In Silico Structure Prediction of Human Fatty Acid Synthase–Dehydratase: A Plausible Model for Understanding Active Site Interactions

Arun John[1], Vetrivel Umashankar[1], A. Samdani[1], Manoharan Sangeetha[2,3], Subramanian Krishnakumar[2] and Perinkulam Ravi Deepa[3]

[1]Centre for Bioinformatics, Kamalnayan Bajaj Institute for Research in Vision and Ophthalmology, Vision Research Foundation, Sankara Nethralaya, Chennai, India. [2]Larsen and Toubro Department of Ocular Pathology, Kamalnayan Bajaj Institute for Research in Vision and Ophthalmology, Vision Research Foundation, Sankara Nethralaya, Chennai, India. [3]Department of Biological Sciences, Birla Institute of Technology and Science, Pilani, Rajasthan, India.

ABSTRACT: Fatty acid synthase (FASN, UniProt ID: P49327) is a multienzyme dimer complex that plays a critical role in lipogenesis. Consequently, this lipogenic enzyme has gained tremendous biomedical importance. The role of FASN and its inhibition is being extensively researched in several clinical conditions, such as cancers, obesity, and diabetes. X-ray crystallographic structures of some of its domains, such as β-ketoacyl synthase, acetyl transacylase, malonyl transacylase, enoyl reductase, β-ketoacyl reductase, and thioesterase, (TE) are already reported. Here, we have attempted an *in silico* elucidation of the uncrystallized dehydratase (DH) catalytic domain of human FASN. This theoretical model for DH domain was predicted using comparative modeling methods. Different stand-alone tools and servers were used to validate and check the reliability of the predicted models, which suggested it to be a highly plausible model. The stereochemical analysis showed 92.0% residues in favorable region of Ramachandran plot. The initial physiological substrate β-hydroxybutyryl group was docked into active site of DH domain using Glide. The molecular dynamics simulations carried out for 20 ns in *apo* and *holo* states indicated the stability and accuracy of the predicted structure in solvated condition. The predicted model provided useful biochemical insights into the substrate–active site binding mechanisms. This model was then used for identifying potential FASN inhibitors using high-throughput virtual screening of the National Cancer Institute database of chemical ligands. The inhibitory efficacy of the top hit ligands was validated by performing molecular dynamics simulation for 20 ns, where in the ligand NSC71039 exhibited good enzyme inhibition characteristics and exhibited dose-dependent anticancer cytotoxicity in retinoblastoma cancer cells *in vitro*.

KEYWORDS: human fatty acid synthase, dehydratase, comparative modeling, molecular docking, molecular dynamics simulation

CITATION: John et al. *In Silico* Structure Prediction of Human Fatty Acid Synthase–Dehydratase: A Plausible Model for Understanding Active Site Interactions. *Bioinformatics and Biology Insights* 2016:10 143–154 doi: 10.4137/BBI.S38317.

TYPE: Original Research

RECEIVED: April 01, 2016. **RESUBMITTED:** July 04, 2016. **ACCEPTED FOR PUBLICATION:** July 09, 2016.

ACADEMIC EDITOR: Thomas Dandekar, Associate Editor

PEER REVIEW: Nine peer reviewers contributed to the peer review report. Reviewers' reports totaled 2,415 words, excluding any confidential comments to the academic editor.

FUNDING: The work described in this article was supported by a research project grant from the Department of Biotechnology, Ministry of Science & Technology, New Delhi, for program support on Retinoblastoma (project no BT/01/CEIB/11/V/16, dated 08/05/2012). The authors confirm that the funder had no influence over the study design, content of the article, or selection of this journal.

COMPETING INTERESTS: Authors disclose no potential conflicts of interest.

CORRESPONDENCE: dipa.bits@gmail.com; drvus@snmail.org

COPYRIGHT: © the authors, publisher and licensee Libertas Academica Limited. This is an open-access article distributed under the terms of the Creative Commons CC-BY-NC 3.0 License.

Paper subject to independent expert blind peer review. All editorial decisions made by independent academic editor. Upon submission manuscript was subject to anti-plagiarism scanning. Prior to publication all authors have given signed confirmation of agreement to article publication and compliance with all applicable ethical and legal requirements, including the accuracy of author and contributor information, disclosure of competing interests and funding sources, compliance with ethical requirements relating to human and animal study participants, and compliance with any copyright requirements of third parties. This journal is a member of the Committee on Publication Ethics (COPE). Provenance: the authors were invited to submit this paper.

Published by Libertas Academica. Learn more about this journal.

Introduction

Type I fatty acid synthase (FASN), the cytosolic enzyme that synthesizes fatty acids, is present as a single integrated protein complex in animals.[1] The molecular weight of this multifunctional enzyme is 270 kDa and is arranged in head-to-head orientation forming a homodimer. Structurally, each monomer unit has N-terminal domains consisting of β-ketoacyl synthase, acetyl transacylase/malonyl transacylase, and dehydrase (DH), while the C-terminal domains are enoyl reductase, β-ketoacyl reductase, thioesterase, and acyl carrier protein (ACP). Catalytic activity of the enzyme is initiated when a saturated acyl moiety is attached to the cysteine group of β-ketoacyl synthase and the malonyl moiety is attached to 4-phosphopantetheine group of ACP, which acts as a flexible arm in substrate transfer.[2] In prokaryotes and plants, Type II FASN catalyzes fatty acid biosynthesis and is present as individual proteins.

DH domain functions by removing the hydroxyl group from the β-OH group from the fatty acid substrate carried by ACP to the different domains for the biosynthetic process producing fatty acids. DH domain shows conservation across species. DH domain contains conserved *hot dog fold* in which five-turn alpha helix acts as *sausage* covered by seven-stranded antiparallel beta-sheet as *bun*. The catalytic diad HIS^{878} and ASP^{1031} were found to be highly conserved in this DH domain.[3]

DH domains are classified into two different types, wherein animal FASN DH domain falls under the Type I category containing two nonidentical pseudo subunits. Type I DH domains contain the N-terminal region called DH1

containing the conserved HIS[878] residue involved in the abstraction of proton during reaction, whereas another important residue is ASP[1031], which is involved in the abstraction of hydroxyl from the β position and is present at another pseudo subunit of C-terminal region called DH2.[4,5] The GLN[1035] anchored to ASP[1031] present in the DH2 region helps in maintaining the catalytic diad at confined position for the proper enzymatic activity. Mutation at GLN[1035] led to a reduction in dehydratase (DH) enzyme activity; while mutation at ASP[1031] reduced the enzyme activity.[3] The tunnel with V-shaped pocket present between two long helices from each dimeric subunits makes the tunnel space accessible to the substrate elongation during further cycles of fatty acids synthesis. This tunnel space plays important role in DH activity depending on the chain length of the substrate, that is, higher substrate chain length will have higher activity.

The Type II bacterial FASN shows only very low sequence identity in some areas with that of mammalian counterpart but shares similar hot dog folds, which is closely related with bacterial fabA, fabZ DH,[6,7] and eukaryotic pseudodimeric enzymes.[8] Maier et al.[9] fitted the electron density map of mammalian FASN (porcine; 4.5 Å resolution) to that of dimeric *Escherichia coli* fabA (protein data bank [PDB] ID: 1MKA) and found that actual size of hot dog fold was much more than expected at sequence level, thereby reducing the interdomain region. The Type II DH domain in the bacteria and other species was found to contain two identical dimers having similar active site folds and tunnels as in the case of Type I DH domains. Two active sites were found in the bacterial DH proteins, while the second loop containing active site His residue was substituted with TRP and LYS, and then 30 amino acids truncation at N-terminus were found in mammalian pseudodomains.[9]

In the present study, we have compared FASN–DH sequence in humans with that in other organisms and identified the sequence region corresponding to DH. We have predicted the three-dimensional (3D) structure of the DH domain of human FASN (hFASN), using comparative modeling method that is reported in the previous studies on other proteins.[10,11] The predicted model was subjected to molecular docking using physiological substrate, and the stability of protein was confirmed using molecular dynamics (MD) simulation in *apo* and *holo* conditions. Finally, high-throughput virtual screening (HTVS) was carried out targeting the DH domain using Glide against National Cancer Institute (NCI) database of chemical ligands for identifying potential lead inhibitory molecules of FASN.

