

# In Silico Structural Modeling and Binding Site Analysis of Cerebroside Sulfotransferase (CST): A Therapeutic Target for Developing Substrate Reduction Therapy for Metachromatic Leukodystrophy

Nivedita Singh\* and Anil Kumar Singh



Cite This: *ACS Omega* 2024, 9, 10748–10768



Read Online

ACCESS |



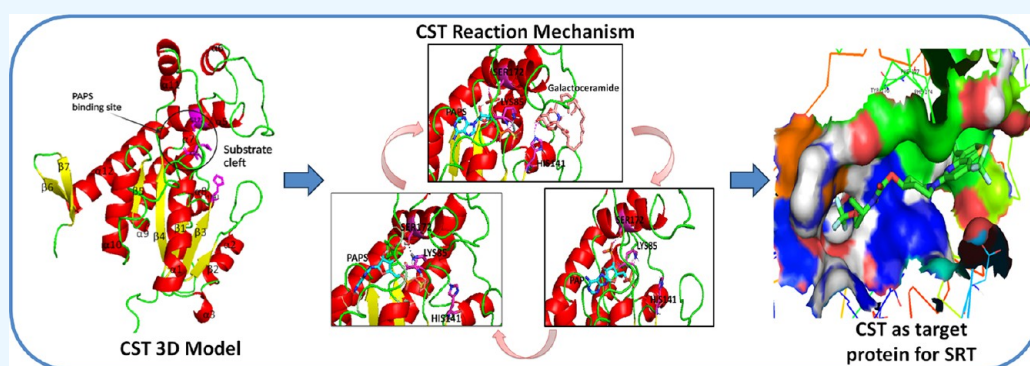
Metrics & More



Article Recommendations



Supporting Information



**ABSTRACT:** Cerebroside sulfotransferase (CST) is emerging as an important therapeutic target to develop substrate reduction therapy (SRT) for metachromatic leukodystrophy (MLD), a rare neurodegenerative lysosomal storage disorder. MLD develops with progressive impairment and destruction of the myelin sheath as a result of accumulation of sulfatide around the nerve cells in the absence of its recycling mechanism with deficiency of arylsulfatase A (ARSA). Sulfatide is the product of the catalytic action of cerebroside sulfotransferase (CST), which needs to be regulated under pathophysiological conditions by inhibitor development. To carry out in silico-based preliminary drug screening or for designing new drug candidates, a high-quality three-dimensional (3D) structure is needed in the absence of an experimentally derived three-dimensional crystal structure. In this study, a 3D model of the protein was developed using a primary sequence with the SWISS-MODEL server by applying the top four GMEQ score-based templates belonging to the sulfotransferase family as a reference. The 3D model of CST highlights the features of the protein responsible for its catalytic action. The CST model comprises five  $\beta$ -strands, which are flanked by ten  $\alpha$ -helices from both sides as well as form the upside cover of the catalytic pocket of CST. CST has two catalytic regions: PAPS (-sulfo donor) binding and galactosylceramide (-sulfo acceptor) binding. The catalytic action of CST was proposed via molecular docking and molecular dynamic (MD) simulation with PAPS, galactosylceramide (GC), PAPS-galactosylceramide, and PAP. The stability of the model and its catalytic action were confirmed using molecular dynamic simulation-based trajectory analysis. CST response against the inhibition potential of the experimentally reported competitive inhibitor of CST was confirmed via molecular docking and molecular dynamics simulation, which suggested the suitability of the CST model for future drug discovery to strengthen substrate reduction therapy for MLD.

## 1. INTRODUCTION

Metachromatic leukodystrophy (MLD) is a rare lysosomal genetic disorder characterized by demyelination of the central and peripheral nervous system, leading to developmental delay, motor dysfunction, loss of communication, and, in severe cases, death.<sup>1,2</sup> Globally, one in 40,000–100,000 population is affected by MLD.<sup>3,4</sup> The underlying cause of MLD is deficiency or dysfunctionality of the arylsulfatase A (ARSA) enzyme, which leads to the accumulation of its substrate, sulfatides.<sup>3,5–8</sup> Sulfatides are sulfated glycolipids that are present in nerve cells and contribute to maintaining the integrity of the myelin

sheath, which acts as an insulator of neurons for the unhindered transduction of neural signals.<sup>4,9,10</sup> Accumulation of sulfatide in oligodendrocyte and Schwann cells results in progressive

**Received:** December 3, 2023

**Revised:** January 26, 2024

**Accepted:** January 31, 2024

**Published:** February 23, 2024



demyelination and subsequent neurological dysfunction with degradation of normal physiological transmission of the electrical impulse between nerve cells and axonal loss in CNS and PNS.<sup>3,7</sup>

The availability of reliable therapeutic options is a major challenge for lysosomal storage disorders including MLD. Currently used therapies including enzyme replacement therapy, gene therapy, chaperone therapy, etc. are under different stages of clinical trials in the case of MLD. Hematopoietic stem cell transplant and enzyme replacement therapy are approved therapies for MLD; however, their success varies on a case-to-case basis primarily because of the uncertainty of the disease variant, mutation types, and the stage of disease at the time of transplant.<sup>4,6,11–16</sup> Above all, the high cost associated with these treatment options deters a larger population from taking advantage of them. Substrate reduction therapy (SRT), which has been shown to be effective in other lysosomal storage disorders, may also be a better treatment strategy in MLD.<sup>8,17,18</sup>

SRT is a promising oral medication technique in which the inhibition of the protein responsible for the biosynthesis of the substrate of the dysfunctional protein is targeted so that the pathophysiological condition arising due to the accumulation of the substrate can be countered.<sup>8,19,20</sup> In MLD, cerebroside sulfotransferase (CST; EC 2.8.2.11) is a target enzyme for the development of SRT to regulate sulfatide accumulation.<sup>21</sup> CST is the final enzyme in the sulfatide biosynthesis process. CST is a Golgi membrane protein that catalyzes the transfer of the sulfuryl group from the cosubstrate (3'-phosphoadenosine-5'-phosphosulfate (PAPS)) to the substrate (galactocerebroside or galactosylceramide (GC)), yielding galactocerebroside sulfate and adenosine-3'-5'-biphosphate (PAP) as products. The *cst* gene with a size of roughly 22 kb is located on chromosome 22 (22q12.2) and comprises eight exons with six noncoding and two coding exons. The CST cDNA with 1791 bp encodes a 423-amino acid protein with a type II transmembrane topology and two conserved N-linked glycosylation sites at ASN66 and ASN312, which may have a crucial role in catalytic activities.<sup>4,22</sup> Despite the availability of sufficient CST sequence information, the critical and most challenging part of CST-based SRT research is the unavailability of a three-dimensional structure of a protein, which is essential to decipher its functional mechanism for drug discovery.

With computational advancement, a huge pull of data is available related to natural as well as synthetic bioactive molecules, which gives strength to the preliminary-stage drug discovery process with *in silico*-based drug screening.<sup>23–27</sup> The three-dimensional (3D) structure of a protein provides important information about protein molecular arrangement, which becomes the basis for inhibitor development.<sup>4,28</sup> So far, the 3D structure of CST through an experimental crystallization technique has not been generated. This information gap hinders the development of inhibitors for establishing a therapeutic potential of SRT in MLD. Thus, dependency on computational modeling can be a way forward. The major challenge in computational modeling is the availability of templates and the accurate prediction of model functionality. Lack of similarity among sulfotransferases increases these complications. Thus, a conserved region that facilitates the catalytic action of sulfotransferases (STs) can be used for reliable homology modeling as SRT is an effort toward the development of competitive inhibitors which can regulate the catalytic function of the target protein. This study aims to generate a three-

dimensional computational model for CST for inhibitor development to reduce the load of sulfatide accumulation under inappropriate recycling and balance the myelin sheath in order to treat metachromatic leukodystrophy.

## 2. MATERIALS AND METHODS

**2.1. Hardware and Software.** The entire computational study was performed in the High Performance Super Computing facility, Param Shivay of 837 TFLOPS capacity with Intel(R) Xeon(R) Gold 6148 CPU @ 2.40 GHz and 40 CPU per node. In the present study, we utilized software including AutoDock MGL Tool 1.5.7, GROMACS 2023, PyMOL, Discovery studio, Chimera 1.17, visual molecular dynamics (VMD), origin2023b, GraphPad Prism 8.0, Adobe PhotoShop 6.0, etc. and web tools including SWISS-MODEL, PubChem, SOPMA, SAVES v6.0, NCBI, ExPASy ProtParam, PDB-Sum, ProQ, ProSA-web, ModLoop, LigPlot, DoGSiteScorer, and CastP.

**2.2. CST Protein Sequence Retrieval and Sequence Alignment.** There is no crystal structure available for the protein cerebroside sulfotransferase (CST), although its sequence is available after cloning and characterization.<sup>29</sup> To perform computational modeling, the CST sequence from *Homo sapiens* was retrieved in the FASTA format from the UniProt (<https://www.uniprot.org/>) database. PSI-BLAST (position-specific iterated-basic local alignment search tool) was performed using the query sequence to generate alignment between the extracted CST protein sequence (as a “query”) and the protein sequence within the database (as “subject” sequences). PSI-BLAST is a method for identifying the best possible homologue of a protein based on the *E*-value (Expect value). A total of 36 protein sequences were obtained. Multiple sequence alignment (MSA) was then performed with Clustal Omega, and subsequently, phylogenetic analysis was carried out. The ExPASy ProtParam tool was used to analyze the physical and chemical characteristics by computation of molecular weight, theoretical PI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index, and grand average of hydropathicity (GRAVY) of the CST protein. Based on the sequence similarity search, the secondary structure was predicted. PSIPREP and SOPMA protein structure servers were employed to get information about the secondary structure of the CST protein. PSIPREP employs two feed-forward neural networks with the help of PSI-BLAST output.

**2.3. Three-Dimensional Modeling with Model Refinement and Validation.** The structural information on CST so far revealed that CST exists in living systems in the homodimeric form with an independent catalytic domain in each monomer. Each monomer contains three domains: N-terminal CST domain I (1–63 amino acid residues), which is a transmembrane domain; the catalytic domain (71–324 amino acid residues); and the C-terminal domain (325–423).<sup>4</sup> As a comparative modeling prediction strategy, we used the SWISS-MODEL server to construct the 3D structure of a protein of interest from *H. sapiens* with three different X-ray crystal structures as the structural templates. Multiple templates were used to avoid the template structure dependency, as well as to consider the probable conformational change of CST. Based on their GMQE values, the maltose-binding protein heparan sulfate 2-*O*-sulfotransferase from *Gallus gallus* (PDB ID: 3f5f.1 chain A), the maltose-binding periplasmic protein heparan sulfate 2-*O*-sulfotransferase 1 with bound heptasaccharide and

PAP (PDB ID: 4ndz.1 chain A), and the maltose-binding protein heparan sulfate 6-*O*-sulfotransferase isoform 3 fusion protein (PDB ID: 5t03.1. chain A) were chosen as templates.

The SAVES v6.0 (Structure Validation Server: <https://saves.mbi.ucla.edu/>) web server was employed as a common platform for quality checks, error assessment, and model suitability for further research. For some portions of the protein structure with low sequence similarity, “threading” or “ab initio” was applied. Modloop was used to refine the loop region. The model was validated through PROCHECK stereochemical assessment, the ERRAT quality factor, the Ramachandran plot, Verify3D, and ProQ analysis. In the Ramachandran plot, plotting of dihedral angles  $\varphi$  against  $\psi$  was used to predict the possible conformational changes of amino acids in the protein structure. Prosa-web (<https://prosa.services.came.sbg.ac.at/prosa.php>) was employed to recognize the error in the 3D structure through the z-score.

**2.4. Structure Analysis.** It is useful to investigate the molecular surface to understand the function of proteins because specific intermolecular interactions such as electrostatic and hydrophobic interactions are important in molecular recognition. Molecular surfaces of the protein data bank-registered protein structures are available in the eF-site database, which calculates molecular surfaces by the MSP program and electrostatic potentials by solving the Poisson–Boltzmann equations using the SCB program. As the molecular surfaces and electrostatic potentials for the template proteins were obtained from the eF-site database, for the newly constructed models, the eF-surf web server was used to calculate the same. eF-surf calculates the molecular surface for the uploaded PDB structures in the same way as the eF-site database does for the RCSB-submitted protein pdb file. Further, the eF-seek server was used to determine the molecular function of the CST protein. eF-seek is a web server to search for similar ligand binding sites for the newly constructed model based on the clique search algorithm.

**2.5. Binding Site Pocket Analysis.** Identifying the binding pocket or active site of the protein is important for understanding the molecular and biological functions of proteins. To analyze the favorable binding site, a sequence alignment tool in PyMol was used. For this, the reference template (3f5f) bound with the PAP ligand was used. Interactions between the protein and the substrate and cosubstrate were plotted using AutoDock and visualized using PyMol. Binding site validation was also carried out using web tools, including LigPlot, DoGSiteScorer, and CastP.

**2.6. Molecular Docking Simulation.** Molecular docking was used to determine the mechanism of action of the CST protein based on the interaction study with PAPS, PAP, and/or GC with the model CST 3D structure. For ligand docking, AutoDock 4.2.6 was used. A grid box of  $90 \times 90 \times 90$  was created in  $x$ ,  $y$ , and  $z$  coordinates with a spacing of  $0.253 \text{ \AA}$  and centered based on the backbone atoms of the key identified active site residues in the substrate site and/or cosubstrate sites. In the AutoDock MGL tool, protein.pdbqt, ligand.pdbqt, grid.gpf, and docking.dpf files were prepared. Further, a Gasteiger charge was assigned to the prepared protein structure file in the PDBQT format. Subsequently, docking was performed in a Linux system with docking parameters including 100 GA run, 300 population size, 27,000 maximum number of generations, and 25,000,000 maximum number of evaluations by applying the Lamarckian genetic algorithm and the gradient-based local search method. From the output docking.dlg file, the

best docked conformation of the ligand with the lowest binding energy was selected, and the corresponding protein–ligand complex was generated using custom python scripts and pdb-tools (PyMol, Discovery Studio). Thereafter, we analyzed the protein–ligand docked complexes and identified the substrate and PAPS binding site residues and the noncovalent interactions between protein residues and the substrate. The high-throughput molecular docking was carried out in the HPC Linux system with the help of scripts for generating the grid.glg file and the docking.dlg file.

**2.7. Molecular Dynamics Simulations.** Molecular dynamics simulations were carried out to optimize and establish the thermal stability of the protein constructed by homology modeling as well as to study the stability of catalytic complexes.<sup>30</sup> GROMACS 2020 version was used with the Charmm 27 force field. As the prepared structure functionally falls under the globular protein category, it was placed at the center of a dodecahedron box with a minimum distance of  $1.2 \text{ \AA}$  from the box edge to apply periodic boundary conditions (PBCs) and minimize the edge effect. The TIP3P water model was used to solvate the system, and thereafter, the system was neutralized by the addition of chloride ( $\text{Cl}^-$ ) ions. Then, the system was energy-minimized using the steepest descent algorithm, with the energy minimization tolerance set at  $100 \text{ kJ mol}^{-1} \text{ nm}^{-1}$ . The system was then subjected to a 1000 ps (1 ns) NVT equilibration simulation with a 2 fs time step and the temperature set at 300 K. Subsequently, the system was subjected to a 1000 ps (1 ns) NPT simulation with a 2 fs time step to equilibrate the pressure of the system to 1 bar. The bond lengths were constrained using the LINCS algorithm. Thereafter, a final equilibration MD simulation was performed for 100 ns with a 2 fs time step after removing the position restraint and placing a distance restraint on the  $\text{Cl}^-$  ions. Additionally, for studying the stability of the three-dimensional structure of the protein, MD simulation was carried out at different time scales of 100, 300, and 500 ns. All of the MD run was performed in replications. Finally, the periodic boundary condition was used to allow the solvent to be integrated by applying the leapfrog algorithm with solving the surface effect issue.

**2.8. Trajectory Analysis.** Using the MD simulation data set, trajectory analysis was done with parameters such as root-mean-square deviation (RMSD), root-mean-square fluctuation (RMSF), radius of gyration ( $R_g$ ), hydrogen bond (hb), potential energy, and solvent-accessible surface area (SASA).

