Journal of Molecular Modeling (2018) 24:53
https://doi.org/10.1007/s00894-018-3590-2

ORIGINAL PAPER

Structural and functional characterization of the Vindoline biosynthesis pathway enzymes of *Catharanthus roseus*

Bilal Ahmad[1] · Anindyajit Banerjee[1] · Harshita Tiwari[1] · Shrabasti Jana[1] · Sudeshna Bose[1] · Saikat Chakrabarti[1]

Received: 18 July 2017 / Accepted: 15 January 2018

Abstract

Vinblastine and its related compound vincristine are important mono terpenoid indole alkaloids accumulated in the leaves of *Catharanthus roseus* (Madagascar periwinkle). They serve as major anticancer drugs. Vinblastine is formed by the condensation of vindoline and catharanthine. The vindoline moiety is derived from tabersonine via vindoline biosynthesis pathway. The reaction sequence from tabersonine to vindoline is now well established and the enzymes involved in this pathway are identified. However, to date, the structures of the enzymes involved in the vindoline biosynthesis pathway are not known, leading to limited mechanistic understanding of the substrate binding and catalysis. The purpose of this work is to provide structural insight regarding all the steps of the vindoline pathway via rigorous homology modeling, molecular docking, and molecular dynamics analyses. Substrate and cofactors required for each step were docked onto the computationally built and validated three-dimensional (3D) model of the corresponding enzyme, and the catalytic reaction was analyzed from the structural point of view. Possible binding modes of the substrates and cofactors were generated and corresponding binding residues were identified. Enzyme-substrate models were verified based on structure evaluation methods and molecular dynamics based approaches. Findings of our analysis would be useful in rational designing of these important enzymes aimed toward bio-production of vindoline.

Keywords Vindoline biosynthesis pathway · Enzyme structure · Molecular modeling · Molecular docking · Co-evolution

Introduction

Cancer is a large group of diseases characterized by uncontrolled cell growth leading to death. According to the World Health Organization (WHO), the number of new cases is expected to rise by about 70% over the next two decades. Cancer is a health concern, which has significantly drawn the attention of medical science in recent times. Natural products, especially of plant origin have a potential to aid the treatment of the disease, apart from the other cancer chemotherapy approaches. The indole alkaloids are the major class of alkaloids of plant origin that have shown promising anti-cancer and anti-tumor effects [1, 2].

Catharanthus roseus is a rich source of such monoterpenoid indole alkaloids (MIA). These secondary metabolites are known for their actions against cancer, but the concentration of these when extracted is not very significant for medical use. So, the knowledge about the biosynthesis of the alkaloids may be used for their artificial synthesis [3, 4]. Vinblastine and vincristine are condensed from two MIA moieties, catharanthine and vindoline. Dimericbis-indole alkaloids are formed by condensation of catharanthine and vindoline to form 3′,4′-anhydrovinblastine by the peroxidase enzyme α-3′,4′-anhydrovinblastine synthase (PRX1) [5] further leading to vinblastine and ultimately to vincristine [6]. The complicated structure of these metabolites makes the total chemical synthesis difficult and costly. Hence, in depth understanding of the biosynthesis of catharanthine and vindoline will be helpful for

Bilal Ahmad and Anindyajit Banerjee contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00894-018-3590-2) contains supplementary material, which is available to authorized users.

✉ Saikat Chakrabarti
saikat@iicb.res.in

[1] Department of Structural Biology and Bioinformatics Division, Council of Scientific and Industrial Research, CSIR-Indian Institute of Chemical Biology, Kolkata, WB, India

Published online: 13 February 2018

high yield commercial production of these crucial metabolites. The vindoline biosynthesis pathway (Fig. 1) is a six step pathway in which tabersonine is converted into the final product vindoline. The reaction sequence from tabersonine to vindoline is well established, the enzymes involved in the catalysis of the different steps have been identified and the respective amino acid sequences of all the enzymes are determined. However, the three dimensional (3D) structures of the enzymes involved in the vindoline biosynthesis pathway are not yet known. In this work, we provide structural insight for the enzymes involved in vindoline synthesis by developing a 3D model structure of each enzyme. The 3D structures of the enzymes were utilized to generate plausible enzyme-substrate complex models using molecular docking analysis. Probable modes of binding of substrates and cofactors were generated and corresponding binding residues were identified. The validity of the enzyme models and their mode of binding to the corresponding substrates were further verified based on standard structure evaluation methods as well as using molecular dynamics based semi-empirical approaches. The findings of this study could be very useful in large-scale biosynthesis of vindoline involving protein engineering and designing protocols.

Materials and methods

Sequence identification and conservation analysis of the important sites

Sequences of the enzymes were identified after analyzing all the reaction steps of the pathway. Sequences were collected from UniProt, a comprehensive resource for protein sequence and annotation data [7] and BRENDA [8] database that represents information of enzymes, enzyme ligands, and catalytic properties.

Orthologs for each enzyme were collected via BLAST [9] search against NCBI non redundant sequence database using filtration criteria of E-value ≤1e-05, sequence identity ≥50%, and subject (ortholog) protein length coverage ≥70%. Orthologs were selected only from the plant source. The ClustalOmega [10] program was used to create a multiple sequence alignment of each enzyme belonging to vindoline pathway.

The AL2CO [11] program was used to calculate the conservation index for each position for a given multiple sequence alignment. For this analysis independent count

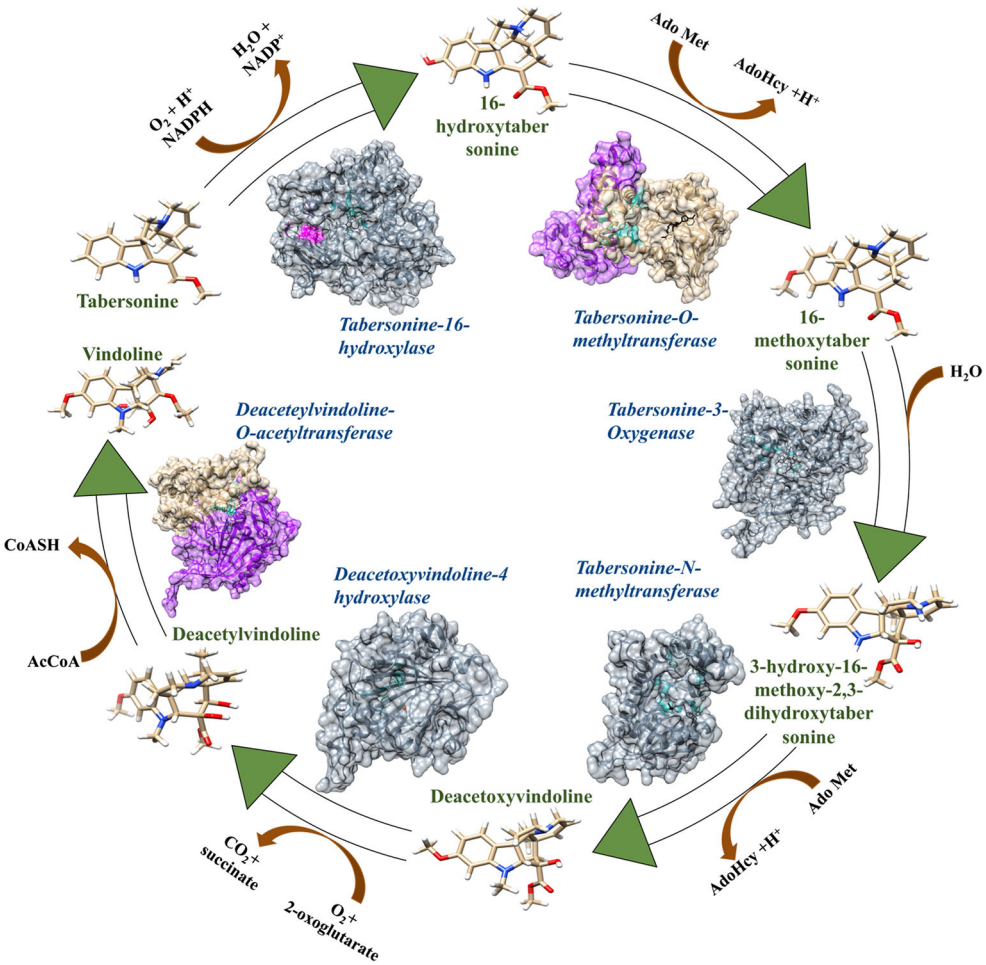

Fig. 1 Circular flowchart of vindoline biosynthesis pathway where three-dimensional (3D) model of each substrate is shown in stick model and 3D model of each enzyme is shown in surface representation

(sequence weighting scheme), matrix based sum-of-pair (conservation calculation method) measure scoring scheme was used to calculate evolutionary conservation of each site or column in the alignment. Conservation index (CI) was converted to Z-scores for comparison.

Template selection and homology modeling

Fold predictions of the enzymes were done using three web servers namely, HHpred [12], Phyre2 [13], and PSIpred [14]. Since different servers subject query to different prediction approaches, the consensus fold prediction approach was used for consistency in results. Top hits were analyzed based on sequence identity, similarity of enzyme class and the source between the query (*C. roseus* sequences), and the predicted template sequences.