Materials and Methods

The complete experimental 3D structure of hFASN–DH has not yet been elucidated. Therefore, a computational modeling approach was used to predict the closest plausible 3D structure of hFASN–DH domain. The amino acids corresponding to DH domain was retrieved from UniProtKB/SWISS-PROT server (UniProt ID: P49327). The sequence position was identified as 312 amino acids (~824–1135). The retrieved sequence was subjected to BLASTP analysis against PDB database to find structurally similar template of more than 35% to perform comparative modeling. The crystal structure of porcine FASN (*Sus scrofa*, PDB ID: 2VZ8) was used as structural template for homology modeling. Previously, we had reported a similar computational approach for deciphering the plausible structure for hFASN enoyl reductase and β-ketoacyl reductase catalytic domain.[12]

Homology modeling of DH domain. Automated homology structure prediction software (MODELLER 9v7) scripts were used to generate sequence alignment between the target (hFASN–DH) and template (PDB ID: 2VZ8).[13] A total of 1000 initial models were generated using the aligned target–template file using MODELLER 9v7 scripts. Discrete optimized protein energy score was used to identify the best model from predicted initial structures.[14] The selected model was then subjected to further processing using *Protein Preparation Wizard* module with Maestro 9.3 (Schrödinger, LLC) for clearing any structural defects, optimizing for different protonation states, flipping and complete energy minimization using optimized potentials for liquid simulation (OPLS) 2005.

Model validation. Structural Analysis and Verification Server (SAVES) was used to validate the structural features of predicted model (http://nihserver.mbi.ucla.edu/SAVES). The overall quality and backbone conformation of torsion angles (phi [Φ] and psi [ψ]) were determined using Ramachandran plot analysis using PROCHECK,[15] which checks the geometrical orientation of each amino acid residues with respect to stereochemical parameters derived from experimentally proven high-resolution structures. ERRAT algorithm[16] was used to refine and rectify nonbonded interactions and to stabilize the interactions. Further MODREFINER algorithm[17] and MODELLER 9v7 scripts were used in building and refining the loop regions and removing the steric clashes. Additionally, the side chain torsional angle clashes (atomic bumps) were removed using WHATIF program.[18] Finally, ProQ algorithm was used to determine the overall quality of predicted model.[19] PDBSUM server was used to understand the complete topology and secondary structure formation of DH domain.[20]

MD simulation of homology models. Following previously reported optimized MD simulation protocols used in anticancer targets,[21,22] the predicted homology model for hFASN–DH domain was subjected to MD simulation using Desmond package (Desmond Molecular Dynamics System, version 3.1; D.E. Shaw Research; Maestro-Desmond Interoperability Tools, version 3.1; Schrödinger, LLC) with OPLS 2005 force field.[23] The starting structure was prepared using Protein Preparation Wizard (macro models) to ensure the correctness of assigning bond orders and bond lengths and to optimize by different protonation states and flipping

histidines, arginine, and glutamine. A system was setup with predefined water model as solvent (Simple Point Charge) in cubic box with periodic condition 10 Å × 10 Å × 10 Å distance. The complete system was neutralized by adding sodium ions. Steepest descent and the limited-memory Broyden–Fletcher–Goldfarb–Shanno algorithms were used in obtaining the local minimum for complete system. A constant temperature of 300 K and a pressure of 1 atm was used in complete simulation process using Berendsen thermostat and Berendsen barostat, respectively. Heavy atoms in solute were iterated with 2000 steps with convergence threshold as 1.0 kcal/mol/Å. After refining and obtaining equilibrium, Smooth Particle Mesh Ewald methods were used for handling long-range coulombic interactions using SHAKE algorithm.[24] The fully built system was set for simulation through 20 ns with NPT ensemble, where the trajectory sampling was done on every 1.0 ps. Furthermore, the root mean square deviation (RMSD) was calculated, and 2D graph was plotted (time – *X*-axis and deviation – *Y*-axis) for understanding the stability of *apo* and *holo* forms. Similarly, the radius of gyration (Rg) was calculated to understand the compactness of the complex and conformational changes. Additionally, root mean square fluctuation was calculated to understand the residue-wise changes and binding efficacy.

Electrostatic maps and active site prediction. Electrostatic potential maps were generated to understand the overall charge distribution on protein surface. A grid-based Poisson–Boltzmann equation using Maestro 9.3 (Maestro, version 9.3; Schrödinger, LLC) was used to calculate the overall charge. This map helps in understanding the interaction of substrate or small molecule with proteins. Pasta et al.[3] have done a multiple sequence alignment of DH domain and found that HIS^{878} from first pseudo subunit and ASP^{1031} of second subunit, which is anchored by GLN^{1035}), are highly conserved across different organisms and act as catalytic site.

Substrate docking and postdocking simulation. The initial biochemical substrate of FASN–DH is a beta hydroxyl butyryl molecule, and its structure was built using 2D draw (Maestro, version 9.3; Schrödinger, LLC). This four-carbon substrate was prepared as single, low energy 3D structure using LigPrep Module (LigPrep 2.6; Schrödinger, LLC), which also checks for various ionization states, tautomers, stereochemistry, and ring conformation. The prepared substrate (beta hydroxyl butyryl molecule) was used for molecular docking to corresponding hFASN–DH domain using Glide 5.8 (Ligand–Receptor docking software; Schrödinger, LLC). A grid box was set covering the volume of active site region with van der Waals radius scaling to 1.0 with partial cut-off of 0.25 to soften the potential for nonpolar parts of receptor, where other atoms are free of scaling. Glide standard precision mode, where the ligand sampling was set flexible, was used for the docking of substrate. Ligands were set to select only less than 300 atoms and less than 50 rotatable bonds with van der Waals scaling factor of 0.8 with partial cut-off of 0.15. Out of the 1000 poses generated per docking run, at least 10 poses per ligand were selected. The pose selection criterion was a minimum threshold of 0.5 kcal/mol in the postdocking minimization analysis. Glide score, an empirical scoring function, was used to find the affinity and binding mode of substrate, which uses OPLS 2005 force field as scoring function. The best docked substrate conformation for DH domain was subjected to MD simulation for 20 ns using previously mentioned simulation parameters and ensembles to understand the stability of the enzyme–substrate complexes.

Ligand preparation and HTVS. The complete NCI database compounds (as listed in May 2012) was optimized using LigPrep (Schrödinger, LLC), which uses Epik 2.0 with pH ranges 7.0 ± 2.0 to prepare each input ligand with various ionization states, tautomers, and ring conformation. Finally, a single energy minimized low energy conformer of each compound was obtained using OPLS 2005 force field. These compounds were docked into binding site of DH domains using HTVS protocols from Schrödinger. The optimized modeled structure was used to generate receptor grid using Grid-Based Ligand Docking with Energetics (Glide) to perform docking of NCI compounds to active site residues HIS^{878}, ASP^{1031}, and GLN^{1035}. The prepared ligands were filtered using different filter criteria. For the initial filtration, Lipinski rules and removal of reactive functional groups were implemented. HTVS protocol from Glide was used which efficiently enriches million compounds (50% of highly accurate and ranked compounds), to the standard precision mode for reliably docking hundreds of ligands with high accuracy (highly reliable 50% of compounds), to the extra precision (XP) mode, where further false positives elimination is accomplished by more intense sampling and more precised scoring. Thus, removing the false positives and highly efficient ligands with high scoring function (10%) were identified as best hits.