RMSD measures the structural change of CST at  $t$  simulation time concerning the structure at  $t = 0$ . In eq 1,  $x_i^{\text{ref}}$  is the coordinate of the reference structure,  $x_i(t)$  is the coordinate of the protein structure at time  $t$ , and  $N$  is the number of atoms in the protein.

$$\text{RMSD}(t) = \sqrt{\frac{1}{N} \sum_i^N \|x_i(t) - x_i^{\text{ref}}\|^2} \quad (1)$$

RMSF reads the fluctuation in the protein residue and also helps us understand the region in the protein with different degrees of fluctuation. RMSF measures the deviation in a particular residue with respect to the reference over time during the simulation.<sup>31</sup> In RMSF in eq 2,  $T$  is the time over which fluctuation needs to be measured.  $x_i$  is the coordinate of particle  $i$ .  $\bar{x}$  is the reference position of particle  $i$ .

>NP\_001305033.1 galactosylceramide sulfotransferase isoform 1 [Homo sapiens]

MLPPQKKPWESMAKGLVLGALFTSFLLLVYSYAVPPLHAGLASTTPEAAASCSPPALEPEAVIRANGSAG  
ECQPRRNI VFLKTHKTASSTLLNILFRFGQKHRLKFAFPNGRNDFDYPTFFARSLVQDYRPGACFNIICN  
HMRFH YDEVRGLVPTNAIFITVLRDPARLFESSFHYFGPVVPLTWKLSAGDKLTEFLQDPDRYDYPNGFN  
AHLRNL LFFDLGYDNSLDPSSPQVQEHILEVERRFHLVLLQEYFDES LVLKDLLCWELEDVLYFKLNA  
RRDSPVPRLSGELYGRATAWNMLDShLYRHFNASFWRKVEAFGRERMAREVAALRHANERMRTICIDGGH  
AVDAAAIQDEAMQPWQPLGTKSILGYNLKKSIGQRHAQLCRRMLTPEIQYLMDLGANLWVTKLWKFIRDF  
LRW

Figure 1. Amino acid sequence of the CST protein.

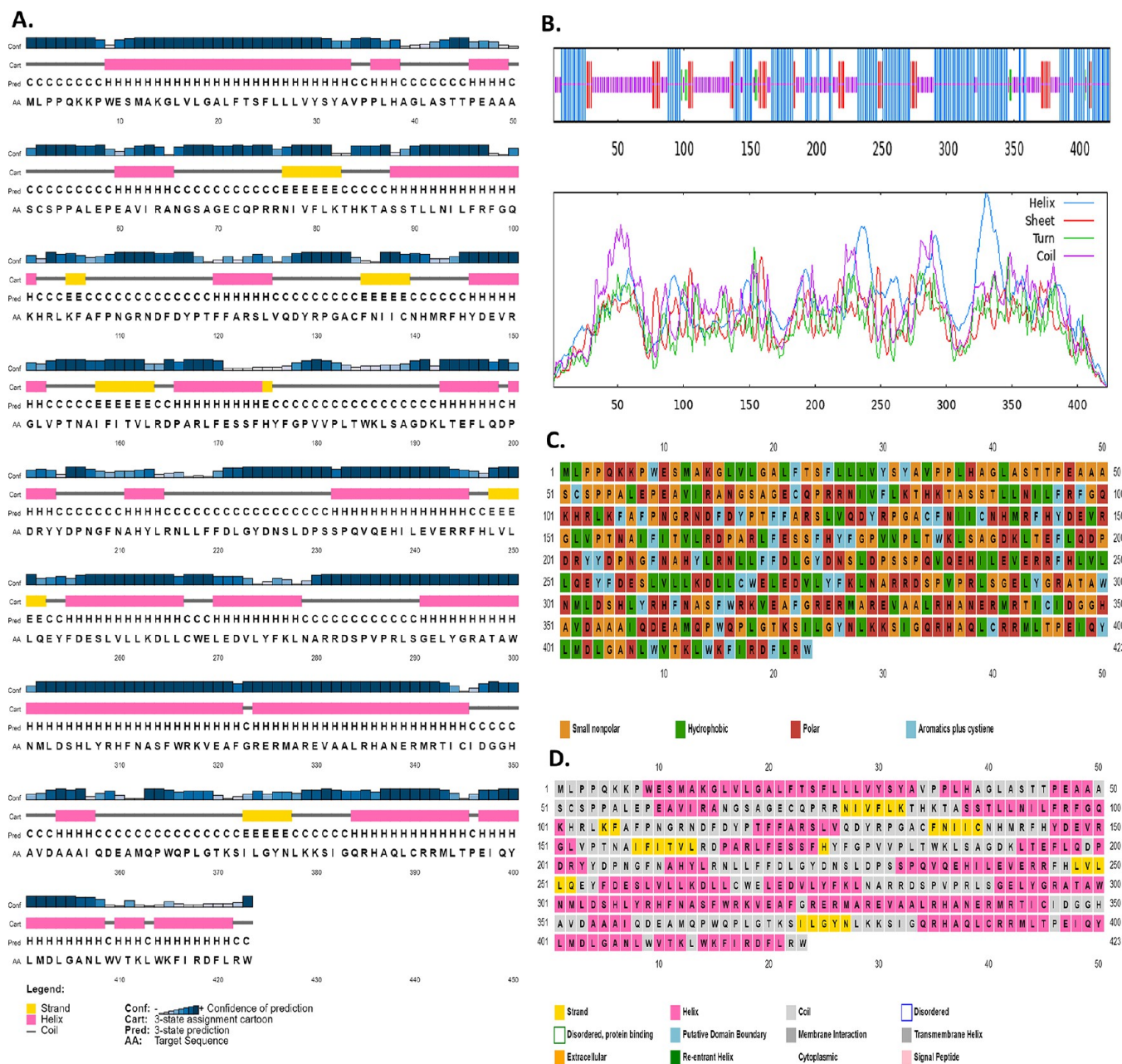


Figure 2. (A) Secondary structure of CST by PSIPRED. (B) Secondary structural types of the CST protein generated by SOPMA:  $\alpha$  helix (blue), extended strand (red),  $\beta$ -turn (light-green), and random coil (purple). (C) Amino acid type of CST protein. (D) Sequence plot of the secondary structure of CST in PSI-Pred.

**Table 1. Comparison of the Top 4 Ranked Models of the CST Protein Generated by the SWISS-MODEL Server and Stereochemical Analysis through the PROCHECK Web Server and SWISS-MODEL Assessment Indicators**

stereochemical analysis	model 1	model 2	model 4	model 3
overall quality factor	85.5469	85.2321	68.6992	68.595
template	3f5f.1.A	4ndz.1.A	5t03.2.A	5t03.1.A
QMEANDisCo Global	0.45 ± 0.05	0.44 ± 0.05	0.44 ± 0.05	0.44 ± 0.05
sequence identity	15.71%	16.03%	17.43%	17.43%
GMQE	0.31	0.31	0.33	0.33
QMEAN Z-score	−6.15	−6.50	−5.18	−5.50
residues in most favored regions [A, B, L] (%)	202 (84.2%)	203 (84.6%)	190 (83.7%)	196 (86.3%)
residues in additional allowed regions [a, b, 1, p] (%)	35 (14.6%)	31 (12.9%)	31 (13.7%)	26 (11.5%)
generously allowed regions [~a, ~b, ~1, ~p]	1 (0.4%)	3 (1.2%)	4 (1.8%)	3 (1.3%)
residues in disallowed regions [XX] (%)	2 (0.8%)	3 (1.2%)	2 (0.9%)	2 (0.9%)
number of non-glycine and non-proline residues (%)	240	240	227	227
numbers of end-residues (excl Gly and Pro)	2	2	2	2
number of glycine residues	12	11	11	11
number of proline residues	14	14	14	14
total no. of residues	268	267	254	254
all Ramachandrans	12 labeled residues (out of 266)	16 labeled residues (out of 265)	13 labeled residues (out of 252)	18 labeled residues (out of 252)
χ <sup>1</sup> -χ <sup>2</sup> plots	3 labeled residues (out of 183)	10 labeled residues (out of 183)	5 labeled residues (out of 175)	6 labeled residues (out of 177)
side chain params	5 better	5 better	5 better	5 better
residual properties	max.deviation: 8.3 bad contacts: 0 bond len/angle: 5.9	max.deviation: 19.7 bad contacts: 0 bond len/angle: 6.2	max.deviation: 6.3 bad contacts: 1 bond len/angle: 7.2	max.deviation: 4.5 bad contacts: 1 bond len/angle: 8.3
G-factors	dihedrals: −0.31 covalent: −0.12 overall: −0.22	dihedrals: −0.39 covalent: −0.16 overall: −0.28	dihedrals: −0.36 covalent: 0.21 overall: −0.28	dihedrals: −0.39 covalent: −0.36 overall: −0.35
planar groups	85.3% within limits 14.7% highlighted 7 off graph	82.2% within limits 17.8% highlighted 6 off graph	83.7% within limits 16.3% highlighted 11 off graph	79.7% within limits 20.3% highlighted 13 off graph

$$\text{RMSF} = \sqrt{\frac{1}{T} \sum_{i=1}^T (x_i - \bar{x})^2} \quad (2)$$

$R_g$  examines the compactness of the protein and changes in the state of complex formation.<sup>32</sup> In  $R_g$  in eq 3,  $r_i$  displays the coordinate of atom  $i$ ,  $r_{\text{center}}$  is the center of mass, and  $N$  is the number of protein atoms.

$$R_g = \sqrt{\frac{1}{N} \sum_{i=1}^N |r(i) - r_{\text{center}}|^2} \quad (3)$$

Protein SASA analysis was used to study the stability of the protein folding. SASA describes the area of a protein exposed to interact with solvent molecules.<sup>33–35</sup> A lower value of SASA indicates greater compactness.<sup>35</sup>

Principal component analysis (PCA), a multivariate statistical data reduction technique, was applied to examine the movement of the protein and the conformational subspace of the complex to understand the dynamic stable behavior of the protein.<sup>36–40</sup> PCA plots revealed different cluster formations. The PCA plot was drawn using the MD data obtained by the GROMACS MD run. We applied `gmx covar` and `gmx analogue` utilities of GROMACS to get the covariance matrix of the protein backbone ( $C\alpha$  atoms) and its diagonalization. The covariance matrix yields a set of eigenvalues and corresponding eigenvectors to determine the principal components. These eigenvectors dictate the nature of the transformation of the

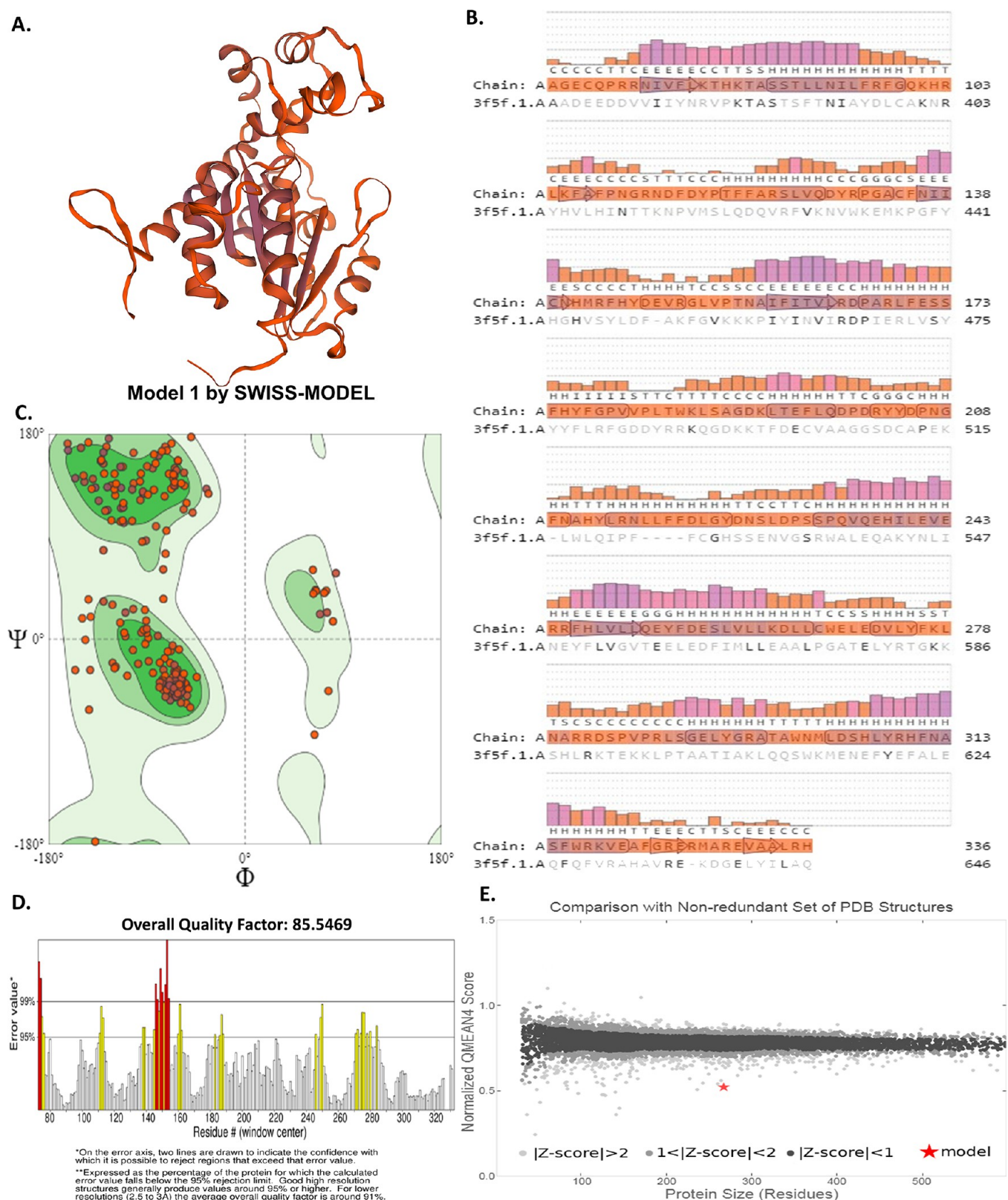
protein over the simulation time, and the eigenvalue determines the magnitude of transformation.<sup>41</sup> In eq 4 of covariance,  $N$  is the number of atoms in the protein and  $x$  and  $y$  are dimensions.

$$\text{cov}_{x,y} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{N - 1} \quad (4)$$

Based on PCA-led reaction coordinates, a free energy landscape (FEL) was derived to understand the energy distribution of the protein folding during the simulation. FEL signifies the stability of the apoprotein and protein in the form of the complex. Using the `gmx sham -f PC1PC2.xvg -ls FES.xpm`, the `xpm` file was generated, which then produced an FEL visualization file using the python command (`python2.7 xpm2txt.py -f FES.xpm -o free-energy-landscape.dat`). The Origin-Pro 2023b graphical representation tool was used to visualize the free energy landscape. The density distribution graph was obtained using the `densmap` Linux script (`gmx densmap -f md_noPBC.xtc -s md.tpr -aver z -bin 0.05-dmin 0`), followed by the `densmap` visualization script (`gmx xpm2ps -f densmap.xpm -o densmap.eps -rainbow blue`). Adobe Photoshop 6.0 was used for visualization of the `eps` file of the `densmap` for density distribution analysis.

### 3. RESULTS AND DISCUSSION

**3.1. Gene Selection, Sequence Retrieval, and Secondary Structure Analysis.** In this study, the goal of generating a CST computational model is to investigate the sulfatide-



reduction potential of drug candidates with the prospect of regenerating the myelin sheath. For this, the protein sequence was retrieved from the Uniport database, containing 423 amino

acid residues consisting of 1791 bp with a molecular mass of 48764.15 Da.<sup>42</sup> The protein sequence was then validated with the sequence deposited in the NCBI database under Accession

no. NP\_001305033.1, gene ID “9514”, and gene identifier GI: 970259838. The protein is reported as a type II transmembrane topology and has two potential N-linked glycosylation sites at ASN66 and ASN312 for full enzyme activity.<sup>22,43</sup> The detailed sequence of the CST protein is provided in Figure 1.

To understand the structural composition and arrangement of CST proteins, a deep analysis of primary, secondary, and tertiary structures is imperative. Accuracy in secondary structure production is particularly crucial in generating an appropriate 3D model of the target protein as well as in obtaining sufficient information about the functionality of the protein. For analyzing the secondary structure of the CST protein, the self-optimized prediction method with alignment (SOPMA) program was used.<sup>44</sup> Based on secondary structures of 64 aligned proteins, the SOPMA program revealed that the protein comprised  $\alpha$  helices (193, 45.63%), extended strands (48, 11.35%),  $\beta$ -turns (83, 1.89%), and random coils (174, 41.13%), while the Pi helix,  $3_{10}$  helix,  $\beta$  bridge, bend region, and ambiguous states were almost absent. PSIPRED 3.2 divulged that the CST protein comprised polar, nonpolar, aromatic, and hydrophobic amino acid residues in its structure. The details of the secondary structure of CST proteins are listed in Figure 2.