Alignments between the query and the selected template sequences were generated by the HHpred [12] program followed by careful manual inspection and adjustment to further improve the quality of the alignments. MODELLER version 9.16 [15] was utilized for backbone generation, loop, side chain modeling, and cofactor addition. Initially, one hundred 3D models were generated and were ranked on the basis of energy parameters like DOPE score. Structures were validated using Ramachandran plot parameters calculated by RAMPAGE [16], fold compatibility by VERIFY_3D [17] and ERRAT [18] validation packages. Certain selected models were further refined and loop modeling was performed in MODELLER 9.16 [15] to predict most plausible 3D models with least stereo-chemical violations. The homology models were visualized and analyzed in chimera 1.11 [19]. Since, cofactors are important for enzymaticreactions, heme was added to the selected best models of tabersonine-16-hydroxylase and tabersonine-3-oxygenase and coordinated with the corresponding residues.

Similarly, SAH, Fe^{++} and acetyl-CoA coordinates were co-modeled with 3-hydroxy-16-methoxy-2,3-dihydrotabersonine-N-methyltransferase, tabersonine-4-hydroxylase, and deacetylvindoline-O-acetyltransferase, respectively using MODELLER [15] ligand modeling option. If the template PDB contains a ligand/cofactor or other non-protein residues (e.g., DNA or RNA, or anything marked as HETATM in the PDB file) then MODELLER [15] can transfer this into the respective, generated model. This is done by using the BLK ('.') symbol type in the, alignment (both in the template(s) and the model sequence) to copy the ligand residue(s) as a rigid body into the model. MODELLER [15] always reads PDB residues in the order they are written in the PDB file, so if one has a ligand at the end of the template PDB file, one needs to put the '.' symbol at the end of the sequence in the alignment too.

Similarly, the co-factor can also be modeled along with the protein co-ordinates of the PDB (hence, we call it co-modeled) by using a single '.' symbol in the model sequence of the alignment. This must be aligned with a suitable cofactor in the template sequence. MODELLER [15] builds restraints on these ligands to keep their geometry and environment reasonably similar to the template, by restraining some intra-ligand, inter-ligand, and ligand-protein distances calculated based on the template structure [20].

These models were further filtered based on DOPE score and were validated using the previously mentioned structure validation tools. The alignment file used as input in MODELLER [15] between the target and the template of each enzyme is provided as supplementary materials S1-S6 along with the coordinate files of the 3D model structures shown in Figs. 2 I-III and 3 I-III.

Molecular docking and selection of predicted interaction pose

The modeled *C. roseus* structures were superimposed onto their respective template crystal structures and sites equivalent to the active and substrate binding sites of the crystal complex were considered as seeding residues for molecular docking. Docking of *C. roseus* vindoline biosynthesis pathway enzymes with their respective substrates were guided with this a prior information while performing molecular docking via multiple programs namely, GOLD [21, 22] and FlexX [23]. Top 100 solutions from both GOLD [21, 22] and FlexX [23] were generated and were further subjected to manual selection based on the suitability of the enzymatic reaction. The manual selection and/or elimination of the likely/unlikely solutions were done based on the mode and orientation of the ligands' functional group with respect to the critical active sites residues. In other words, the spatial proximity and orientation of the docked atoms and active site residues that are required for the enzymatic function were considered to select the most likely docking poses. Only those solutions were considered for further selection where the reaction centers are located within reasonable distances and suitable angles from the cofactor molecules, which facilitate the oxidation/hydroxylation/methylation reactions.

The highest scoring solution (based on *Chemscore*) among the manually selected group of GOLD [21, 22] solutions was considered as the most likely solution. The GOLD [21, 22] solution selected by the above-mentioned protocol was considered as a representative pose for the enzyme substrate complex and was compared with the top five FlexX [23] solutions that are most suitable for the enzymatic reaction to take place. Enzyme-substrate binding energy and affinity of each representative docked pose were calculated using MOPAC [ΔG (kcal mol^{-1})] [24] and Cyscore [25].

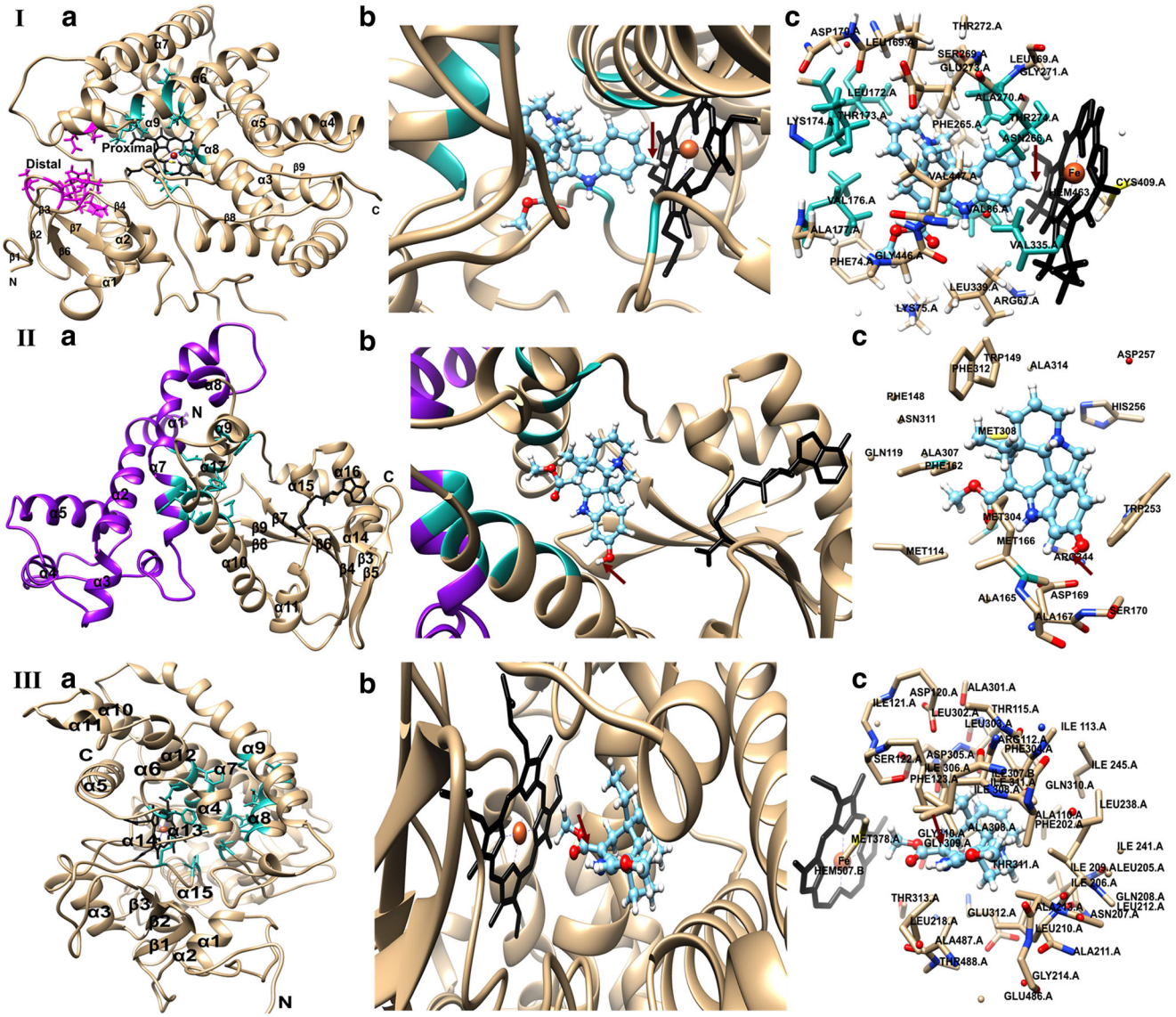

Fig. 2 Panel I: (**a**) Three-dimensional (3D) model of tabersonine-16-hydroxylase with heme cofactor (black) containing the Fe^{+++} ion (orange red). Two proposed active sites of the enzymes (distal and proximal) are shown in magenta and cyan, respectively. (**b**) Close up view of the tabersonine-16-hydroxylase docked with substrate tabersonine (ball and stick model) where the residue that undergoes hydroxylation is marked by red arrow. (**c**) Interactive residues within 5 Å distance from the docked substrate are shown in tan color. Coordinate file is provided as supplementary materials (Fig. 2I.pdb). Panel II: (**a**) 3D model of enzyme tabersonine-O-metyltransferase with ligand SAH shown in black color. N-terminal domain (purple) resides the substrate binding pocket (cyan green) and C-terminal domain (tan) contains the SAH binding pocket. (**b**) Tabersonine-O-methyltransferase docked with substrate 16-hydroxytabersonine (ball and stick model) where red arrow shows the center for O-methylation. (**c**) Interactive residues within 5 Å distance from the docked substrate are shown in tan color. Coordinate file is provided as supplementary materials (Fig. 2II.pdb). Panel III: (**a**) Modeled structure of enzyme tabersonine-3-oxygenase with cofactor heme (black) containing the Fe^{+++} (orange) ion is in the center of the porphyrin ring. Substrate binding pocket is shown in cyan. (**b**) Close up view of tabersonine-O-methyltransferase docked with substrate 16-hydroxytabersonine (ball and stick) where red arrow shows the hydroxylation center in the substrate. (**c**) Interactive residues within 5 Å distance from the docked substrate are shown in tan color. Coordinate file is provided as supplementary materials (Fig. 2III.pdb)

Molecular dynamics study of the apo-enzyme and substrate bound enzyme complexes

Both the substrate unbound protein structures (apo-structures) and the ligand/substrate bound (bound-structure) protein complex derived from the docking analysis were further considered for molecular dynamics (MD) study using Desmond [26] program for 100 ns (ns) at 300 K temperature. The OPLS_2005 force field parameters [27–31] were used for the simulation.