Absorption, distribution, metabolism, and excretion studies of ligands. Favorable absorption, distribution, metabolism, and excretion (ADME) properties are essential for pharmacological/clinical applications of identified ligands. Therefore, the ligands were evaluated for key physicochemical properties that are critical for ADME properties like molecular weight, solvent accessible area (SASA), computed dipole moment (dipole), predicted blood/brain partition coefficient, predicted central nervous system (CNS) activity, estimated number of hydrogen bond acceptor and hydrogen donor, van der Waals force field surface area on polar nitrogen and oxygen and volume to filter screened compounds using QikProp (Schrödinger, LLC).

Cell viability assessment by 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide assay. Cell viability was determined by 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) assay in ligand NSC71039-treated RB cells (WERI-RB-1). WERI-RB-1 cells (RIKEN BioResource Center) were maintained in RPMI 1640 medium (Sigma Aldrich), supplemented with 10% fetal

A

```
Query  1     PRGTPLISPLIKWDHSLAWDVPAAEDFPNGSGSPSAAIYNIDTSSESPDHYLVDHTLDGR  60
             PRGTPLISP IKWDHS AWDVP+A DFP+GS   S A+Y  D S ESPDHYLVDH +DGR
Sbjct  824   PRGTPLISPHIKWDHSQAWDVPSAADFPSGSSCSSVAVYKFDVSPESPDHYLVDHCIDGR  883

Query  61    VLFPATGYLSIVWKTLARALGLGVEQLPVVFEDVVLHQATILPKTGTVSLEVRLLEASRA  120
             VLFP TGYL + WKTLARAL   +E+ PVVFEDV LHQATILPKTGTVSLEVRLLEAS A
Sbjct  884   VLFPGTGYLWLTWKTLARALSQNLEETPVVFEDVTLHQATILPKTGTVSLEVRLLEASHA  943

Query  121   FEVSE-NGNLVVSGKVYQWDDPDPRLFDHPESPTP-NPTEPLFLAQAEVYKELRLRGYDY  178
             FEVS+ NG+L+ SGKVYQW+ PDP+LFD   +  P + T    L+Q +VYK+LRLRGYDY
Sbjct  944   FEVSDSNGSLIASGKVYQWESPDPKLFDTRAAVDPADSTAEFRLSQGDVYKDLRLRGYDY  1003

Query  179   GPHFQGILEASLEGDSGRLLWKDNWVSFMDTMLQMSILGSAKHGLYLPTRVTAIHIDPAT  238
             GP FQ +LE+ LEG+ GRL W D+WVSF+D ML MSIL   + GLYLPTR T+I IDP T
Sbjct  1004  GPFFQLVLESDLEGNRGRLQWNDSWVSFLDAMLHMSILAPGQLGLYLPTRFTSIRIDPVT  1063

Query  239   HRQKLYTLQDKAQVADVVVSRWLRVTVAGGVHISGLHTESAPRRQQEQQVPILEKFCFTP  298
             HRQKLYTLQD  Q ADVVV R L   VAGG    G H+  APRR QE   PILEKFCFTP
Sbjct  1064  HRQKLYTLQDTTQAADVVVDRNLNTVVAGGALFLGAHSSVAPRRPQEHLKPILEKFCFTP  1123

Query  299   HTEEGCLSERAALQ  312
             H E GCL+   ALQ
Sbjct  1124  HVESGCLAGNTALQ  1137
```

B

Figure 1. (**A**) Target–template alignment of hFASN–DH domain with mammalian FASN. (**B**) The 3D structure of hFASN–DH domain, helices shown in red, beta sheets in cyan, and turns in green color using maestro.

bovine serum (Sigma Aldrich), 20 µg/mL streptomycin, and 20 U/mL penicillin (Invitrogen), at 37 °C in 5% CO_2 atmosphere. For the viability assay, the cells were seeded in 96-well plates at a density of 1×10^4 cells/well. After 24 hours, the cells were treated with ligand NSC71039 (dissolved in ethanol) in a dosage range of 2.5–150 µg/100 µL well and cultured for 48 hours. Cell proliferation was determined by adding MTT (5 mg/mL) (Invitrogen), incubating the cells at 37 °C further for 4 hours, and then the precipitate was solubilized by adding 200 µL/well DMSO (Sigma-Aldrich) and shaken for 10 minutes. Absorbance in each well was measured at a wavelength of 490 nm with a microplate reader (Bio-Tek ELX800). Triplicate values were obtained from at least three independent experiments. Values were expressed as percentage of viable cells in ligand-treated versus control RB cells (test OD/control OD × 100). The 50% inhibitory concentration (IC_{50}) of the ligand was calculated using polynomial regression analysis using Microsoft Excel.

Results and Discussion

Target–template alignment. As per reports in Ref. 9, the hFASN–DH domain region spans from ~824 to 1135 of total 2500 amino acids per monomer. These sequence regions were retrieved from UniProt ID (P49327). Hence, homology-based comparative modeling was performed to obtain the protein model. BLASTP was used to identify similar templates for query sequence. The template was based on query coverage and similarity score. Accordingly, mammalian (porcine, *S. scrofa*) hFASN domain (Chain A) (PDB ID: 2vz8, 3.2 Å resolution, 1635–1860 region, 70% similarity, and 99% query coverage) was chosen for DH domain (Fig. 1A).

Homology modeling. The obtained crystal structure of mammalian FASN (PDB ID 2VZ8, from (*S. scrofa*)) was used as template for homology modeling of hFASN–DH domain. The pair-wise sequence alignment was generated using MODELLER 9v7 between target and template sequences. The best model was selected based on discrete optimized protein energy score from 1000 models generated for Human DH domain. Furthermore, the predicted model had similar secondary structure (folds and loops) similar to the template. The backbone alignment score was generated to understand the root mean deviation of predicted model using maestro (Maestro, version 9.3; Schrödinger, LLC) superimposing with respective templates. The generated alignment score was 0.1 for DH domain. The predicted structure of human DH, therefore, showed hot dog folds with two pseudo subunits linked by interlinking region. The catalytic active residues (HIS^{878} and ASP^{1031}) are in place from the first and second pseudo subunits, respectively. These active site residues are anchored by GLN^{1035}, which also falls in second pseudo subunit (Fig. 1B).

Secondary structure analysis. The secondary structure motif information was analyzed using PDBsum server (PROMOTIF program) for the selected model (http://www.ebi.ac.uk/pdbsum/). The overall secondary structure elements of DH domain include 32.1% of residues forming strands, 19.6% forming alpha helices, 3.8% forming 3–10 helix and 44.6% of residues forming other structures. The overall topology and secondary structure information of predicted model is shown in Figure 2.

Model validation. The best predicted model was subjected to different model refinement and evaluation tools to understand the inter-atomic configuration, backbone and side chain interactions of different amino acids. The initial Ramachandran plot for DH domain showed 1.5% residues in disallowed region. Further refinement of the predicted model was done using WHAT IF program to remove the atomic bumps. MODELLER 9v7 scripts was used for loop modeling and loop refinement and MODREFINER algorithm for removing the inter-steric clashes. The final Ramachandran plot showed no residues in disallowed region and an overall quality of 92.1% in allowed region (Fig. 3).