**3.2. Three-Dimensional Model Building Using SWISS-MODEL.** In the absence of experimentally determined protein 3D structures, computational modeling plays a cost-effective role in structure-based drug discovery.<sup>45–48</sup> In this study, for 3D model generation, a fully automated protein structure homology modeling server SWISS-MODEL was used. One of the advantages of SWISS-MODEL over other modeling software is that it follows the proper natural assembly of the template protein, expels all anticorrosive amino residues, and further adjusts the target-template arrangement in DeepView program.<sup>49,50</sup> One important feature of SWISS-MODEL is ProMod3, a central script-based platform for increasing the accuracy of the produced model as well as for quality estimation based on QMEANDisCo Global values. For this study, using the FASTA sequence of CST, the SWISS-MODEL server discovered 91 templates from its library by the HHblits method, which is profiled by hidden Markov models (HMMs), the fastest sequence search tool. The predicted protein was selected based on its GMQE score and identity score. The higher numbers of QMEAN Z-score and GMQE scores affirmed the higher reliability of the generated models. Among them, the four most favorable templates selected were 3f5f.1.A, 4ndz.1.A, 5t03.1.A, and 5t03.2.A using global model quality estimation (GMQE), a method of quality estimation that calculates both the target–template alignment and the template structure with a score in the range of 0–1 based mainly on coverage percentage. Further, based on the four templates, four models were generated by the SWISS-MODEL server. The detailed comparison is given in Table 1.

Based on comparative data with overall quality factors, model 1 was selected, which used template 3f5f.1.A, a *H. sapiens* maltose-binding periplasmic protein, heparan sulfate 2-O-sulfotransferase 1. Modeling of the CST protein based on the predicted template structure followed the thread method close to the 3D structure of a standard protein. The top model obtained from SWISS-MODEL is depicted in Figure 3A. SWISS-MODEL generated a 3D model for amino acid residues between 69 and 336 out of a total of 423 amino acids in the CST FASTA sequence. The sequence alignment of the amino acids used for CST model development with the corresponding reference sequence in the template showed 15.71% sequence

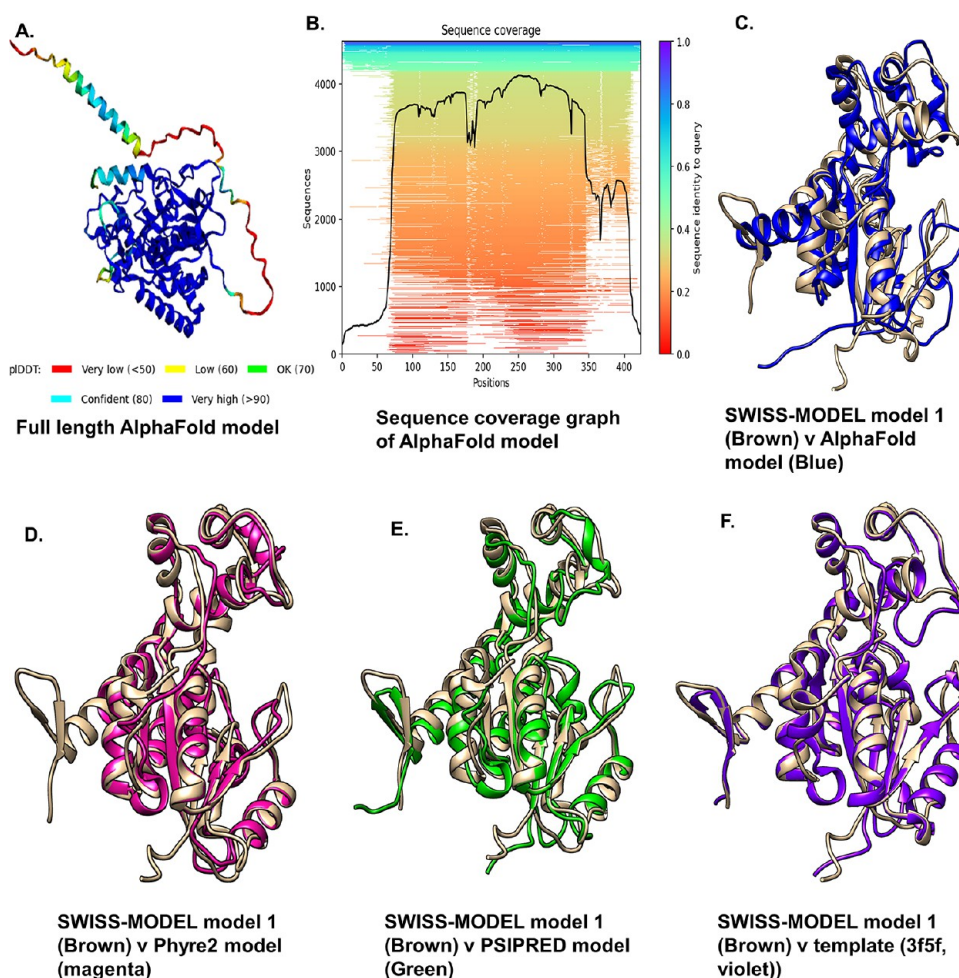
identity as illustrated in Figure 3B. The superimposition of the model protein over the template was checked through PyMol and UCSF Chimera, which was seemingly favorable. The missing residues at the N-terminal (1–68) and C-terminal (337–423) do not impact the catalytic activity of the protein as the N-terminal amino acid (1–68) is involved in the formation of the transmembrane domain with sufficient information for Golgi localization,<sup>51</sup> while the missing residues at the C-terminal amino acid (337–423) are mainly involved in dimerization. The sequence of amino acid residues comprising a binding pocket or cavity site defines its physicochemical properties, which, along with its form and position in a protein, describe its functionality. The characterization of the model generated by SWISS-MODEL is depicted in Figure 3C–E. The model was further subjected to comparison with the top models generated by other model builders.

### 3.3. Comparison of SWISS-MODEL-Predicted CST Model 1 with Other Models.

**Table 2. Comparative Analysis of Top Models from Different Model Builders**

properties	SWISS-MODEL	AlphaFold 2.0	Phyre2	PSIPRED/Domserf model
reference template	3f5f.A	Not mentioned	3f5f.A	3f5f.A
coverage of amino acids	69–336	1–423	78–315	71–323
ERRAT quality score	85.54	85.81	30.8696	45.3061
verify3D	79.48%	74.94%	73.11%	69.17%
Ramachandran plot				
most favorable region	84.2%	86.1%	82.2%	86.3%
additionally allowed region	14.6%	9.1%	15.0%	12.8%
generously allowed regions (%)	0.4%	3.8%	1.4%	0.4%
disallowed regions (%)	0.8%	1.1%	1.4%	0.4%

based predicted model 1 was compared with models generated by other model builders including AlfaFold 2.0, Phyre2.0, and PSIPRED 3.2. Table 2 presents a comparative analysis of models from different model builders to select the best predicted model for future studies. AlphaFold 2.0 generated five models with full length of the CST protein with amino acid residues 1–423, and ranking was based on the IDDT positioning of amino acids.<sup>52</sup> The top-ranked model was selected for comparison with the model from SWISS-MODEL (Figure 4A). In the AlfaFold sequence coverage graph, the sequence coverage lies mainly between 65 and 350 as shown in Figure 4B. Thus, the N-terminal end and C-terminal ends of the full-length AlphaFold model were not considered for structural comparison as these regions were also missing in the SWISS-MODEL, the Phyre2 model, and the PSIPRED model. Interestingly, the heparan sulfate 2-*o*-sulfotransferase protein PDB structure (code: 3f5f) was used as a common template in SWISS-MODEL-derived model 1, the Phyre2 model, and the PSIPRED model, while in the AI-based AlphaFold model, although the template used for modeling was not mentioned, the structural similarity with the other three models suggests the usage of a common template for model building. In SAVE v6.0 overall quality analysis, the SWISS-MODEL server-predicted model 1 was found to be the



**Figure 4.** Comparative study with the AlphaFold 2.0 model. (A) Full-length AlphaFold model. (B) Sequence coverage graph in the AlphaFold model. (C) Alignment of the AlphaFold model (in blue) and the SWISS-MODEL model (in brown). (D) Alignment of SWISS-MODEL model 1 (in brown) with the Phyre2 model (in magenta). (E) Alignment of model 1 (in brown) with the PSIPRED model (in green). (F) Alignment of model 1 (in brown) and the template protein, 3f5f (in violet).

best model, while the AlphaFold model also performed well. Structural alignment of model 1 with the AlphaFold model in Chimera 1.17 showed a fair degree of similarity in the helix region, while the major divergences were found in the  $\beta$ -strands and loop regions. In the SWISS-MODEL-predicted model 1, five parallel  $\beta$ -strands were present, while in the AlphaFold model, only three  $\beta$ -strands present at the corresponding position were completely overlapped. Additionally, the PAPS binding site (discussed in detail in a later section) remained intact in both SWISS-MODEL and AlphaFold 2 models, suggesting that both models had adhered to catalytic functional integrity. The comparison of SWISS-MODEL with AlphaFold 2 is represented in Figure 4C. Surprisingly, despite having low quality scores, the models built by Phyre2 and PSIPRED showed better structural similarity with model 1 from SWISS-MODEL as shown in Figure 4(D,E). Structural comparison with the template revealed that SWISS-MODEL-predicted model 1 was better aligned over the template than the other models (Figure 4F). Based on comparative studies, we selected SWISS-MODEL-predicted model 1 for the present study.

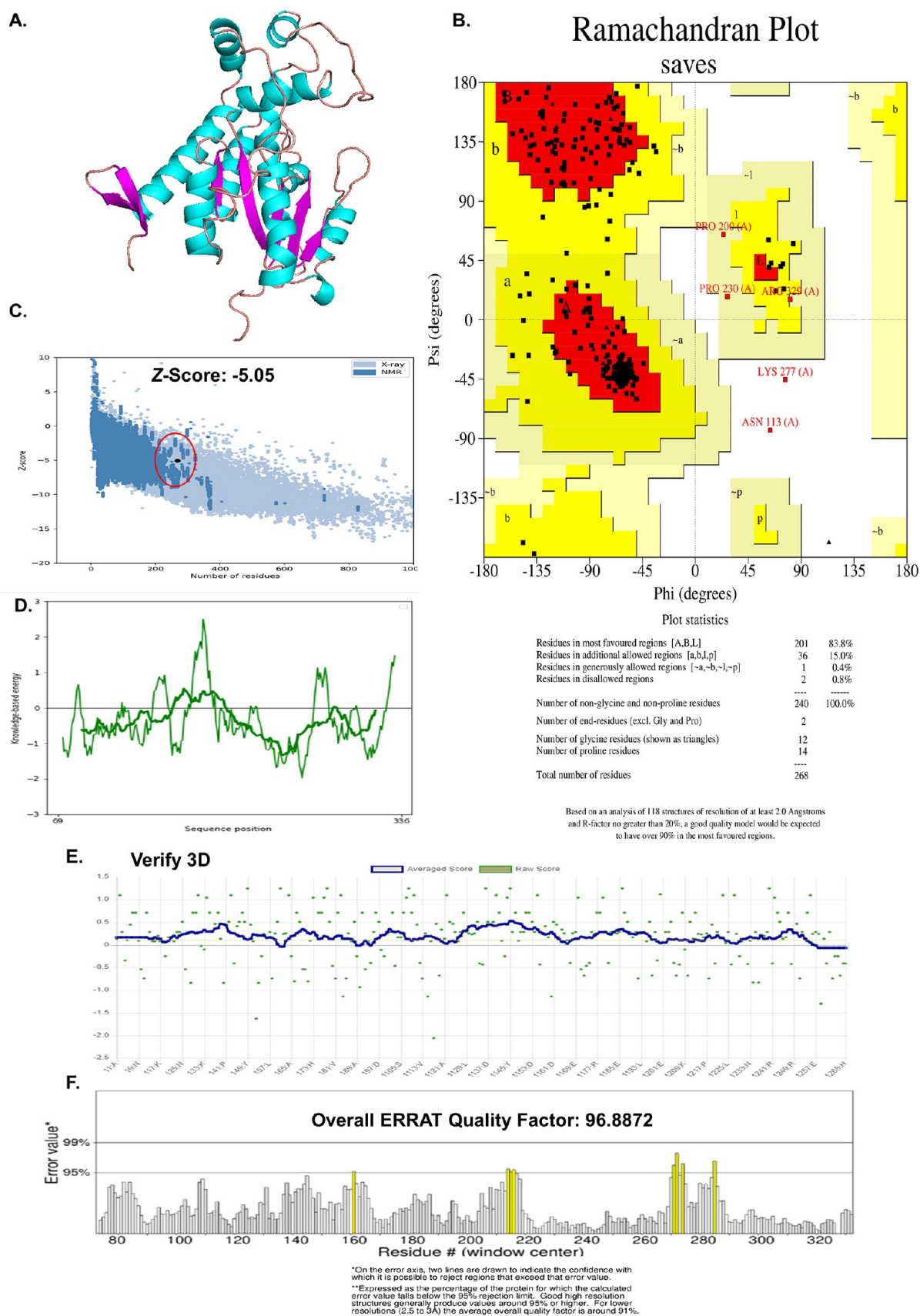
**3.4. Model Refinement and Validation of the Best Model.** Model refinement is an important section for generating a good and reliable model. The validation of model 1 from the SWISS-MODEL server was performed to identify the stereo-

chemical properties of amino acid residues of CST protein in allowed and disallowed regions. To check the quality of the top CST model obtained from SWISS-MODEL, ERRAT analysis was performed, which gave a quality score of 85.5469 with a possible rejected region between 145 and 155 amino acid residues that considerably spanned the loop region, as shown in Figure 3D. Therefore, the model was further subjected to refinement using the ModLoop web server. Subsequently, the validation of the refined CST model was completed by various online servers and web tools to ensure the accuracy and dependability of the predicted model. The structural validation of the CST refined model is depicted in Figure 5.

**3.4.1. Ramachandran Plot of the Refined 3D CST Model.** The Ramachandran plot of the CST model was generated using PROCHECK and MolProbity. The Ramachandran plot revealed that 99.2% of the residues were in the most favored, additionally allowed region, generously allowed region, and only 0.8% residues were in the disallowed region, which strongly support CST entry as a good model. Stereochemical analysis of the model revealed that the model was highly reliable in the distribution of  $\varphi$  and  $\psi$  angles. The Ramachandran plot and the related stereochemical data are depicted in Figure 5B.

**3.4.2. ProSA and ProQ Analyses.** Protein structure analysis (ProSA) is a web tool with a large data pool, and it was used for





**Figure 5.** Model quality development. (A) Final refined model of the CST protein. (B) Ramachandran plot for the final model. (C, D) ProSA z-score assessment and energy plot. (E) Verify3D graph of the model. (F) ERRAT quality assessment of the model.

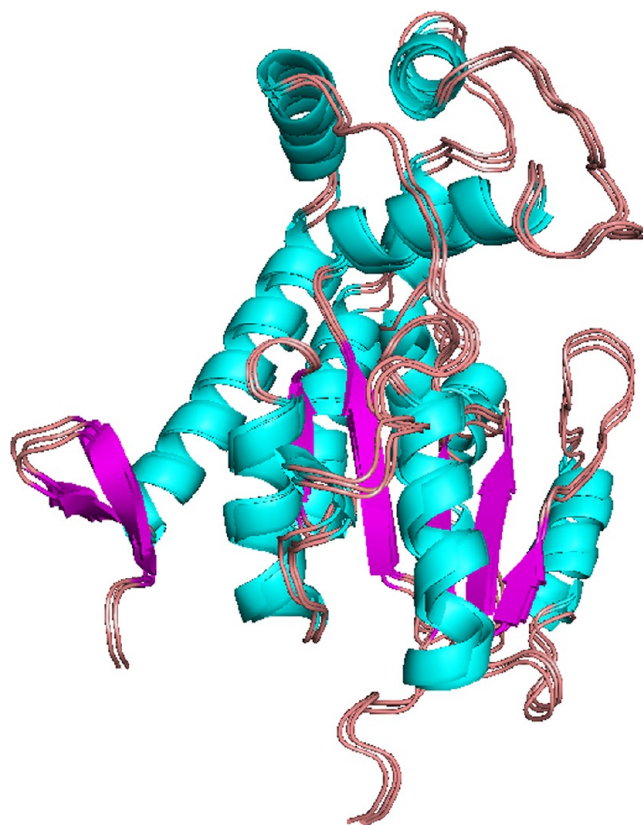
refinement and validation of the model protein by checking the 3D model of the protein structure for possible errors.<sup>53</sup> ProSA utilizes the atomic coordinates of the protein structure in a pdb format. Using a scoring function derived from empirical information based on the distances between two amino acid types in the PDB and following a Boltzmann distribution, ProSA calculated a z-score that indicates the overall quality of the model. The z-score of the model was  $-5.05$ , which falls within the range of all protein chains of relative sequence length in the Protein Data Bank determined by X-ray crystallography and nuclear magnetic resonance spectroscopy (Figure 5C,D). Further, the model was subjected to ProQ analysis, which is a neural network-based predictor for ensuring the quality of the model. It helped to understand the structural features of model proteins based on LGscores and MaxSub scores. The LG score of the final models was 6.488, which falls in the “extremely good model” category, while the predicted Max score was 0.635, which falls in the “very good model” category. Once a protein model is prepared, it is always essential to check the overall quality factor.

