J Mol Model (2009) 15:203–221
DOI 10.1007/s00894-008-0395-8

ORIGINAL PAPER

Phylogenetic analysis, homology modelling, molecular dynamics and docking studies of caffeoyl–CoA-*O*- methyl transferase (CCoAOMT 1 and 2) isoforms isolated from subabul (*Leucaena leucocephala*)

Nataraj Sekhar Pagadala · Manish Arha · P. S. Reddy · Ranadheer Kumar · V. L. Sirisha · S. Prashant · K. Janardhan Reddy · Bashir Khan · S. K. Rawal · P. B. Kavi Kishor

Received: 7 August 2008 / Accepted: 3 October 2008 / Published online: 2 December 2008

Abstract Caffeoyl coenzyme A *O*-methyltransferase (CCoAOMT) is an important enzyme that participates in lignin biosynthesis especially in the formation of cell wall ferulic esters of plants. It plays a pivotal role in the methylation of the 3-hydroxyl group of caffeoyl CoA. Two cDNA clones that code CCoAOMT were isolated earlier from subabul and in the present study; 3D models of CCoAOMT1 and CCoAOMT2 enzymes were built using the MODELLER7v7 software to find out the substrate binding sites. These two proteins differed only in two amino acids and may have little or no functional redundancy. Refined models of the proteins were obtained after energy minimization and molecular dynamics in a solvated water layer. The models were further assessed by PROCHECK, WHATCHECK, Verify_3D and ERRAT programs and the results indicated that these models are reliable for further active site and docking analysis. The refined models showed that the two proteins have 9 and 10 α-helices, 6 and 7 β-sheets respectively. The models were used for docking the substrates CoA, SAM, SAH, caffeoyl CoA, feruloyl CoA, 5-hydroxy feruloyl CoA and sinapyl CoA which showed that CoA and caffeoyl CoA are binding with high affinity with the enzymes in the presence and absence of SAM. It appears therefore that caffeoyl CoA is the substrate for both the isoenzymes. The results also indicated that CoA and caffeoyl CoA are binding with higher affinity to CCoAOMT2 than CCoAOMT1. Therefore, CCoAOMT2 conformation is thought to be the active form that exists in subabul. Docking studies indicated that conserved active site residues Met58, Thr60, Val63, Glu82, Gly84, Ser90, Asp160, Asp162, Thr169, Asn191 and Arg203 in CCoAOMT1 and CCoAOMT2 enzymes create the positive charge to balance the negatively charged caffeoyl CoA and play an important role in maintaining a functional conformation and are directly involved in donor-substrate binding.

Keywords Caffeoyl–CoA 3-*O*-methyl transferase · Docking · Modelling · *S*-adenosyl homocysteine

N. Sekhar Pagadala · P. S. Reddy · R. Kumar · V. L. Sirisha · S. Prashant · P. B. Kavi Kishor (✉)
Department of Genetics, Osmania University,
Hyderabad 500 007, India
e-mail: pbkavi@yahoo.com

M. Arha · B. Khan · S. K. Rawal
Plant Tissue Culture Division, National Chemical Laboratory,
Pune 411 008, India

K. Janardhan Reddy
Department of Botany, Osmania University,
Hyderabad 500 007, India

Introduction

Subabul (*Leucaena leucocephala* L.) is a leguminous tree species used mainly for pulpwood. The deposition of a heterogeneous plant polymer lignin in specialized cell walls, has allowed successful land colonization by tracheophytes. To endure the negative pressure generated from transpiration, lignin provides the mechanical strength to the walls of tracheary elements (TEs). Lignin also renders the walls of maturing tracheary elements indigestible by hydrolytic enzymes released during autolysis of xylo-

genesis. Further, deposition of lignin in the walls of sclerenchyma cells adds physical toughness and chemical durability to the wall. It may also deter feeding by herbivores. Lignin is often deposited at the sites of wounding or pathogen invasion, which may provide a physical barrier for protection of adjacent tissues from further damage. Lignin contributes up to 15% to 35% of the dry weight of wood [1–2], and is considered to be dehydrogenatively polymerized from the monolignols like p-coumaryl, coniferyl and sinapyl alcohols. It is known that these monolignols are synthesized through the phenylpropanoid metabolism. These monolignols differ structurally by the methoxyl group at the 3C and 5C positions of the aromatic ring. Therefore, the enzymatic steps involved in the methoxylation of hydroxycinnamic acids are highly critical in the synthesis of different monolignols, thus influencing lignin composition. Also, the roles of methoxylation in determining lignin composition were amply demonstrated in transgenic plants with alterations in the expression of genes involved in methoxylation [3–5]. Since the first elucidation of the phenylpropanoid biosynthetic pathway, the methylation step is thought to be carried out by caffeic acid O-methyl-transferase (COMT) using free acid forms of hydroxycinnamates as substrates. As suggested by Neish (1968), methylation might also occur on the ester forms of hydroxycinnamic acids. The first evidence for a possible involvement of caffeoyl-CoA 3-*O*-methyl transferase (CCoAOMT) in lignin biosynthesis came from the study of xylogenesis in the *Zinnia* system [6]. It was found that the activity of an *O*-methyl transferase that uses both caffeoyl CoA and 5-hydroxyferuloyl CoA as substrates increased concomitantly with the timing of lignification during *in vitro* differentiation of tracheary elements. The expression of CCoAOMT gene was shown to be induced during lignification in both *in vitro* tracheary elements and lignifying tissues of *Zinnia* stems. Besides *Zinnia* and parsley [6–7], in a number of other plants such as forsythia, tobacco, tomato, alfalfa, soybean, and pine [8–12], the association of CCoAOMT was shown with lignification. In aspen, CCoAOMT activity was shown to be seasonally regulated during wood formation [13]. Thus, enough evidence exists for the role of CCoAOMT during lignification.

Like other enzymes involved in lignin biosynthesis, CCoAOMT is also thought to be a gene family in many plants [14]. However, it is difficult to determine the specific roles of each isoenzyme in plants because many potential substrates and multitude of isozymes exist. CCoAOMT has not yet been purified and the x-ray crystallographic structure does not exist in literature. This being an important enzyme in lignin biosynthesis, homology modelling in combination with molecular dynamics simulations and docking studies provide a powerful approach in understanding the structure-function relationships of isoforms of this enzyme. In order to understand the structural basis of the high degree of specificity of CCoAOMT isoforms for hydroxycinnamoyl-CoA esters and the structure and function of each isoenzyme in *Leucaena leucocephala*, three-dimensional models of CCoAOMT isoforms were constructed using homology-modelling methods. Docking studies were carried out with the substrate *S*-adenosyl homocystiene to find out the accurate conformation and orientation of protein with the substrate.

Materials and methods

Sequence analysis

Protein sequences of the CCoAOMT1 and CCoAOMT2 from subabul (*Leucaena leucocephala*) were aligned with the related family of gene sequences using clustalX [15] software. Percentage of identity and similarity of the query with the family of sequences was analyzed using GENEDOC software [16]. The phylogenetic tree for these sequences was produced by TREEVIEW software [17] along with the bootstrap values predicted using NJPLOT software [18].

Computational methods for building three dimensional structure

The 3D models of caffeoyl-CoA 3-*O*-methyl transferase (CCoAOMT1 and CCoAOMT2) were built by homology modelling based on high-resolution crystal structures of homologous proteins. A basic local alignment search tool (BLAST) search for the sequence similarities with several members of the CCoAOMT family was used for selecting the 3D models of the closest homologues available in the Brookhaven Protein Data Bank (PDB). The gene sequences of CCoAOMT1 (DQ431233) and CCoAOMT2 (DQ431234) were obtained from National Center for Biotechnology Information (NCBI) GenBank database. The BLAST search [19] resulted in three-reference proteins, including crystal structure of *Medicago sativa* feruoyl and caffeoyl coenzyme A 3-*O*- methyltransferase [20], crystal structure of human catechol-*O*-methyltransferase domain containing 1 in complex with *S*-adenosyl-l-methionine and crystal structure of putative *O*-methyltransferase from *Bacillus halodurans*. The crystal structure of caffeoyl CoA-*O*-methyltransferase from *Medicago sativa* was obtained earlier by Ferrer et al. [20] and provided a new understanding of the substrate preferences. The above three proteins exhibited a high level of sequence identity with CCoAOMT. The coordinates of crystal structure of alfalfa feruoyl coenzyme A 3-O-methyltransferase and

alfalfa caffeoyl coenzyme A 3-O-methyltransferase were used as templates to build the initial models of CCoAOMT1 and CCoAOMT2 by pair-wise sequence alignment using clustalX software [15], based on the Needleman-Wunsch algorithm [21]. The 3D models of CCoAOMT1 and CCoAOMT2 were generated by the automated homology modelling software MODELLER 7v7 (http://salilab.org) on windows operating environment [22]. This program is used for comparative protein structure modelling that optimally satisfies spatial restraints which includes (i) homology-derived restraints on the distances and dihedral angles in the target sequence extracted from its alignment with the template structures (ii) stereo chemical restraints such as bond length and bond angle preferences, obtained from the CHARMM-22 molecular mechanics force field [23] (iii) statistical preferences for dihedral angles and non-bonded interatomic distances, obtained from a representative set of known protein structures [24] and (iv) optional manually curated restraints, such as those from NMR spectroscopy, rules of secondary structure packing, cross-linking experiments, fluorescence spectroscopy, image reconstruction from electron microscopy, site-directed mutagenesis and intuition. The spatial restraints are expressed as probability density functions (pdfs) for the features restrained. The pdfs restrain C^{α}-C^{α}distances, main-chain N-O distances, main-chain and side-chain dihedral angles. The 3D model of the protein was obtained by optimization of the molecular pdf such that the model violates the input restraints as little as possible. The molecular pdf was derived as a combination of pdfs restraining individual spatial features of the whole molecule. The optimization procedure is a variable target function method that applies the conjugate gradients algorithm to positions of all non-hydrogen atoms. This model building procedure is similar to structure determined by NMR spectroscopy.

Molecular dynamics simulations

The structure with the least modeller objective function, obtained from the modeller was improved by molecular dynamics and equilibration methods using Nano Molecular Dynamics (NAMD 2.5) software [25] and Chemistry of Harvard Molecular Modelling (CHARMM27) force field for lipids and proteins [26–28] along with the TIP3P model for water [29]. The simulations began with a 100,000-step minimization of the designed side chains and solvent to remove any bad contacts. A cut off of 12 Å (switching function starting at 10 Å) for van der Waals interactions was assumed. An integration time step of 2 fs was used, permitting a multiple time-stepping algorithm [30–31] to be employed in which interactions involving covalent bonds were computed every time step. Short-range non-bonded interactions were computed every two-time step, and long-range electrostatic forces were computed every four-time steps. The pair list of the non-bonded interaction was recalculated every ten-time steps with a pair list distance of 13.5 Å. The short-range non-bonded interactions were defined as van der Waals and electrostatic interactions between particles within 12 Å. A smoothing function was employed for the van der Waals interactions at a distance of 10 Å. The protein backbone, unmutated side chains, and crystallographic water were fixed for this minimization. The backbone atoms were harmonically constrained with a restraining constant of 10.0 kcal/mol/A°, and the systems were heated to 300 K over the course of 6 ps at constant volume. The simulations were equilibrated for 500 ps with NPT ensemble (1 atm, 300 K) while the harmonic constraints were gradually turned off. With no harmonic constraints, the simulations ran for 3 ns in the NPT ensemble using Langevin dynamics at a temperature of 300 K with a damping coefficient of γ=5 ps^{-1} [32]. Pressure was maintained at 1 atm using the Langevin piston method with a piston period of 100 fs, a damping time constant of 50 fs, and a piston temperature of 300 K. Non-bonded interactions were smoothly switched off from 10 to 12 A°. The list of non-bonded interactions was truncated at 14 A°. Covalent bonds involving hydrogen were held rigid using the SHAKE algorithm, allowing a 2 fs time step. Periodic boundary conditions were included for the above studies. Atomic coordinates were saved every 1 ps for the trajectory analysis during the last 2 ns of MD simulation. CHARMM27 [33] force-field parameters were used in all simulations in this study. The molecular dynamics studies were performed with a periodic boundary conditions in the NPT ensemble at T=310 K with temperature coupling and constant pressure (P= 1atm). The SHAKE algorithm [34] was applied to fix all covalent bonds containing a hydrogen atom, a time step of 2 fs was used, and the non-bonded pair list was updated every 10 steps. The particles mesh Ewald (PME) method [35] was used to treat long-range electrostatic interactions. A residue-based cutoff of 10 Å was applied to the non-covalent interactions. During the molecular dynamics simulation, the coordinates of the simulated protein were saved every 1 ps. The structure having nearer to average energy was taken and again minimization and equilibration were performed using the same calculations given above. Finally, the graph was drawn by taking root mean square deviation (RMSD) of structures generated during minimization and equilibration methods on X-axis with time in ps on Y-axis. Structure with least RMSD difference between the structures generated was used for further studies. All hydrogen atoms were included during the calculation. In this step, the quality of the initial model was improved.