Stereochemical parameters for the modeled proteins were measured using PROCHECK G-factor, which is a measure of the proper dihedrals and covalent bond orders and gives an overall log-odd score. This suggests the quality of structure

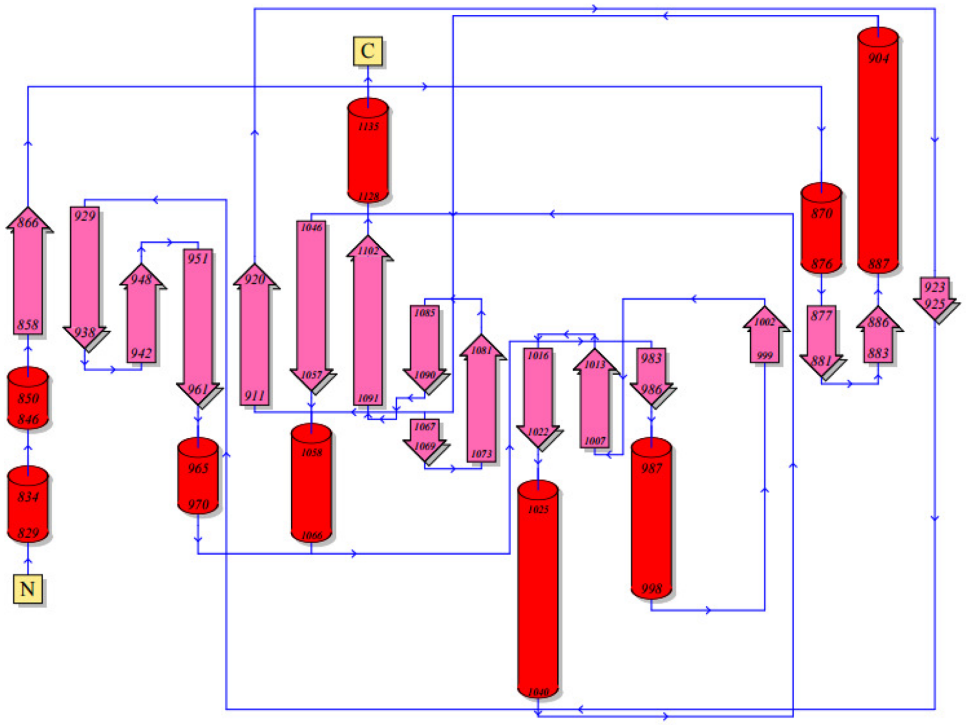

Figure 2. Overall topology of DH domain from N to C terminus using PDBSum.

prediction and ranges between 0 and 0.1. Negative G-factor score indicates the improper conformation of the residues, while the higher positive scores indicate proper conformation. In the present case of the predicted model of hFASN–DH domain, the overall log-odd score was found to be 0.06, which indicates the high plausibility of the structure with negligible low probable conformations in residues. ERRAT tool was used to analyze the nonbonded interactions, and the overall qual-

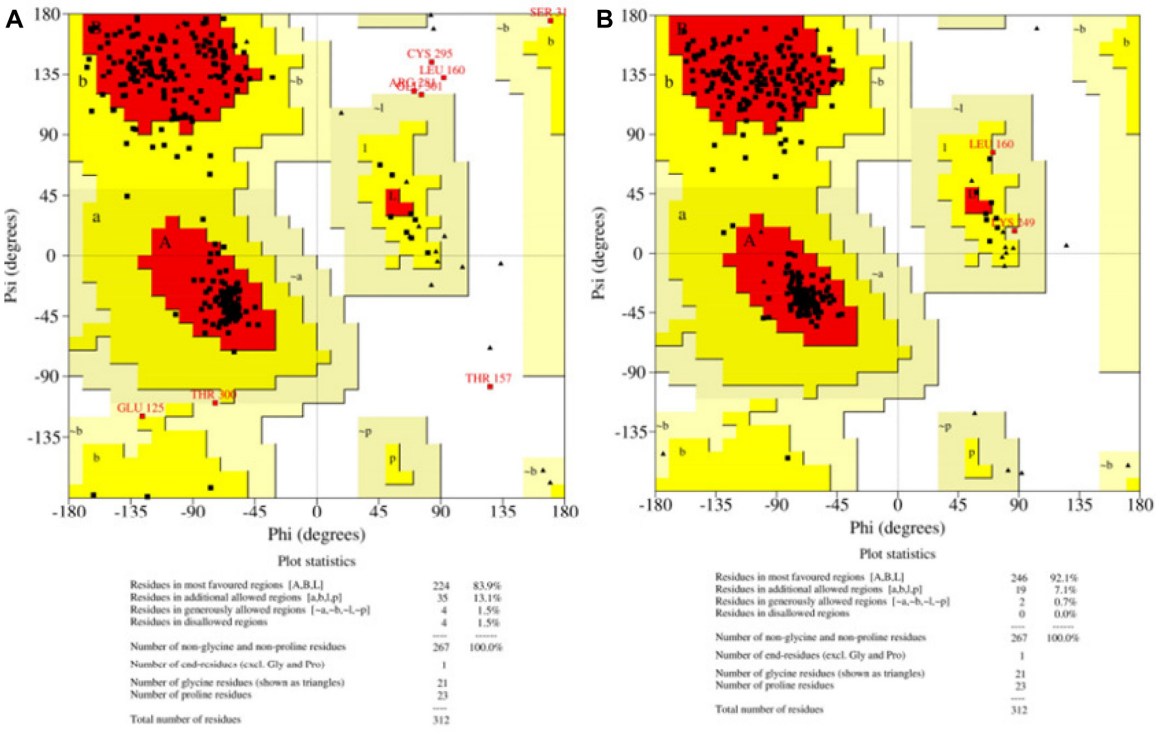

Figure 3. Ramachandran plots for the modeled DH domain (**A**, before refinement; **B**, after refinement).

ity score for predicted structure of DH domain was found to be 86.8 and the Verify3D score of 73.80, which is considered as a standard for good models. The standard quality score for ERRAT prediction is to be above 50.[15] To validate further, the modeled structure was also assessed for its quality by using ProQ server. The results indicate that the predicted LG score 4.101 (>4: extremely good model) and the predicted max subscore 0.185 (>0.1: fairly good model) of DH domain were to be in the acceptable range of good models. All these validations collectively confirm the correctness of secondary structure assignment for the predicted plausible model.

MD simulation of Apo DH domain. MD simulation of DH domain (Apo form) revealed that the predicted model remained stable in 20 ns production run. The total system was built for simulation with 4371 atoms in cubic box with using 13250 Simple Point Charge water as solvent. The system was neutralized by adding 8 Na^+ and equilibrated before production run. The physiological temperature and pressure was maintained as 300 K and 1 atm throughout the simulation process. The total energy and the potential energy were found to be −477,196.42 and −586,776.71 kJ/mol. The trajectory was analyzed for 20 ns and found to remain stable throughout simulation process. The RMSD was plotted for 20 ns simulation and remains stable from 10 ns to the completion of 20 ns simulation with minimal standard deviation (Fig. 6D). Similarly, the compactness of the predicted structure was analyzed using radius of gyration plot. The protein remained compact throughout the simulation process within the limit of 1 Å (Fig. 9A), thereby not requiring extended production run.

Electrostatic potential and active site prediction. Electrostatic potential maps, solvent excluded, solvent accessible were also predicted for the modeled structure. These maps help in understanding the overall charge distribution and solvent accessibility of the predicted model (Fig. 4A and B). This in turn helps in designing suitable ligands to active site cavity of predicted model. Previous studies point out that DH domain has conserved catalytic active site region formed by HIS^{878} from first pseudo subunit and ASP^{1031} of second subunit which is anchored by GLN^{1035} among all species.[9] Hence, the receptor grid box was set incorporating these residues along with other active site residues for docking.