**3.4.3. SAVE v6.0 Quality Analysis.** For application-based in silico studies, it was necessary to ensure the accuracy of the 3D structure to determine the usefulness of our generated 3D model. Verify3D showed 83.21% of the residues with a score  $\geq 0.1$  in the 3D/1D profile (Figure 5E), which validated CST as a good-quality model because more than 80% of the residues scored  $\geq 0.1$  in the 3D/1D profile. The ERRAT quality factor of the refined model increased to 96.8872 (Figure 5F), which now corresponds to the expected value of a high-quality model. The model validation confirmed the “good quality” of the CST model and its suitability for future research.

**3.5. Molecular Dynamics Simulation Interpretation of the Newly Generated CST Model.** MD simulation was carried out to understand the dynamic behavior of each protein molecule in the presence of water and ions. We performed molecular dynamics simulation of the CST model at the time scales of 100, 300, and 500 ns to optimize the stability of the CST model protein. The model maintained its stability in different time spans as depicted in Figure 6. For this study, the Linux environment was used. Additionally, with 500 ns MD simulation data, we performed RMSD, RMSF,  $R_g$ , intramolecular hydrogen bond, SASA, FEL, PCA, and density distribution analysis to analyze the temporal behavior of atoms in the model.

**3.5.1. RMSD and RMSF.** The deviations were calculated on the CST model defined from the starting point of the simulation to the end frame through a consistent mapping of RMSD (root-mean-square deviation) and RMSF (root-mean-square fluctuation) plots with a time scale of 500 ns. The MD trajectory was analyzed by the RMSD (root-mean-square deviation) computation of  $C_\alpha$ . In a 500 ns MD simulation, the molecular system reached a steady state after 60 ns of early fluctuations as shown in Figure 7A. Between the time interval of 100 and 500 ns, there were no significant fluctuations observed in the collected data. The average RMSD value was 0.5 nm. The simulation result showed that the protein model underwent negligible structural changes during simulation, as depicted in Figures 6 and 7A.

The root-mean-square fluctuation (RMSF) of the model protein is presented in Figure 7B. RMSF characterizes the mobility of particular residues in the model structure. In the RMSF plot of the model, the peak centered on the residue LYS187 showed a high RMSF, imitating its position in the most flexible part of the loop spans between 175 and 190, which is

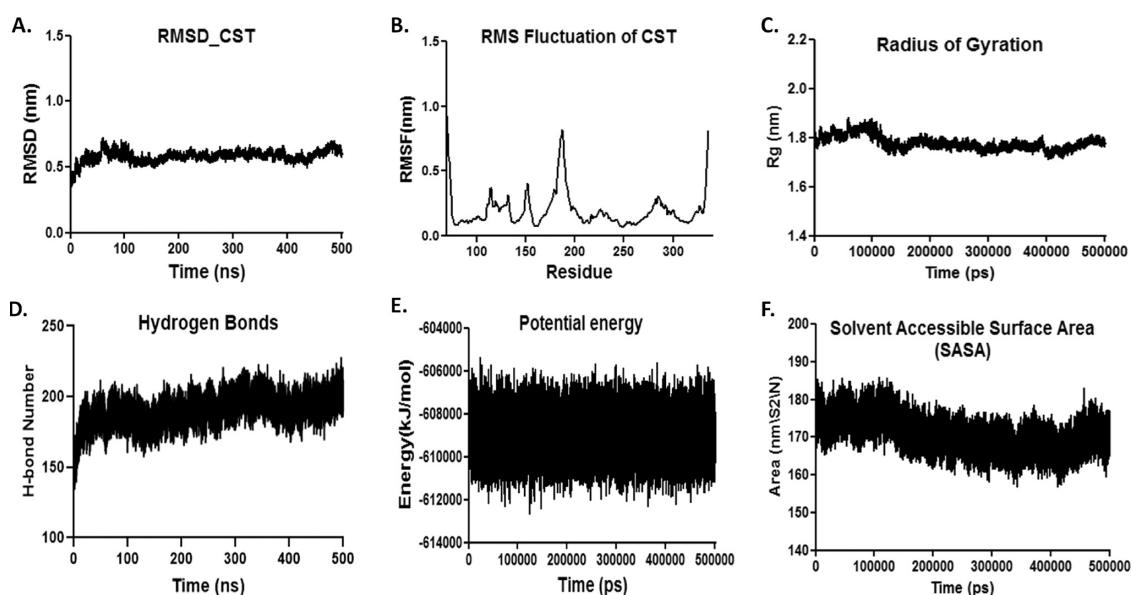


**Figure 6.** MD simulation and structural stability validation at 100, 300, and 500 ns time scales. The water molecule has been removed for a clearer view. Secondary structures are represented in different colors:  $\alpha$ -helix in cyan,  $\beta$ -strands in magenta, and loop region in tint.

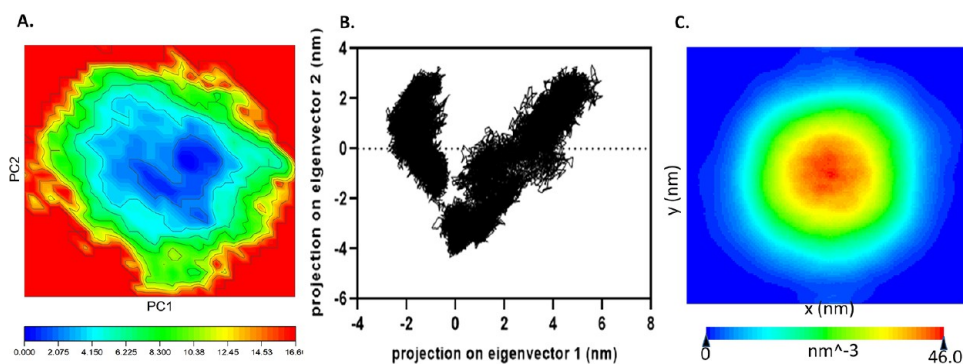
probably a significant loop region playing a critical role in the formation of active sites and interaction with the substrate/ligand. As expected, significant fluctuations were found between the CST model structure and the template structure as there is less identity between the two. Fluctuations at some residues may be relaxed due to relaxation in the force field-induced homology model.

**3.5.2.  $R_g$  SASA, and Hydrogen-Bond Analyses.** The radius of gyration ( $R_g$ ) analysis of the CST model showed the distribution of atoms of proteins around its axis.  $R_g$  represents the mass-weighted root-mean-square distance of atoms from their respective centers of mass.  $R_g$  is an indicator of 3D model compactness.<sup>54,55</sup> The calculated average  $R_g$  for the CST protein was found to be nominal with the value of 1.8 nm throughout the simulation time, strongly supporting the stability and compactness of the model structure (Figure 7C).

Solvent-accessible surface area (SASA) analysis measured the surface area of the CST model protein that could be directly accessed by solvent molecules since the surface area exposed to the solvent has a tendency to self-fix. SASA analysis is related to the structure–function relationship of the protein and its residues.<sup>56</sup> SASA was helpful in understanding the ligand binding potential of the CST model protein. Under SASA analysis, a hydrogen-bonding network between the side chain of surface amino acid residues and solvent molecules was calculated. As shown in Figure 7F, the SASA values of the CST model protein were in the small range of 160–180 nm<sup>2</sup> in a time frame of 500 ns, suggesting the structural integrity of the protein throughout the simulation period.



**Figure 7.** Optimization of the homology model by MD simulations with a time frame of 500 ns. (A) Root-mean-square deviation (RMSD) values of the CST protein. (B) Root-mean-square fluctuation (RMSF) values of particular atoms in backbones. (C) Radius of gyration. (D) Number of hydrogen-bond formations at the time scale of 500 ns. (E) Potential energy occupied by the CST model. (F) Solvent-accessible surface area (SASA) of the CST model in 500 ns.



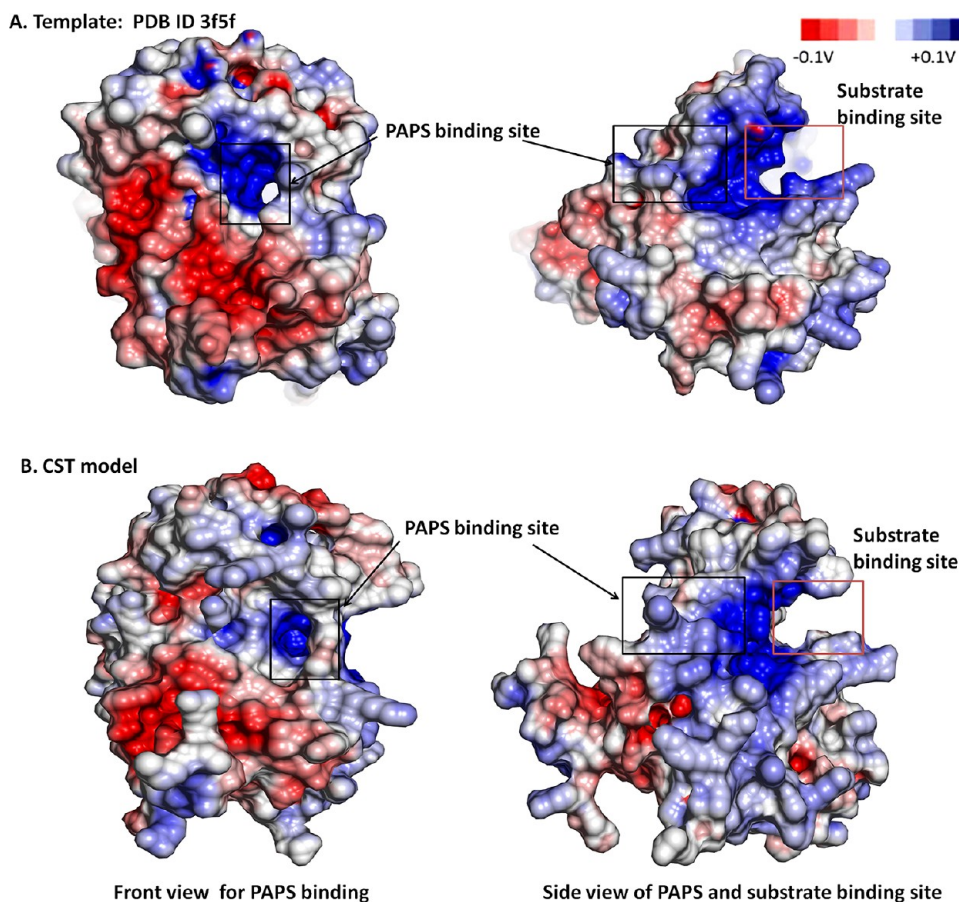
**Figure 8.** (A) Free energy landscape represents the folding behavior and direction of motion of the CST model protein, blue spots represent the energy minima and the most favorable conformations, and yellow spots represent unfavorable conformations. (B) Dynamic energy fluctuation plot between two projections on eigenvectors 1 and 2 generated for the modeled structure. (C) Density distribution analysis plot of the CST model to understand the atomic orientation using the densmap script.

The stability of the model protein was further determined by analyzing in detail the hydrogen-bond formation among atoms of the macromolecule. Being a weak bond, hydrogen bonding ensures the maximum flexibility of the protein to get into a possible stable conformation. The intramolecular hydrogen-bond calculation was performed for the CST model protein, which comprised 4423 atoms with 409 donors and 788 acceptors. The average number of H-bonds in the model protein per time frame of 500 ns was 190.267 out of 161146 possible. The stable numbers of intramolecular hydrogen bonds in Figure 7D showed no significant alteration and divulged strong contacts between the residues in the protein.

The total energy consumption was  $-493,311$  kJ/mol for the CST model protein with a total drift of  $-144,387$  kJ/mol over the 500 ns time scale. The average kinetic energy of the model protein was  $115,642$  kJ/mol with a total drift of  $1.37191$  (kJ/mol), revealing the achievement of the equilibrium stage. The conserved energy of the CST model was  $679,636$  kJ/mol, while the potential energy was  $-608,953$  kJ/mol with a total drift of

$-145,759$  kJ/mol. Figure 7E highlights the stability of the homology model through analysis of different MD parameters.

**3.5.3. Free Energy Landscape (FEL), Principal Component Analysis (PCA), and Density Distribution Analysis of CST Model.** To analyze the accurate folding behavior of the model protein by understanding the availability of the conformational landscape to the CST model, free energy landscape (FEL) analysis of the protein becomes essential.<sup>57,58</sup> FEL is considered fundamental in the study of protein folding.<sup>57,59,60</sup> Total energy consumption on the landscape determines the positioning of the protein and stability. The global free energy minimum for the CST model protein was  $16.60$  kJ/mol. Figure 8A shows the two-dimensional representation of the free energy landscape from simulated atomic trajectories of the CST protein using PC1 and PC2. The dark-blue area shows the energy minima and the most favorable conformational region in the model protein, while the yellow spots represent the unfavorable conformations. The wide area covered with energy minima and very little conformational region occupied by yellow spots suggests the conformational stability in the model protein.



**Figure 9.** Surface electrostatic potential distribution of (A) the template heparan sulfate 2-*O*-sulfotransferase 1 (PDB ID: 3f5f) and (B) the CST model generated by the SWISS-MODEL using the Ef-Surf web server.

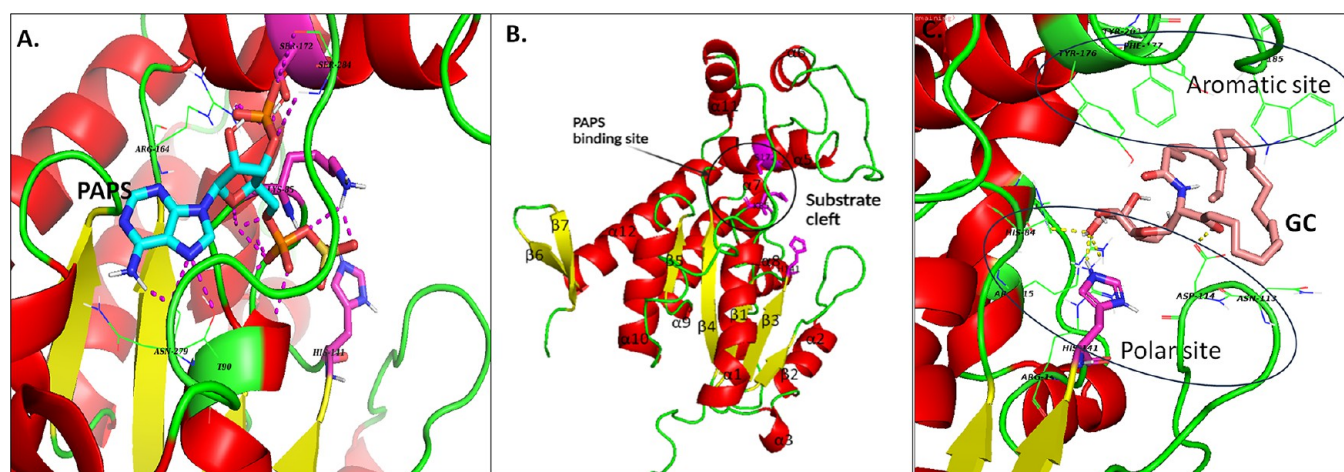
Further, the PCA data revealed that clusters were well-defined, occupying a minimum subspace on the protein. Applying PCA to a protein trajectory is reflected as “Essential Dynamics”, which extracts essential motion required for model stabilization.<sup>61</sup> The eigenvector range of  $-3$  to  $5.5$  nm<sup>2</sup> in eigenvector 1 and  $-4$  to  $3.5$  nm<sup>2</sup> in eigenvector 2 for the CST protein showed the restricted space with a well-defined internal motion behavior, which was essential for model stabilization as illustrated in Figure 8B. The MD trajectory analysis suggests two states of a protein as seen by two clusters in a scattered plot.

