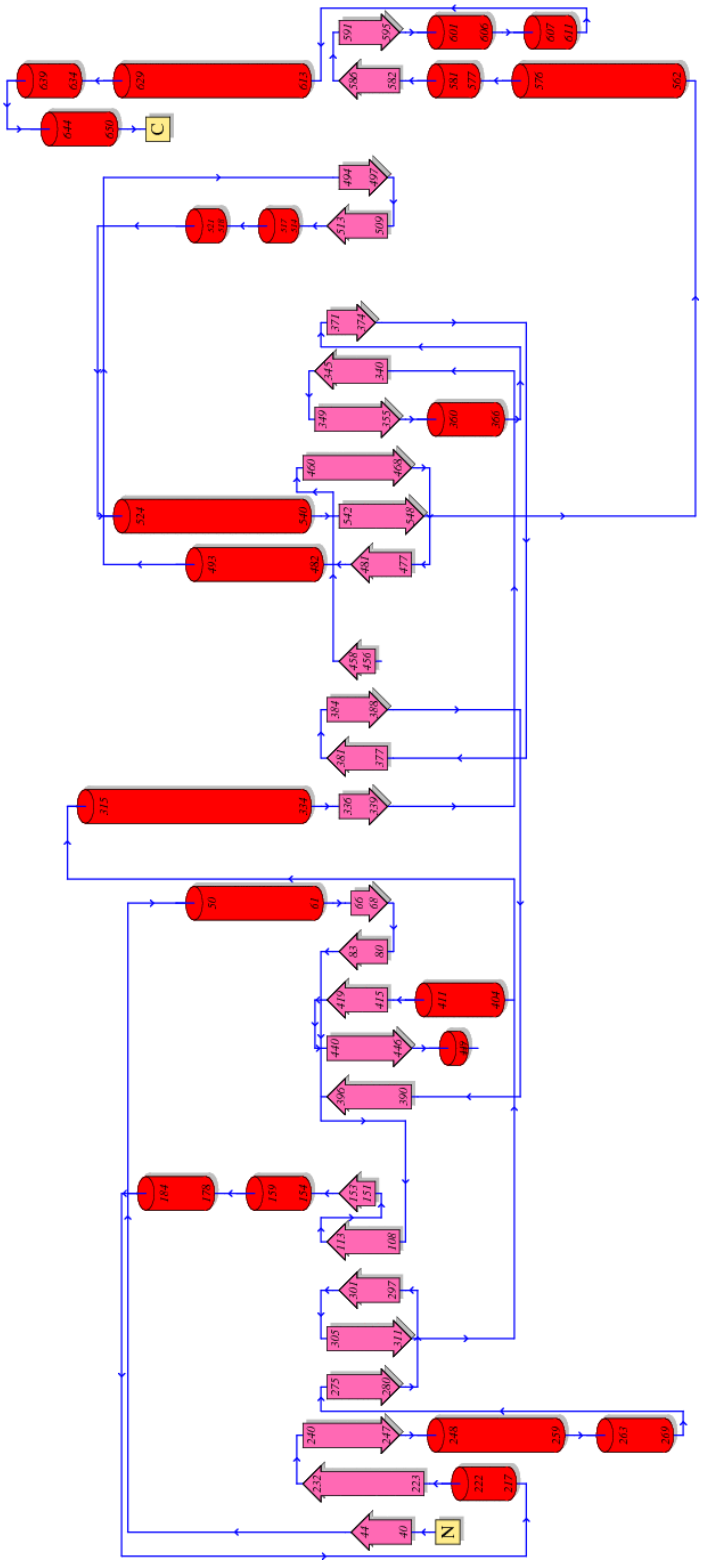

Fig. 6. Topology of predicted structure shown in distinct domains.