Substrate docking and postdocking simulation. The β-hydroxybutyryl group was docked to modeled DH domain using Glide 5.8 (Schrödinger, LLC). The four-carbon β-hydroxybutyryl is the initial substrate for DH domain during the seven-cycle synthesis of 16-carbon palmitate. LigPrep 2.6 (Schrödinger, LLC) was used to build the four-carbon substrate and was subsequently docked to respective active site. The docking score for β-hydroxybutyryl-DH domain was −4.67 kcal/mol, and the docking energy was −14.07 kcal/mol. The oxygen group of the substrate shows strong hydrogen bonding with GLN^{1035} backbone atom, which assists in anchoring ASP^{888} that in turn is involved in substrate β-OH quenching. The substrate also forms hydrogen bond interactions with ALA^{888} and ASP^{1031} side chain residues, which are the main hydroxyl abstracting residues from the substrate during the catalytic reaction. The substrate also fitted well between the two nonidentical pesudodimeric domains along the helix tunnel. These interactions collectively infer that substrate to be docked well within the main catalytic diad residues that are involved in the enzymatic process, indicating the proper and stable docking of the substrate. Furthermore, few weak interactions were also observed in the active cavity bound with the substrate (Fig. 5A).

Now to understand the stability of holo enzyme and the most favorable binding mode with its substrate, MD simulation was carried out for 20 ns using Desmond. The protein showed negative charge of −8 and the substrate with 0. Therefore, the system was built and neutralized by adding eight Na^+ ions at 12.162 mM concentration. The total energy and the potential energy were found to be −665,529.48 and −816,135.84 kJ/mol. RMSD graphs were plotted to understand the changes in active site region of the protein after substrate docking. The RMSD mean of 2.689 Å with an SD of 0.298 Å was observed during the 20 ns simulation (Fig. 6A). The results showed no significant increase in deviation till the completion of 20 ns. Similarly, the substrate also remained intact with the protein throughout 20 ns simulation showing lesser deviation when comparing

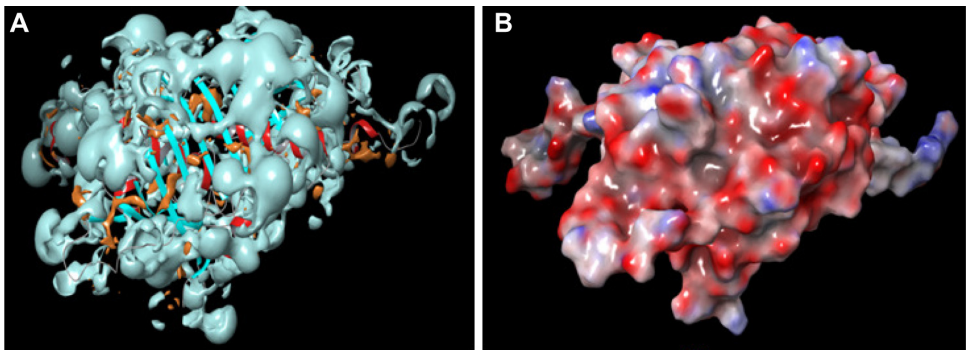

Figure 4. Overall charge distribution and solvent accessibility prediction. (**A**) Total solvent accessibility of modeled protein; cyan, hydrophobic; red, hydrophilic regions. (**B**) Electrostatic potential graph of modeled protein in different colors: red, negative; blue, positive; white, neutral.

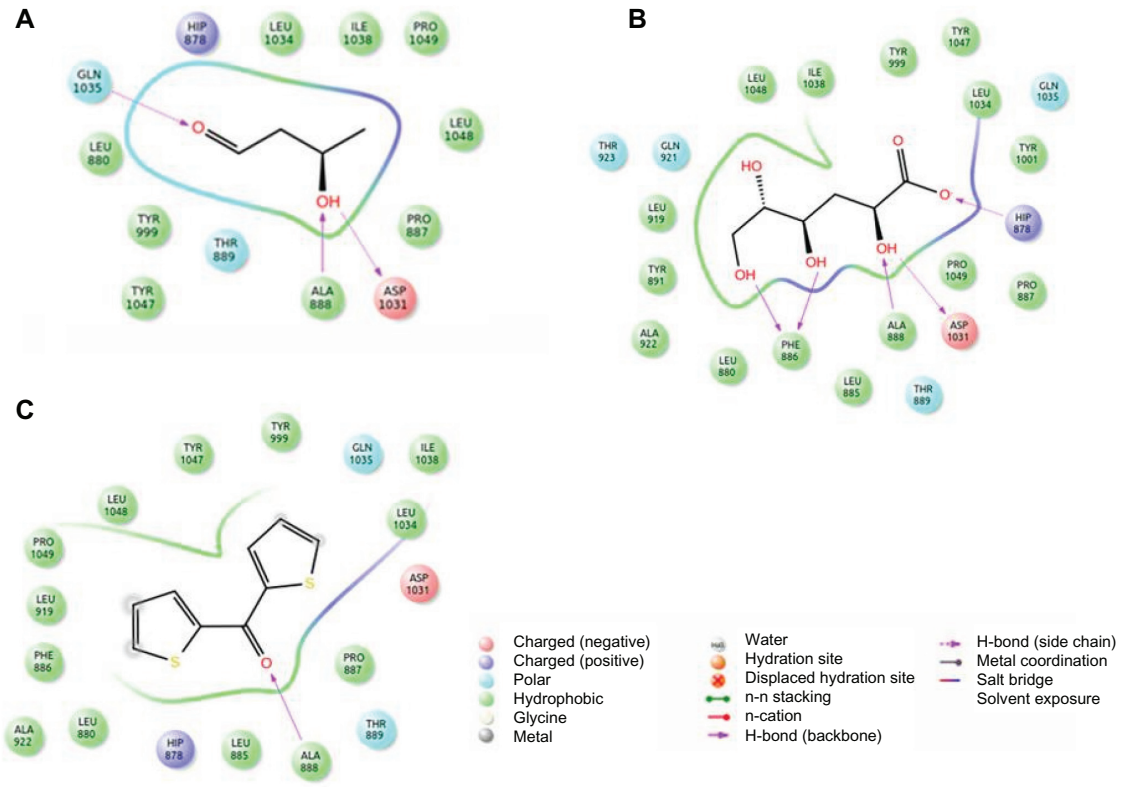

Figure 5. (**A**) Docked 2D interaction map of β-hydroxybutyryl substrate with DH domain. Amino acids are mentioned with three letter codes, and arrow indicates hydrogen bonding with GLN[1035], ALA[888], and ASP[1031]. (**B**) 2D interaction diagram of NSC71039 with protein arrow indicates H-bond with PHE[886], ALA[888], ASP[1031], and HIS[878]. (**C**) 2D interaction diagram of NSC201397 with protein arrow indicates H-bond with ALA[888].