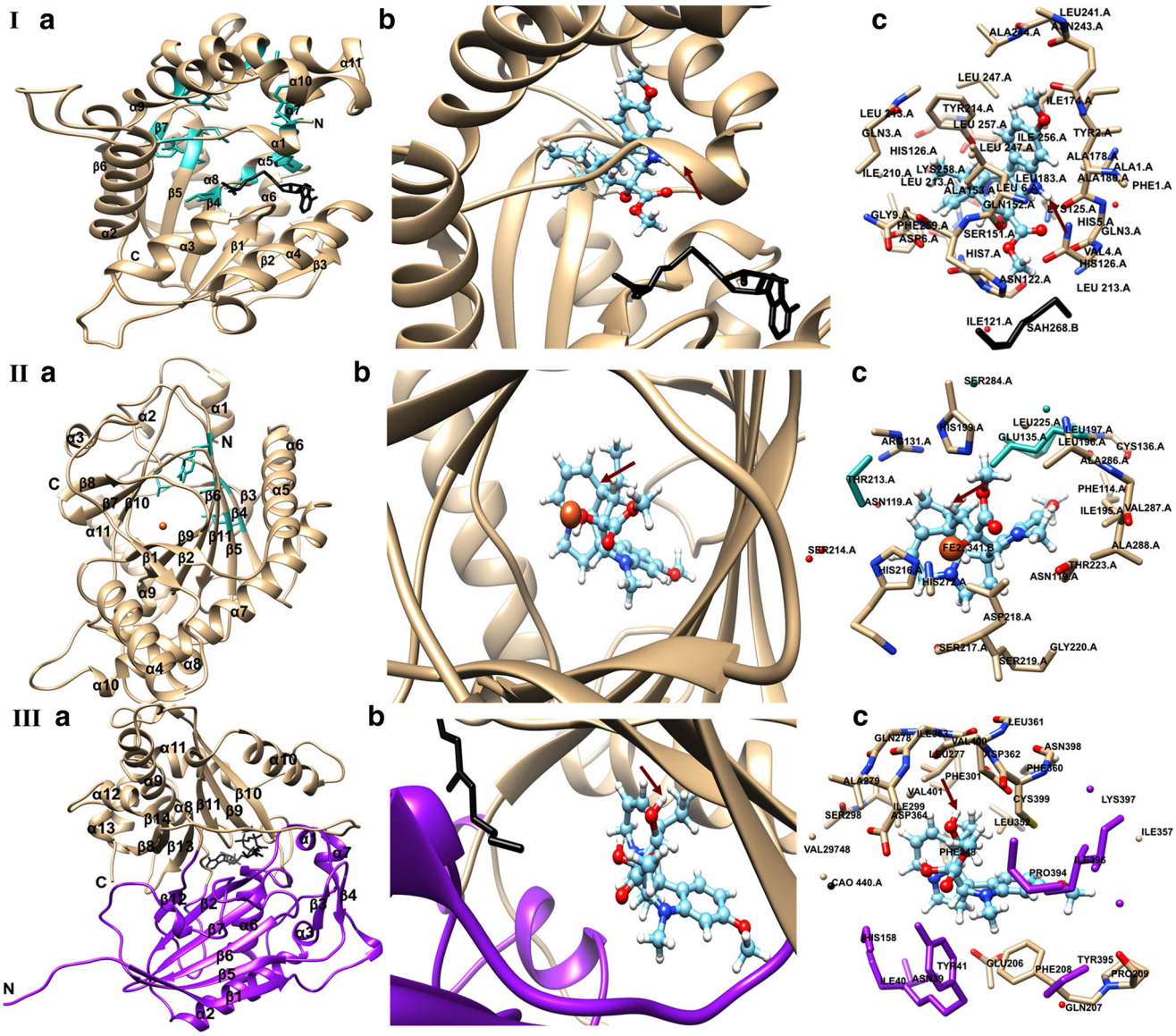

Fig. 3 Panel I: (**a**) 3D model of enzyme tabersonine-N-methyltransferase with SAH (black) where substrate binding sites are shown in cyan. (**b**) Tabersonine-N-methyltransferase docked with substrate 3-hydroxy-16-methoxy 2,3-dihydrotabersonine (ball and stick) and the N-methylation center is shown by red arrow. (**c**) Interactive residues within 5 Å distance from the docked substrate are shown in cyan color. Coordinate files provided as supplementary materials (Fig. 3I.pdb). Panel II: (**a**) Modeled structure of enzyme deacetoxyvindoline-4 hydroxylase having β jellyroll topology with cofactor Fe^{++} (orange red) and substrate binding cavity marked in cyan. (**b**) Deacetoxyvindoline-4-hydroxylase docked with deacetoxyvindoline (ball and stick) and the red arrow depicts the hydroxylation center. (**c**) Interactive residues within 5 Å distance from the docked substrate are shown in cyan color. Coordinate file is provided as supplementary materials (Fig. 3II.pdb). Panel III: (**a**) 3D model of enzyme deaceteylvindoline-O-acetyltransferase with cofactor acetyl-CoA (black). The structure consists of domain1 (purple) and domain2 (tan) with different topology. (**b**) Deaceteylvindoline-O-acetyltransferase docked with deacetylvindoline (ball and stick) where the red arrow points to the acetylation center of the substrate. (**c**) Interactive residues within 5 Å distance from the docked substrate are shown in tan and purple colors. Coordinate file is provided as supplementary materials(Fig. 3III.pdb)

Optimized potential for liquid simulations (OPLS) have a large number of variants, which include OPLS_AA, OPLS_UA, OPLS_2001, OPLS_2005, OPLS3. OPLS is developed by William L. Jorgensen at the Yale University Department of Chemistry. Monte Carlo statistical mechanics simulations are used to setup non-bonded parameters using computing thermodynamic and structural properties for 34 pure organic liquids including alkanes, alkenes, alcohols, ethers, acetals, thiols, sulfides, disulfides, aldehydes, ketones and amides. The force-field is capable of handling diverse functional groups of the organic molecules. Parameter assignment on the atom types are performed by matching the molecular fragments which are stored as string characters using the notation; SMILES/SMARTS language. The SMARTS pattern-matching algorithm is based on the Lewis structure of the molecules. The atomic number, bond orders, and the formal

charges of the atoms are derived from the defined Lewis structure. However, in absence of the atom types present in Lewis structure, initially the probable formal charges are guessed based on the type and number of bonds connecting the atoms. This process continues unless a valid Lewis structure is obtained. More detailed information on the OPLS_2005 parameters is provided in the supplementary material S7.

The system was solvated with TIP3P [32] water and further the ions were added to neutralize the whole system. Orthorhombic periodic boundary box with defined shape and size of the simulation box buffered at 10 Å distances were specified during the system preparation. Further, each system was considered for minimization. Relaxation of the system was performed under NPT ensemble using the default protocol of Desmond, which consists of a total of eight stages among which there are two minimization steps (restrained and unrestrained) and four short simulations (equilibration phase) steps before starting the actual production run. A hybrid method combining the steepest decent and the limited-memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS) algorithm was used to minimize the system. The temperature of NPT-simulations was regulated with Nose-Hoover Chain thermostat [33] with 1.0 ps (ps) relaxation-time and the pressure was regulated with Martyna-Tobias-Klein barostat [34] with isotopic coupling and 2.0 ps relaxation time. The final simulation run was carried out under normal pressure and temperature (NPT) condition for 100 ns at 300 K temperature and 1 atmospheric pressure with a time step of 2 fs. The RESPA [35] integrator was used to integrate the equations of motion. The trajectories were saved at the interval of 5.0 ps.

The molecular dynamic simulation trajectory files were considered for further analysis. The Simulation Quality Analysis (SQA) program was used to calculate the protein energies over the simulation period. Further, the in-house perl program was used to calculate the distance deviation of the active site residues from the initial time frame over the phase of simulation time of 100 ns for each interval of 10 ns.

Comparison of structural variation of the active sites between the apo and the bound complex structure

The spatial distance analysis of the binding site residues for each enzyme structure provides the average deviation of the active site residues over the different time frame during the MD simulation studies. In order to calculate the spatial distances among the different binding residues, initially we considered the structures of 10th ns, 20th ns, 30th ns, 40th ns, 50th ns, 60th ns, 70th ns, 80th ns, 90th ns and 100th ns of both apo and ligand-bound structures. The distance of the active site residues of each state of structure is compared with the 0th ns structure using an in-house perl program to determine its respective active site residual deviations. In the case of the apo-structure, the deviations of the active site residues over (10th, 20th, 30th...100th ns) different time frames are calculated in comparison to the 0th ns apo structure. Likewise, deviations of the active site residues of the ligand-bound protein structures are also calculated. Further, the difference of deviation of each active site residue of the bound structure with respect to the apo-structure of each enzyme is also calculated. The overall deviation values for all the active sites from all the intermediate structures (10th to 100th) range from (>= − 8 to >8). Overall percentage of deviation of the active site residues was calculated by dividing the number of active site residues within each bin of deviation range by the total number of active sites.

Suppose for one active site AS_1, the distance of the active site residues at i^{th}ns (10thns, 20thns, 30^{th}ns, etc.) with respect to the initial structure (0^{th} ns) for apo structure is X_i and for bound structure is Y_i, where $i = 10\ (10)\ 100$.