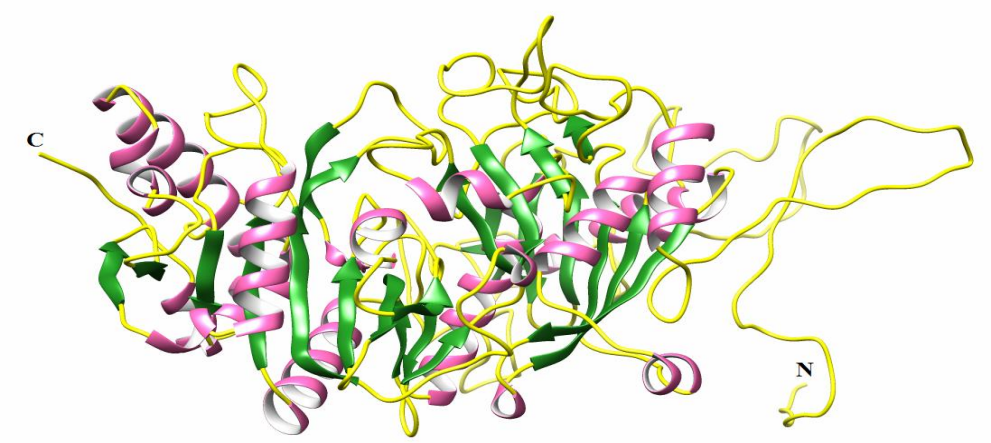

Fig. 7. Homology model of PEPCK protein (NCBI accession: gi|XP_003612927.1) from *Medicago truncatula* solid ribbon representation of hypothetical model colored by its secondary structure element.

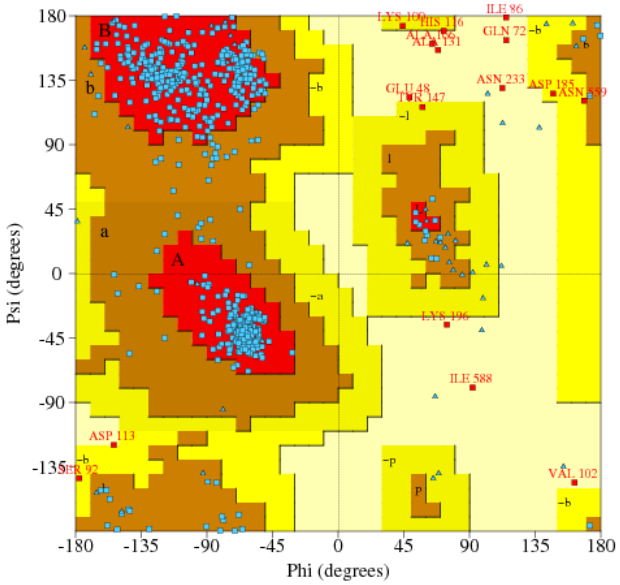

Fig. 8. Ramachandran plot of the modeled PEPCK protein (NCBI accession: gi|XP_003612927.1). The plot was calculated by PROCHECK program

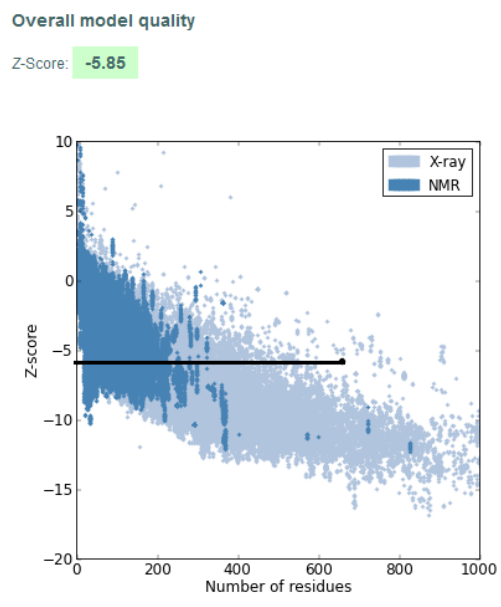

Fig. 9. The Z-score of -5.85 indicated on the graph represent the overall quality of the modeled PEPCK protein.