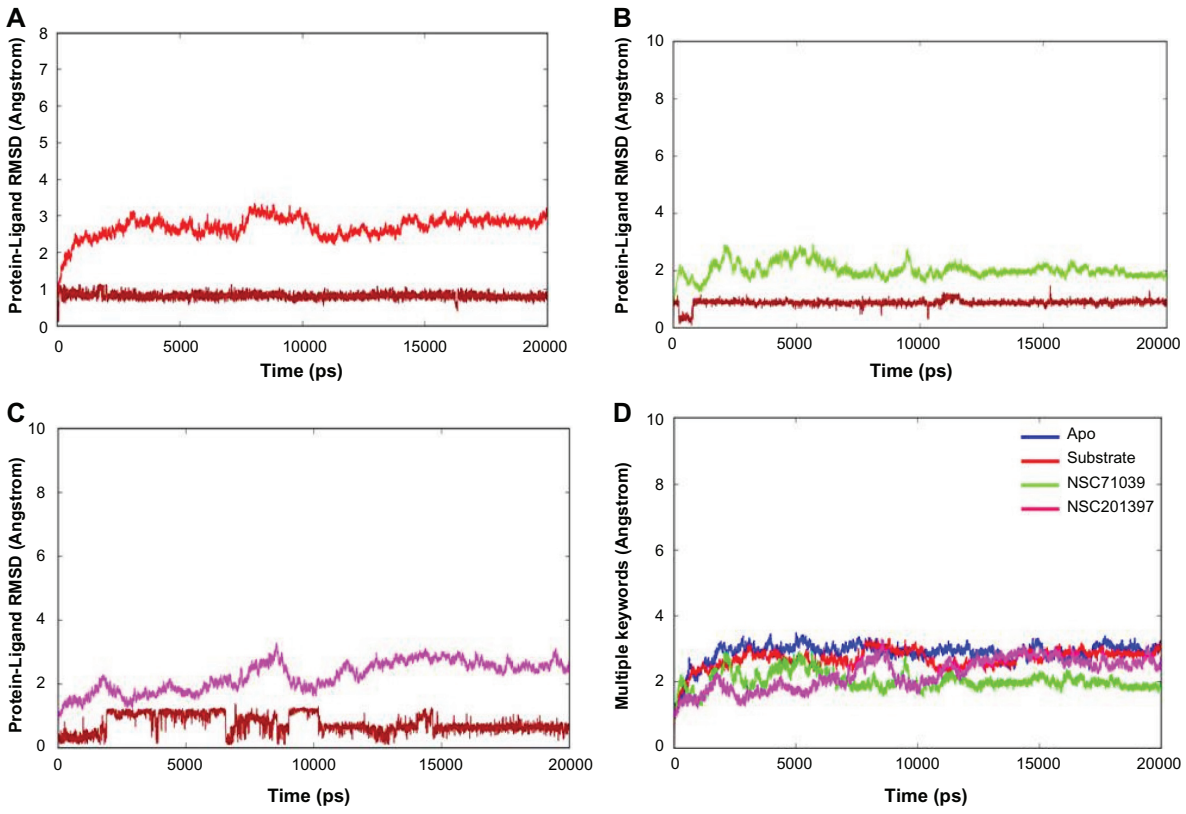

Figure 6. RMSD backbone plots for (**A**) β-hydroxybutyryl substrate–protein complex, (**B**) NSC71039–protein complex, and (**C**) NSC201397–protein complex. Respective ligands are shown in brown color. (**D**) Multiline RMSD plot for Apo (blue), substrate complex (red), NSC71039–protein complex (green), and NSC201397–protein complex (magenta).

with the *apo* form (Fig. 6D). This implies that the substrate has fitted well in its actual active site confining to the proper holo structure with the substrate complex. The substrate showed very good H-bond interaction with the important catalytic diad residue ASP[1031] (Fig. 5A). The radius of gyration was also calculated for the complex structures having 0.699 Å deviation with an SD of 0.123 Å, which showed the highly compact nature of these proteins during the entire 20 ns of MD simulation (Fig. 9B) Furthermore, the docked complexes showed very minimal deviation from their original conformations obtained from apo protein and were within the limit of 1 Å.

All these findings strongly suggested the high stability of predicted DH domain. Presubstrate binding and postsubstrate binding generated RMSF graphs that clearly showed the decrease in overall flexibility of the residues in *holo* condition (Fig. 10B). The RMSF analysis further revealed that no residues went beyond the value calculated for *apo* enzyme. All these results imply good quality of the predicted model, which can be used to understand structure–function relationships and new ligand identification. The backbone H-bond interaction with ALA[888] was observed with substrate and maintained for 92% of simulation time of 20 ns. Furthermore, side chain H-bond interaction with ASP[1031] was observed for 99% of simulation time and also with TYR[999]. Additionally, the presence of few hydrophobic interactions with the residues present in the tunnel region indicated the proper docking and placement of the substrate (similar to native state conformation at the start of the simulation period) (Figs. 7A and 8A).

HTVS and postdocking analysis of anti-cancer ligands. Postoptimization using LigPrep, the 231,617 compounds available in the NCI database were sequentially docked against human fatty acid DH domain using Glide. HTVS protocol was used to screen the ligands, where finally 153 top hit ligands were obtained. Subsequently, the ligands were ranked based on Glide-docking score, XP-H bond score, and binding free energy (Table 1). The top hit ligand NSC71039 had a docking score of –7.1 kcal/mol and MM-GBSA score of –35.275 kcal/mol, which had higher binding affinity when compared with the substrate. Similarly, NSC-201397showed good binding affinity than substrate with a XP score of –5.86 kcal/mol.

The ligand NSC71039 showed backbone H-bond with PHE[886], ALA[888], ASP[1031], and side-chain H-bond with HIS[878]. The ligand docked well with the important catalytic diad residues HIS[878] and ASP[1031] involved in substrate proton and β-OH quenching (Fig. 5B). The inhibitor has also fitted well into the helix tunnel residues mediated by hydrophobic interaction. The ligand NSC201397 showed backbone H-bond with ALA[888] and hydrophobic interactions with the tunnel residues (Fig. 5C).

Table 1. The XP G score and binding free-energy values of top two ligands.

NSC NUMBER	XP G SCORE (kcal/mol)	XP-HB SCORE (kcal/mol)	PRIME/MM-GBSA (kcal/mol)
71039	–7.18488	–4.24545	–35.275
201397	–5.86534	–0.7	–44.123

Table 2. ADME properties predicted using QikProp for the top hit ligands.

ADME PROPERTIES	NSC71039	NSC201397
Molecular weight	180.157	194.266
Volume	594.105	619.945
SASA	377.816	390.715
Dipole	3.611	3.117
HB donors	5	0
HB Acceptors	8	2
Qlog BB	–2.062	0.235
PSA	130.088	27.853
CNS	–2	1
Metab	4	2
Human Oral Absorption	2	3
Rule of five	0	0
Rule of three	1	0
Qlog HERG	–1.268	–4.452

Note: Descriptors' recommended ranges are as follows: molecular weight (130.0–725.0), volume (500–2000), SASA (300.0–1000.0), dipole (1.0–12.5), HB donors (0–6), HB acceptors (2–20), *Q* log BB (–3.0 to 1.2), PSA (7–200), CNS (–2[inactive] and +2[active]), Metab (1–8), human oral absorption (–1.5 to 1.5), rule of five (maximum is 4), rule of three (maximum is 3), *Q* log HERG (concern below –5).

ADME analysis. The important pharmacological and other physicochemical properties for the top hit ligands were performed using the QikProp. The results are shown in Table 2.

The ligand NSC71039 has predicted CNS value of –2 indicating inactivity to CNS, and good human oral absorption quality has been predicted. Many other parameters like metabolism, QlogBB, and QlogHERG found to be within the optimum range. The ligand has 0 violation in passing Lipinski's rule of 5 indicating a very good plausible drug candidate. Even though the second ligand NSC201397 showed 0 violation with Lipinski's and Jorgensen's rules, it showed high predicted activity to CNS. However, the other descriptors were favorable, such as no effect on HERG K^+ channel, minimal possible metabolic reactions, and good bioavailability on oral consumption.

MD simulation studies for the best inhibitors–enzyme complex. To analyze the stability of the enzyme-ligand complex, MD simulation were performed for 20 ns using Desmond. The simulation was carried out for ligand NSC71039 and NSC201397.