During MD simulations, density distribution analysis of the molecular coordinates of the CST model is an essential step to analyze the atomic density and changes in atomic orientation. A stable density area was observed for the CST model protein with a minimum value of  $46.0$  nm<sup>-3</sup>. The density area analysis confirmed that the CST structure was stable with minimum energy. Figure 8C depicts the density distribution graph of the CST model.

**3.6. Surface Analysis of the CST Model.** The molecular surface forms the boundary of the protein and acts as an interface for interaction with other molecules. The molecular surface of the modeled CST protein was analyzed using an eF-surf web server.<sup>62</sup> The surface potentials of the model protein are depicted from anionic (red) to cationic (blue) to hydrophobic area (yellow). The PAPS binding site of CST was inferred from the alignment of the CST model and the reference template protein, heparan sulfate 2-*O*-sulfotransferase 1 (PDB code: 3f5f\_A chain). Electrostatic potentials and hydrophobic properties were calculated by eF-Surf on the Connolly polygon vertex

surface active site for molecular recognition to search the corresponding aligned binding site residues. The PAPS and substrate binding sites were found to be more cationic than any other region in the model structure, facilitating the binding of PAPS and the substrate to their respective positions. Later, in the reaction mechanism section, we will learn about the strong binding affinity of PAPS in the PAPS binding site, and the electrostatic interaction probably played a crucial role in the binding of PAPS (negatively charged) to the PAPS binding site (highly cationic) in the target protein. However, compared with the template structure, the electrostatic potential of the CST model protein was relatively less in terms of surface area for the binding ligand as shown in Figure 9.

**3.7. Binding Pocket Detection.** To understand the molecular and biological function and physiochemical characteristics of the protein, the identification of a favorable binding site is important for ligand positioning.<sup>63–65</sup> To date, literature on sulfotransferase catalytic activities suggests that LYS, SER, and HIS form a ternary complex.<sup>66–68</sup> Despite the great level of structural variation among sulfotransferases, sequence alignment of the CST model with the template (PDB: 3f5f) showed the presence of LYS85, SER172, and HIS141 in CST with the formation of a ternary complex to facilitate the binding of PAPS. PAPS acts as a cosubstrate of the CST protein and is conserved in sulfotransferases.<sup>69,70</sup> The main core of the CST model protein is composed of an  $\alpha/\beta$  motif that comprises five parallel  $\beta$ -strands surrounded by  $\alpha$ -helices on both sides and a conserved helix running on the top of the fold. The PAPS binding site constitutes the strand-loop-helix and strand-turn-helix motif.



**Figure 10.** Proposed computational 3D model of the CST protein: (A) cosubstrate PAPS binding site of CST; (B) representing the catalytic sites with the sulfuryl group donor PAPS binding site and the sulfuryl group acceptor substrate binding site. Secondary structures are represented in different colors:  $\alpha$ -helices in red,  $\beta$ -strands in yellow, and loop regions in green. (C) Substrate binding site of CST with key amino acid residues flanked around the substrate galactocerebroside (GC).

The 5'-phosphosulfate-binding (PSB)-loop runs parallel to helix 5 of the strand-turn-helix. This was further validated by some computational tools such as DoGSiteScorer and CastP, which also suggested a similar binding site for the cosubstrate, PAPS. Furthermore, to understand the unique amino acid residues in CST, we performed a structural alignment study of CST in PyMol with other sulfotransferases including human estrogen sulfotransferase (PDB: 1HY3). Besides the conserved residues LYS85 and SER172 at the PAPS binding site, the CST model showed some unique residues that differentiate it from other sulfotransferases. In the PAPS binding site of CST, ARG164, ASN279, SER284, and ARG282 were distinctively present at the mouth of the PAPS binding site. Since cosubstrate PAPS is a conserved active site in sulfotransferases, there might be the possibility of side effects with other sulfotransferases or PAPS-dependent proteins if inhibitors target this site.

Besides conserved binding sites for PAPS among sulfotransferases, CST showed a uniqueness in amino acid composition in the substrate binding site. The CST substrate binding site is broadly divided into two subpockets: polar site and aromatic site, which are opposite to each other in the active site. The polar site was flanked by HIS141, HIS82, ARG143, ARG215, and LYS85, while the aromatic site was dominated by PHE170, TRY176, PHE177, TYR203, and TRP185 in  $\alpha 5$  and the neighboring loop region. In the substrate binding site, HIS84, ARG143, ASP114, TYR173, TYR176, PHE177, TRP185, TYR203, and ARG215 are uniquely present. While preliminary drug screening or designing new drug candidates, these unique residues at the substrate binding site can be targeted to develop specific CST inhibitors. Figure 10 represents the binding sites for both the substrate and the cosubstrate.

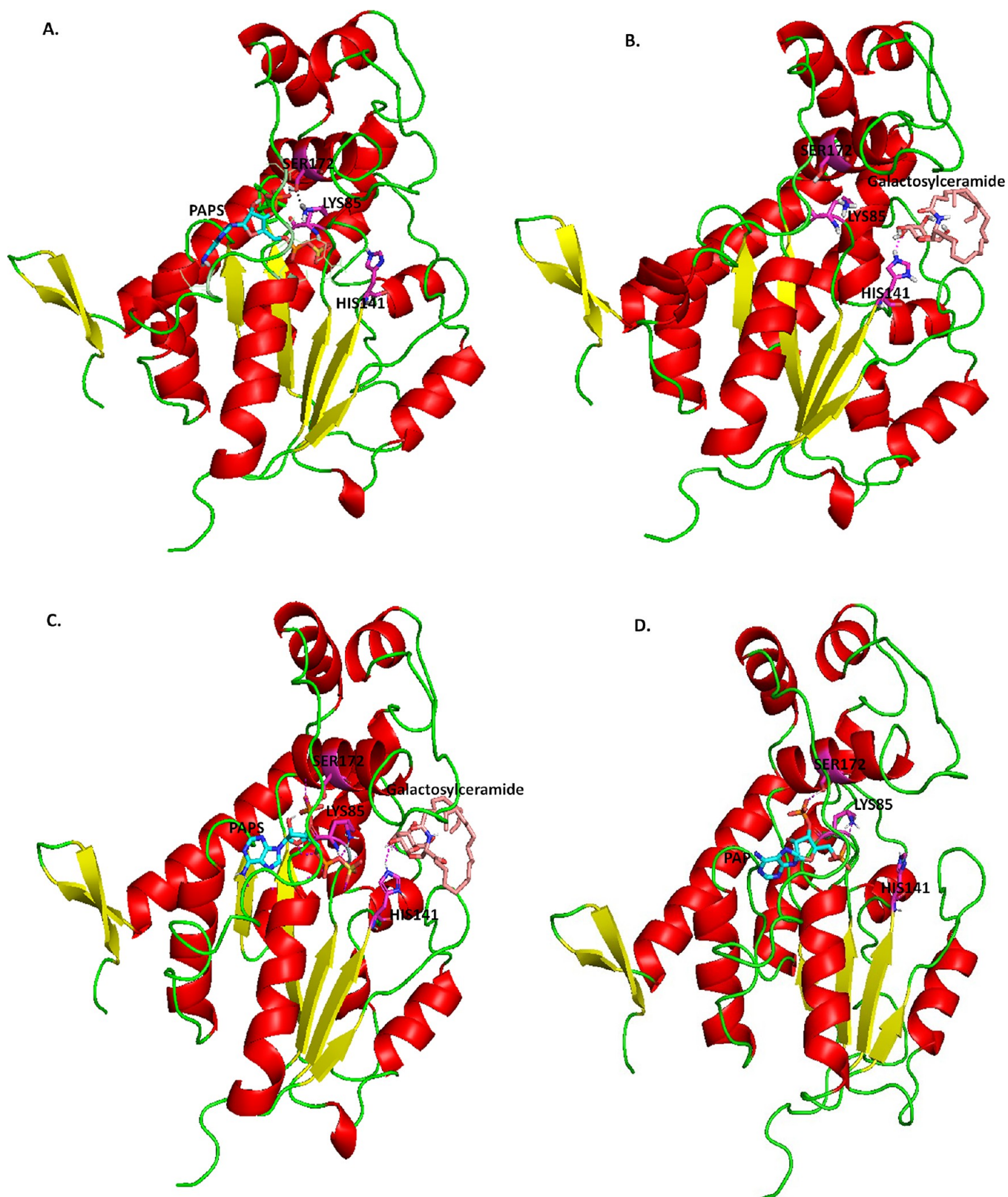
### 3.8. Insight into the Active Site and Catalytic Reaction

**Mechanism of CST.** Sulfotransferases, in general, catalyze the transfer of a sulfuryl group from PAPS to a substrate through an in-line ternary displacement reaction mechanism in which the conserved amino acid residues found for catalytic activities are LYS, SER, and HIS, which form a ternary complex and facilitate binding of PAPS and its action as a sulfuryl donor.<sup>66,71,72</sup> According to the reported mechanism of action of sulfotransferases in general, LYS is an important conserved amino acid residue that allows HIS in the ternary complex to attack the

bridging oxygen in PAPS and facilitate sulfuryl transfer and leave a negative charge on the bridging oxygen, which then attracts the side-chain nitrogen of LYS to switch from SER to the bridging oxygen and thus facilitate sulfate dissociation. Like other sulfotransferases, cerebroside sulfotransferase also acts on the sulfuryl moiety of 3'-phosphoadenosine-5'-phosphosulfate (PAPS) and breaks it from PAPS and transfers it to the sugar chain, yielding PAP and sulfated glycan. In CST, LYS85, SER172, and HIS141 are the catalytic residues present at the corresponding position in the PAPS binding site.

#### 3.8.1. Derivation of the Mechanism of Action of CST through Substrate-Protein Interaction Using Molecular Docking and Molecular Dynamics Simulation.

In our studies, based on a structural similarity study with the reference sulfotransferase protein heparan sulfate 2-O-sulfotransferase 1 and molecular docking studies, we determined that the newly generated CST homology model contains LYS85 and SER172 as conserved amino acid residues with HIS141. SER172 on helix 5 acts as a key catalytic residue and interacts with 3'-phosphate of PAPS. In the PAPS binding state, the side-chain nitrogen of the main conserved residue LYS85 interacts with the side chain of conserved SER172, and thus, this side chain-side chain interaction facilitates the bridging oxygen between the 5' phosphate and sulfuryl group to orient toward HIS141. HIS141 acts as a catalytic base that removes the proton from the 3'-OH group in the galactocerebroside substrate and thus converts it into a strong nucleophile, which then attacks the sulfur atom of PAPS, which in turn puts a negative charge on the bridging oxygen. The negatively charged bridging oxygen then attracts the positively charged side-chain nitrogen of LYS85. This leads the side-chain nitrogen of LYS85 to switch from SER172 to the bridging oxygen and thus facilitates the dissociation of the sulfuryl group from PAPS and its transfer to the galactocerebroside substrate. Figure 12 describes the proposed catalytic reaction mechanism of CST using the comparative molecular docking analysis under the cosubstrate PAPS, the substrate galactocerebroside, and the product PAP binding state. The positioning of SER172 and LYS85 with HIS141 is important in determining the catalytic action of the CST protein via triad formation, which facilitates the proper positioning of PAPS to act as a sulfuryl group donor for the acceptor substrate,



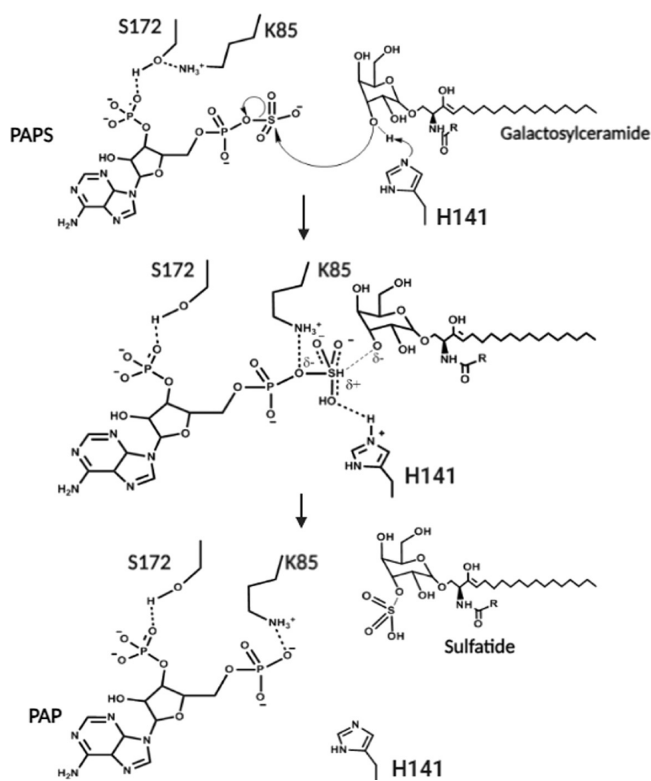
**Figure 11.** Representation of the catalytic mode of the acceptor and donor molecule in the active site of CST through molecular docking and molecular dynamic simulation: (A) CST bound with PAPS in the cosubstrate binding site, (B) CST bound with galactosylceramide in the substrate binding site, (C) CST bound with both PAPS and galactosylceramide in their respective binding sites, and (D) CST with the end product, PAP.

galactocerebroside, or galactosylceramide. To check the catalytic process in CST, independent interaction studies of the CST model protein with the cosubstrate PAPS, the substrate

GC, and the product PAP were carried out (Figure 11). The comparative docking and subsequent molecular dynamics studies support existing knowledge. The docking simulation

was performed in Autodock 4.2 to study the (i) interaction between CST (protein) and PAPS (cosubstrate), (ii) interaction between CST and GC, (iii) interaction of CST with PAPS and GC, and (iv) interaction between CST and PAP (product).

In the PAPS-CST binding complex (Figure 11A), the proton of the –OH group of SER172 interacted with the oxygen of the 3′-phosphate moiety, while the oxygen atom of the –OH group of SER172 interacted with the side-chain nitrogen of LYS85. This suggested the initial stage of CST catalytic action. The lowest free energy of binding for PAPS binding was –17.95 kcal/mol, which supports a very strong binding. In the galactocerebroside or galactosylceramide (GC) binding mode (Figure 11B), in the substrate cleft, conserved HIS141 interacted with GC by forming a hydrogen bond with its 3′OH group with the free energy of binding –2.57 kcal/mol. Interestingly, one more HIS84 facilitated this interaction and supported the faster removal of the proton from 3′–OH of galactocerebroside. In the CST-PAPS-GC binding complex (Figure 11C), both the substrate and cosubstrate occupied their respective binding sites. Interestingly, CST under the PAPS-GC binding mode showed a shift and orientation of the side-chain nitrogen of LYS85 from SER172 to the bridging oxygen and formed a hydrogen bond with the bridging oxygen. This suggests that substrate binding in the CST active site takes place after PAPS binding to facilitate the catalytic action of the CST protein. Finally, in the PAP binding mode (Figure 11D), the interactions with catalytic residues were found similar to PAPS with the lowest free energy of binding –13.95 kcal/mol. The interaction poses are illustrated in Figure 11. Based on the structural similarity with the template and molecular docking and molecular dynamic analysis, we proposed the reaction



**Figure 12.** Proposed reaction mechanism of CST based on molecular docking and molecular dynamics studies of CST with PAPS and the substrate.

mechanism for CST catalytic action, which is represented in Figure 12.