Validation of CCoAOMT1 and CCoAOMT2

The least energy structure obtained from the homology modelling was solvated with solvent water molecules and was energy-minimized to make it suitable for performing molecular dynamics (MD) simulation to relax the loops and side chains (see below for the details of the MD simulation procedure). The simulated 3D model was evaluated for its stereochemical quality by Ramachandran's map using PROCHECK [36] and environment profile using ERRAT graph (Structure Evaluation Server) [37]. The residue packing and atomic contact analysis was performed by using the Whatif program [38] to identify bad packing of side chain atoms or unusual residue contacts. The software WHATCHECK [39] was used to obtain the Z-score of Ramachandran's plot. Secondary structures of proteins were analyzed and compared by the Swiss Protein Databank Viewer (SPDBV) software [40]. The protein models were used for the identification of active site and for docking the substrate with the enzyme.

Active site identification

The binding pockets of CCoAOMT1 and CCoAOMT2 from subabul (*Leucaena leucocephala*) were identified using cavity module of SPDBV software suite and also based on structure-structure comparison. Computed Atlas of Surface Topography of Proteins (CASTP) program [41] was used for identifying and characterizing protein active sites, binding sites, and functional residues located on protein surfaces and voids buried in the interior of proteins by measuring concave surface regions on three-dimensional structures of proteins. It also measures the area and volume of pocket or void by solvent accessible surface model (Richards' surface) and by molecular surface model (Connolly's surface). It can also be used to study surface features and functional regions of proteins.

Docking studies of CCoAOMT1 and CCoAOMT2

The substrates, including all hydrogen atoms, were built and optimized with Chemsketch software suite. Extremely Fast Rigid Exhaustive Docking (FRED) version 2.1 was used for docking studies (Open Eye Scientific Software, Santa Fe, NM). This program generates an ensemble of different rigid body orientations (poses) for each compound conformer within the binding pocket and then passes each molecule against a negative image of the binding site. Poses clashing with this 'bump map' are eliminated. Poses surviving the bump test are then scored and ranked with a Gaussian shape function. We defined the binding pocket using the ligand-free protein structure and a box enclosing the binding site. This box was defined by extending the size of a ligand by 4 Å (add box parameter of FRED). This dimension was considered here appropriate to allow, for instance, compounds larger than the co-crystallized ones to fit into the binding site. One unique pose for each of the best-scored compounds was saved for the subsequent steps. The compounds used for docking were converted in 3D with OMEGA (Open Eye Scientific Software, Santa Fe, NM). To this set, the coenzyme (generation of multiconformer with Omega) corresponding to the modelled protein was added. It is an implementation of multiconformer docking, meaning that a conformational search of the ligand is first carried out, and all relevant low-energy conformations are then rigidly placed in the binding site. This two-step process allows only the remaining six rotational and translational degrees of freedom for the rigid conformer to be considered. The FRED process uses a series of shape-based filters and the default scoring function is based on Gaussian shape fitting [42].

Results

Sequence analysis

cDNA clones of *CCoAOMT1* and *CCoAOMT2* were isolated from subabul (GenBank Accession numbers DQ431233 and DQ431234 respectively) by Rawal and his co-workers at the National Chemical Laboratory, Pune, earlier. In the present study, the phylogenetics and 3D structures of these two enzymes are modelled. The BLAST search against the deduced amino acid sequences of CCoAOMT1 and CCoAOMT2 from subabul resulted in the identification of 30 sequences from different species as shown in Fig. 1. Sequence analysis using GENEDOC software showed that most of the negatively charged amino acids (glutamic acid and aspartic acid) are highly conserved within this family of sequences. Amino acid residues DNTL appeared as highly conserved in the CCoAOMT family of proteins. Percentage of identity and similarity of the query with these sequences showed that CCoAOMT1 and CCoAOMT2 are closely related to *Medicago sativa* sequences with the percentage identity of 91% and 83% and similarity of 96% and 90% respectively. CCoAOMT1 and CCoAOMT2 showed 85% and 78% identity with AtCAMT4, 83% and 75% with PcCAMT, 83% and 75% with McCAMT, 79% and 72% with PtaCAMT_PINTA, 85% and 78% with ZeCAMT, 75% and 68% with ZmCAMT1, 73% and 67% with ZmCAMT2, 55% and 50% with AtCAMT3, 52% and 47% with S1CAMT, 49% and 45% with AtCAMT1, 47% and 43% with AtCAMT2, 50% and 45% with PKCAMT, 89% and 79% with

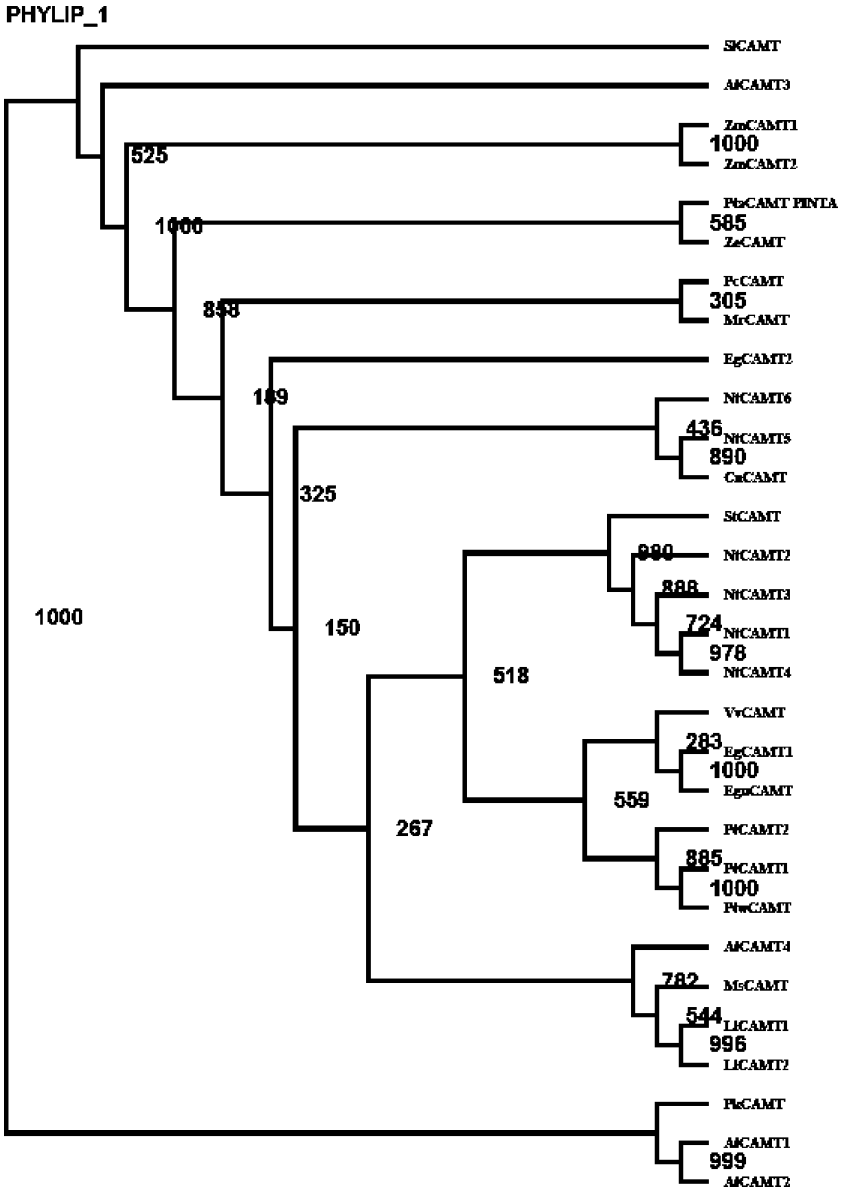


Fig. 1 Phylogenetic tree of CCoAOMT gene family (11 subfamilies). Programs used were clustalX for alignments, and graphical output was produced by TREEVIEW. Values indicate the number of times of 1,000 bootstraps that each branch topology was found during bootstrap analysis

NtCAMT5, 82% and 74% with CnCAMT, 87% and 80% with EgCAMT2, 87% and 80 NtCAMT1, 87% and 79% with NtCAMT4, 87% and 79% with NtCAMT2, 87% and 80% with NtCAMT3, 88% and 80% with StCAMT, 89% and 82% with VvCAMt, 89% and 81% with PtCAMT1, 89% and 81% with PtwCAMT, 89% and 82% with PtCAMT2 and 89% and 82% with NtCAMT6. Phylogenetic analysis of the CCoAOMT family revealed 12 subfamilies. Subfamilies 2, 3, 4, and 5 fell under the major subfamily with 15 sequences, where CCoAOMT1 and CCoAOMT2 are closely related to each other falling under subfamily 5. Minor families 5, 6, 7, 8, 9, 10 and 11 fell under one subfamily where 6, 10 and 11 are too divergent showing separate branches in the phylogenetic tree as shown in Fig. 1.