The protein-ligand NSC71039 complex showed the protein charge of –8 and ligand charge of –1. So the system was equilibrated by adding nine Na^+ ions of concentration

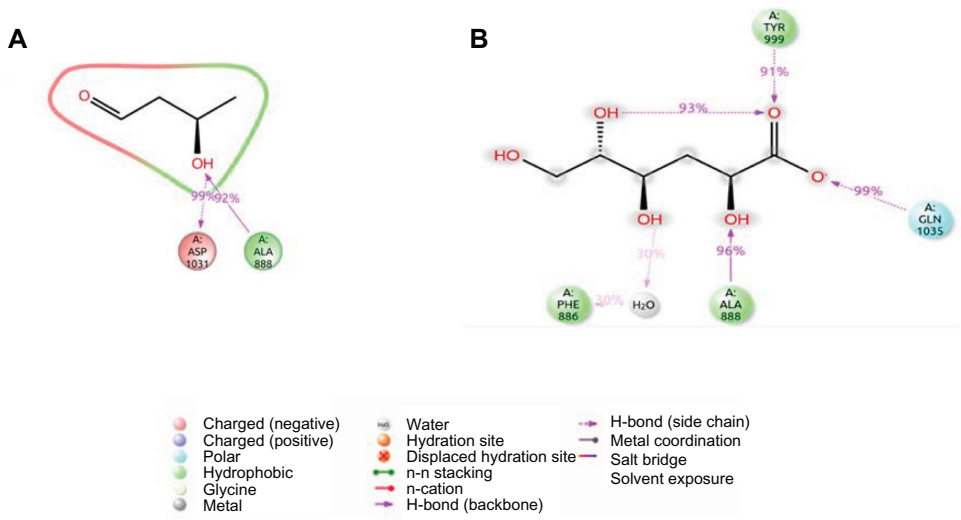

Figure 7. (**A**) 2D interaction diagram of β-hydroxybutyryl substrate with protein during simulation period. The arrow indicates the hydrogen bonding with ASP[1031] and ALA[888]. (**B**) 2D interaction diagram of ligand NSC71039 with protein during simulation period. The arrow indicates the hydrogen bonding with TYR[999], GLN[1035], and ALA[888].

12.35 mM. Energy minimization and dynamic simulation were performed for 20 ns. The RMSD backbone deviation of 2.94 Å was observed, and followed by the mean RMSD of 2.02 Å with an SD of 0.4 Å was observed during 20 ns simulation period for the DH protein (Fig. 6B). The ligand RMSD was found to have lesser deviation with the protein backbone indicating that protein and ligand are in stable conformation without much deviation during simulation period. The interaction diagram of ligand with the protein residues during simulation indicates that 91% H-bond interaction with the side chain of TYR[999], 99% side chain H-bond with GLN[1035], which anchors one of the main catalytic diad residues ASP[1031] in the native state for the proper catalytic activity, and 96% H-bond with backbone of ALA[888] (Fig. 7B).[5] Ligand contact map also indicated that ionic

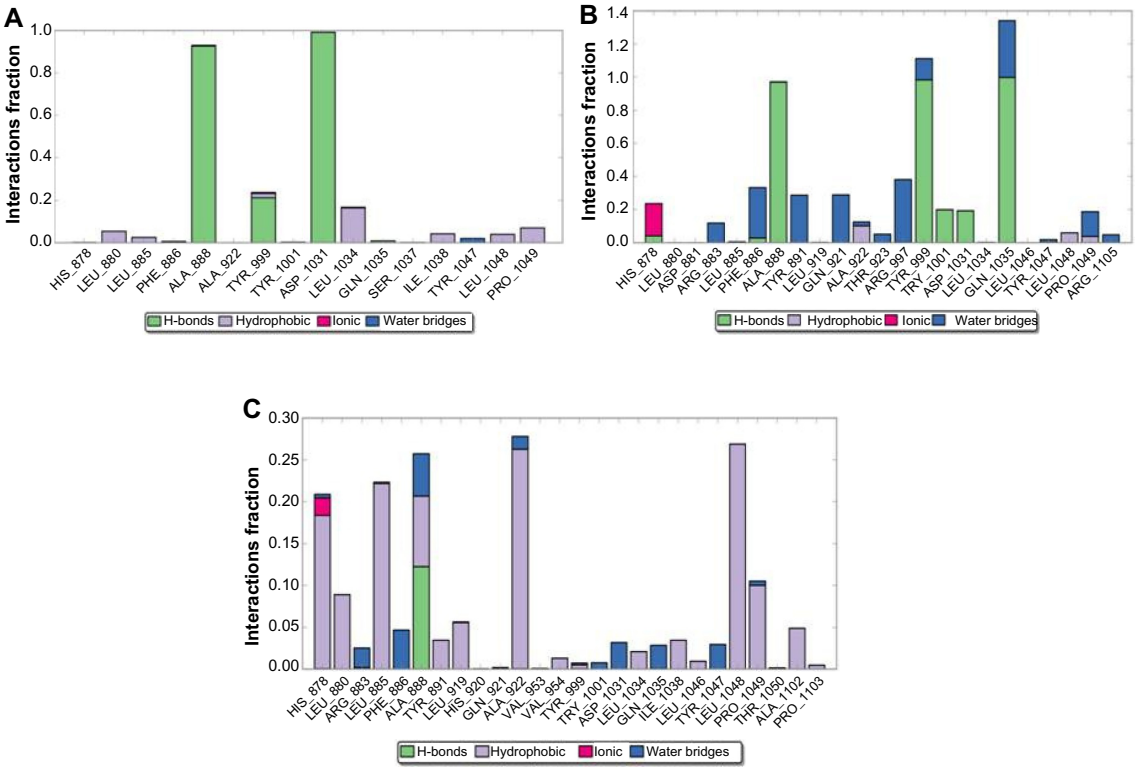

Figure 8. Timeline representations for the (**A**) β-hydroxybutyryl substrate, (**B**) NSC71039, and (**C**) NSC201397 interactions with the DH domain. The bar indicates the fold of interaction fraction and contacts (H bonds, hydrophobic, ionic, water bridges) with the residues in DH protein.

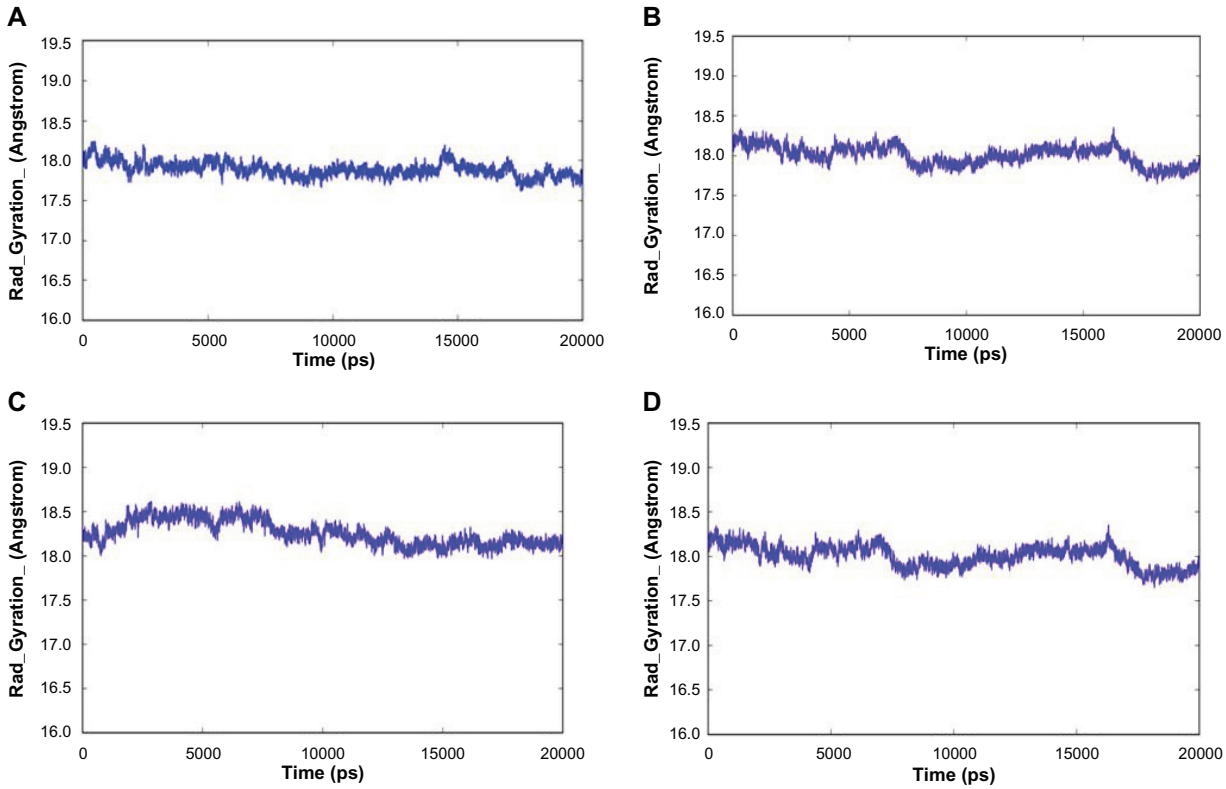

Figure 9. Radius of gyration showing protein compactness for (**A**) Apo, (**B**) β-hydroxybutyryl substrate bound protein complex, (**C**) NSC71039–protein complex, and (**D**) NSC201397–protein complex showing less than 1 Å deviation during simulation period.