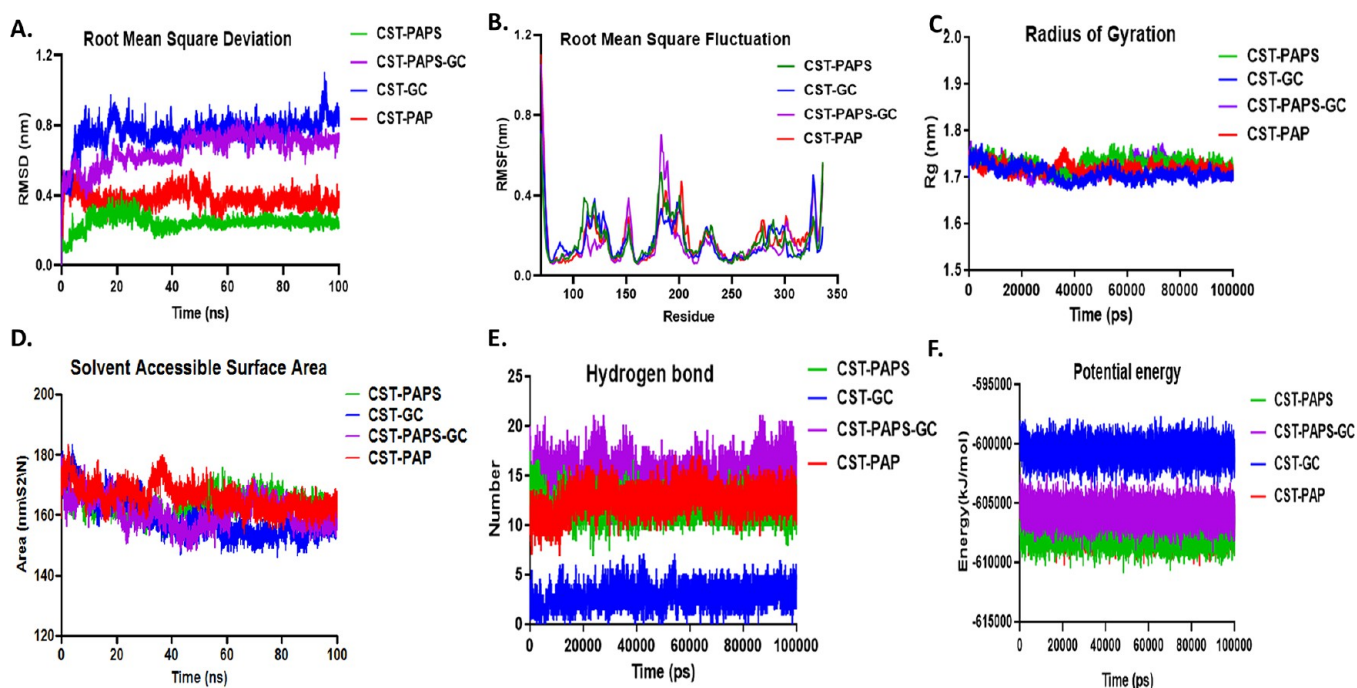
**3.8.2. Validation of the Catalytic Action of CST Using MD Simulation.** The stability of the complexes (CST-PAPS, CST-GC, CST-PAPS-GC and CST-PAP) was derived by MD simulation studies using GROMACS in a Linux environment to analyze the MD trajectories of each complex. Parameters studied were RMSD, RMSF, radius of gyration, solvent-accessible surface area, hydrogen bond, principal component analysis, free energy landscape, and density distribution analysis as shown in Figure 13. The MD simulation study was to understand the dynamic behavior of CST in the presence of water and ions toward the PAPS or/and galactocerebroside substrate.

In RMS deviation, the average RMSD values were 0.2, 0.8, 0.7, and 0.35 nm for CST-PAPS, CST-GC, CST-PAPS-GC, and CST-PAP complexes, respectively. The plateau of the average RMSD is directly proportional to the stability of the complex. The graph portrayed that the CST-PAPS complex reached equilibrium after 35 ns, while the CST-GC complex achieved equilibrium after 30 ns. Small final fluctuations indicated that the entire system with the protein–substrate complex was stabilized and the substrate remained securely present in the substrate binding site pocket. The complex of CST with both PAPS and GC is the intermediary or transition phase during the catalytic reaction mechanism because the enzyme needs to occupy PAPS as well as the substrate in their respective sites and thus adjustment toward equilibrium took a relatively longer time. The CST-PAPS-GC complex achieved stability after 45 ns. The CST-PAP complex showed stability after 55 ns of simulation. Since PAP is the end product of the CST catalytic action, it was no longer required by the enzyme; hence, its binding efficiency toward the protein was also not much stronger. The RMSD graphical representation in Figure 13A reaffirms the stability of PAPS and GC binding in their respective binding sites in the CST model protein.

RMS fluctuation of four complexes from their time-average value was found stable. In the loop region of 177–206 residues, the fluctuation was visible in the CST-PAPS complex, the CST-GC complex, and the CST-PAP complex, while the CST-PAPS-GC complex was found to be the most fluctuating complex relative to the model apoprotein (Figure 13B). Since the CST-PAPS-GC complex represents the transition phase, the kind of flexibility probably denotes the transition-related instability. Positioning of LYS85 and SER172 was in similar pattern and within the flexible range of nominal residue displacement in a protein complex with PAPS, PAPS-GC, and PAP, while the fluctuation in the CST-GC complex was significantly in a different manner. This suggests that PAPS binding led to GC binding via loop modification as reported in an earlier work.<sup>73</sup>

The radius of gyration analysis of the model CST protein in different complexes showed the overall dimensional stability and compactness between the residues and the bonding pattern, in turn portraying a direct proportion to the protein volume. Compared to the free protein (~1.8 nm), the lower value  $R_g$  (~1.7 nm) during complex formation showed increasing rigidity; however, the similar trend in  $R_g$  in all complexes suggests the stability of CST in complex formation (Figure 13C).

Then, the SASA analysis clearly illustrated that the exposed surface area of the model protein was little impacted under different complex formations (Figure 13D) when comparing the SASA of the complexes with that of the free protein. The lower



**Figure 13.** MD simulation for CST with PAPS, GC, and PAP. (A) RMSD, (B) RMSF, (C)  $R_g$ , (D) hydrogen bonds, (E) potential energy, and (F) solvent-accessible surface analysis (SASA).

value of SASA in CST-ligand complexes indicated the contracted nature of the CST protein under complex formation. Consistent with our earlier discussion, a similar SASA graph projection with time in CST-PAPS and CST-PAPS-GC complexes suggested that the cosubstrate PAPS bound first to CST and facilitated the binding of GC in its respective substrate binding pocket. This also indicated that during complex formation with the cosubstrate and substrate during the transition stage of the reaction mechanism, the protein remained structurally stable, whereas little deviation was observed when the CST bound with GC alone.

Following the SASA analysis, the hydrogen-bonding pattern between the side chains of the exposed amino acid residues with water molecules highlighted that CST bound with PAPS and GC did not impact much the overall solvent accessibility. This also suggested that the binding of PAPS and GC in their respective sites in CST facilitated smooth catalytic action rather than disturbing the protein structure. As expected, when GC bound to the CST-PAPS complex under CST-PAPS-GC complex formation, intermolecular hydrogen bonds increased to accommodate the substrate (Figure 13E).

Subsequently, the energy calculation showed the lesser energy requirement of the CST-PAPS complex (608,000 kJ/mol) as compared to the complex in the presence of GC (600,000 kJ/mol). This also highlights the relative stability of CST-PAPS binding due to a conserved and defined active site pocket for PAPS binding. The binding of PAPS further facilitates the binding of GC, which can be seen with the reduction of the potential energy requirement for CST when bound with PAPS and GC in their respective sites (Figure 13F).

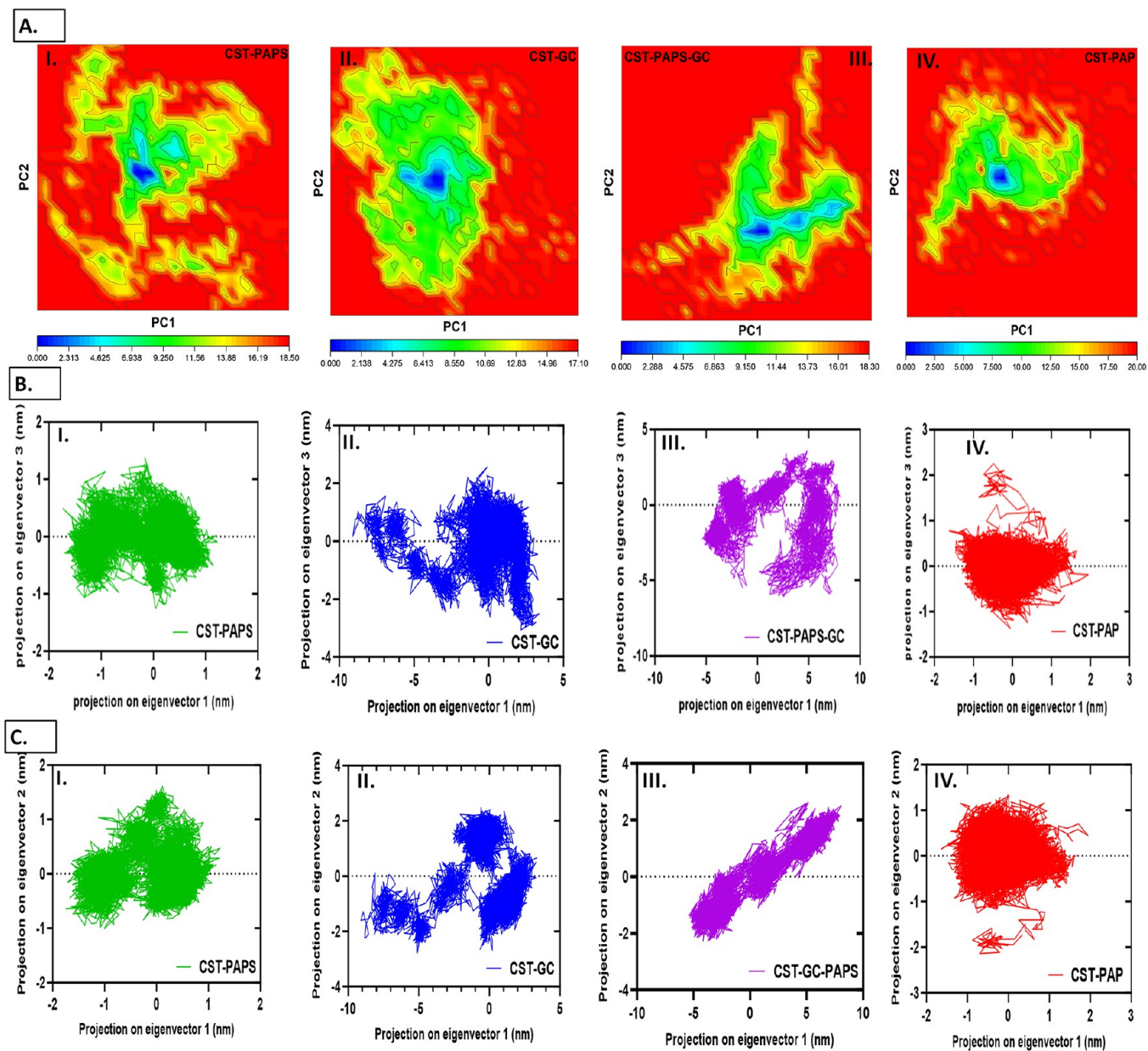
To further understand the energy requirement of complexes, we performed a free energy landscape study that described the changes in the protein folding behavior under different protein–ligand interactions. The protein attaining a specific position can be distinguished by the total energy on the landscape. The free energy surface (FES) of the ligand binding process leads to

defining certain structural states. The global free energy minima were 18.50, 17.10, 18.30, and 20.0 kJ/mol for CST-PAPS, CST-GC, CST-PAPS-GC, and CST-PAP, respectively (Figure 14A). Multiple energy minima in GROMACS revealed the stable conformation of the CST model under different complex conditions during the catalytic action of the protein. The free energy landscape analysis also suggested the stable folding pattern of the CST under different interactions.

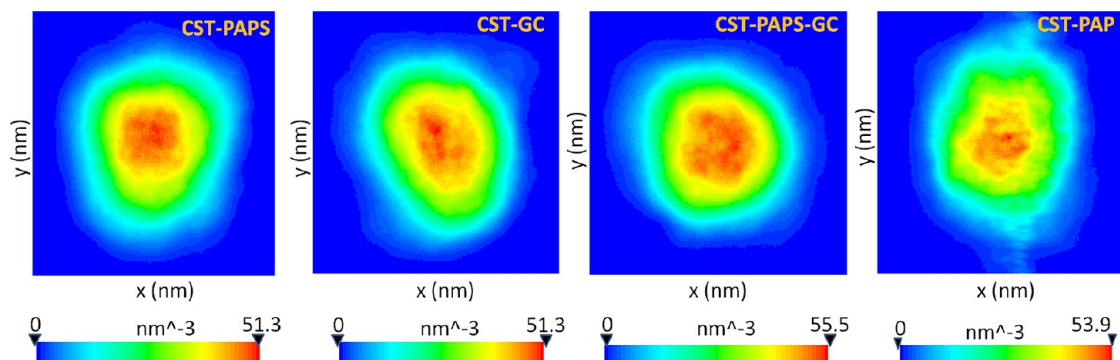
Further, the conformational subspace of the complex was examined to understand the dynamic behavior of the CST in four complexes by PCA techniques. The displacement of atomic fluctuation and motions in the structures in the form of the complex is represented by the plot between eigenvectors 1 and 2 (Figure 14B) or between eigenvectors 1 and 3 (Figure 14C). The CST-PAPS complex eigenvector 1 values range from  $-1.5$  to  $1$  nm, eigenvector 2 values range from  $-1.0$  to  $1.5$ , and eigenvector 3 values range from  $-1.0$  to  $1.0$  nm (Figure 14B(I),C(I)). The value ranged from  $-8$  to  $3$  nm in eigenvector 1, from  $-2.7$  to  $2.5$  for eigenvector 2, and from  $-3$  to  $2$  nm for eigenvector 3 in the CST-GC complex (Figure 14B(II),C(II)). The range was from  $-5$  to  $8$  nm in eigenvector 1, from  $-2$  to  $2.5$  for eigenvector 2, and from  $-6$  to  $4$  nm in eigenvector 3 in the CST-PAPS-GC complex (Figure 14B(III),C(III)). The final CST-PAP complex showed an eigenvector range of 1 from  $-1.5$  to  $2$ , while eigenvector 2 ranged from  $-2$  to  $1$ , and eigenvector 3 ranged from  $-1.0$  to  $2$  nm (Figure 14B(IV),C(IV)). Thus, comparing eigenvector plots highlighted that binding of GC led to significant periodic transition in the three-dimensional structure of the CST protein than binding of CST alone of either PAPS or PAP. The eigenvector ranges of all four complexes highlighted the restricted conformational space, which led to stabilization through a well-defined internal motion behavior at the atomic level.

Furthermore, density distribution analysis of molecular coordinates obtained from MD simulation was performed using a densmap script to understand atomic density, atomic

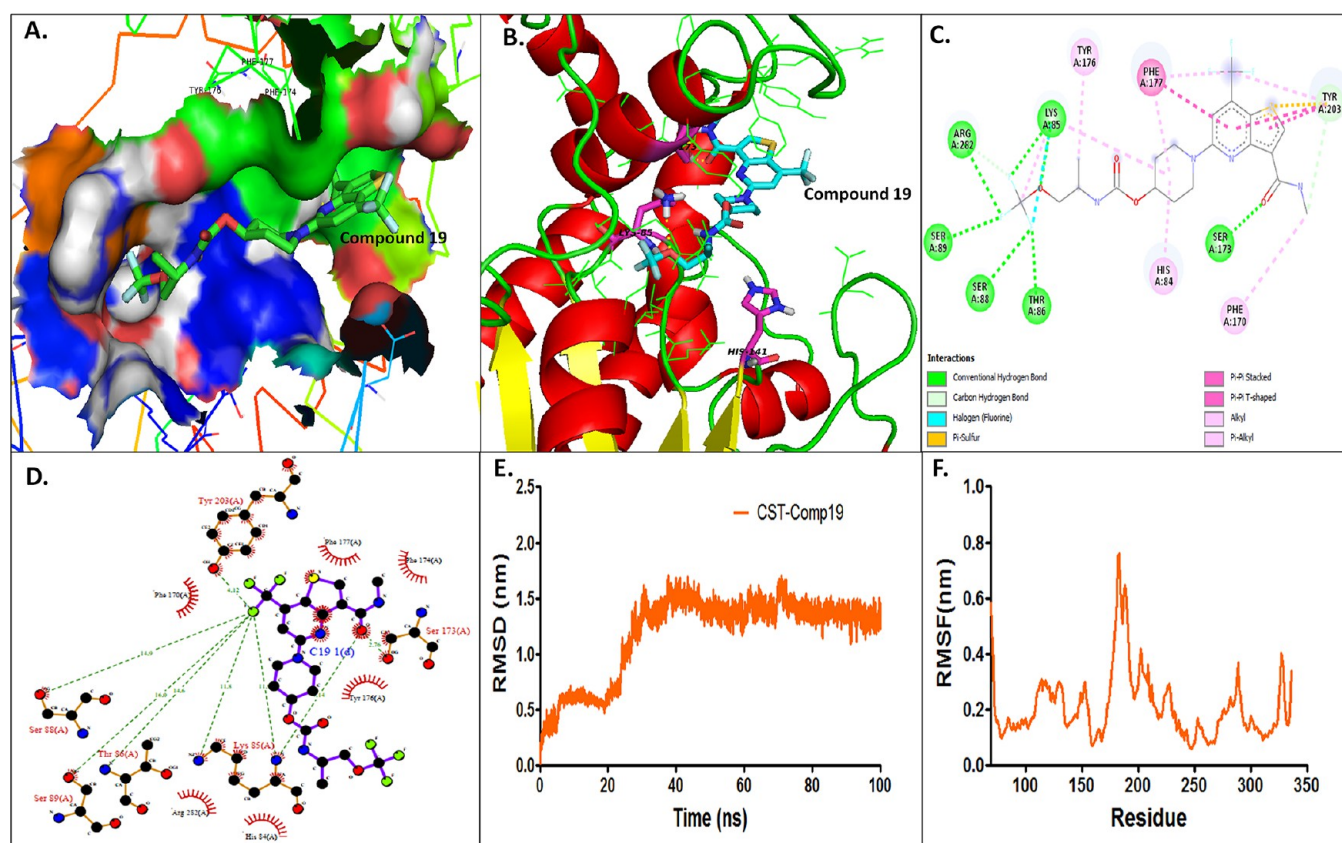




**Figure 14.** (A) Free energy landscape representing the folding behavior and direction of motion in the presence of PAPS, GC, PAPS + GC, and PAP. (B) Dynamic energy fluctuation plot with projections of eigenvectors 3 on 1 generated for the modeled structures showing the conformational space of  $C\alpha$ -atoms in the presence of PAPS, GC, PAPS + GC, and PAP. (C) Projection of the first two eigenvectors 1 and 2 for the CST complex with PAPS, GC, PAPS + GC, and PAP.