Average deviation with respect to the initial structures (X_0 and Y_0) for AS_1 is:

$$Avg\_Deviation_{AS_1} = \frac{1}{10}\sum_{i=10}^{100}[(X_0 - X_i) - (Y_0 - Y_i)], \forall i = 10\ (10)\ 100.$$

Similarly, deviations for other active site residues ($AS_2..AS_n$) have been calculated.

Root mean square deviation (RMSD) analysis determines overall structural deviation over the period of simulation time. We performed the RMSD trajectory analysis for each modeled structure; both apo and the ligand bound proteins. The stability of the apo and the ligand-bound protein structures are projected via the RMSD analysis.

However, in order to estimate the stability of the binding site residues, we have also performed root mean square fluctuation (RMSF) analysis. RMSF analysis estimates the residue wise fluctuations which measures the displacement of a particular atom, or group of atoms considering all the ensemble structures within the simulation.

Results and discussion

The following sections describe the results obtained from molecular modeling of each enzyme and the corresponding substrate. Further, the probable binding mode of each substrate to its enzyme aided by molecular docking and dynamics studies is also discussed.

Tabersonine-16-hydroxylase

Three-dimensional (3D) structure of tabersonine-16-hydroxylase, the first enzyme in the vindoline biosynthesis pathway was modeled based on the template steroid 21-hydroxylase (P450c21) (PDB ID: 3QZ1) (Fig. 2 Ia). Fe^{+++} is associated

with heme bound to this cytochrome P450-dependent enzyme, tabersonine-16-hydroxylase. The Fe before reaction is in Fe^{+++} state and is changed to Fe^{++} when the catalysis is carried out by the enzymes. The heme cofactor was also modeled along with the tabersonine-16-hydroxylase (Fig. 2 Ib) and the energy-minimized structure of the modeled enzyme was further validated using the standard validation protocol (Table 1). The structure of the enzyme tabersonine-16-hydroxylase exhibits the typical cytochrome P450 fold. The tertiary structure consists of major α-helices, which are similar to those found in most other cytochrome P450s. There are only two groups of β-sheets with three β-strands in the structure instead of four groups of sheets with 10 β-strands as observed in most of the other P450s. Like the template structure, the 3D model of tabersonine-16-hydroxylase also contains two probable substrate binding pockets. In order to determine the most suitable binding pocket for the substrate tabersonine, the 3D model of the tabersonine was used to dock onto the two probable binding pockets of tabersonine-16-hydroxylase using docking software like GOLD version 5.2.2 [21, 22] and FlexX version 3.1.4 [23]. Higher docking scores were observed for the site proximal (average *Chemscore*: 32.02) to the bound heme cofactor compared to the site distal (average *Chemscore*: 27.33) to heme. This indicates that the proximal binding site is perhaps energetically more favorable for binding with tabersonine. Moreover, for the conversion of tabersonine to 16-hydroxytabersonine, proximity of the substrate to heme is necessary as the Fe^{+++} of heme might facilitate the oxidoreduction reaction. After analyzing 100 poses generated by the GOLD [21, 22] software, the most plausible solution was selected based on higher docking scores (*Goldscore*, *Chemoscore*, and Δ*G of docking*) and the relative position of reaction center from Fe^{+++} of heme. The best GOLD derived solution was further compared for similarity in docking score and orientation with the top five FlexX [23] solutions (Table 2). The selected model of tabersonine-16-hydroxylase and the substrate complex possessed virtual binding energy (represented by ΔG) and binding affinity (represented by Cyscore [25]) similar to that of the top five FlexX [23] solutions (Table 2). The probable interactive residues along with their type of interaction within 5 Å distance from the docked substrate are shown in Fig. 2Ic and Table S1.

Tabersonine-O-methyltransferase

The second enzyme of the pathway, tabersonine-O-methyltransferase was modeled based on the crystal structure of *Medicago truncatula* isoflavonoid-O-methyltransferase [PDB ID: 1ZG3] [36]. The methyl donor ligand SAM/SAH was also modeled within its binding domain containing a Rossmann fold that is generally shared within most of the families of plant SAM-dependent small molecule methyltransferases. The final 3D model of tabersonine-O-methyltransferase with SAH was selected based on the structure validation results (Table 1). SAH resides within a cleft formed by α14, β3, β4, and β6 and the connecting loops (Fig. 2IIa). Overall structure contains two distinct domains where the N-terminal domain resides in the substrate binding pocket and the C-terminal domain contains SAH binding surfaces. Hence, it was expected that 16-hydroxytabersonine may bind at the N-terminal domain.

The 3D model of the 16-hydroxytabersonine was docked onto the N terminal phenolic substrate binding pocket using GOLD [21, 22] and FlexX [23]. The best docking pose shown in Fig. 2IIb was selected based on average *Goldscore*, *Chemoscore*, and Δ*G of docking*. 16-hydroxytabersonine fits well in the phenolic substrate binding cavity. The binding of 16-hydroxytabersonine is achieved through a complementary binding pocket formed by a hydrophobic surface in the Y-shaped active site. This cavity consists of hydrophobic side chains, a large fraction of which are aromatic and methionine residues. In particular, Met-166 and Met-308 may act like a clamp to constrain the orientation of the docked substrate. Methylation of 16-hydroxytabersonine to 16-methoxytabersonine is possible in the selected GOLD [21, 22] docking pose where the methyl group of the substrate is in close proximity to the active sites and the orientation of the substrate is energetically optimized and quite similar to those observed in the top five FlexX [23]solutions (Table 2). The

Table 1 Fold prediction and validation of structural models of the vindoline biosynthesis pathway

Query structure	Reference (PDB ID)	Sequence identities	Sequence similarity	RMSD value (Å)	Ramachandran plot			VERIFY_3D	ERRAT-score
					Favored	Allowed	Outlier		
Tabersonine-16-hydroxylase	3QZ1	27%	39.2%	0.481	92%	7.2%	0.9%	69.26%	55.507
Tabersonine-O-methyltransferase	1ZG3	48%	82%	0.515	94.0%	5.2%	0.9%	77.78%	75.510
Tabersonine-3-oxygenase	3SWZ	25%	39.8%	0.376	91.5%	7.5%	1.0%	68.97%	48.889
Tabersonine-N-methyltransferase	3BUS	27%	42.7%	0.482	96.2%	2.3%	1.5%	82.77%	60.618
Deacetoxyvindoline-4 hydroxylase	1GP6	28%	49.4%	0.345	94.1%	4.1%	1.8%	82.06%	57.273
Deaceteylvindoline-O-acetyltransferase	2BGH	31%	51.7%	0.429	92.7%	4.6%	2.7%	73.35%	51.515

Table 2 Comparison of docking score and interaction energy of GOLD vs FlexX docking solutions and RMSD between these solutions

Enzyme	Best GOLD-docked solution					Most likely FlexX-docked cluster (best 5 solutions)		
	Goldscore	*Chemscore*	Normalized *ChemScore*	ΔG (kcal mol^{-1})	Cyscore	Average RMSD (Å)	Average ΔG (kcal mol^{-1})	Average Cyscore
Tabersonine-16-hydroxylase	45.27	33.43	1	−6.43	−3.304	6.94 ± 0.323	−7.77 ± 0.537	−3.320 ± 0.050
Tabersonine-O-methyltransferase	33.06	26.21	0.885	−6.98	−3.306	4.764 ± 0.694	−6.17 ± 0.685	−2.551 ± 0.237
Tabersonine-3-oxygenase	23.17	27.66	0.847	−8.91	−4.693	6.986 ± 1.213	−7.82 ± 1.261	−2.748 ± 0.223
Tabersonine-N-methyltransferase	25.41	19.19	0.804	−6.04	−3.416	5.522 ± 0.098	−9.15 ± 0.499	−3.169 ± 0.302
Deacetoxyvindoline-4 hydroxylase	27.95	25.09	0.767	−6.65	−3.421	3.617 ± 0.376	−6.47 ± 0.791	−2.119 ± 0.216
Deaceteylvindoline-O-acetyltransferase	24.07	26.1	1	−7.93	−4.508	6.536 ± 1.220	−8.15 ± 0.197	−2.211 ± 0.106

probable interactive residues within 5 Å distance from the substrate are shown in Fig. 2 IIc and Table S2.

Tabersonine-3-oxygenase

Tabersonine-3-oxygenase was modeled based on the best template structure, cytochrome P450 17A1 [PDB ID: 3SWZ] [37]. Fe^{+++} is associated with heme bound to tabersonine-3-oxygenase, which is also a cytochrome P450. The Fe before reaction is in Fe^{+++} state and is changed to Fe^{++} when the catalysis is carried out by the enzymes. The 3D model consists of 21α-helices and five sheets with 12 β-strands, and three β-hairpins. Similar to the first enzyme of the pathway, the heme cofactor was modeled along with the tabersonine-3-oxygenase (Fig. 2 IIIa) and the energy minimized structure of the modeled enzyme was further validated using the standard validation protocol (Table 1). The substrate, 16-hydroxytabersonine was docked onto the 3D model of tabersonine-3-oxygenase and the best docking pose was selected based on *Chemscore, Goldscore*, and Δ*G of docking*(Fig. 2IIIb). The selected pose shows close proximity of the hydrolysable acetyl group of 16-hydroxytabersonine to the Fe^{+++} of the heme. The substrate binding cavity is primarily lined with hydrophobic atoms of Val114, Thr115, Ser122, Phe123, Ile209, Leu210, Leu212, Arg242, Leu317, and Ala487 residues. The probable interacting residues within 5 Å distance of the docked substrate are shown in Fig. 2IIIc while the types of interaction are listed in Table S3.