Homology modelling of CCoAOMT1 and CCoAOMT2 enzymes

Two cDNA clones of caffeoyl coenzyme A 3-*O*-methyltransferases were earlier isolated (GenBank Accession numbers DQ431233 and DQ431234) from subabul. The two clones (each 244 amino acids long) differed from each other by two amino acids (isoleucine in place of lysine and lysine in place of arginine (K 46 I and K 133 R). Usually, a high level of sequence identity should guarantee more accurate alignment between the target sequence and template structure. The BLAST search resulted in the identification of the crystal structures of *Medicago sativa* feruoyl coenzyme A 3-*O*-methyltransferase, human catechol-*O*-methyltransferase and the putative *O*-methyltransferase

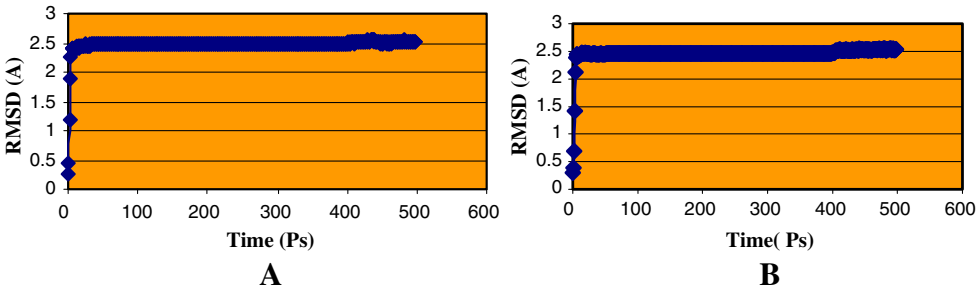


Fig. 2 **a** and **b** Calculated RMSD graphs of molecular dynamics simulations of CCoAOMT1 (**a**) and CCoAOMT2 (**b**) using NAMD software. Time (ps) is taken on X-axis and RMSD on Y-axis

from *Bacillus halodurans*. These three proteins exhibited a high level of sequence identity with CCoAOMT. The identity of these three reference proteins with CCoAOMT1 and CCoAOMT2 enzymes were found as 92%, 36% and 29% respectively. In the following step, 1SUS and 1SUI were chosen as reference structures for modelling CCoAOMT1 and CCoAOMT2. Coordinates from the reference protein (1SUS and 1SUI) to the Structurally Conserved Regions (SCRs), structurally variable regions (SVRs), N-termini and C-termini were assigned to the target sequence based on the satisfaction of spatial restraints. All side chains of the model protein were set by rotamers. Of the 20 structures calculated for the same target (CCoAOMT1 and CCoAOMT2) and the template (1SUS and 1SUI), the one with the lowest value of the MODELLER objective function was selected as the best model for CCoAOMT1 and CCoAOMT2. These initial models generated were refined by molecular dynamics and the graphs are drawn by taking time in ps on X-axis and RMSD (Å) on Y-axis (Figs. 2a and b). From these graphs, it was found that the RMSD of Cα for the structures generated are highly stable upto 400 ps, then increases and become stable at 500 ps of molecular dynamics. The structures having least RMSD of Cα generated were used for further analysis. The final stable structures of CCoAOMT1 (Fig. 3a) and CCoAOMT2 (Fig. 3b) enzymes have 9 and 10 α-helices, 6 and 7 β-sheets. The β-sheets appeared in parallel joined by α-helices at both the ends in both the enzymes. In CCoAOMT1, β1 is between H3 and H4, β2 is between H4 and H5, β3 is between H6 and H7, β4 is between H7 and H8, β5 and β6 are parallel to each at the C terminus of the enzyme (Fig. 3a). In CCoAOMT2, β1 is between H4 and H5, β2 is between H5 and H6, β3 is between H7 and H8, β4 is between H8 and H9, β5 is between H9 and H10, β6 and β7 are parallel to each other at the C terminus (Fig. 3b). These beta sheets between different α-helices play an important role in conformational changes of the protein forming Rossman fold in the core regions of the protein.

Validation of CCoAOMT1 and CCoAOMT2 enzymes

The geometry of the final refined models was evaluated with Ramachandran's plot calculations computed with the PROCHECK program. This revealed that the backbone φ and ϕ dihedral angles of CCoAOMT1 and CCoAOMT2

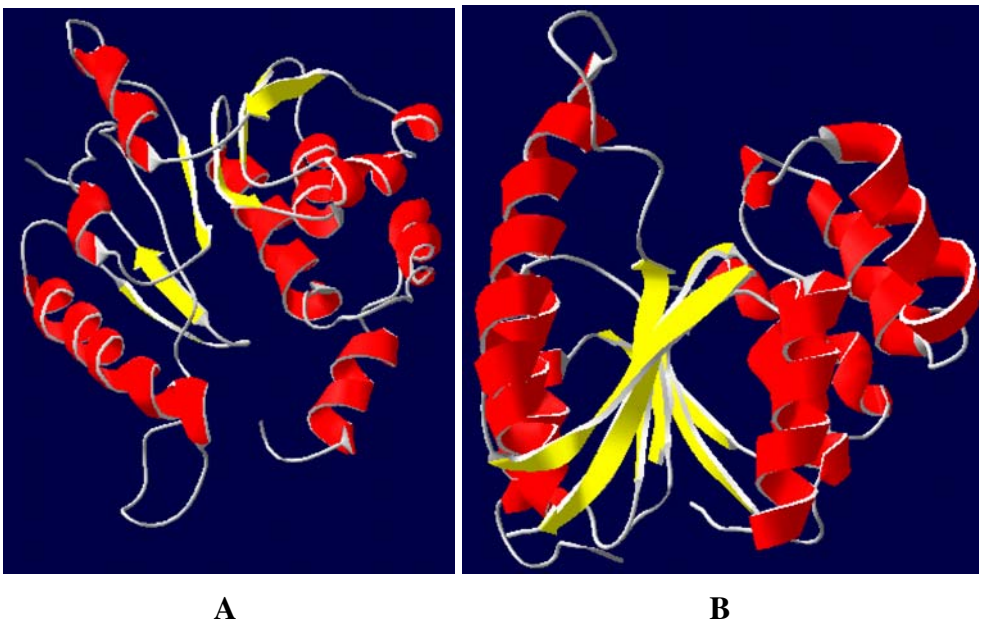


Fig. 3 **a** and **b** Final 3D structure of CCoAOMT1 (**a**) and CCoAOMT2 (**b**) enzymes. The structure is obtained by energy minimization and equilibration over the last 25,000 runs with 50 ps of molecular dynamics simulation. α-helices are represented in red and beta sheets in yellow coloured ribbons

are 84.0% and 85.9%, 13.6 and 9.9, 1.7 and 1.0 of the residues are located within the most favourable, additionally allowed, and generously allowed regions, respectively of the Ramachandran's plot as shown in Figs. 4a and b and summarized in Table 1. This good stereochemical quality is not surprising for the high sequence identity (92%) between the template and the target, which are shown in Fig. 1. The RMSD for covalent bonds relative to the standard dictionary was 0.036 Å and - 0.149 Å and for the covalent angles were 4.0 degrees and 6.2 degrees. Totally, 96.9% and 97.7% of the residues were in favoured and allowed regions. The overall PROCHECK G-factor was - 0.89 and - 0.14 and the environment profiles of VERIFY_3D above zero (Figs. 4c and d) and overall quality factors of 78.3 and 88.7 in the ERRAT graphs indicate acceptable protein environment (Figs. 4e and f). Further evaluation of the final models of CCoAOMT1 and CCoAOMT2 with Whatif program predicted the RMS Z score of backbone-backbone contacts as -1.73, -0.82, backbone-side chain contacts as -2.10, -1.65, side chain-backbone contacts as -3.06, -2.76, side chain-side chain contacts as -1.35 and -1.52. Moreover, the evaluation of the structural integrity of the final models of CCoAOMT1 and CCoAOMT2 showed a Z score of -2.37 and -1.82, which is less than the normal value of 2.0.

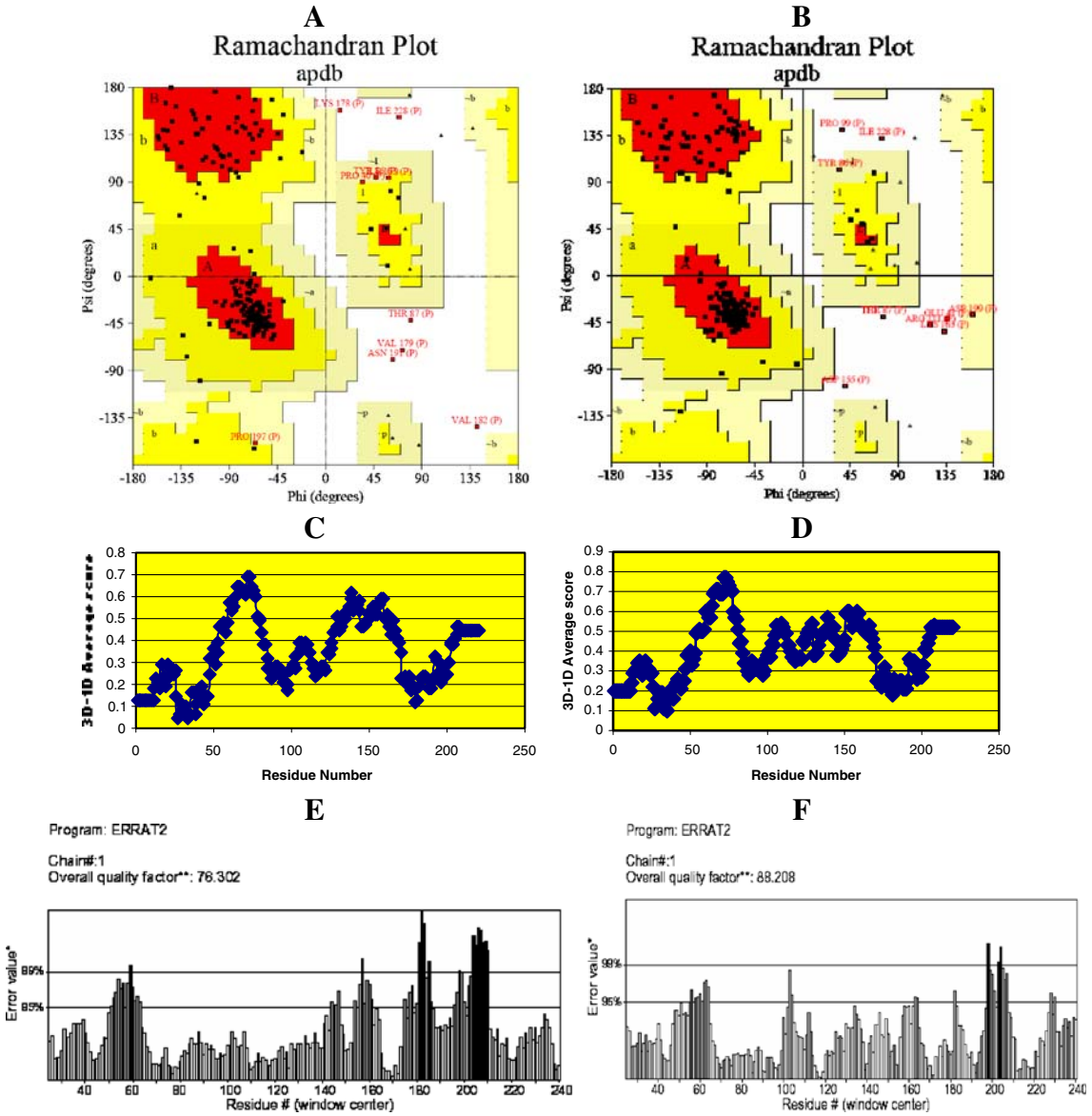


Fig. 4 **a** and **b** The Ramachandran's plot calculations of CCoAOMT1 (**a**) and CCoAOMT2 (**b**) were carried out using PROCHECK server. The figures showed that all the residues are in core regions of the Ramachandran's plot. **e** and **f** The 3D profiles of CCoAOMT1 (**c**) and CCoAOMT2 (**d**) models were verified using verify_3D server. Overall compatibility score above zero indicates residues are reasonably folded. The 3D profiles of CCoAOMT1 (**e**) and CCoAOMT2 (**f**) models were verified using ERRAT server

Table 1 % of residues falling in the core region of the Ramachandran's plot

	CCoAOMT1	CCoAOMT2
% of residue in most favoured regions	84.0	85.9
% of residue in the additionally allowed zones	13.6	9.9
% of residue in the generously allowed regions	1.7	1.0
% of residue in disallowed regions	0.7	3.1
% of non-glycine and non-proline residues	100.0	100.0

Therefore, it is believed that the final refined models are good for further analysis. This value falls in the acceptable range for a valid structure. It is generally accepted that if the Whatif score is below -5, the model is certainly of low quality and if it is above -2, it is recommended as a good structure. This trend continued even in the data obtained with the validation program WHATCHECK in which the Z-scores of bond lengths, bond angles, omega angle restraints, side-chain planarity, improper dihedral distribution inside/outside distribution for CCoAOMT1 are 1.871, 1.605, 1.366, 4.783, 2.228, 1.087 and CCoAOMT2 are 5.714, 1.756, 1.836, 3.312, 4.735, 1.062 and are positive (positive is better than average). In all likelihood, the backbone conformation (BBC) and inside/outside distribution (IOD) parameters are nearer to crystal structure values predicting that the structures are highly reliable for further studies as shown in Table 2.