**Figure 15.** Density distribution analysis plotted to understand the atomic orientation using the densmap script for CST-PAPS, CST-GC, CST-PAPS-GC, and CST-PAP.



**Figure 16.** Protein–inhibitor interaction. (A) Surface view of compound **19** interaction. (B) Docking poses of the binding interaction of molecules in the active site of the CST model protein and compound **19**. (C) 2D ligand–protein interaction. (D) PDBsum protein–ligand interaction. (E, F) MD simulation: RMSD and RMSF plots.

orientation, and distribution of the CST model under different complex conditions.<sup>74</sup> Stable density area can be seen in the PAPS binding site, the GC binding site, and the PAPS–GC complex with values of 51.3, 51.3, and 55.5 nm<sup>−3</sup>, respectively. The comparative density distribution analysis also confirms the better stability of CST when PAPS occupies the PAPS binding site and GC occupies the GC binding site with a minimum value of 55.5 nm<sup>−3</sup> as shown in Figure 15, whereas a lower density was noticed when CST binds either with PAPS or with GC. Thus, a comparison of density distribution affirmed the high impact structural transition of CST under the bound form of both the substrate and cosubstrate than binding alone with PAPS, GC, or PAP. This structural transition probably led by catalytic action of protein.

**3.9. Validation of Response of the CST Model toward the Reported Competitive Inhibitor through *In Silico* Inhibition Studies.** Based on the uniqueness of the active site composition and stability study, the substrate binding site was considered an inhibitor binding site for the development of competitive inhibitors. The electrostatic potential of the substrate binding site was found to be very high; therefore, those inhibitors would be considered potential competitive inhibitors that meet the electrostatic potential of the substrate and probably would be substrate analogues. To understand the potential of the developed CST model to respond effectively against inhibitors, we used a molecular docking approach to test compound **19**, which was claimed by Thurairatnam et al. as a competitive inhibitor for the CST substrate.<sup>7</sup> Docking of compound **19** at the CST substrate site showed its better occupancy in the active site pocket and the higher interaction

potential than the substrate GC in terms of binding energy to make it a competitive inhibitor toward CST. The lowest free energy of binding for compound **19** was −7.17 kcal/mol, while for the substrate, it was −2.57 kcal/mol. The fluorinated moiety of compound **19** formed hydrogen bonds with LYS85, SER88, SER89, THR86, and ARG282, while an aromatic fragment of the compound stacked toward aromatic residues including PHE177, TYR203, and HIS84 via pi–pi stacking, thus occupying the substrate active site pocket tightly. MD simulation further confirmed the stability of the ligand–protein binding, which is depicted in Figure 16. After 40 ns, the complex started showing equilibrium. *In silico* inhibition study with a known inhibitor thus projects the suitability of the CST model for future *in silico*-based drug screening to strengthen the substrate reduction therapy for MLD.

## 4. CONCLUSIONS

Genetic studies suggest that the role of ARSA is essential in the regulation of the sulfatide level and subsequent maintenance of the myelin sheath and overall communication channel. In the absence of ARSA-led sulfatide recycling, accumulation of sulfatide in lysosomes takes place, which can be regulated by CST inhibition via inhibitor development. This treatment strategy is considered substrate reduction therapy. CST is a rate-limiting enzyme in sulfatide biosynthesis. In SRT, CST is targeted to develop a negative feedback mechanism. For preliminary drug screening against CST, it is logical to use a large pool of available data sets of compounds. In this regard, the 3D computational structure of the CST protein is the first

requisite. In this study, we developed a 3D model of CST protein using SWISS-MODEL. Subsequently, comparing the model with the best model from other model builders confirmed the quality of SWISS-MODEL over others. Model refinement and dynamic evaluation optimized the model on different time scales. The final model showed the acceptable distribution of the  $\phi$  and  $\Psi$  angle. Subsequently, we performed MD simulation of the model protein to study the stability and dynamism of the model protein using RMSD, RMSF,  $R_g$ , hydrogen bond, SASA, PCA, and FEL parameters. To use the CST model for drug screening, it was important to understand the molecular and functional dynamics of the model by deciphering its catalytic mechanism. To do this, we performed molecular docking study using PAPS, GC, and PAP as the ligand, which revealed the probable catalytic reaction mechanism of the CST protein. Further, MD simulation revealed that binding of PAPS facilitates the binding of the substrate and both share their defined binding sites. Like other sulfotransferases, the PAPS binding site is conserved in the CST and is primarily identified by the conserved ternary positioning of three key catalytic residues, LYS85, SER172, and HIS141. Since PAPS is a conserved region in sulfotransferases in general, using the PAPS binding site for inhibitor development may cause side effects. Targeting unique amino acid residues at the substrate binding site can facilitate the development of specific and potent inhibitors for CST. The stability of the substrate binding site in MD trajectory analysis validates the suitability of the substrate binding pocket as a target site in the CST model for inhibitors. Thus, the present study is crucial in the context of developing substrate reduction therapy against MLD by providing the most needed 3D model of the CST protein and proposing its catalytic mechanism to understand the binding pattern and active site composition in order to develop specific and potent CST inhibitors.

## ■ ASSOCIATED CONTENT

### Data Availability Statement

The deposited FASTA sequence of CST was retrieved from the NCBI database with Accession no. NP\_001305033.1, gene ID "9514", and gene identifier (GI: 970259838). The CST model is provided in the.pdb format.

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acsomega.3c09462>.

Physicochemical characteristics of the CST protein; best hits obtained by PSI-BLAST with the CST sequence; multiple sequence alignment; phylogenetic tree for the query sequence; top 5 AlphaFold models; Ramachandran plot for the top AlphaFold model, PSI-PRED model, and Phyre2 model; and chemical structure of compound **19** (PDF)

CST model (PDB)

## ■ AUTHOR INFORMATION

### Corresponding Author

Nivedita Singh – Department of Dravyaguna, Faculty of Ayurveda, Institute of Medical Sciences, Banaras Hindu University, Varanasi 221005 Uttar Pradesh, India;  
✉ [orcid.org/0000-0003-4049-2313](mailto:orcid.org/0000-0003-4049-2313); Email: [s.nivedita@bhu.ac.in](mailto:s.nivedita@bhu.ac.in), [niv234@gmail.com](mailto:niv234@gmail.com)

### Author

Anil Kumar Singh – Department of Dravyaguna, Faculty of Ayurveda, Institute of Medical Sciences, Banaras Hindu University, Varanasi 221005 Uttar Pradesh, India

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acsomega.3c09462>

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

N.S. acknowledges the Institute of Eminence, Banaras Hindu University, Government of India, for providing fellowship and research grant for the ongoing project under the Malaviya Post Doctoral Fellowship scheme with Grant ID: R/Dev/G/6031/IoE/MPDFs/61698.

## ■ REFERENCES

- (1) Lamichhane, A.; Cabrero, F.-R. Metachromatic Leukodystrophy. *Front. Lysosomal Storage Disor. Treat.* **2023**, 153–162.
- (2) Gieselmann, V. Metachromatic Leukodystrophy: Genetics, Pathogenesis and Therapeutic Options. *Acta Paediatr.* **2008**, 97 (457), 15–21.
- (3) Shaimardanova, A.-A.; Chulpanova, D.-S.; Solovyeva, V.-V.; Mullagulova, A. I.; Kitaeva, K.-V.; Allegrucci, C.; Rizvanov, A.-A. Metachromatic Leukodystrophy: Diagnosis, Modeling, and Treatment Approaches. *Front. Med.* **2020**, 7, No. 576221, DOI: [10.3389/fmed.2020.576221](https://doi.org/10.3389/fmed.2020.576221).
- (4) Singh, N.; Singh, A.-K. A Comprehensive Review on Structural and Therapeutic Insight of Cerebroside Sulfotransferase (CST) - An Important Target for Development of Substrate Reduction Therapy against Metachromatic Leukodystrophy. *Int. J. Biol. Macromol.* **2024**, 258, No. 128780.
- (5) Takahashi, T.; Suzuki, T. Role of Sulfatide in Normal and Pathological Cells and Tissues. *J. Lipid Res.* **2012**, 53 (8), 1437–1450, DOI: [10.1194/jlr.R026682](https://doi.org/10.1194/jlr.R026682).
- (6) Dali, C.-i.; Barton, N.-W.; Farah, M.-H.; Moldovan, M.; Månsson, J.; Nair, N.; Dunø, M.; Risom, L.; Cao, H.; Pan, L.; Sellos-Moura, M.; Corse, A. M.; Krarup, C. Sulfatide Levels Correlate with Severity of Neuropathy in Metachromatic Leukodystrophy. *Ann. Clin. Transl. Neurol.* **2015**, 2 (5), 518–533, DOI: [10.1002/acn3.193](https://doi.org/10.1002/acn3.193).
- (7) Thurairatnam, S.; Lim, S.; Barker, R.-H.; Choi-Sledeski, Y.-M.; Hirth, B.-H.; Jiang, J.; Macor, J.-E.; Makino, E.; Maniar, S.; Musick, K.; Pribish, J.-R.; Munson, M. Brain Penetrable Inhibitors of Ceramide Galactosyltransferase for the Treatment of Lysosomal Storage Disorders. *ACS Med. Chem. Lett.* **2020**, 11 (10), 2010–2016.
- (8) Babcock, M.-C.; Mikulka, C.-R.; Wang, B.; Chandriani, S.; Chandra, S.; Xu, Y.; Webster, K.; Feng, Y.; Nelvagal, H.-R.; Giaramita, A.; Yip, B.-K.; Lo, M.; Jiang, X.; Chao, Q.; Woloszynek, J.-C.; Shen, Y.; Bhagwat, S.; Sands, M.-S.; Crawford, B.-E. Substrate Reduction Therapy for Krabbe Disease and Metachromatic Leukodystrophy Using a Novel Ceramide Galactosyltransferase Inhibitor. *Sci. Rep.* **2021**, 11 (1), No. 14486, DOI: [10.1038/s41598-021-93601-1](https://doi.org/10.1038/s41598-021-93601-1).
- (9) Blomqvist, M.; Zetterberg, H.; Blennow, K.; Månsson, J.-E. Sulfatide in Health and Disease. The Evaluation of Sulfatide in Cerebrospinal Fluid as a Possible Biomarker for Neurodegeneration. *Mol. Cell. Neurosci.* **2021**, 116, No. 103670, DOI: [10.1016/j.mcn.2021.103670](https://doi.org/10.1016/j.mcn.2021.103670).
- (10) Barnes-Vélez, J.-A.; Aksoy Yasar, F.-B.; Hu, J. Myelin Lipid Metabolism and Its Role in Myelination and Myelin Maintenance. *Innovation* **2022**, 4 (1), No. 100360, DOI: [10.1016/j.xinn.2022.100360](https://doi.org/10.1016/j.xinn.2022.100360).
- (11) Rosenberg, J.-B.; Kaminsky, S.-M.; Aubourg, P.; Crystal, R.-G.; Sondhi, D. Gene Therapy for Metachromatic Leukodystrophy. *J. Neurosci. Res.* **2016**, 94 (11), 1169–1179.
- (12) Groeschel, S.; Köhl, J.-S.; Bley, A.-E.; Kehrer, C.; Weschke, B.; Döring, M.; Böhringer, J.; Schrum, J.; Santer, R.; Kohlschötter, A.

Krägeloh-Mann, I.; Möller, I. Long-Term Outcome of Allogeneic Hematopoietic Stem Cell Transplantation in Patients with Juvenile Metachromatic Leukodystrophy Compared with Nontransplanted Control Patients. *JAMA Neurol.* **2016**, *73* (9), 1133–1140.

(13) Page, K.-M.; Stenger, E.-O.; Connelly, J.-A.; Shyr, D.; West, T.; Wood, S.; Case, L.; Kester, M.; Shim, S.; Hammond, L.; Hammond, M.; Webb, C.; Biffi, A.; Bambach, B.; Fatemi, A.; Kurtzberg, J. Hematopoietic Stem Cell Transplantation to Treat Leukodystrophies: Clinical Practice Guidelines from the Hunter's Hope Leukodystrophy Care Network. *Biol. Blood Marrow Transplant.* **2019**, *25* (12), e363–e374.

(14) Wolf, N.-I.; Breur, M.; Plug, B.; Beerepoot, S.; Westerveld, A.-S.-R.; van Rappard, D.-F.; de Vries, S.-I.; Kole, M.-H.-P.; Vanderver, A.; van der Knaap, M.-S.; Lindemans, C.-A.; van Hasselt, P.-M.; Boelens, J.-J.; Matzner, U.; Gieselmann, V.; Bugiani, M. Metachromatic Leukodystrophy and Transplantation: Remyelination, No Cross-correction. *Ann. Clin. Transl. Neurol.* **2020**, *7* (2), 169–180, DOI: 10.1002/acn3.50975.

(15) Kurtzberg, J. Gene Therapy Offers New Hope for Children with Metachromatic Leukodystrophy. *Lancet* **2022**, 399 (10322), 338–339.

(16) Hironaka, K.; Yamazaki, Y.; Hirai, Y.; Yamamoto, M.; Miyake, N.; Miyake, K.; Okada, T.; Morita, A.; Shimada, T. Enzyme Replacement in the CSF to Treat Metachromatic Leukodystrophy in Mouse Model Using Single Intracerebroventricular Injection of Self-Complementary AAV1 Vector. *Sci. Rep.* **2015**, *5*, No. 13104, DOI: 10.1038/srep13104.

(17) LeVine, S. M.; Tsau, S. Substrate Reduction Therapy for Krabbe Disease: Exploring the Repurposing of the Antibiotic D-Cycloserine. *Front. Pediatr.* **2022**, *9*, No. 807973, DOI: 10.3389/fped.2021.807973.

(18) Komada, N.; Fujiwara, T.; Yoshizumi, H.; Ida, H.; Shimoda, K. A Japanese Patient with Gaucher Disease Treated with the Oral Drug Eliglustat as Substrate Reducing Therapy. *Case Rep. Gastroenterol.* **2022**, *15* (3), 838–845.

(19) Shemesh, E.; Deroma, L.; Bembi, B.; Deegan, P.; Hollak, C.; Weinreb, N.-J.; Cox, T.-M. Enzyme Replacement and Substrate Reduction Therapy for Gaucher Disease *Cochrane Database Syst. Rev.* **2015** Vol. **2015** (3) CD010324 DOI: 10.1002/14651858.CD010324.pub2. PMID: 25812601; PMCID: PMC8923052.