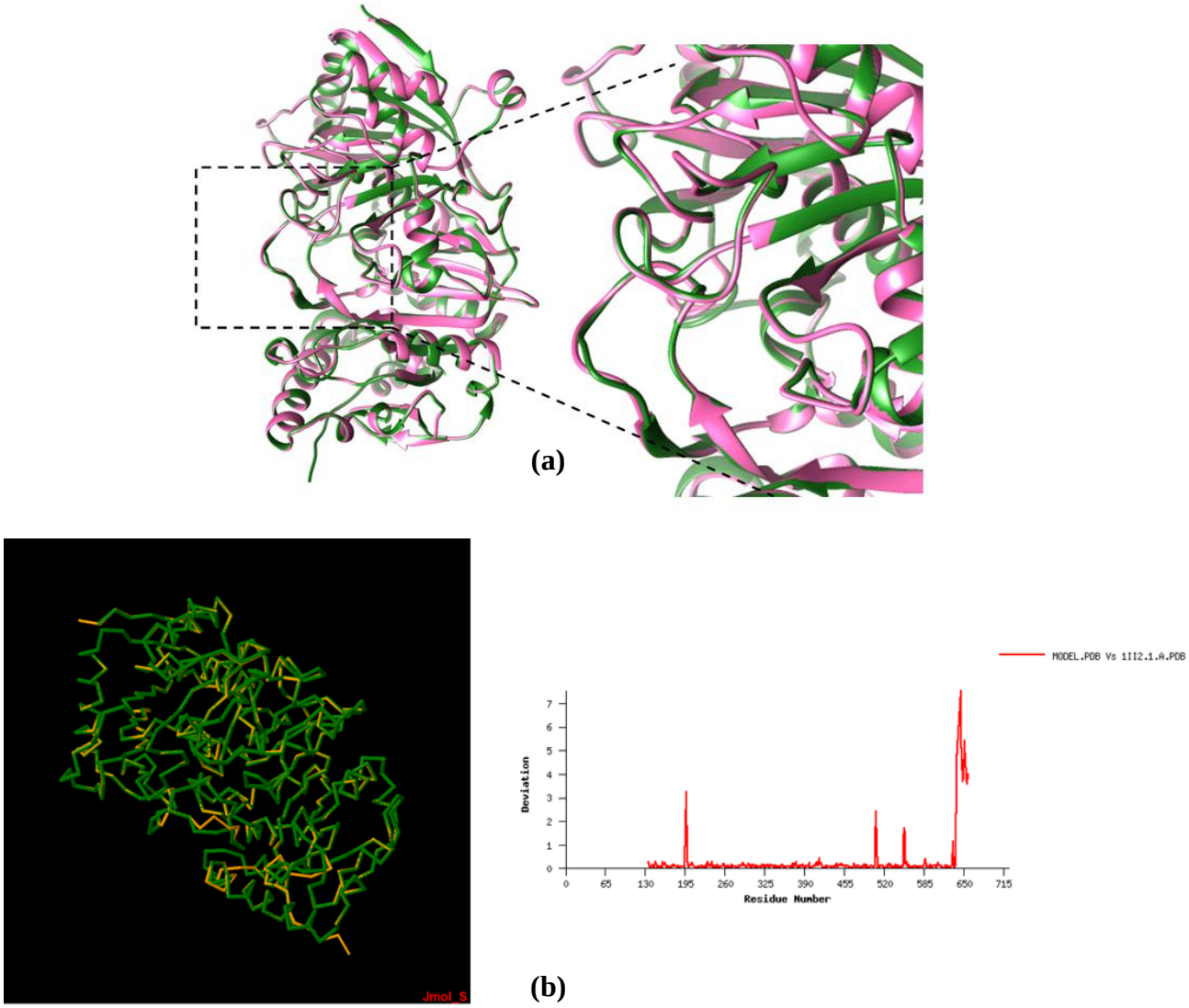

Fig. 10. Superimposition between template and model protein (a) Backbone showing the superimposition of template (*forest green color*) and the refined model (*hot pink color*) and showing the enlarge view of superimposition (b) Showing 3dSS generated image (Jmol view) and graphical representation of RMSD value of amino acid residues of template and model protein structure.