Secondary structure prediction

Amino acid sequences of template, final refined models of CCoAOMT1 and CCoAOMT2 proteins were aligned using SPDBV based on the superimposition of their 3D structures. Given their PDB files, secondary structures were also analyzed and compared by the SPDBV software suite (http://www.expasy.org/spdbv). The secondary structures of template and final CCoAOMT1 and CCoAOMT2 enzymes appeared highly conserved and showed close similarity to the whole structures of template (1SUS and 1SUI) indicating that final structures are reliable (Figs. 2a and b). The aligned proteins of CCoAOMT1 and CCoAOMT2 with the templates contain a 9 and 10 stranded α- helices with 6-stranded β-sheets in CCoAOMT1 and 7 stranded β-sheets in CCoAOMT2 (Figs. 5a and b). Furthermore, in spite of several amino acid differences in the primary structures of CCoAOMT1 and CCoAOMT2, their secondary structures turned out to be identical except for one α-helix at Trp55, Asn56, and Ile57 and one β-sheet at Val182, Ile183, Gly184, Tyr185, Asp186, and Asn187 as shown in Fig. 5c. In fact, from the structure-structure comparison, it was found that α-helices 1, 2, 6 and 9 contain 3, 3, 2 and 8 residues longer than CCoAOMT1 and β-sheets (of CCoAOMT2) 2, 4, 6 and 7 contain 1, 1, 6, and 6 residues longer than of CCoAOMT1. These secondary structures of CCoAOMT1 were also compared with the templates and found that α-helices 1, 2, 5, 6 and 8 contain 1, 3, 3, 1 and 3 residues lesser than the template 1SUS and β-sheets of CCoAOMT1 1, 2, 3, 5 and 6 contains 2, 1, 2, 4 and 4 residues lesser than that of template. The comparison of secondary structures of CCoAOMT2 with the template 1SUI showed that α-helices 1 and 10 contain 1 and 1 and β-sheets of CCoAOMT2 1, 3 and 4 contain 2, 2, and 1 residue lesser than the template 1SUI. These domains exhibit a core α/β Rossmann-fold topology in which 6 and 7 parallel β-sheets are flanked on each side by α-helices that provide the binding site for SAM/SAH [43]. This core structure is highly conserved among the CCoAOMT family members, despite relatively low residue identity between these enzymes. These refined models were used further for active site and docking analysis.

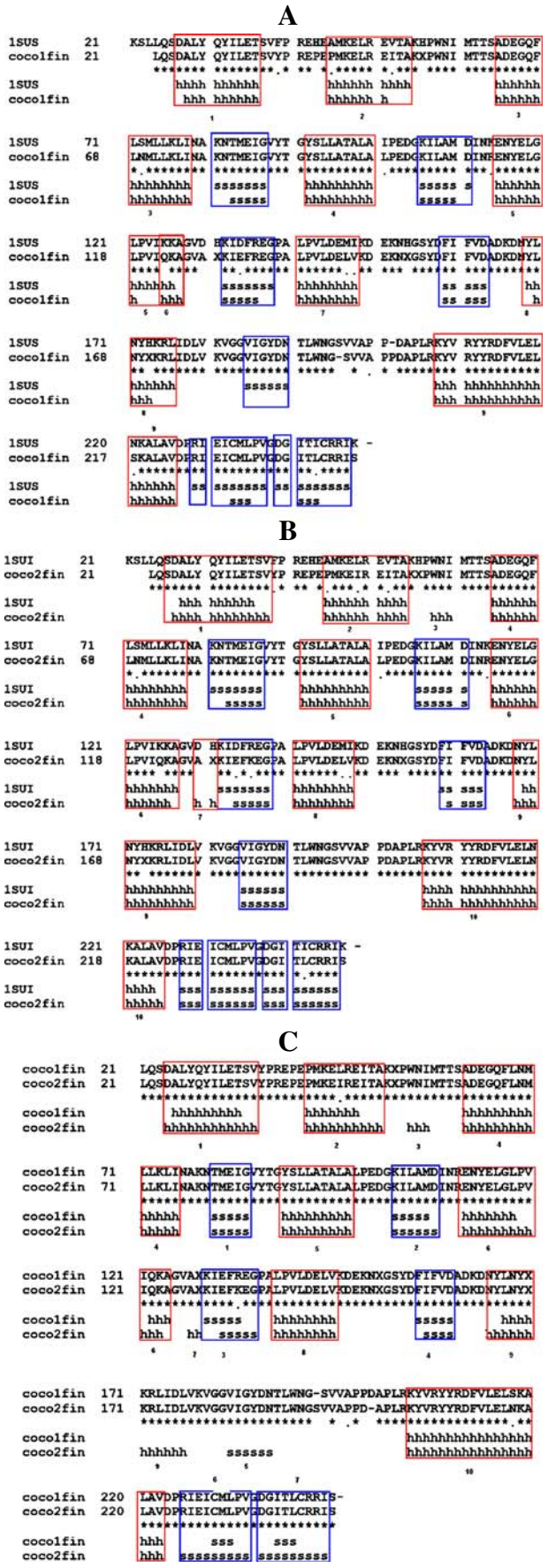
Active site identification of CCoAOMT1 and CCoAOMT2 enzymes

Once the final model was built, the possible binding sites of CCoAOMT1 and CCoAOMT2 were searched based on the

Table 2 WHATCHECK Z-scores for quality assessment and statistical analysis of CCoAOMT1 and CCoAOMT2

	Structure *Z*-score							RMS Z-scores		
Structure	PQ	RPA	χNR	BBC	BL	BA	ΩR	SCP	IDD	IOD
CCoAOMT1	-2.371	-3.390	-3.110	-10.894	1.871	1.605	1.366	4.783	2.228	1.087
CCoAOMT2	-1.816	-1.531	-3.369	-10.439	5.714	1.756	1.836	3.312	4.735	1.062

PQ, second-generation packing quality; RPA, Ramachandran plot appearance; χNR, χ^{-1}/χ^{-2} rotamer normality; BBC, backbone conformation; BL, bond lengths; BA, bond angles; Ω, omega angle restraints; SCP, side-chain planarity; IDD, improper dihedral distribution; and IOD, inside/outside distribution.



◀ **Fig. 5** **a**, **b** and **c** Secondary structure alignment of CCoAOMT1 and CCoAOMT2 with the template 1SUS and 1SUI were predicted using SPDBV software suite (**a** and **b**). **c** represents secondary structure alignment of CCoAOMT1 and CCoAOMT2

CASTP Server and structural comparison of the template and the models built. In this study, active sites were searched to identify protein active sites and binding sites by locating cavities in the CCoAOMT1 and CCoAOMT2 structures. When the search was complete, the largest site was automatically displayed on the structure as shown in Figs. 6a and b. It appeared that CCoAOMT1 and CCoAOMT2 and their templates 1SUS and 1SUI are well conserved in both sequence and structure, hence, their biological function may be identical. In fact, from the structure-structure comparison of template, and from final refined models using SPDBV program (Mate *et al* 1999), it was found that the residues in active sites, Met58, Thr60, Val63, Glu82, Gly84, Ser90, Asp160, Asp162, Thr169, Asn191 and Arg203 of CCoAOMT1 and CCoAOMT2 are highly conserved and also with the active site of templates (Figs. 7a and b). Hence, the results were used to guide the protein-ligand docking experiments.

Superimposition of 1SUS and 1SUI with CCoAOMT 1 and CCoAOMT 2 enzymes

The structural superimposition of Cα trace of template, CCoAOMT1 and CCoAOMT2 are shown in the Figs. 7a and b respectively. The weighted RMSD of Cα trace between the template and initial models of CCoAOMT1 and CCoAOMT2 generated from the MODELLER was 0.22 Å and 0.2 Å. Cα trace between the templates, 1SUS and 1SUI with the final refined models of CCoAOMT1 and CCoAOMT2 was 0.72 Å and 0.4 Å with a difference of 0.5 Å and 0.2 Å between initial and final refined models. The RMSD of Cα trace between initial and final refined models of CCoAOMT1 and CCoAOMT2 was also calculated and found as 0.69 Å and 0.34 Å. The RMSD of Cα trace of active sites between initial and final refined models of CCoAOMT1 and CCoAOMT2 with the templates 1SUS and 1SUI was calculated and found as 0.45 Å and 0.17 Å with a difference of 0.28 Å with the template. The RMSD of Cα trace of active sites of final refined models of CCoAOMT1 and CCoAOMT2 was found as 0.54 Å. Further, RMSD of Cα trace of active sites of initial and final refined models of CCoAOMT1 and CCoAOMT2 were found as 0.46 Å and 0.15 Å. These studies show that RMSD difference of 0.31 Å was responsible for conformation changes in the active sites of CCoAOMT1 and CCoAOMT2, which makes the substrate to bind in the active site of the enzymes. The two protein models built (CCoAOMT1 and CCoAOMT2) were used for docking the substrates and products.

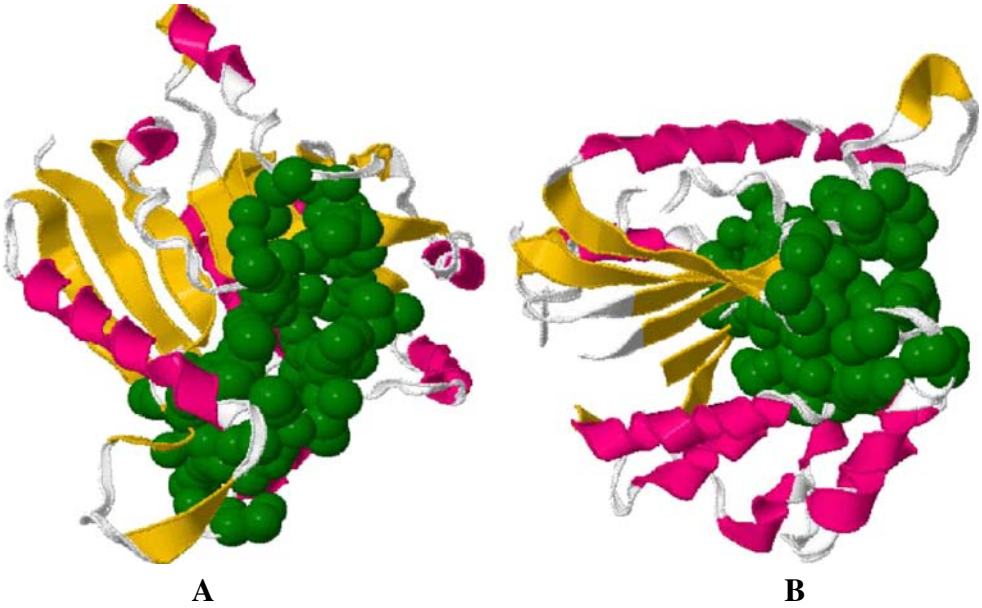


Fig. 6 **a** and **b** Active sites of CCoAOMT1 (**a**) and CCoAOMT2 (**b**) predicted using CASTp server

Docking the substrates with the active site of CCoAOMT1in the absence of SAM

Docking the substrates and products *S*-adenosyl methionine (SAM), *S*-adenosyl-L-homocysteine (SAH), caffeoyl CoA, feruoyl CoA, 5-hydroxyferuloyl CoA, sinapyl CoA with CCoAOMT1 was performed using FRED v 2.1, which is based on rigid body shape-fitting [42] (Open Eye Scientific Software, Santa Fe, NM). By automatic docking, protein-ligand conformations were analyzed for each substrate and products in the active site of CCoAOMT1 as shown in Fig. 8. The strongest binding interaction, characterized by ligand binding energies, was found in the case of caffeoyl CoA with a higher binding affinity than SAM (Fig. 8b) and sinapyl CoA (Fig. 8e) with a total score of -650.35 (chemgauss score of -78.94, chemscore of 50.92, PLP score of -39.11, screen score of -93.39 and shapeguass score of -428.07) as shown in Table 3 (the more negative the value, better the fit). From Table 6 it was found that caffeoyl CoA is binding with Lys163, Thr209 and Asp108 with three hydrogen bonding interactions (Fig. 8d). Also it was found that SAM is binding with higher affinity with the total score of -622.66 than sinapyl CoA (-491.72) with four hydrogen bonding interactions with Thr59, Ser61, Asp235 and Arg203 and sinapyl CoA is binding with four hydrogen bonding interactions with Asp108, Ile109, Thr209 and Asp164. Figure 8c shows that SAH is binding with four