(20) Platt, F.-M.; Jeyakumar, M. Substrate Reduction Therapy. *Acta Paediatr.* **2008**, *97* (457), 88–93.

(21) Li, W.; Zech, I.; Gieselmann, V.; Müller, C.-E. A Capillary Electrophoresis Method with Dynamic PH Junction Stacking for the Monitoring of Cerebrosidase Sulfotransferase. *J. Chromatogr. A* **2015**, *1407*, 222–227.

(22) Roeske-Nielsen, A.; Buschard, K.; Månson, J.-E.; Rastam, L.; Lindblad, U. A Variation in the Cerebrosidase Sulfotransferase Gene Is Linked to Exercise-Modified Insulin Resistance and to Type 2 Diabetes. *Exp. Diab. Res.* **2009**, *2009*, No. 429593.

(23) Singh, N.; Saravanan, P.; Thakur, M. S.; Patra, S. Development of Xanthine Based Inhibitors Targeting Phosphodiesterase 9A. *Lett. Drug Des. Discovery* **2017**, *14* (10), 1122–1137, DOI: 10.2174/1570180813666161102125423.

(24) Singh, N.; Malik, A.-H.; Iyer, P.-K.; Patra, S. Diversifying the Xanthine Scaffold for Potential Phosphodiesterase 9A Inhibitors: Synthesis and Validation. *Med. Chem. Res.* **2021**, *30* (6), 1199–1219, DOI: 10.1007/s00044-021-02722-9.

(25) Agamah, F. E.; Mazandu, G. K.; Hassan, R.; Bope, C. D.; Thomford, N. E.; Ghansah, A.; Chimusa, E. R. Computational/in Silico Methods in Drug Target and Lead Prediction. *Briefings Bioinf.* **2020**, *21* (5), 1663–1675.

(26) Chang, Y.; Hawkins, B. A.; Du, J. J.; Groundwater, P. W.; Hibbs, D. E.; Lai, F. A Guide to In Silico Drug Design. *Pharmaceutics* **2023**, *15* (1), No. 49, DOI: 10.3390/pharmaceutics15010049.

(27) Singh, N.; Patra, S.; Patra, S. Identification of Xanthine Derivatives as Inhibitors of Phosphodiesterase 9A through in Silico and Biological Studies. *Comb. Chem. High Throughput Screening* **2018**, *21* (7), 476–486, DOI: 10.2174/1386207321666180821100713.

(28) Kuhlman, B.; Bradley, P. Advances in Protein Structure Prediction and Design. *Nat. Rev. Mol. Cell Biol.* **2019**, *20* (11), 681–697.

(29) Honke, K.; Tsuda, M.; Koyota, S.; Wada, Y.; Iida-Tanaka, N.; Ishizuka, I.; Nakayama, J.; Taniguchi, N. Molecular Cloning and Characterization of a Human  $\beta$ -Gal-3'-Sulfotransferase That Acts on Both Type 1 and Type 2 (Gal $\beta$ 1–3/1–4GlcNAc-R) Oligosaccharides. *J. Biol. Chem.* **2001**, *276* (1), 267–274.

(30) Shi, Y.-X.; Li, S.-H.; Zhao, Z.-P. Molecular Simulations of the Effects of Substitutions on the Dissolution Properties of Amorphous Cellulose Acetate. *Carbohydr. Polym.* **2022**, *291*, No. 119610.

(31) Martínez, L. Automatic Identification of Mobile and Rigid Substructures in Molecular Dynamics Simulations and Fractional Structural Fluctuation Analysis. *PLoS One* **2015**, *10* (3), No. e119264, DOI: 10.1371/journal.pone.0119264.

(32) Lobanov, M.-Y.; Bogatyreva, N.-S.; Galzitskaya, O.-V. Radius of Gyration as an Indicator of Protein Structure Compactness. *Mol. Biol.* **2008**, *42* (4), 623–628.

(33) Ali, S.; Hassan, M.; Islam, A.; Ahmad, F. A Review of Methods Available to Estimate Solvent-Accessible Surface Areas of Soluble Proteins in the Folded and Unfolded States. *Curr. Protein Pept. Sci.* **2014**, *15* (5), 456–476.

(34) Savojardo, C.; Manfredi, M.; Martelli, P.-L.; Casadio, R. Solvent Accessibility of Residues Undergoing Pathogenic Variations in Humans: From Protein Structures to Protein Sequences. *Front. Mol. Biosci.* **2021**, *7*, No. 626363.

(35) Ghahremanian, S.; Rashidi, M.-M.; Raeisi, K.; Toghraie, D. Molecular Dynamics Simulation Approach for Discovering Potential Inhibitors against SARS-CoV-2: A Structural Review. *J. Mol. Liq.* **2022**, *354*, No. 118901.

(36) Maisuradze, G.-G.; Liwo, A.; Scheraga, H.-A. Principal Component Analysis for Protein Folding Dynamics. *J. Mol. Biol.* **2009**, *385* (1), 312–329, DOI: 10.1016/j.jmb.2008.10.018.

(37) David, C.-C.; Jacobs, D.-J. Principal Component Analysis: A Method for Determining the Essential Dynamics of Proteins. In *Protein Dynamics, Methods in Molecular Biology*; Springer, 2014; Vol. 1084, pp 193–226.

(38) Hayward, S. A Retrospective on the Development of Methods for the Analysis of Protein Conformational Ensembles. *Protein J.* **2023**, *42* (3), 181–191.

(39) Cossio-Pérez, R.; Palma, J.; Pierdominici-Sottile, G. Consistent Principal Component Modes from Molecular Dynamics Simulations of Proteins. *J. Chem. Inf. Model.* **2017**, *57* (4), 826–834.

(40) Islam, R.; Parves, M. R.; Paul, A.-S.; Uddin, N.; Rahman, M.-S.; Mamun, A. Al.; Hossain, M.-N.; Ali, M.-A.; Halim, M.-A. A Molecular Modeling Approach to Identify Effective Antiviral Phytochemicals against the Main Protease of SARS-CoV-2. *J. Biomol. Struct. Dyn.* **2020**, *39* (9), 3213–3224, DOI: 10.1080/07391102.2020.1761883.

(41) Sakamoto, K.; Kayanuma, M.; Inagaki, Y.; Hashimoto, T.; Shigeta, Y. In Silico Structural Modeling and Analysis of Elongation Factor-1 Alpha and Elongation Factor-like Protein. *ACS Omega* **2019**, *4* (4), 7308–7316.

(42) Honke, K.; Tsuda, M.; Hirahara, Y.; Ishii, A.; Makita, A.; Wada, Y. Molecular Cloning and Expression of cDNA Encoding Human 3'-Phosphoadenylylsulfate:Galactosylceramide 3'-Sulfotransferase. *J. Biol. Chem.* **1997**, *272* (8), 4864–4868.

(43) Eckhardt, M.; Fewou, S.-N.; Ackermann, I.; Gieselmann, V. N-Glycosylation Is Required for Full Enzymic Activity of the Murine Galactosylceramide Sulphotransferase. *Biochem. J.* **2002**, *368* (Pt 1), 317–324.

(44) Geourjon, C.; Deleage, G. SOPMA: Significant Improvements in Protein Secondary Structure Prediction by Consensus Prediction from Multiple Alignments. *Bioinformatics* **1995**, *11*, 681–684.

(45) Schmidt, T.; Bergner, A.; Schwede, T. Modelling Three-Dimensional Protein Structures for Applications in drug Design. *Drug Discovery Today* **2014**, *19* (7), 890–897, DOI: 10.1016/j.drudis.2013.10.027.

- (46) Haddad, Y.; Adam, V.; Heger, Z. Ten Quick Tips for Homology Modeling of High-Resolution Protein 3D Structures. *PLoS Comput. Biol.* **2020**, *16* (4), No. e1007449.
- (47) Kuhlman, B.; Bradley, P. Advances in Protein Structure Prediction and Design. *Nat. Rev. Mol. Cell Biol.* **2019**, *20* (11), 681–697.
- (48) Singh, R.; Sledzieski, S.; Bryson, B.; Cowen, L.; Berger, B. Contrastive Learning in Protein Language Space Predicts Interactions between Drugs and Protein Targets. *Proc. Natl. Acad. Sci. U.S.A.* **2023**, *120* (24), No. e2220778120.
- (49) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F.-T.; De Beer, T.-A.-P.; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T. SWISS-MODEL: Homology Modelling of Protein Structures and Complexes. *Nucl. Acid Res.* **2018**, *46* (W1), W296–W303.
- (50) Schwede, T.; Kopp, J.; Guex, N.; Peitsch, M.-C. SWISS-MODEL: An Automated Protein Homology-Modeling Server. *Nucl. Acid Res.* **2003**, *31* (13), 3381–3385.
- (51) Yaghoofam, A.; Sorkalla, T.; Häberlein, H.; Gieselmann, V.; Kappler, J.; Eckhardt, M. Cerebroside Sulfotransferase Forms Homodimers in Living Cells. *Biochemistry* **2007**, *46* (32), 9260–9269.
- (52) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S.-A.-A.; Ballard, A.-J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A.-W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583–589.
- (53) Wiederstein, M.; Sippl, M.-J. ProSA-Web: Interactive Web Service for the Recognition of Errors in Three-Dimensional Structures of Proteins. *Nucleic Acid Res.* **2007**, *35*, W407–W410, DOI: [10.1093/nar/gkm290](https://doi.org/10.1093/nar/gkm290).
- (54) Liu, T.; Wang, Z. MASS: Predict the Global Qualities of Individual Protein Models Using Random Forests and Novel Statistical Potentials. *BMC Bioinform.* **2020**, *21* (4), No. 246, DOI: [10.1186/s12859-020-3383-3](https://doi.org/10.1186/s12859-020-3383-3).
- (55) Zha, M.; Wang, N.; Zhang, C.; Wang, Z. Inferring Single-Cell 3D Chromosomal Structures Based on the Lennard-Jones Potential. *Int. J. Mol. Sci.* **2021**, *22* (11), No. 5914, DOI: [10.3390/ijms22115914](https://doi.org/10.3390/ijms22115914).
- (56) Durham, E.; Dorr, B.; Woetzel, N.; Staritzbichler, R.; Meiler, J. Solvent Accessible Surface Area Approximations for Rapid and Accurate Protein Structure Prediction. *J. Mol. Model.* **2009**, *15* (9), 1093–1108.
- (57) Maisuradze, G.-G.; Liwo, A.; Scheraga, H.-A. Relation between Free Energy Landscapes of Proteins and Dynamics. *J. Chem. Theory Comput.* **2010**, *6* (2), 583–595, DOI: [10.1021/ct9005745](https://doi.org/10.1021/ct9005745).
- (58) Tavernelli, I.; Cotesta, S.; Di Iorio, E.-E. Protein Dynamics, Thermal Stability, and Free-Energy Landscapes: A Molecular Dynamics Investigation. *Biophys. J.* **2003**, *85* (4), 2641–2649.
- (59) Banushkina, P.-V.; Krivov, S.-V. High-Resolution Free Energy Landscape Analysis of Protein Folding. *Biochem. Soc. Trans.* **2015**, *43* (2), 157–161.
- (60) Amir, M.; Mohammad, T.; Kumar, V.; Alajmi, M.-F.; Rehman, M.-T.; Hussain, A.; Alam, P.; Dohare, R.; Islam, A.; Ahmad, F.; Hassan, M.-I. Structural Analysis and Conformational Dynamics of STN1 Gene Mutations Involved in Coat plus Syndrome. *Front. Mol. Biosci.* **2019**, *6*, No. 41, DOI: [10.3389/fmolb.2019.00041](https://doi.org/10.3389/fmolb.2019.00041).
- (61) David, C.-C.; Jacobs, D.-J. Principal Component Analysis: A Method for Determining the Essential Dynamics of Proteins. In *Protein Dynamics, Methods in Molecular Biology*; Springer, 2014; Vol. 1084, pp 193–226.
- (62) Kinoshita, K.; Murakami, Y.; Nakamura, H. EF-Seek: Prediction of the Functional Sites of Proteins by Searching for Similar Electrostatic Potential and Molecular Surface Shape. *Nucleic Acid Res.* **2007**, *35*, W398–W402, DOI: [10.1093/nar/gkm351](https://doi.org/10.1093/nar/gkm351).
- (63) Guo, Z.; Li, B.; Cheng, L.-T.; Zhou, S.; McCammon, J.-A.; Che, J. Identification of Protein-Ligand Binding Sites by the Level-Set Variational Implicit-Solvent Approach. *J. Chem. Theory Comput.* **2015**, *11* (2), 753–765.
- (64) Du, X.; Li, Y.; Xia, Y. L.; Ai, S. M.; Liang, J.; Sang, P.; Ji, X. L.; Liu, S. Q. Insights into Protein–Ligand Interactions: Mechanisms, Models, and Methods. *Int. J. Mol. Sci.* **2016**, *17* (2), No. 144, DOI: [10.3390/ijms17020144](https://doi.org/10.3390/ijms17020144).
- (65) Singh, N.; Patra, S. Phosphodiesterase 9: Insights from Protein Structure and Role in Therapeutics. *Life Sci.* **2014**, *106* (1–2), 1–11, DOI: [10.1016/j.lfs.2014.04.007](https://doi.org/10.1016/j.lfs.2014.04.007).
- (66) Pedersen, L.-C.; Petrotchenko, E.; Shevtsov, S.; Negishi, M. Crystal Structure of the Human Estrogen Sulfotransferase-PAPS Complex. *J. Biol. Chem.* **2002**, *277* (20), 17928–17932.
- (67) Liu, C.; Sheng, J.; Krahn, J.-M.; Perera, L.; Xu, Y.; Hsieh, P.-H.; Dou, W.; Liu, J.; Pedersen, L.-C. Molecular Mechanism of Substrate Specificity for Heparan Sulfate 2-O-Sulfotransferase. *J. Biol. Chem.* **2014**, *289* (19), 13407–13418.
- (68) Gesteira, T.-F.; Pol-Fachin, L.; Coulson-Thomas, V.-J.; Lima, M.-A.; Verli, H.; Nader, H.-B. Insights into the N-Sulfation Mechanism: Molecular Dynamics Simulations of the N-Sulfotransferase Domain of Ndst1 and Mutants. *PLoS One* **2013**, *8* (8), No. e70880, DOI: [10.1371/journal.pone.0070880](https://doi.org/10.1371/journal.pone.0070880).
- (69) Dudas, B.; Toth, D.; Perahia, D.; Nicot, A.-B.; Balog, E.; Miteva, M.-A. Insights into the Substrate Binding Mechanism of SULT1A1 through Molecular Dynamics with Excited Normal Modes Simulations. *Sci. Rep.* **2021**, *11* (1), No. 13129, DOI: [10.1038/s41598-021-92480-w](https://doi.org/10.1038/s41598-021-92480-w).
- (70) Hirschmann, F.; Krause, F.; Papenbrock, J. The Multi-Protein Family of Sulfotransferases in Plants: Composition, Occurrence, Substrate Specificity, and Functions. *Front. Plant Sci.* **2014**, *5*, No. 556, DOI: [10.3389/fpls.2014.00556](https://doi.org/10.3389/fpls.2014.00556).
- (71) Negishi, M.; Pedersen, L.-G.; Petrotchenko, E.; Shevtsov, S.; Gorokhov, A.; Kakuta, Y.; Pedersen, L.-C. Structure and Function of Sulfotransferases. *Arch. Biochem. Biophys.* **2001**, *390* (2), 149–157.
- (72) Gesteira, T.-F.; Pol-Fachin, L.; Coulson-Thomas, V.-J.; Lima, M.-A.; Verli, H.; Nader, H.-B. Insights into the N-Sulfation Mechanism: Molecular Dynamics Simulations of the N-Sulfotransferase Domain of Ndst1 and Mutants. *PLoS One* **2013**, *8* (8), No. e70880.
- (73) Dudas, B.; Toth, D.; Perahia, D.; Nicot, A.-B.; Balog, E.; Miteva, M.-A. Insights into the Substrate Binding Mechanism of SULT1A1 through Molecular Dynamics with Excited Normal Modes Simulations. *Sci. Rep.* **2021**, *11* (1), No. 13129, DOI: [10.1038/s41598-021-92480-w](https://doi.org/10.1038/s41598-021-92480-w).
- (74) Hema, K.; Ahamad, S.; Joon, H.-K.; Pandey, R.; Gupta, D. Atomic Resolution Homology Models and Molecular Dynamics Simulations of Plasmodium Falciparum Tubulins. *ACS Omega* **2021**, *6* (27), 17510–17522.