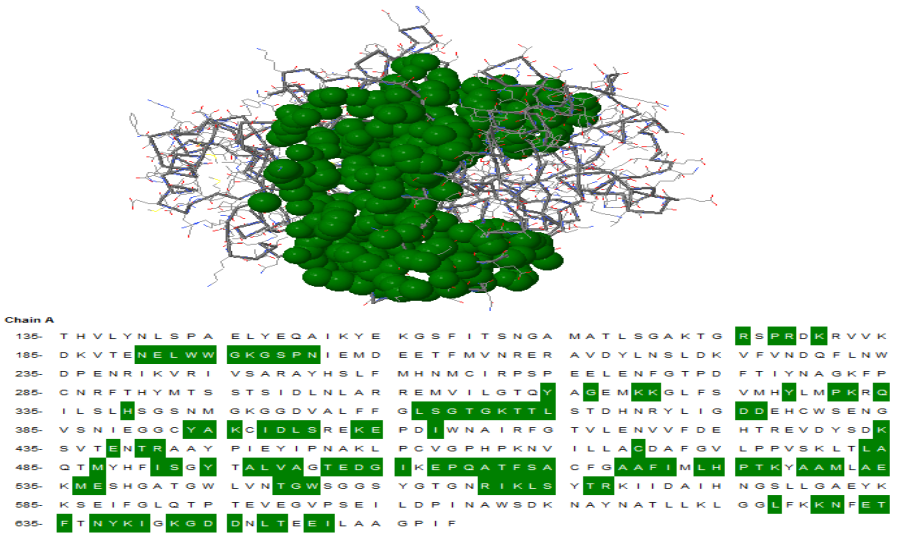

Fig. 11. Pocket identification through castp server highlighted with green color, residues involved in construction of pocket

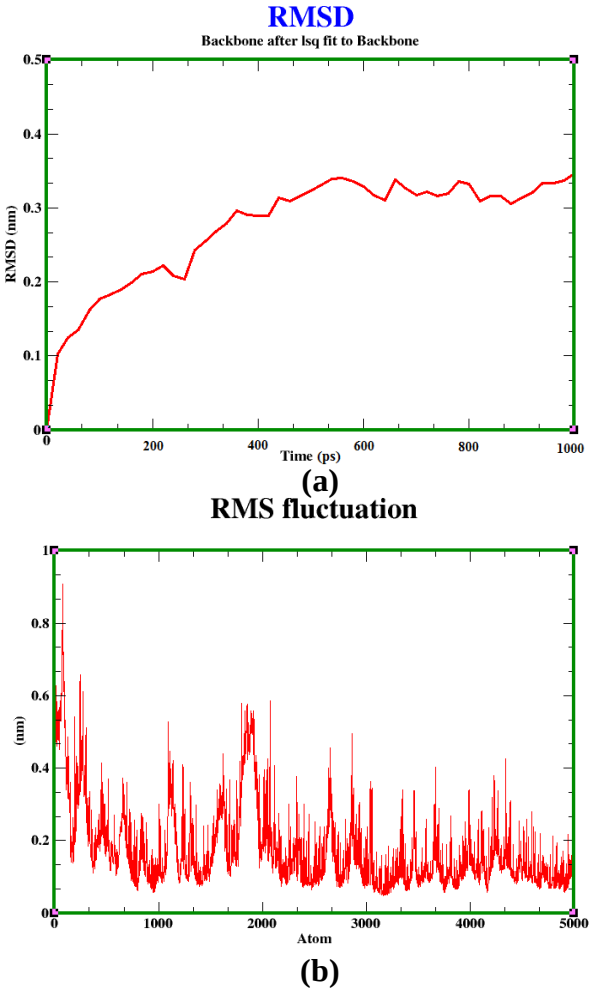

Fig. 12. The RMSD and RMSF graph of the modeled PEPCK protein during MD simulation. (a) RMSD of backbone C atom of the modeled PEPCK protein structure. (b) RMSF analysis of amino acid residues of the modeled PEPCK protein structure.

Table 1. Retrieved sequences from NCBI/Enterz with their accession number

S.No.	Name of organisms	Accession number
1.	*Medicago truncatula*	gi|XP_003612927.1
2.	*Populus trichocarpa*	gi|EEE91317.1
3.	*Cucumis sativus*	gi|AAM00814.1
4.	*Morus notabilis*	gi|EXC01153.1
5.	*Leavenworthia alabamica*	gi|AGN12791.1
6.	*Nicotiana tabacum*	gi|AIL30509.1
7.	*Zea mays*	gi|BAA36483.1
8.	*Solanum lycopersicum*	gi|AAG01894.2
9.	*Arabidopsis thaliana*	sp|Q9T074

Table 2. Physiochemical characteristics computed using Expasy's ProtParam tool

Seq. ID	Seq. Len.	MW	pI	(+)R	(-)R	EC	II	AI	Gravy
XP_003612927.1	658	72475.0	6.73	70	73	77280	35.20	75.91	-0.395
1II2	524	58733.3	8.63	68	63	64330	36.17	78.38	-0.280

Table 3. Multiple sequence alignment-scoring table complied through ClustralW2.