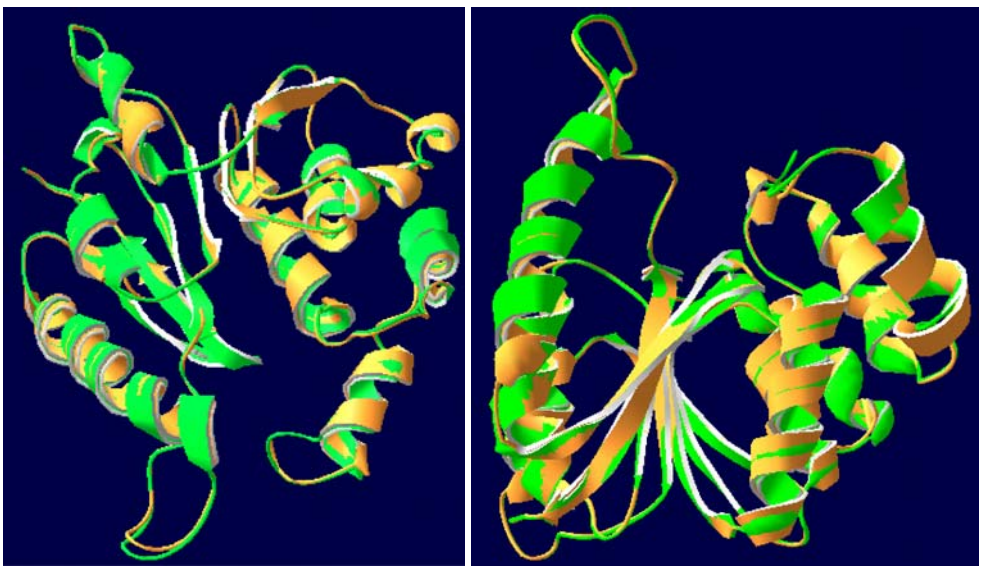


Fig. 7 **a** and **b** Superimposition of Cα trace (represented in orange) of CCoAOMT1 (**a**) and CCoAOMT2 (**b**) templates (represented in green colour)

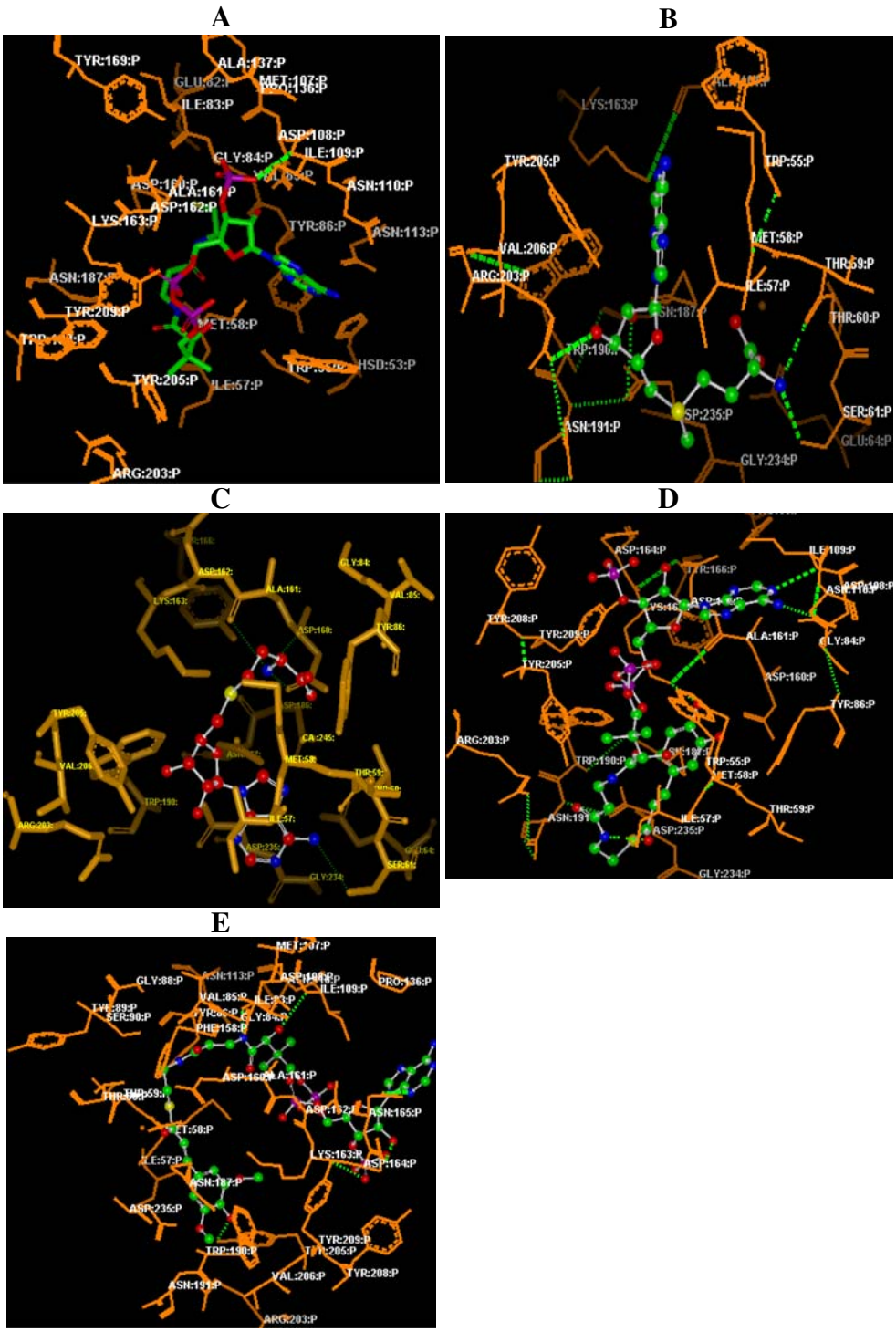


Fig. 8 Binding of CoA (**a**), SAM (**b**), SAH (**c**), caffeoyl CoA (**d**), sinapyl CoA (**e**) in the active site of CCoAOMT1 enzyme. Substrates and products are represented in ball and stick model and residues are labelled in white colour. Protein is represented in orange colour

Table 3 The total energies of Chemguass, Chem score, PLP and Shapeguass scores of the best docked conformations of substrates and products against CCoAOMT1 in the absence of SAM

Substrate	Chemgauss score	Chem score	PLP score	Screen score	Shapeguass score	Total score
CoA	-78.23	31.86	5.09	-40.42	-637.47	-719.17
SAM	-55.38	-6.71	-39.11	-93.39	-428.07	-622.66
SAH	-57.58	-3.27	-38.11	-86.86	-411.52	-597.34
Caffeoyl CoA	-78.94	50.92	31.61	17.82	-671.76	-650.35
Sinapyl CoA	-82.96	73.97	74.44	75.58	-632.75	-491.72

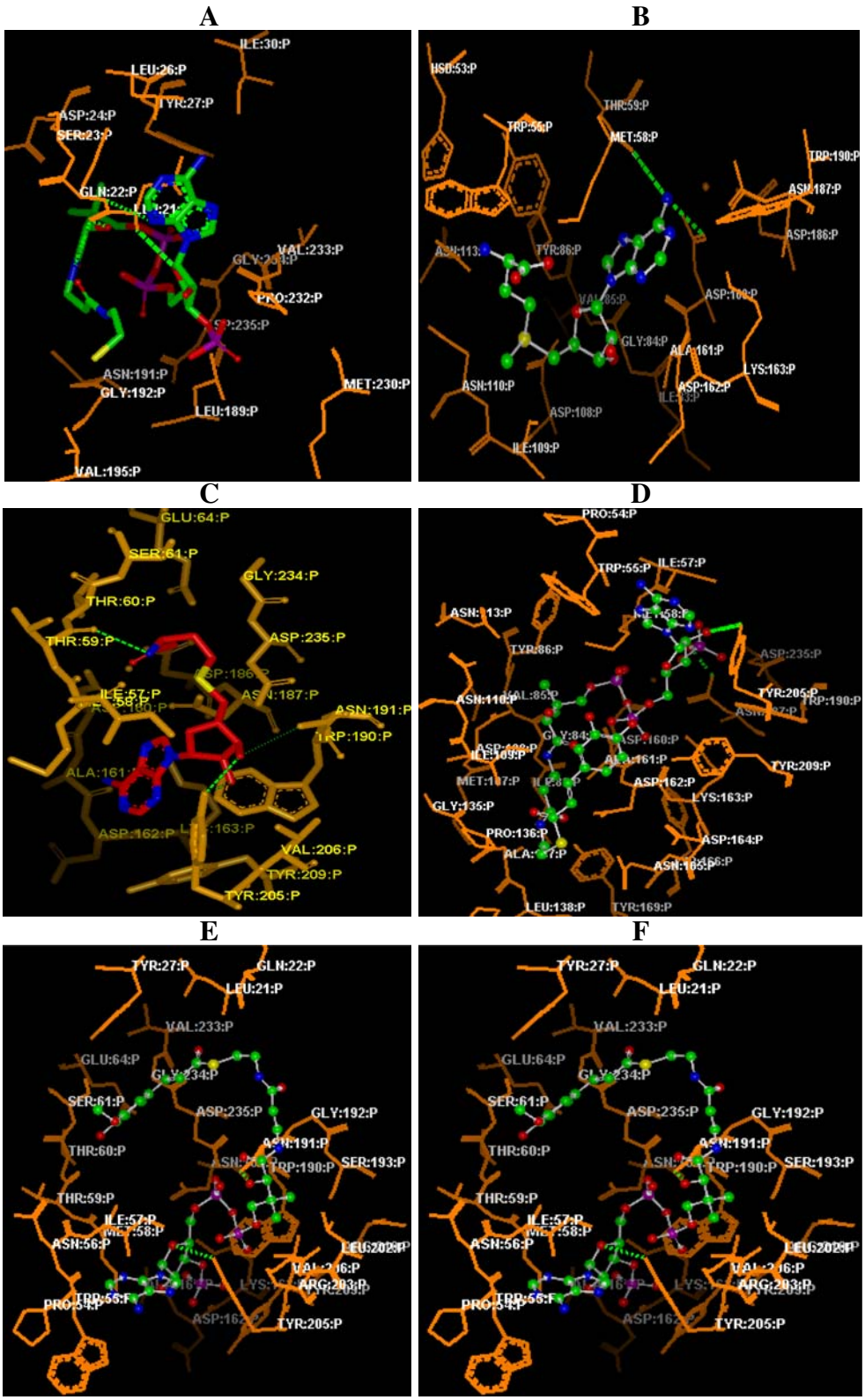


Fig. 9 Binding of CoA (**a**), SAM (**b**), SAH (**c**), caffeoyl CoA (**d**), feruoyl CoA (**e**), sinapyl CoA (**f**) and 5-hydroxyferuloyl CoA (**g**) in the active site of CCoAOMT2 enzyme in the absence of SAM. Substrates and products are represented in ball and stick model and residues are labelled in white colour. Protein is represented in orange colour

hydrogen-bonding interactions with Ser61, Asp162 and Asp160. This shows that caffeoyl CoA and sinapyl CoA are binding with the same residues in the active site of the protein. It appears that substrates and products bind with the same residues in the active site of the protein in the absence of the SAM as shown in Table 6. The binding scores of SAM, SAH and sinapyl CoA are shown in Table 3.