Tabersonine-N-methyltransferase

The structure of the enzyme tabersonine-N-methyltransferase (Fig. 3 Ia) consists of 11α-helices and seven β-strands. The domain exhibits a common tertiary structure consisting of a core α/β Rossmann fold, which is characteristic of all other AdoMet-dependent MTases, and a unique α-helical cap that forms the top of the active site cavity. The β-sheet contains a single antiparallel strand (β7) and packs against three proximal (α1–α3) and two distal (α5 and α7) helices. The enzyme was modeled based on the rebeccamycin-4'-O-methyltransferase RebM in complex with S-adenosyl-l-homocysteine (SAH) [PDB ID:3BUS] [38] was further used to dock the energy minimized structure of the substrate 3-hydroxy-16-methoxy-2,3-dihydrotabersonine. On manual examination of the poses, it was found that the pose shown in (Fig. 3 Ib) could be most realistic, as it possesses higher average *Goldscore*, *Chemoscore*, and Δ*G of docking* and is reasonably similar to the top five docking complexes suggested by FlexX [23]. Similarly, in this pose the distance between the methyl group donor and receiver was also found to be lowest among the docked complexes. The substrate binding is dominated by hydrophobic, electrostatic, and vander Waals interactions involving residues His7, Ser22, Gln23, Asp24, Val26, Ile27, Ile30, Lue57, Gly58, Thr60, Tyr63, Tyr84, Tyr88, Glu220, and Tyr224. Only a few putative hydrogen bonds participate in anchoring the substrate, potentially derived from residues His5 and Glu220. The residues marked in Fig. 3 Ic are the probable interactive residues within the 5 Å region of the substrate, whereas Table S4 lists the types of probable interactions.

Deacetoxyvindoline-4-hydroxylase

The structure of enzyme deacetoxyvindoline-4-hydroxylase was modeled based on the 3D template structure of anthocyanidin synthase from *Arabidopsis thaliana* [PDB ID: 1GP6 [39]. In the case of deacetoxyvindoline-4-

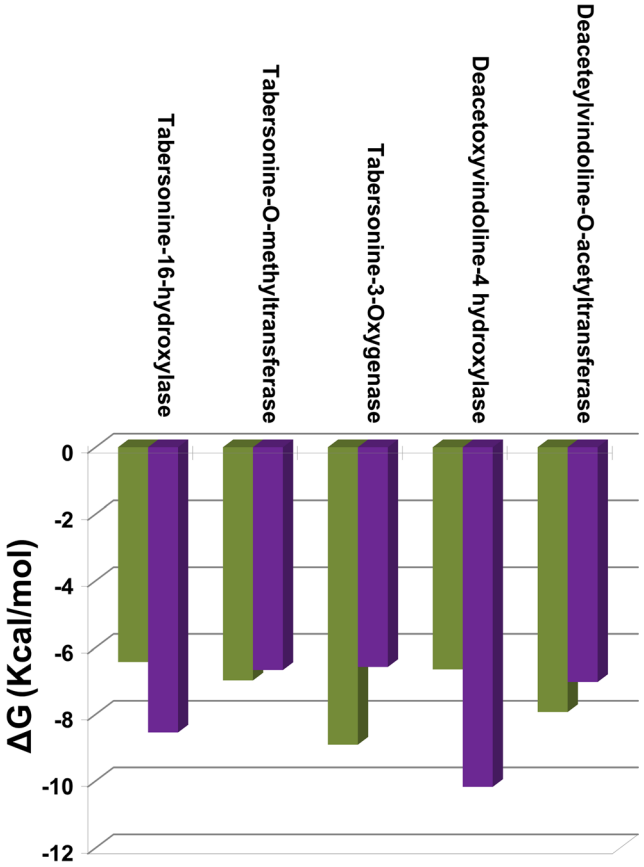

Fig. 4 Enzyme-substrate binding energy [ΔG (kcal mol^{-1})] of each representative docked pose was calculated using MOPAC[31]. These computationally derived ΔG values (shown in green bars) were compared with ΔGs (shown in purple bars) calculated from the experimentally derived K_m values of tabersonine-16-hydroxylase, tabersonine-O-methyltransferase, tabersonine-3-oxygenase, deacetoxyvindoline-4-hydroxylase, and deaceteylvindoline-O-acetyltransferase, respectively. K_m values for the respective enzymes were collected from the Braunschweig Enzyme Database (BRENDA). Experimentally derived binding constants (K_m) were converted to ΔG_{exp} using the formula $\Delta G = -RTlnK_m$, where, $R = 0.0019872041$ cal K^{-1} mol^{-1} and $T = 297$ Kelvin. Km value for tabersonine-N-methyltransferase was not available in BRENDA

hydroxylase where Fe is involved, the Fe is in +2 oxidation state which is required because deacetoxyvindoline-4-hydroxylase belongs to a oxygenase subfamily that uses ferrous iron (Fe^{++}) as a cofactor. The reaction also suggests that Fe^{++} is primarily responsible for the reactivation of the deacetoxyvindoline-4-hydroxylase [40]. The modeled 3D structure of deacetoxyvindoline-4-hydroxylase consists of 13 β-strands, where eight of them (β4- β11) form a β jellyroll or double stranded β helix topology (Fig. 3 IIa). The long α-helix (α7) of the jellyroll forms the structural backbone. Figure 3IIb provides the most likely pose of the docked substrate, deacetoxyvindoline. The enzyme hydroxylates deacetoxy vindoline to deacetyl-vindoline. In the docked pose (Fig. 3 IIb and IIc) the deacetoxy vindoline is surrounded by hydrophobic residues, including Arg131, Glu135, Ile195, His199, His216, Asp218 Ser219, Gly220, Ala286, and Gly290. Gly220, Ala286, and Gly290 are on the "inside" face of a C-terminal helix (α11 in Fig. 3 IIa) forming a lid on one edge of the active site. In this pose, the hydroxyl group of the substrate is in the vicinity of Fe^{++} and forms a probable hydrogen bond with the catalytic residue His216. The probable interactive residues along the types of interaction within 5 Å distance from substrate are shown in Fig. 3 IIc and Table S5.

Deaceteylvindoline-O-acetyltransferase

The final enzyme of this pathway was modeled based on the crystal structure of vinorine synthase from *Rauvolfia serpentine* [PDB ID: 2BGH] [41]. The model structure of the deacetylvindoline-o-acetyltransferase (Fig. 3IIIa) contains 14 β-strands and 13 α-helices and consists of two approximately equal-sized domains. Domain 1 contains a pair of strands on the surface of the enzyme at one end of the central sheet. Domain 2 contains six helices and a mixed β-stranded sheet. Domain 1 and domain 2 share a backbone fold albeit with a different topology. The cofactor acetyl-CoA was co-modeled and is located near the docked substrate (Fig. 3 IIIb). The substrate binding is dominated by hydrophobic, electrostatic, and vander Waals interactions involving residues Asn39, Ile40, Tyr41, His158, Glu206, Gln207, Phe208,Pro209, Leu277, Gln278, Ala279, Val297, Ser298, Ile299, Leu352, Leu361, Asp362, Ile363, Asp364, Phe380, Pro394, Tyr395, Ile396, Asn398, Cys399, and Val400. Table S6 and Fig. 3 IIIb list the types of probable interactions between the docked substrate and these residues.

Rationalization of the docking poses and predicted interacting sites

The representative docking pose predicted for each enzymatic step was subjected to critical manual scrutiny based on the suitability of the enzymatic reaction. The manual selection and/or elimination of the likely/unlikely solutions were done based on the mode and orientation of the ligands' functional group with respect to the critical active sites residues. In other words, the spatial proximity and orientation of the docked atoms and active site residues that are required for the enzymatic function were considered to select the most likely docking poses. Only those solutions were considered for further selection where the reaction centers (marked by red arrow in Figs. 2 and 3) are located within reasonable distances and suitable angles from the cofactor molecules, which facilitate the oxidation/hydroxylation/methylation reactions. The highest scoring solution (based on *Chemscore*) among the manually selected group of GOLD [22, 23] solutions was considered as the most