	M.truncatula	*P.trichocarpa*	*C.sativus*	*M.notabilis*	*L.alabamica*	*N.tabacum*	*Z.mays*	*S.lycopersicum*	*A.thaliana*
M.truncatula	-	-	-	-	-	-	-	-	-
P.trichocarpa	83.13	-	-	-	-	-	-	-	-
C.sativus	78.88	77.86	-	-	-	-	-	-	-
M.notabilis	81.46	81.48	80.51	-	-	-	-	-	-
L.alabamica	82.52	81.93	76.27	82.01	-	-	-	-	-

N.tabacum	81.16	82.02	81.87	82.78	82.48	-	-	-	-
Z.mays	79.18	77.86	75.08	79.28	78.53	79.76	-	-	-
S.lycopersicum	82.37	82.18	81.42	83.69	82.18	93.5	80.36	-	-
A.thaliana	83.28	82.38	76.27	81.86	97.02	82.48	78.23	81.87	-

Table 4. Percentage of amino acid sequence forming secondary structure using SOPMA prediction server

S.No.	Name of Enzyme	Accession No.	α helix %	β-turns %	Extended strands %	Random coil %
1	PEPCK	XP_003612927.1	30.55	7.90	22.49	39.06

Table 5. Ramachandran plot statistics with Verify_3D and ProSA

Protein	Most favoured regions	Add. Allowed regions	Gene. Allowed regions	Disallowed regions	Gly residues	Pro residues	Verify_3D	ProSA Z-score
XP_003612927.1	479 (84 %)	73 (13.2 %)	5 (0.9 %)	11 (1.90 %)	54	32	68.24 %	-5.85
1II2	813 (89.7 %)	90 (9.9 %)	3 (0.3 %)	0 (0.0 %)	75	52	-	-

ACKNOWLEDGEMENT

We are grateful to Director, FRI University and Uttaranchal College of Science and Technology, Dehradun, INDIA for providing the necessary facilities to this pursue the above research work.

REFERENCES

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E., (2000). The Protein Data Bank. Nucleic Acids Res. **28**: 235–242.

Binkowski, T.A., Naghibzadeh, S. and Liang, J., (2003). CASTp: Computed Atlas of Surface Topography of proteins. Nucleic Acids Res. **31**: 3352–3355.

Borland, A.M., Técsi, L.I., Leegood, R.C. and Walker, R.P., (1998). Inducibility of crassulacean acid metabolism (CAM) in Clusia species; physiological/biochemical characterisation and intercellular localization of carboxylation and decarboxylation processes in three species which exhibit different degrees of CAM. Planta **205**: 342–351.

Burnell, J.N. and Hatch, M.D., (1988). Photosynthesis in phosphoenolpyruvate carboxykinase-type C4 plants: photosynthetic activities of isolated bundle sheath cells from Urochloa panicoides. Arch. Biochem. Biophys. **260**, 177–186.

Carnal, N., Agostino, A. and Hatch, M., (1993). Photosynthesis in Phosphoenolpyruvate carboxykinase-type C4 plants: Mechanism and regulation of C4 acid decarboxylation in bundle sheath cells. Arch. Biochem. Biophys. **306**: 360–367.

Chen, Z.H., Walker, R.P., Técsi, L.I., Lea, P.J. and Leegood, R.C., (2004). Phosphoenolpyruvate carboxykinase in cucumber plants is increased both by ammonium and by acidification, and is present in the phloem. Planta **219**, 48–58.

De Beer, T.A.P., Berka, K., Thornton, J.M. and Laskowski, R.A., (2014). PDBsum additions. Nucleic Acids Res. **42**.

DeLano, W. (2002). Pymol Molecular Graphics System: An open-source molecular graphics tool. CCP4 Newsl. Protein Crystallogr.