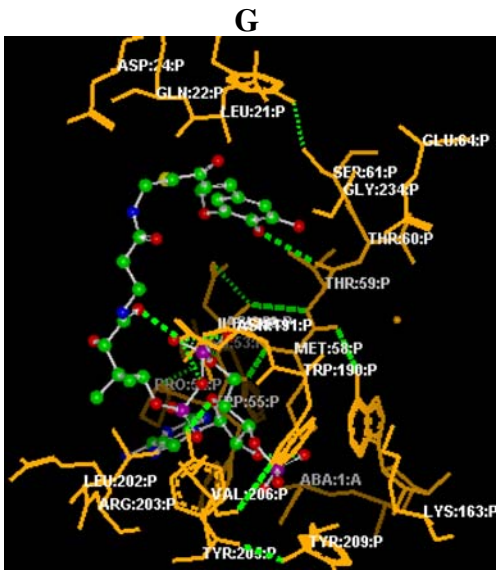


Fig. 9 (continued)

Docking the substrates with the active site of CCoAOMT2 in the absence and presence of SAM

To understand the interaction between CCoAOMT1 and CCoAOMT2 enzymes with the substrates and products, SAM (Fig. 9a), SAH (Fig. 9b), caffeoyl CoA, feruoyl CoA, 5-hydroxyferuloyl CoA, sinapyl CoA, substrate-enzyme complexes were generated using the OPENEYE software suite in the absence and presence of SAM as shown in Figs. 9 and 10. It is evident from these two figures that the SAM, caffeoyl CoA, feruoyl CoA, 5-hydroxy feruloyl CoA, sinapyl CoA are bound at the center of the active site, and are stabilized by hydrogen bonding interactions. Tables 4, 5 and 6 show the binding scores including chemgauss, chem, PLP, screen score and shapeguass for all the residues in the active site of enzyme-substrates and enzyme-products complex in the absence and presence of SAM. In the absence of SAM, enzyme-caffeoyl CoA complex has a large, favourable total binding score of -934.29 (chemguass score of -94.99, chemscore of 38.93, PLP score of -1.87, screen score of -77.00 and shapeguass score of -799.36) for CCoAOMT2. The binding scores of

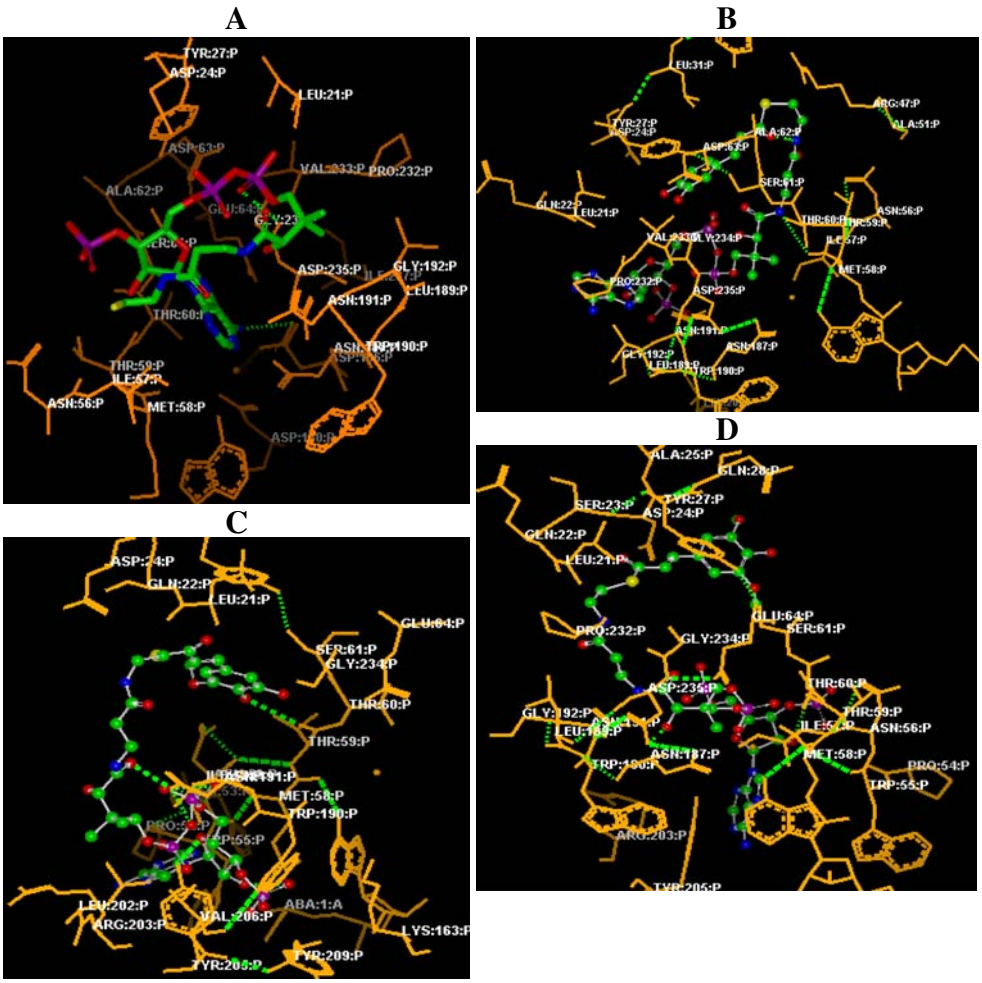


Fig. 10 Binding of CoA (**a**), caffeoyl CoA (**b**), 5-hydroxyferuloyl CoA (**c**), sinapyl CoA (**d**) in the active site of CCoAOMT2 enzyme in the presence of SAM. Substrates and products are represented in ball and stick model and residues are labelled in white colours. Protein is represented in orange colour

Table 4 The total energies of Chemguass, Chem score, PLP and Shapeguass scores of the best-docked conformations of substrates and products against CCoAOMT2 in the absence of SAM

Substrate	Chemgauss score	Chem score	PLP score	Screen score	Shapeguass score	Total score
CoA	-61.43	13.17	-21.50	-74.54	-537.27	-681.57
SAM	-56.13	-0.78	-32.24	-101.17	-414.71	-605.03
SAH	-56.05	-8.84	-41.34	-108.61	-439.21	-654.05
Caffeoyl CoA	-94.99	38.93	-1.87	-77.00	-799.36	-934.29
Ferruoyl CoA	-64.62	114.68	160.58	194.36	-615.25	-210.25
5-hydroxy feruoyl CoA	-63.71	88.28	103.71	114.18	-587.16	-344.7
Sinapyl CoA	-77.33	73.66	85.45	78.80	-637.43	-476.85

feruoyl CoA, 5-hydroxy feruloyl CoA and sinapyl CoA are shown in Table 4. Through the interaction analysis, it was found that Met58, Thr59, Asp108, Tyr205, Asn191, Asn187, Ile57, Asp162, Pro54, Thr59, Asp235, Leu21, and Ser61 of CCoAOMT2 were important anchoring residues for the substrates and products and are the main contributors to the substrate-product interactions (Fig. 9b). The hydrogen bonds present in enzyme-substrate and products complex along with their distances and angles are listed in Table 7. It is clear from Table 7 that O51, O38, O45, O40 atoms of caffeoyl CoA are binding with OH, ND2, O, and O atoms of Tyr205, Asn187, Ile57, and Asp162 (Fig. 9d). Oxygen and nitrogen atoms O41, O43, O50, N33 and N38 of feruoyl CoA are interacting with oxygen atoms of Pro54, Thr59, Asp235, Asn191 and nitrogen atoms of Leu21 (Fig. 9e). This reveals that caffeoyl CoA and feruoyl CoA are not binding with the same residues but conformational changes are occurring in the active site of the protein converting the substrate to product (Fig. 9d). The hydrogen bonding interactions also show that while atoms of 5-hydroxyferuoyl CoA, O47, O50 and N33 are binding with ND2, O and N atoms of Asn191, Pro54, and Ser61 residues (Fig. 9g), atoms of sinapyl CoA, N38, N39, O44, and O55 are binding with O, OD2, and ND2 atoms of Met58, Asp108 and Asn187 (Fig. 9f). This indicates that 5-hydroxyferuloyl CoA and sinapyl CoA are not binding with the same residues in the active site of the protein. It is evident from Figs. 9b and c that SAM is binding with Met58 and Asp160 but SAH is binding with Thr59, Tyr205 and Asn191. Thus, it appears that the active site of the protein undergoes conformational changes. In the presence of SAM, caffeoyl CoA is binding with the total score of -597.53, 5-hydroxy feruoyl CoA with -334.58 and sinapyl CoA with -332.82 as shown in Table 5 along with their individual docking scores (chemguass, chem score, PLP, screen and shapeguass scores). These substrates and products, caffeoyl CoA, 5-hydroxy feruloyl CoA and feruoyl CoA are binding with Met58, Pro54, Asn56, Ile57, Thr59, Asn191, and Thr205 as shown in Table 8. It appears from the table that atoms O45 and N37 of caffeoyl CoA are binding with O and OD1 of Met58 and Asn191 (Fig. 10b), and atoms O40, O47 and O58 of feruoyl CoA are binding with oxygen atoms of Pro54, Pro57 and OD1 of Asn57 (Fig. 10c). Atoms of 5-hydroxy feruloyl CoA O50, O51, O57, O58 and N38 are binding with oxygen, OH and OD1 groups of Thr59, Thr205 and Asn191 (Fig. 10d). These results show that caffeoyl CoA and SAM are binding with the same oxygen atom and with the same residue Met58. From the Figs. 8 and 9, it is apparent that significant key residues in the active site of the model are determined based on the interaction energies of the substrate and products with residues in the active site of the enzyme CCoAOMT2. This identification, compared with a definition based on the distance from the substrate clearly shows the relative significance for every residue. Though the interaction energy does not include the contribution from the water or the extended enzyme structure, this preliminary data along with the list of hydrogen bond interactions between the enzyme and the active site residues indicate that Met58, Pro54, Asn56, Ile57, Thr59, Asp160, Asn191, Thr205 are more preferred residues in binding of the substrate and products.