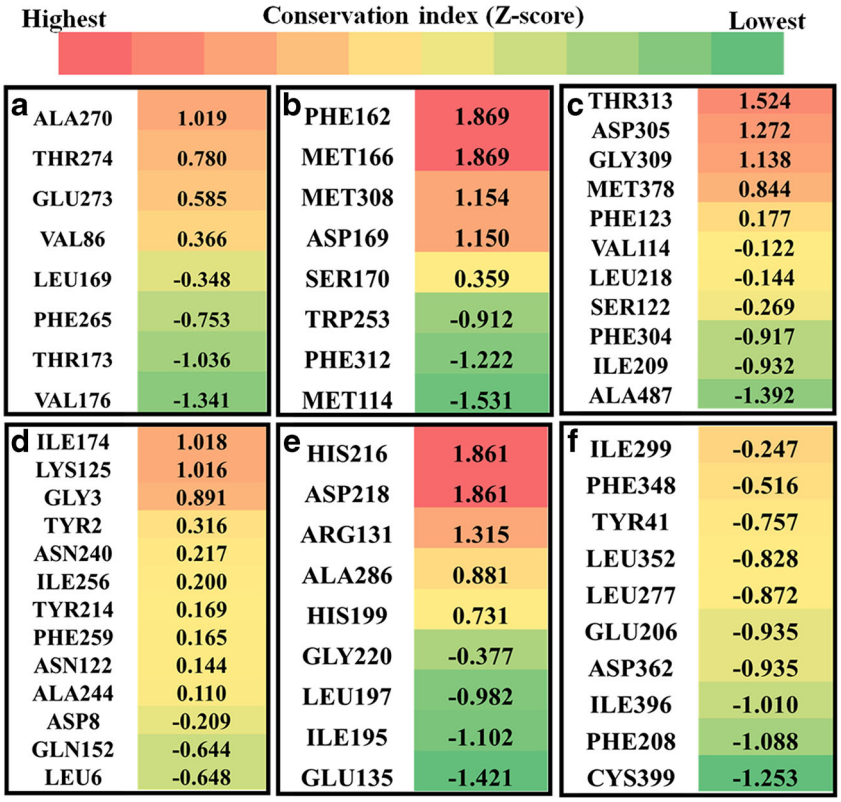

Fig. 5 Conservation index (shown in Z-score scale) of the identified substrate binding sites of the enzymes (**a**) tabersonine-16-hydroxylase, (**b**) tabersonine-O-methyltransferase, (**c**) tabersonine-3-oxygenase, (**d**) tabersonine-N-methyltransferase, (**e**) deacetoxyvindoline-4-hydroxylase, and (**f**) deaceteylvindoline-O-acetyltransferase, respectively

likely solution. However, the selected representative docking pose was further compared with the five most likely docking solutions obtained from another program called FlexX [24]. Reasonable similarity was observed between the predicted representative pose and the best FlexX [24] solutions for each of the enzyme-substrate complexes (Table 2). Enzyme-substrate binding energy [ΔG] [25] and affinity [Cyscore] [26] calculated for the representative pose are also similar and comparable with that obtained from the best FlexX [24] solutions (Table 2). Similarly, the enzyme-substrate binding energies [ΔG] [25] obtained from the computational complexes are found to be comparable to the ΔGs calculated from the experimentally derived K_m values (Fig. 4). These observations add reliability to the selected representative docking poses. Further, it was also observed that the binding energy and affinity calculated for the representative pose are comparable and/or better to a known protein complex bound with vinblastine and vindoline moiety (Table S7).

Substrate binding sites surrounding the docked substrate (<=5 Å) could be involved in critical hydrogen bonding and other non-bonded interactions as shown in Tables S2-S6. Conservation pattern of the predicted substrate binding sites were estimated in order to gain evolutionary support for their involvement in binding and catalysis. Figure 5 suggests a relatively higher conservation index of the predicted substrate binding sites for most of the enzymes in the vindoline biosynthesis pathway. For example, 50%, 62.5%, 45%, 77%, and 55% of the predicted substrate binding sites from tabersonine-16-hydroxylase, tabersonine-O-methyltransferase, tabersonine-3-oxygenase, tabersonine-N-methyltransferase, and deacetoxyvindoline-4-hydroxylase, respectively possess positive conservation index Z-scores. However, the deaceteyl vindoline-O-acetyl transferase enzyme family was found to be more diverse, and the predicted substrate binding sites do not possess positive conservation index Z-scores.

Further, the protein-ligand complexes were subjected to molecular dynamics simulations to check the variability of the binding properties over simulation time.

The trajectory analysis for each modeled structure; both apo and the ligand bound protein structures are performed. The stability of the apo and the ligand-bound protein structures are projected as RMSD plots in panel As of Figs. 6 and 7. RMSD profiles of each of the apo and the ligand-bound complex of enzyme structures indicate the equal or higher stability of the ligand-bound structures over their apo-modeled structures.

However, in order to estimate the stability of the binding site residues, we have also performed RMSF analysis and the plots are provided in panel Bs and Cs of Figs. 6 and 7. The active site residues for both the apo and the ligand-bound structures are marked over the plot and the average RMSF of the active site residues of both apo and the ligand bound protein structures are plotted in panel C in each series. This exercise indicates that the active site residues of the ligand-bound structures are either having lesser or similar atomic fluctuations in comparison to the apo-structure of each enzyme.

The distances of the binding residues with respect to the corresponding docked ligand were calculated and the

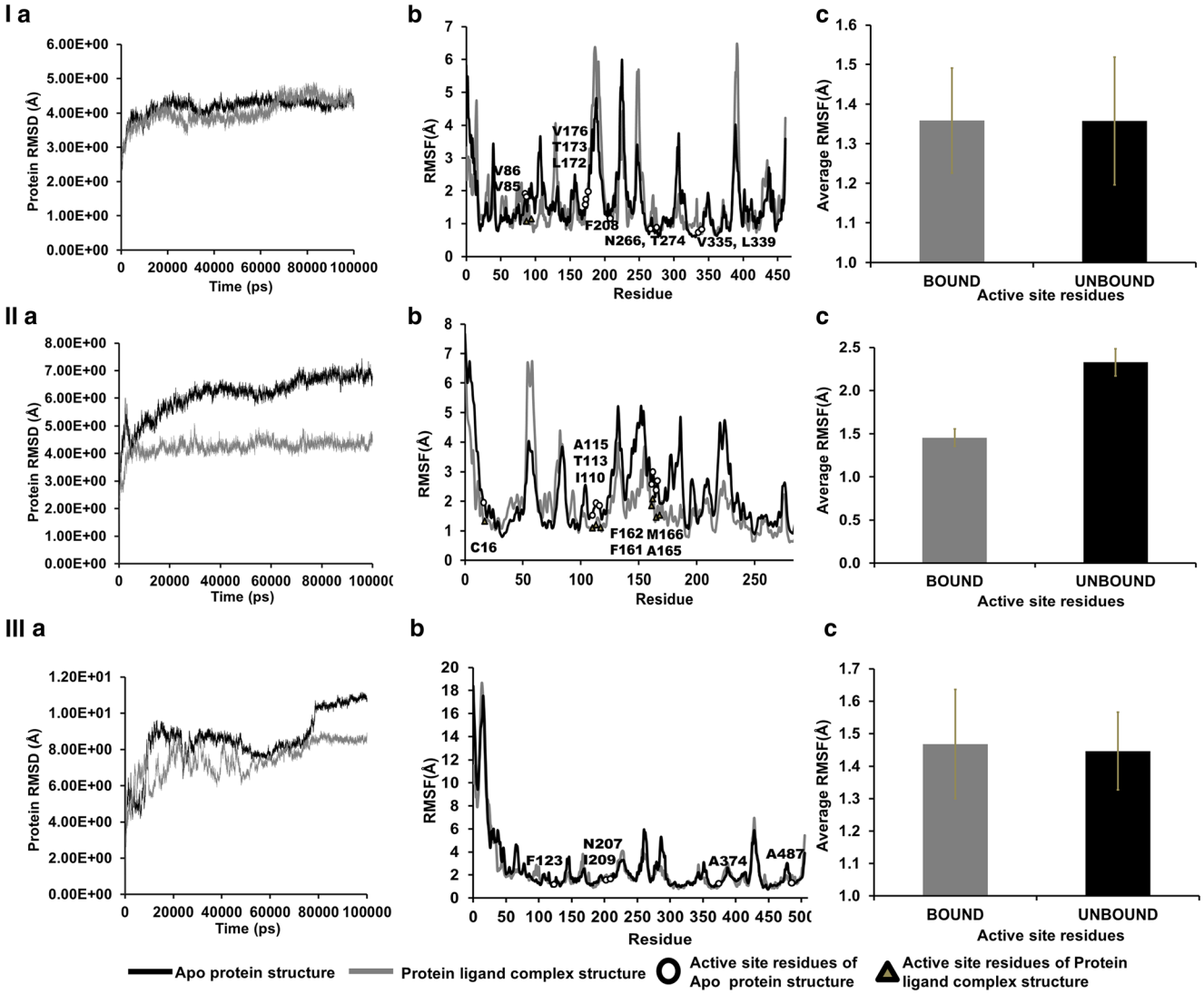

Fig. 6 Panel I: (**a**) Protein RMSD plot of both apo and ligand bound protein structure of enzyme tabersonine-16-hydroxylase (**b**) The RMSF plots between apo and ligand bound protein structure of enzyme tabersonine-16-hydroxylaseare plotted and the active site residues are highlighted onto it. (**c**) The comparative analysis of the average RMSF of the active site residues of apo and ligand bound protein structure of enzyme tabersonine-16-hydroxylase are represented. Panel II: (**a**) Protein RMSD plot of both apo and ligand bound protein structure of enzyme tabersonine-O-metyltransferase. (**b**) The RMSF plots between apo and ligand bound protein structure of enzyme tabersonine-O-metyltransferaseare plotted and the active site residues are highlighted onto it. (**c**) The comparative analysis of the average RMSF of the active site residues of apo and ligand bound protein structure of enzyme tabersonine-O-metyltransferase are represented. Panel III: (**a**) Protein RMSD plot of both apo and ligand bound protein structure of enzyme tabersonine-3-oxygenase. (**b**) The RMSF plots between apo and ligand bound protein structure of enzyme tabersonine-3-oxygenaseare plotted and the active site residues are highlighted onto it.(**c**) The comparative analysis of the average RMSF of the active site residues of apo and ligand bound protein structure of enzyme tabersonine-3-oxygenase are represented

deviationfrom the initial structure (0^{th} ns) vs final structure (100th ns) was measured. Figure 8a shows that almost 70% of such ligand-interacting residue distances deviates upto 4 Å, whereas 81% deviates only up to 6 Å. This indicates that the initial protein-ligand docked complexes are reasonably stable and most of the predicted interactions between the ligand and protein atoms are reliable. Similarly, fluctuations of the active site residues in the simulated structures extracted at regular intervals of the simulation with respect to the initial structure are calculated. Average deviations of these fluctuations between apo and bound structures are calculated and plotted in Fig. 8b. Almost 70% of the actives/binding sites fluctuate within ±2 Å, indicating the good quality of the 3D models and the docking complexes.