Delgado-Alvarado, A., Walker, R.P. and Leegood, R.C., (2007). Phosphoenolpyruvate carboxykinase in developing pea seeds is associated with tissues involved in solute transport and is nitrogen-responsive. Plant, Cell Environ. **30**: 225–235.

Dittrich, P., Campbell, W.H. and Black, C.C. (1973). Phosphoenolpyruvate carboxykinase in plants exhibiting crassulacean Acid metabolism. Plant Physiol. **52**: 357–361.

Eisenberg, D., Luthy, R. and Bowie, J.U., (1997). VERIFY3D: Assessment of protein models with three-dimensional profiles. Methods Enzymol. **277**: 396–406.

Geourjon, C. and Deléage, G., (1995). SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput. Appl. Biosci. **11**: 681–684.

Hartmann, A.K. (2009). Introduction to Randomness and Statistics, Arxiv preprint arXiv09104545.

Hess, B., Bekker, H., Berendsen, H.J.C. and Fraaije, J.G.E.M., (1997). LINCS: A linear constraint solver for molecular simulations. J. Comput. Chem. **18**: 1463–1472.

John, B. and Sali, A. (2003). Comparative protein structure modeling by iterative alignment, model building and model assessment. Nucleic Acids Res. **31**: 3982–3992.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S. and Madden, T.L. (2008). NCBI BLAST: a better web interface. Nucleic Acids Res. **36**: W5–W9.

Kim, D.J. and Smith, S.M. (1994). Molecular cloning of cucumber phosphoenolpyruvate carboxykinase and developmental regulation of gene expression. Plant Mol Biol **26**: 423–434.

Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr.

Leegood, R.C. and Ap Rees, T. (1978). Phosphoenolpyruvate carboxykinase and gluconeogenesis in cotyledons of Cucurbita pepo. BBA - Enzymol. **524**: 207–218.

Leegood, R.C. and Walker, R.P. (2003). Regulation and roles of phosphoenolpyruvate carboxykinase in plants. Arch. Biochem. Biophys.

Malone, S., Chen, Z.-H., Bahrami, A.R., Walker, R.P., Gray, J.E. and Leegood, R.C. (2007). Phosphoenolpyruvate carboxykinase in Arabidopsis: changes in gene expression, protein and activity during vegetative and reproductive development. Plant Cell Physiol. **48**: 441–450.

Miller, C.J. and Attwood, T.K. (2003). Bioinformatics goes back to the future. Nat. Rev. Mol. Cell Biol. **4**: 157–62.

Miyamoto, S. and Kollman, P.A. (1992). Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. J. Comput. Chem. **13**: 952–962.

Pearson, W.R. (2013). An introduction to sequence similarity ("homology") searching. Curr. Protoc. Bioinforma.

Pellegrini, M. (2001). Computational methods for protein function analysis. Curr. Opin. Chem. Biol.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. J. Comput. Chem. **25**: 1605–12.

Rylott, E.L., Gilday, A.D. and Graham, I.A. (2003). The gluconeogenic enzyme phosphoenolpyruvate carboxykinase in Arabidopsis is essential for seedling establishment. Plant Physiol. **131**: 1834–1842.

Sumathi, K., Ananthalakshmi, P., Roshan, M.N.A.M. and Sekar, K. (2006). 3dSS: 3D structural superposition. Nucleic Acids Res. **34**.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol.

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E. and Berendsen, H.J.C. (2005). GROMACS: fast, flexible, and free. J. Comput. Chem. **26**: 1701–1718.

Walker, R.P. and Chen, Z.H. (2002). Phosphoenolpyruvate carboxykinase: Structure, function and regulation, in: Advances in Botanical Research. Academic Press, pp. 93–189.

Walker, R.P., Chen, Z.-H., Acheson, R.M. and Leegood, R.C. (2002). Effects of phosphorylation on phosphoenolpyruvate carboxykinase from the C4 plant Guinea grass. Plant Physiol. **128**: 165–72.

Wheeler, D.L., Church, D.M., Federhen, S., Lash, A.E., Madden, T.L., Pontius, J.U., Schuler, G.D., Schriml, L.M., Sequeira, E., Tatusova, T.A. and Wagner, L. (2003). Database resources of the national center for biotechnology. Nucleic Acids Res. doi:10.1093/nar/gkg033.