Table 5 The total energies of Chemguass, Chem score, PLP and Shapeguass scores of the best-docked conformations of substrates and products against CCoAOMT2 in the presence of SAM

Substrate	Chemgauss score	Chem score	PLP score	Screen score	Shapeguass score	Total score
CoA	-68.76	18.69	-25.07	-58.79	-528.21	-662.14
Caffeoyl CoA	-58.07	33.76	-10.04	-45.13	-518.05	-597.53
5-hydroxy feruoyl CoA	-52.19	71.11	79.51	104.04	-537.05	-334.58
Sinapyl CoA	-58.07	85.42	102.15	109.32	-571.64	-332.82

Table 6 Hydrogen bond between the substrates and products and active site residues of CCoAOMT1 as deciphered using OPENEYE software in the absence of SAM

Substrate	Hydrogen-bonding		Distance (Å)	Angle (Degrees)	Angle between the atoms
	Ligand atom	Protein atom			
CoA	CoA O35	N Ile 109	1.80	25.01	P45-O35-N
	CoA H27	O Ala161	2.76	2.76	N27-H27-O
SAM	SAM O43	NH2 Arg 203	2.71	45.0	CZ-NH2-O18
SAH	SAH N14	OG Ser 61	3.26	122.9	CA-N-O50
	SAH N	O Asp 162	3.18	116.94	CG-OD2-H52
	SAH N	O Asp 160	2.65	110.06	CG-NO2-O47
	SAH N	OD1 Asp160	2.97	98.32	CG-NO2-O47
Caffeoyl CoA	CcoA O49	NZ Lys163	2.8	54.8	CE-NZ-O49
	CcoA O5	OH Thr209	2.5	39.7	CZ-OH-O39
	CcoA N32	OD1 Asp108	3.4	55.3	N32-OD1-CG
Sinapyl CoA	ScoA N39	O Asp108	2.3	41.39	CG-OD2-N39
	ScoA O53	N Ile109	3.2	66.1	H53-O53-N
	ScoA O41	OH Thr209	2.9	60.6	CZ-OH-O41
	ScoA N52	OD2 Asp162	2.4	32.7	CG-OD2-H52

SAM—-S-adenosyl methionine
SAH—-S -adenosyl homocysteine
CcoA—Caffeoyl CoA
ScoA———-Sinapyl CoA

Docking studies of CoA

Docking studies of CoA in the absence of SAM with CCoAOMT1 show that CoA binds with two hydrogen bonding interactions Ile109 and Ala161 (Fig. 8a) with a total score of -638.8 (chemguass score of -78.23, chem 31.86, PLP 7.09, screen -40.42 and shapeguass -637.47) as shown in Table 3. The binding studies of CoA with CCoAOMT2 revealed that the residue Leu21 is involved with four hydrogen-bonding interactions (Fig. 9a) with the total score of -681.57 (-61.43, 13.17, -21.50, -71.54, -681.51 scores for chemguass, chem, PLP, screen and shapeguass respectively) as shown in Table 4. In the presence of SAM, CoA binds with one hydrogen bonding interaction with Asp186 (Fig. 10a) with the total score of -662.8 (chemguass score of -68.76, chem score of 18.69, PLP score of -25.07, screen score of -58.79 and shapeguass score of -662.14) as shown in Table 5. These studies show that CoA binds with higher affinity with CCoAOMT2 than CCoAOMT1 in the presence and absence of SAM. This reveals that CCoAOMT2 is the active form that exits in the biological systems.

In order to evaluate the validity of the modelled structure, site-directed mutagenesis of CCoAOMT1 and CCoAOMT2 was carried out with the substrate CoA. Based on the molecular structure and reaction mechanisms, metal binding amino acids and hydrophobic pocket were hypothesized to be important for the reaction. Thus, we selected for mutagenesis those amino acids that are thought to be important for the metal binding and hydrophobic pocket formation. Asp162, Asp188, and Asn189, which are predicted to be involved in metal binding, were mutagenized. Lys165 was also mutagenized as it is predicted to assist Asp122 and Asp189 in maintaining the proper configuration of these two metal binding amino acids through hydrogen bonds. Lastly, Tyr205 and Met60 were mutagenized, as both are involved in the formation of hydrophobic pocket and the former is likely to be important for the regioselectivity of CCoAOMT1 and CCoAOMT2. First, site-directed mutations were carried out in the three metal binding amino acids. Mutation of Asp160 to Gly did not change the reactivity of CCoAOMT1 binding with the total score of -712.42 but complete loss of activity was noticed in CCoAOMT2. Mutation of Asp186 to Gly resulted in an almost complete loss of activity in CCoAOMT2 but did not change the reactivity of CCoAOMT1 binding with the score of -722.78. Also, mutation of Asn187 to Gly resulted in the loss of enzymatic activity in CCoAOMT2 but no change was observed in the activity of CCoAOMT1 binding with the total score of -719.45. Lys163 is predicted to have a role in the proper positioning of the two metal binding amino acids, Asp160 and Asn187 through hydrogen bonds. Consistent with our hypothesis, mutation of Lys163 to Leu resulted in complete loss of activity in CCoAOMT2, but binding with the total scores of -747.63 in CCoAOMT1 suggests that the hydrogen bonds between Lys163 and the two metal binding amino acids are indeed critical for the enzymatic activity of CCoAOMT2 than CCoAOMT1.

Table 7 Hydrogen bond between the substrates and products and active site residues of CCoAOMT2 as deciphered using OPENEYE software in the absence of SAM

Substrate	Hydrogen-bonding		Distance (Å)	Angle (Degrees)	Angle between atoms
	Ligand atom	Protein atom			
CoA	CoA O38	N Leu21	2.13	19.89	N-H-O38
	CoA H28	O Leu21	2.29	25.76	C-O-H28
	CoA H40	O Leu21	1.28	45.16	C-O-H40
	CoA O43	N Leu21	2.11	29.69	N-H-O43
SAM	SAM N16	OD1 Asp160	3.3	55.4	CG-NO2-O47
	SAM N16	O Met58	3.2	41.1	CA-O-N33
SAH	SAH N	O THR 59	2.95	26.04	C-O-O50
	SAH O16	OH TYR 205	2.78	29.60	CZ-OH-O51
	SAH O17	OH TYR 205	3.25	95.78	CZ-OH-O51
	SAH O16	ND2 ASN 191	3.31	132.7	C-O-N38
Caffeoyl CoA	CcoA O51	OH Tyr205	2.9	23.1	CZ-OH-O51
	CcoA O38	ND2 Asn 187	2.8	37.1	CG-NO2-O38
	CcoA O45	O Ile 57	3.2	41.6	P57-O45-O
	CcoA O40	O Asp162	2.5	44.0	C-O-O40
Feruoyl CoA	FcoA N33	O Pro54	3.0	48.8	C-O-N33
	FcoA O50	O Thr59	3.3	14.5	C-O-O50
	FcoA O41	O Asp235	2.67	24.0	CG-OD2-O41
	FcoA N38	O Asn 191	2.4	36.4	C-O-N38
	FcoA O43	N Leu21	3.1	15.8	CA-N-O43
5-hydroxy feruoyl CoA	5HFcoA O47	ND2 Asn191	2.3	25.7	CG-NO2-O47
	5HFcoA N33	O Pro54	3.19	29.8	CA-O-N33
	5HFcoA O50	N Ser61	2.4	58.6	CA-N-O50
Sinapyl CoA	ScoA N38	OMet58	3.2		
	ScoA N39	OD2 Asp108	3.3	50.7	CG-OD2-O53
	ScoA O53	OD2 Asp108	2.3	11.9	OD2-N39-H39
	ScoA O44	ND2 Asn 187	3.2	44.7	CG-ND2-O44

SAM—-S-adenosyl methionine
SAH—-S-adenosyl homocysteine
CcoA—Caffeoyl CoA
FcoA—-Feruoyl CoA
5HFcoA—5-hydroxyferuoyl CoA
ScoA———-Sinapyl CoA

But mutation of Lys163 to Gly did not change the reactivity in both CCoAOMT1 and CCoAOMT2 and they are binding with total scores of -756.64, -681.57 respectively (Tables 9 and 10). Among the putative substrate binding amino acids, Tyr205 and Met58 were predicted to form a hydrophobic pocket for the substrates. Based on the result from the mutagenesis, it appears that the two residues have different roles. Met58 is located on the opposite side of Tyr205 and a mutation of met58 with a similar amino acid such as leucine may not alter the reactivity of CCoAOMT1 (binding with the total score of -636.21) but may change CCoAOMT2. Replacing Met58 with a smaller amino acid such as glycine may change the reactivity due to the size of hydrophobic pocket. In CCoAOMT1, mutation of this residue to glycine shows the CoA binding score of -725.68 whereas in CCoAOMT2, CoA binds with the total score of -707.94 (difference of 17.74). This study shows that mutation of Met58 to Gly did not significantly change enzyme reactivity, but mutation of Met58 to Leu reduced the reactivity of CCoAOMT1 and CCoAOMT2. Similarly, mutation of Tyr205 to Gly did not change the reactivity of CCoAOMT1 binding with the total score of -628.44 but changed the regioselectivity of CCoAOMT2. Based on molecular modelling and docking experiments, the reaction mechanism of CCoAOMT1 and CCoAOMT2 could be predicted. First, the flavonoid substrate goes into the substrate-binding pocket formed by hydrophobic amino acids like Met58 and Tyr205; which has a role in determining the regioselectivity of CCoAOMT1 and CCoAOMT2. Once the flavonoid is located in the pocket, the metal ion serves as a base removing a proton from the hydroxyl group of the substrates. Thus, proper configuration of the metal ion in the enzyme is indispensable for the activity. Our *in silico*

Table 8 Hydrogen bond between the substrates and products and active site residues of CCoAOMT2 in the presence of SAM as deciphered using OPENEYE software in the presence of SAM

Substrate	Hydrogen-bonding		Distance (Å)	Angle (Degrees)	Angle between atoms
	Ligand atom	Protein atom			
CoA	CoA H26	OD1 Asp186	2.34	69.77	CG-OD1-H26
Caffeoyl CoA	CcoA N37	O Met58	2.5	42.9	C-O-N37
	CcoA O45	OD1 Asn191	2.9	31.8	CG-OD1-O45
Feruoyl CoA	FcoA O47	O Pro54	3.4	18.6	C-O-O47
	FcoA O40	OD1 Asn 56	3.1	57.3	CG-H4-O40
	FcoA O58	O Ile 57	2.6	20.5	C-O-O58
5-hydroxy feruoyl CoA	5HFcoA O50	O Thr59	3.2	22.3	C-O-O50
	5HFcoA O51	O Thr59	3.2	25.1	C-O-O51
	5HFcoA O57	OH Thr205	3.3	46.9	CZ-OH-O57
	5HFcoA O58	OH Thr205	3.1	35.4	CZ-OH-O58
	5HFcoA N38	OD1 Asn191	2.8	47.1	CG-OD1-N38

SAM—-S-adenosyl methionine
SAH—-S-adenosyl homocysteine
CcoA—Caffeoyl CoA
FcoA—-Feruoyl CoA
5HFCoA—5-hydroxy feruoyl CoA

mutagenesis results clearly demonstrated the importance of these amino acids to the enzymatic activity of CCoAOMT1 and CCoAOMT2.

Discussion

In plants, CCoAOMTs are *S*-adenosyl-L-methionine-dependent *O*-methyltransferases (OMTs) involved in lignin biosynthesis. Plant CCoAOMTs belong to a distinct family of OMTs. CCoAOMT plays a predominant role in the synthesis of guaiacyl lignin and is essential in providing substrates for the synthesis of syringyl lignin. It also plays a pivotal role in the methylation of 3-hydroxyl group of caffeoyl CoA. Further, CCoAOMT-mediated methylation reaction is essential to channel substrates for 5-methoxylation of hydroxycinnamates [44]. It was found that this enzyme is expressed highly in lignifying tissues especially in xylem ray parenchyma [44]. It appears therefore that CCoAOMT-mediated lignin biosynthesis is common among plants and is also an important pathway as pointed out earlier by Ye and Varner (1995). Indeed, suppression of CCoAOMT enzyme in transgenic tobacco caused a decrease in lignin content [5]. However, it is not known whether the two genes are expressed in a tissue specific manner or not. It is also not known if they have any functional redundancy. Our work shows that two genes (isoforms) for CCoAOMT exist in subabul tree.