Conclusions

The three dimensional (3D) structures of the enzymes involved in the vindoline biosynthesis pathway are yet to be

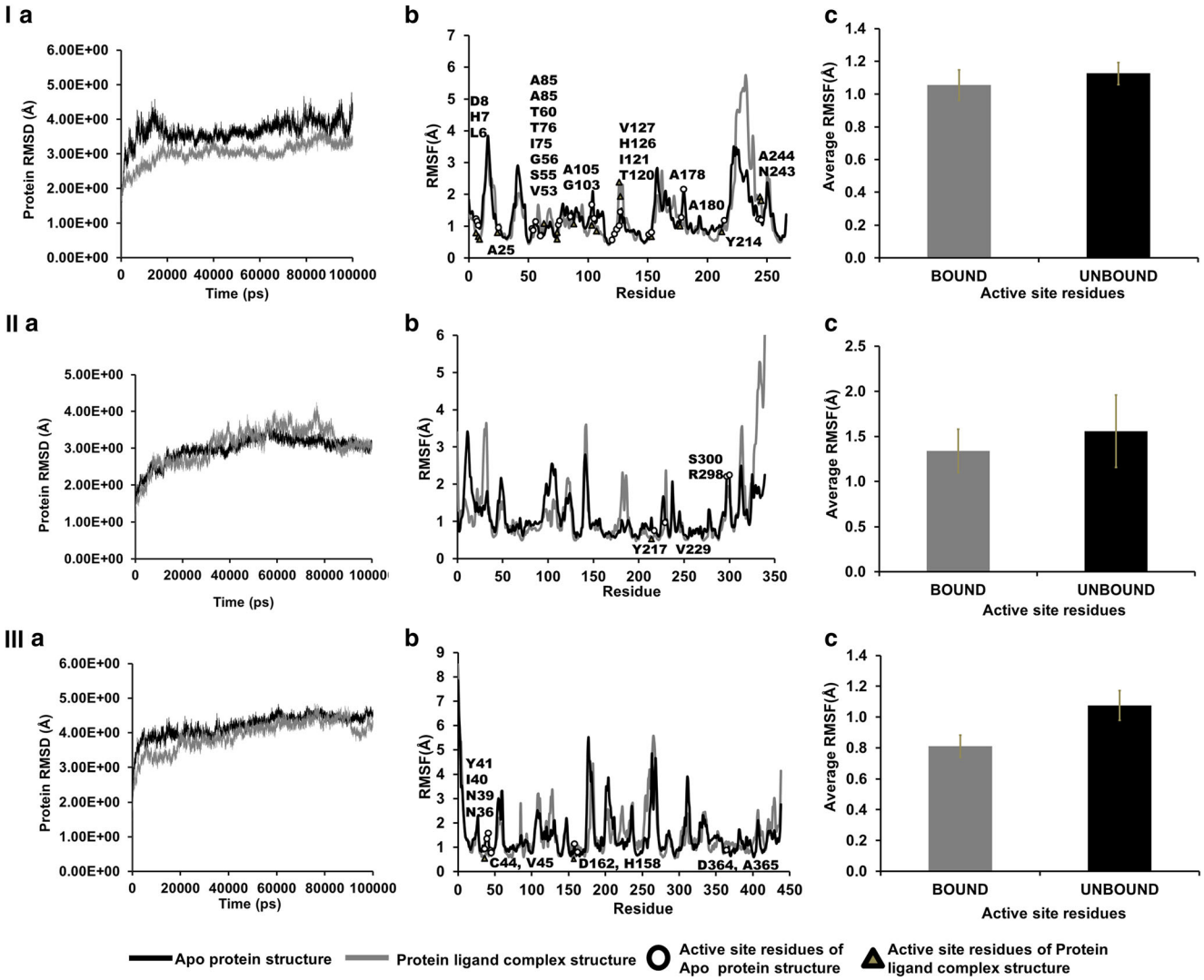

Fig. 7 Panel I: (**a**) Protein RMSD plot of both apo and ligand bound protein structure of enzyme tabersonine-N-methyltransferase (**b**) The RMSF plots between apo and ligand bound protein structureof enzyme tabersonine-N-methyltransferaseare plotted and the active site residues are highlighted onto it. (**c**) The comparative analysis of the average RMSF of the active site residues of apo and ligand bound protein structure of enzyme tabersonine-N-methyltransferase are represented. Panel II: (**a**) Protein RMSD plot of both apo and ligand bound protein structure of enzyme deacetoxyvindoline-4 hydroxylase. (**b**) The RMSF plots between apo and ligand bound protein structure of enzyme deacetoxyvindoline-4 hydroxylaseare plotted and the active site residues are highlighted onto it. (**c**) The comparative analysis of the average RMSF of the active site residues of apo and ligand bound protein structure of enzyme deacetoxyvindoline-4 hydroxylase are represented. Panel III: (**a**) Protein RMSD plot of both apo and ligand bound protein structure of enzyme deaceteylvindoline-O-acetyltransferase. (**b**) The RMSF plots between apo and ligand bound protein structure of enzyme deaceteylvindoline-O-acetyltransferaseare plotted and the active site residues are highlighted onto it. (**c**) The comparative analysis of the average RMSF of the active site residues of apo and ligand bound protein structure of enzyme deaceteylvindoline-O-acetyltransferase are represented

solved using experimental techniques. Hence, the knowledge about the mechanism of substrate binding and catalysis of various steps in this pathway is still limited. In this work, we intend to provide structural insight for the enzymes involved in vindoline synthesis by developing reliable three-dimensional (3D) models. A 3D model structure of each enzyme (total six) was generated and evaluated via rigorous molecular modeling and structure validation protocols, respectively. We also utilized these structures to generate plausible enzyme-substrate complex models using molecular docking analysis. Substrate and cofactors required for each step were docked onto the model of the corresponding enzyme and the probable binding residues were identified. Altogether, for the six steps of the pathway, 61 residues were identified as substrate binding residues out of which ten form probable hydrogen bonds with the respective enzyme, whereas 59 are involved in non-bonded interactions.

The validity of the enzyme models and their mode of binding to the corresponding substrates were verified based on standard structure evaluation methods as well as using semi-

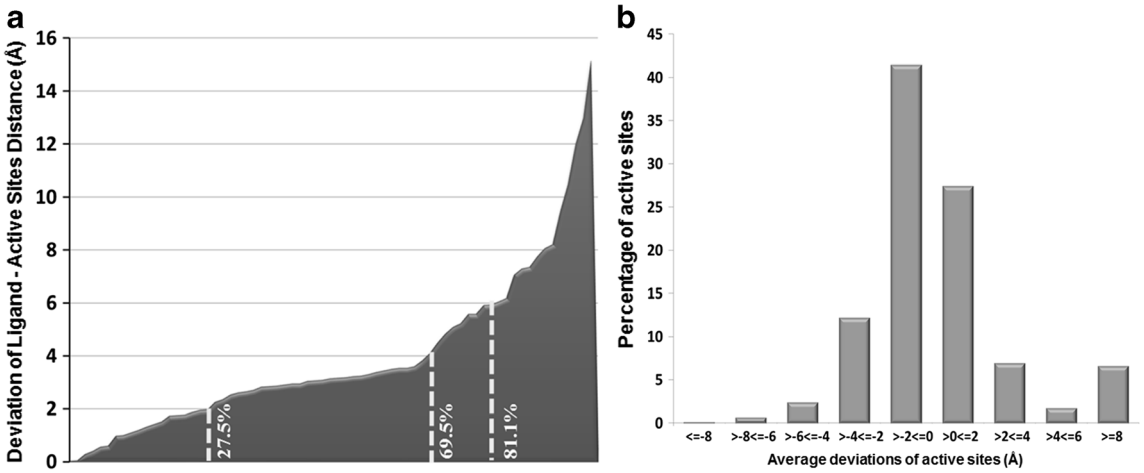

Fig. 8 Molecular dynamics based analyses to check the variability and reliability of the binding sites. The deviations of the ligand-active site distances of the final structure (100th ns) with respect to the initial structure (0th ns) of the simulation is plotted (panel A). Fluctuations of the active site residues in the simulated structures with respect to the initial structure are calculated. Average deviations of these fluctuations between apo and bound structures are calculated and plotted (panel B)

empirical approaches. The sequence identities between the *C. roseus* enzyme sequences and their respective templates range from 25 to 48%, indicating reasonable reliability of the resultant model structures. The 3D models possess acceptable structure validation scores having more than 97–99% of their residues located in allowed regions of the Ramachandran plot (Table 1). Fold compatibility scores estimated by 1D-3D compatibility via the VERIFY_3D [17] program also suggest reliability of the built 3D models.

The most likely pose of each enzyme-substrate complex was identified by manual scrutiny and docking score based selection. Various approaches that evaluate protein-ligand interaction suggest a reasonable range of binding energies and affinities for the proposed enzyme-substrate complexes. However, as there is no reference complex to compare, it is rather difficult to assume whether the complex energies are acceptable or not. In the absence of knowledge regarding the binding affinities of vindoline or its intermediate moieties to their respective enzymes or to any other proteins, we thought it would be appropriate to compare the substrate binding energies [ΔG] [25] and affinities [Cyscore] [26] of each intermediate enzyme-substrate complex with that of vinblastine complexed with tubulin αβ dimmer [PDB ID: 1Z2B] [42]. Vinblastine is formed by condensation of vindoline and catharanthine. In part, it has similarities with vindoline or its precursor moieties; hence, the binding energy related to the vindoline part of the vinblastine complexed with tubulin can be comparable to that derived from vindoline precursors and their respective enzymes. Indeed, we observed that binding energies and affinities calculated from the representative poses are comparable and/or better than the tubulin complexed (Table S7) with vindoline where the catharanthine part was removed from the complex. Molecular dynamics based analyses checked the variability and reliability of the binding properties of the active sites. Acceptable range of deviation for active site ligand distances was observed for the majority of the cases indicating good quality and reliability of the docking complexes. Similarly, stability of the modeled active sites was also examined via fluctuation analysis of the active site in the simulated structures.

Overall, this study provides a basic framework for structural and functional understanding of the enzyme-substrate interaction mediated in vindoline synthesizing enzymes. The ingredients and the findings of this study could be useful in the large-scale biosynthesis of vindoline, which is a primary precursor of vinblastine and vincristine.

Acknowledgments SC acknowledges CSIR-IICB for infrastructural support and Department of Biotechnology for Ramalingaswami fellowship. The work is supported by CSIR network projects funds (HCP002 and BSC0121).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Negi A, Singla R, Singh V (2014) Indole based alkaloid in cancer: an overview. PharmaTutor 2:76–82
2. Noble RL (1990) The discovery of the vinca alkaloids–chemotherapeutic agents against cancer. Biochem Cell Biol 68(12):1344–1351
3. Zu YG, Luo M, Mu FS, Fu YJ, Wang L (2006) Advances on alkaloids in Catharanthusroseus and theirpharmacological activities. Nat Prod Res Dev (18):325–329
4. Qu Y, Easson ML, Froese J, Simionescu R, Hudlicky T, De Luca V (2015) Completion of the seven-step pathway from tabersonine to the anticancer drug precursor vindoline and its assembly in yeast.

Proc Natl Acad Sci U S A 112(19):6224–6229. https://doi.org/10.1073/pnas.1501821112
5. Sottomayor M, Duarte P, Figueiredo R, RosBarcelo A (2008) A vacuolar class III peroxidase and the metabolism of anticancer indole alkaloids in Catharanthusroseus: can peroxidases, secondary metabolites and arabinogalactan proteins be partners in microcompartmentation of cellular reactions? Plant Signal Behav 3(10):899–901
6. O'Connor SE, Maresh JJ (2006) Chemistry and biology of monoterpeneindole alkaloid biosynthesis. Nat Prod Rep 23(4): 532–547. https://doi.org/10.1039/b512615k
7. UniProt: the universal protein knowledgebase (2017). Nucleic Acids Res. 45(D1):D158–D169. https://doi.org/10.1093/nar/gkw1099
8. Schomburg I, Chang A, Ebeling C, Gremse M, Heldt C, Huhn G, Schomburg D (2004) BRENDA, the enzyme database: updates and major new developments. Nucleic Acids Res 32(Database issue): D431–D433. https://doi.org/10.1093/nar/gkh081
9. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215(3):403–410. https://doi.org/10.1016/S0022-2836(05)80360-2S0022-2836(05)
10. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD, Higgins DG (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal omega. Mol Syst Biol 7:539. https://doi.org/10.1038/msb.2011.75
11. Pei J, Grishin NV (2001) AL2CO: calculation of positional conservation in a protein sequence alignment. Bioinformatics 17(8):700–712
12. Soding J, Biegert A, Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res 33(Web Server issue):W244–W248
13. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ (2015) The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10(6):845–858. https://doi.org/10.1038/nprot.2015.053
14. McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. Bioinformatics 16(4):404–405
15. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234(3):779–815. https://doi.org/10.1006/jmbi.1993.1626
16. Lovell SC, Davis IW, Arendall 3rd WB, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC (2003) Structure validation by Calpha geometry: phi,psi and Cbeta deviation. Proteins 50(3):437–450. https://doi.org/10.1002/prot.10286
17. Luthy R, Bowie JU, Eisenberg D (1992) Assessment of protein models with three-dimensional profiles. Nature 356(6364):83–85. https://doi.org/10.1038/356083a0
18. Colovos C, Yeates TO (1993) Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci 2(9):1511–1519. https://doi.org/10.1002/pro.5560020916
19. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF chimera–a visualization system for exploratory research and analysis. J Comput Chem 25(13): 1605–1612. https://doi.org/10.1002/jcc.20084
20. Sali A (1995) Comparative protein modeling by satisfaction of spatial restraints. Mol Med Today 1(6):270–277
21. Jones G, Willett P, Glen RC (1995) Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. J Mol Biol 245(1):43–53
22. Jones G, Willett P, Glen RC, Leach AR, Taylor R (1997) Development and validation of a genetic algorithm for flexible docking. J Mol Biol 267(3):727–748. https://doi.org/10.1006/jmbi.1996.0897
23. Rarey M, Kramer B, Lengauer T, Klebe G (1996) A fast flexible docking method using an incremental construction algorithm. J Mol Biol 261(3):470–489. https://doi.org/10.1006/jmbi.1996.0477
24. James JP, Stewart (2016) MOPAC2016, Stewart Computational Chemistry, Colorado Springs, CO. http://OpenMOPAC.net 2016
25. Cao Y, Li L (2014) Improved protein-ligand binding affinity prediction by using a curvature-dependent surface-area model. Bioinformatics 30:1674–1680
26. Kevin JB,Chow E, Xu H, Dror RO, Eastwood MP, Gregersen BA, Klepeis JL, Kolossváry I, Moraes MA, Sacerdoti FD, Salmon JK, Shan Y, Shaw DE (2006) Scalable algorithms for molecular dynamics simulations on commodity Clusters. Proceedings of the ACM/IEEE conference on supercomputing (SC06), Tampa, FL, 11–17Nov 2006
27. Jorgensen W L,Maxwell D S,Tirado R J (1996) Development and testing of the OPLS all-atomforce field on conformational energetics and properties of organic liquids. J Am Chem Soc (118):11225–11236
28. Kaminski G, Friesner RA, Tirado-Rives J, Jorgensen WL (2001) Evaluation and reparameterizationof the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations onpeptides. J Phys Chem B 105:6474–6487
29. Levy RM (2005) The OPLS_2005 parameters are described in Banks. Integrated Modeling Program, Applied Chemical Theory (IMPACT). J Comp Chem (26): 1752
30. Maxwell S (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. Connecticut 06520–8107
31. McAliley JH, Bruce DA (2011) Development of force field parameters for molecular simulation of polylactide. J Chem Theory Comput 7(11):3756–3767. https://doi.org/10.1021/ct200251x
32. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79:926–935
33. Nose S (1984) A unified formulation of the constant temperature molecular dynamics methods. J Chem Phys (81): 511
34. Martyna GJ, Tobias DJ, Klein ML (1994) Constant pressure molecular dynamics algorithms. J Chem Phys 101:4177–4189
35. Tuckerman M, Berne BJ, Martyna GJ (1992) Reversible multiple time scale molecular dynamics. J Chem Phys 97:1990–2001
36. Liu CJ, Deavours BE, Richard SB, Ferrer JL, Blount JW, Huhman D, Dixon RA, Noel JP (2006) Structural basis for dual functionality of isoflavonoid O-methyltransferases in the evolution of plant defense responses. Plant Cell 18(12):3656–3669. https://doi.org/10.1105/tpc.106.041376
37. DeVore NM, Scott EE (2012) Structures of cytochrome P450 17A1 with prostate cancer drugs abiraterone and TOK-001. Nature 482(7383):116–119. https://doi.org/10.1038/nature10743
38. Singh S, McCoy JG, Zhang C, Bingman CA, Phillips Jr GN, Thorson JS (2008) Structure and mechanism of the rebeccamycin sugar 4′-O-methyltransferase RebM. J Biol Chem 283(33):22628–22636. https://doi.org/10.1074/jbc.M800503200
39. Wilmouth RC, Turnbull JJ, Welford RW, Clifton IJ, Prescott AG, Schofield CJ (2002) Structure and mechanism of anthocyanidin synthase from Arabidopsis Thaliana. Structure 10(1):93–103
40. De Carolis E, De Luca V (1993) Purification, characterization, and kinetic analysis of a 2-oxoglutarate-dependent dioxygenase involved in vindoline biosynthesis from Catharanthusroseus. J BiolChem 268(8):5504–5511
41. Ma X, Koepke J, Panjikar S, Fritzsch G, Stockigt J (2005) Crystal structure of vinorine synthase, the first representative of the BAHD superfamily. J Biol Chem 280(14):13576–13583. https://doi.org/10.1074/jbc.M414508200
42. Gigant B, Wang C, Ravelli RB, Roussi F, Steinmetz MO, Curmi PA, Sobel A, Knossow M (2005) Structural basis for the regulation of tubulin by vinblastine. Nature 435(7041):519–522. https://doi.org/10.1038/nature03566