Amino acid sequence analysis of CCoAOMT1 and CCoAOMT2 revealed that they are closely related to the crystal structure of *Medicago sativa* feruoyl coenzyme A 3-O-methyltransferase and caffeoyl coenzyme A 3-O-methyltransferase of plant OMTs as shown in Fig. 1. Sequence analysis also reveal that CnCAMT, NtCAMT5, NtCAMT6 and EgCAMT2 belong to the subfamily 1, NtCAMT4, NtCAMT1, NtCAMT3, NtCAMT2, StCAMT

Table 9 The total energies of Chemguass, Chem score, PLP and Shapeguass scores of the best-docked conformation of CoA against CCoAOMT1 in the absence of SAM

Residue mutated	Chemguass score	Chem score	PLP score	Screen score	Shapeguass score	Total score
Met58Gly	-77.75	21.84	-12.21	-52.13	-605.43	-725.68
Met58Leu	-64.59	24.31	-11.05	-51.87	-533.01	-636.21
Asp160Gly	-76.82	32.11	5.44	-37.02	-636.13	-712.42
Lys163Gly	-83.32	32.57	-5.00	-76.03	-624.86	-756.64
Lys163Leu	-84.91	34.59	0.36	-64.23	-633.44	-747.63
Asp186Gly	-78.18	31.86	5.42	-39.83	-642.05	-722.78
Asn187Gly	-68.00	29.45	-8.95	-76.68	-595.27	-719.45
Tyr205Gly	-64.51	25.39	-4.81	-62.23	-522.28	-628.44
Asp225Gyr	-78.34	31.86	5.42	-39.83	-642.65	-723.54

Table 10 The total energies of Chemguass, Chem score, PLP and Shapeguass scores of the best-docked conformation of CoA against CCoAOMT2 in the absence of SAM

Residue mutated	Chemguass score	Chem score	PLP score	Screen score	Shapeguass score	Total score
Met58gly	-80.70	30.19	-2.33	-51.65	-603.45	-707.94
Lys163gly	-61.43	13.17	-21.50	-74.54	-537.27	-681.57
Arg225thr	-61.43	13.17	-21.50	-74.54	-537.27	-681.57

belong to subfamily 2, EguCAMT, EgCAMT1 and VvCAMT1 fall under subfamily 3, PtwCAMT, PtCAMT and PtCAMT2 fall under subfamily 4, LlCAMT1, LlCAMT2, MsCAMT and AtCAMT4 fall under subfamily 5, McCAMT, PcCAMT fall under subfamily 6, ZeCAMT, PtaCAMT_PINTA fall under subfamily 7, ZmCAMT1 and 2 fall under subfamily 8, AtCAMT3, SiCAMT fall under subfamily 9 and 10 respectively. Phylogenetic tree showed that AtCAMT1, AtCAMT2 and PkCAMT are too divergent and fall in a separate subfamily. These results revealed that CCoAOMT1 and CCoAOMT2 are closely related to the species *Medicago sativa* in the evolution. To understand the structural and functional characteristics, the 3D structures of CCoAOMT1 and CCoAOMT2 were built by homology modelling and simulations, which revealed their possible interactions and conformational changes with substrates and products. The 3D models showed specific interactions between the substrates caffeoyl CoA, 5-hydroxyferuloyl CoA and products feruoyl CoA and sinapyl CoA and key amino acid residues in the active site and thus these interactions are consistent with all of the previously reported experimental data concerning the catalytic activity of CCoAOMT1 and CCoAOMT2. Our models demonstrated the substrate-binding cavity of the proteins, which was not known previously. Further, the docking results indicated that caffeoyl CoA is binding with high affinity with these isoenzymes, in the presence and absence of *S*-adenosyl methionine. It appeared that caffeoyl CoA is binding with more affinity with CCoAOMT2 than with CCoAOMT1. Our results reveal that CCoAOMT2 is the active conformation that exists in subabul and perhaps in other plant systems too. As is known, hydrogen bonds play an important role for the structure and function of biological molecules, especially for the enzyme catalysis. The specific binding shown in the model is consistent with the previous experimental observation that Thr59, Ser61, Asp235 and Arg203, Asp108, Ile109, Thr209 and Asp164 are important determinant residues in CCoAOMT1. In the absence of SAM, Met58, Thr59, Asp108, Tyr205, Asn191, Asn187, Ile57, Asp162, Pro54, Thr59, Asp235, Leu21 and Ser61 are important for strong hydrogen bonding interactions with the substrate and products in CCoAOMT2. Docking studies of CCoAOMT2 in the presence of SAM predicted that Met58, Pro54, Asn56, Ile57, Thr59, Asn191 and Thr205 are important residues in binding the substrates and products. Our results are consistent with the experimental results obtained from previously characterized plant CCoAOMTs. It appears from the data that Met58, Pro54, Asn56, Ile57, Thr59, Asp160, Asn191, Thr205 are conserved in these two enzymes and in the templates and hence may be important for structural integrity or maintaining the hydrophobicity of the substrate-binding pocket. These residues both in CCoAOMT1 and CCoAOMT2 lie at the N terminal extension, 193 to 211 and 237 to 239 loops and help in binding the CoA linked substrates and determine the specificity of the enzymes. Since caffeoyl CoA and SAM are binding with the same residue Met58 that is very nearer to calcium ion, it is predicted that the environment around the caffeoyl CoA is positively charged due to positively charged sulphur atoms and calcium ions. These results indicate that caffeoyl CoA exists as a negative charge to balance the positive charge around the active site of the protein. Interactions of caffeoyl CoA with calcium ion position the negative charge of caffeoyl CoA to close proximity to methyl group of SAM for converting the substrate to product. Docking studies of CoA with CCoAOMT1 and CCoAOMT2 showed that CoA binds with high affinity with CCoAOMT1 than CCoAOMT2. The role of some of the residues in CCoAOMTs was earlier suggested to be important for enzyme activity [45]. These authors observed that upon mutation of Asp66 and Gln69, there was a reduction in the enzyme activity and this was attributed to the possible role of these residues in substrate binding. Our mutagenesis studies also showed that mutations of Asp160Gly, Lys163Leu, Asp186Gly, Asn187Gly and Tyr205Gly results in complete loss of enzymatic activity in CCoAOMT2 than CCoAOMT1. These studies show that even though CoA binds with high affinity with CCoAOMT1 than CCoAOMT2 both in the mutated and non-mutated forms, the substrates and products bind only with CCoAOMT2. While the mammalian catechol OMTs are monomeric, the plant CCoAOMTs appear to be dimeric in nature as pointed out by Ferrer et al. [20]. However, the overall architectural similarity and the substrate-binding sites of subabul CCoAOMTs relate more to a mammalian catechol OMT rather than plant OMTs. As pointed by Ferrer et al. [20], these two enzyme families perhaps adopt different oligomerization states. The 3D structures created here provide a new understanding of the substrate preferences and the catalytic mechanisms accompanying

CCoAOMT-mediated *O*-methylation of CoA-linked substrates because of high similarity in SAM binding region of methyl transferases despite the differences in structural similarity. The approach might be applicable for the prediction of substrate and regioselectivity of other enzymes, and aid in efforts to engineer new regioselectivities of existing enzymatic reactions. Also, the interactions between the enzyme and the substrate proposed in this study are useful for understanding the potential mechanism of enzyme-substrate binding. Thus, it is concluded that these two genes may have tissue specific expression for methylation reactions associated with lignin biosynthesis in plants.

Acknowledgements The authors are thankful to the CSIR, New Delhi, for financial assistance in the form of a research project (CSIR-NMITLI) on paper and pulp.

References

1. Higuchi T (1998) Kung S-D, Yang S-F (eds) Discoveries in plant biology. World Scientific, Singapore, 233–269
2. Grima-Pettenati J, Goffner D (1999) Plant Sci 145:51–65
3. Meyer K, Sirley AM, Cusumano JC, Bell D Lelong A, Chappel C (1998) Proc Natl Acad Sci USA 95:6619–6623
4. Tsai CJ, Mielke MR, Hu WJ, Podila GK, Chiang VL (1998) Plant Physiol 117:101–112
5. Zhong R, Morrison WH, Negrel J, Ye ZH (1998) Plant Cell 10:2033–2045
6. Ye ZH, Kneusel RE, Matern U, Varner JE (1994) Plant Cell 6:1427–1439
7. Ye ZH, Varner JE (1995) Plant Physiol 108:459–467
8. Ye ZH (1997) Plant Physiol 115:1341–1350
9. Inoue K, Vincent JH, Sewalt G, Balance MNIW, Sturzer C, Dixon RA (1998) Plant Physiol 117:761–770
10. Martz F, Maury S, Pincon G, Legrand M (1998) Plant Mol Biol 36:427–437
11. Kersey R, Inoue K, Schubert KR, Dixon RA (1999) Protoplasma 209:46–57
12. Li L, Osakabe K, Joshi CP, Chiang VL (1999) Plant Mol Biol 40:555–565
13. Meng H, Campbell WH (1998) Plant Mol Biol 38:513–520
14. Maury S, Geoffroy P, Legrand M (1999) Plant Physiol 121:215–224
15. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DJ (1997) Nucleic Acids Res 24:4876–4882
16. Nicholas KB, Nicholas HB (1997) http://www.psc.edu/biomed/genedoc [Online.]
17. Page RDM (1996) Comput Appl Biosci 12:357–358
18. Perrière G, Gouy M (1996) Biochimie 78:364–369
19. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) J Mol Biol 215:403–410
20. Ferrer J, Zubeita C, Dixon RA, Noel JP (2005) Plant Physiol 137:1009–1017
21. Needleman SB, Wunsch CD (1970) J Mol Biol 48:443–453
22. Sali A, Blundell TL (1993) J Mol Biol 234:779–815
23. MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL Jr, Evanseck J, Field M, Fischer JS, Gao J, Guo H, Ha S (1998) J Phys Chem B 102:3586–3616
24. Sali A, Overington JP (1994) Protein Sci 3:1582–1596
25. Kale L, Skeel R, Bhandarkar M, Brunner R, Gursoy A, Krawetz N, Phillips J, Shinozaki A, Varadarajan K, Schulten K (1999) J Comput Phys 151:283
26. MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL Jr, Evanseck J, Field M, Fischer JS, Gao J, Guo H, Ha S, Joseph D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Roux B, Schlenkrich M, Smith J, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M (1992) FASEB J 6:A143–A143
27. MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL Jr, Evanseck J, Field M, Fischer JS, Gao J, Guo H, Ha S, Joseph D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher IWE, Roux B, Schlenkrich M, Smith J, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M (1998a) J Phys Chem B 102:3586–3616
28. Schlenkrich M, Brickmann J, MacKerell AD Jr, Karplus M (1996) A molecular perspective from computation and experiment. In: Merz KM, Roux B (eds) Birkhauser, Boston MA, pp 31–81
29. Jorgensen WL, Chandresekhar J, Madura JD, Impey RW, Klein ML (1983) J Chem Phys 79:926–935
30. Grubmuller H, Heller H, Windemuth A, Schulten K (1991) Mol Sim 6:121–142
31. Schlick T, Skeel R, Brunger A, Kale L, Board JA Jr, Hermans J, Schulten K (1999) Algorithmic challenges in computational molecular biophysics. J Comput Phys 151:9–48
32. Brunger (1992) X-PLOR, Version 3.1: Yale University New Haven CT
33. MacKerell AD Jr, Brooks B, Brooks CL III, Nilsson L, Roux B, Won Y, Karplus M (1998b) In: Schleyer P (ed) The encyclopedia of computational chemistry. John Wiley & Sons, Chichester, UK, pp 271–277
34. Ryckaert J-P, Ciccotti G, Berendsen HJC (1977) J Comput Phys 23:327–341
35. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995) J Chem Phys 103:8577
36. Laskoswki RA, MacArthur MW, Moss DS, Thorton JM (1993) J Appl Cryst 26:283–291
37. Colovos K, Yeates TO (1993) Protein Sci 2:1511–1519
38. Vriend G (1990) J Mol Graph 8:52–56
39. Hooft RWW, Vriend G, Sander C, Abola EE (1996) Nature 381:272
40. Guex N, Peitsch MC (1997) Electrophoresis 18:2714–2723
41. Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y, Liang J (2006) Nucleic Acids Res 34:W116–W118
42. Schulz-Gasch T, Stahl M (2003) J Mol Model 9:47–57
43. Rossmann MG, Moras D, Olsen KW (1974) Nature 250:194–199
44. Zhong RW, Morrison H III, Himmelsbach DS, Poole FL II, Ye ZH (2000) Plant Physiol 124:563–577
45. Hoffmann L, Maury S, Bergdoll M, Thion L, Erard M, Legrand M (2001) J Biol Chem 276:36831–36838