interaction were observed with His^{878}, an important substrate proton quencher and H-bond contacts with ALA^{888}, TYR^{999}, TYR^{1001}, ASP^{1031}, and GLN^{1035} (Fig. 8B). More water bridges mediated ligand and protein interaction stabilization with residues ARG^{883}, PHE^{886}, TYR^{891}, GLN^{921}, ARG^{997}, and PRO^{1049}. NSC71039–enzyme complex RMSF was found to be lesser than 1.2 Å. The RMSF plot indicates that important residues and helix tunnel involved in the substrate binding remains to be less fluctuated (Fig. 10C). The Rg deviation was found to be 0.639 Å with an SD of 0.141 Å indicating that enzyme complex is in good compact state without much deviation during the simulation with the ligand–NSC71039 (Fig. 9C).

Using simulation studies, the second selected inhibitor NSC201397 was studied for its complex with the DH enzyme over 20 ns. The protein had a charge of −8 and ligand with 0 charge, so eight Na^+ ions were added with concentration of 10.977 mM. Energy minimization was carried out and proceeded to dynamic simulation for 20 ns. The RMSD graph for the complex indicates converging after 10 ns throughout the simulation period. The backbone RMSD shows the deviation of 3.2 Å with a mean and SD of 2.243 and 0.474 Å, respectively (Fig. 6C). The ligand RMSD shows lesser deviation with the protein indicating a good binding nature of the ligand within the protein active site without much deviation from its positioning. The ligand contacts with protein indicated that even though the docking score was higher when comparing with the substrate, no stable H-bond interaction >30% was formed with any of the important residue involved in catalysis or also with the tunnel region (Results not shown). Most of the hydrophobic contacts were formed with HIS^{878}, LEU^{885}, TYR^{891}, LEU^{919}, ALA^{922}, LEU^{1048}, PRO^{1049}, and ALA^{1102} (Fig. 8C). These hydrophobic interactions and fewer water bridges might have provided this ligand a docking score more than the substrate. The RMSF of the complex indicated no more fluctuation at the active site residues. The Rg deviation was found to be very much low of 0.616 Å with an SD of 0.098 Å indicating very good compact complex structure (Fig. 9D).

Even though the inhibitor NSC201397 showed good docking scores, RMSD, and protein compactness, it was found that this ligand was not forming any proper interaction with the important catalytic residues. However, in the case of NSC71039, increased hydrophobic interactions in the tunnel region would have provided better binding affinity in comparison with the substrate and also good compactness and RMSD stability.[5] The RMSD deviation for NSC71039 was found to be lesser and formed more stable complex than the substrate (Fig. 6D). It also showed lesser fluctuation on backbone residues during simulation time on comparing with substrate bound conformations (Fig. 10C). Comparatively, the presence of longer aliphatic carbon chain in the ligand NSC71039 (in comparison with other ligand NSC201397 as well as the four-carbon substrate tested here) may have favored better affinity with the DH domain.[3] The ligand NSC71039 also formed good H-bond interactions with the important residues like GLN^{1035} and ionic interaction with HIS^{878} a substrate, proton

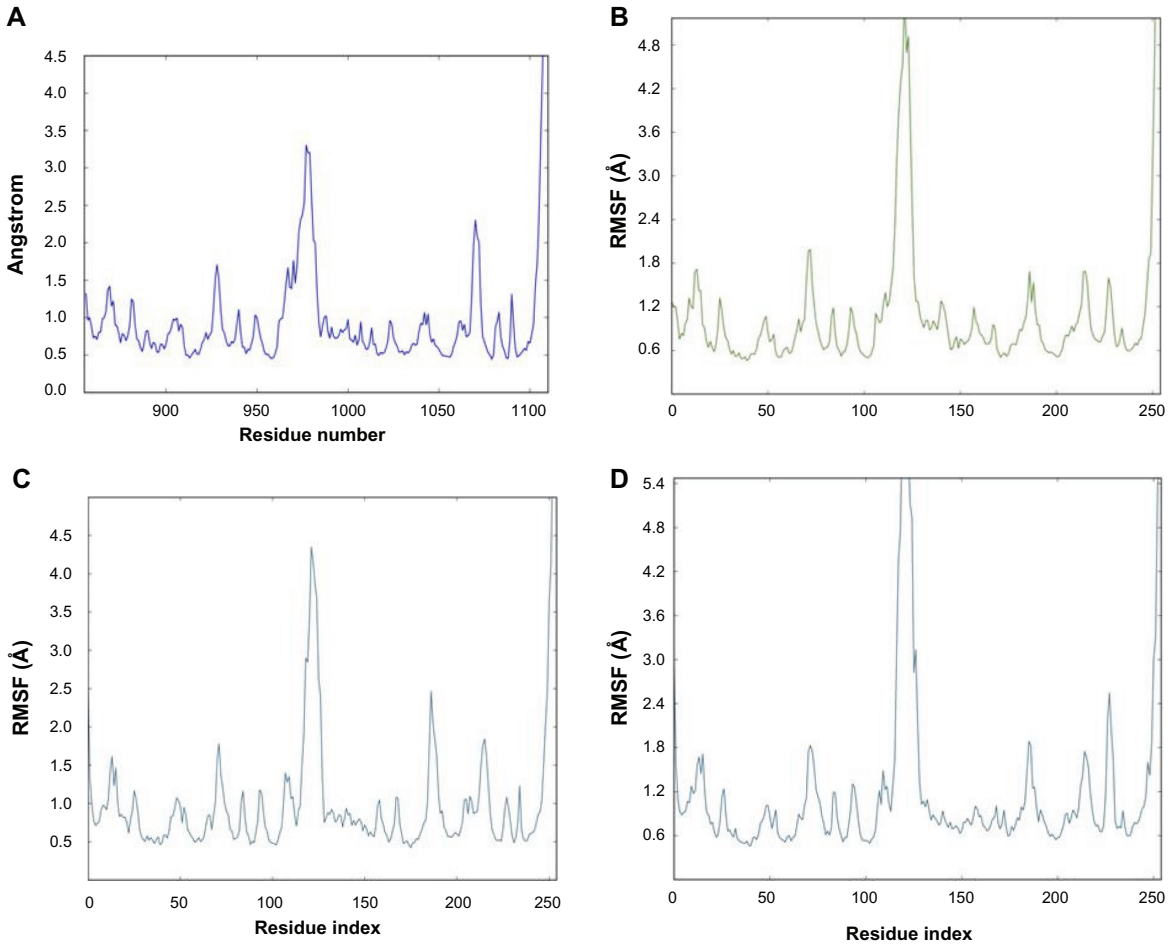

Figure 10. RMSF graph for (**A**) Apo, (**B**) substrate-docked complex, (**C**) NSC71039–protein complex, and (**D**) NSC201397–protein complex during 20 ns simulation period.

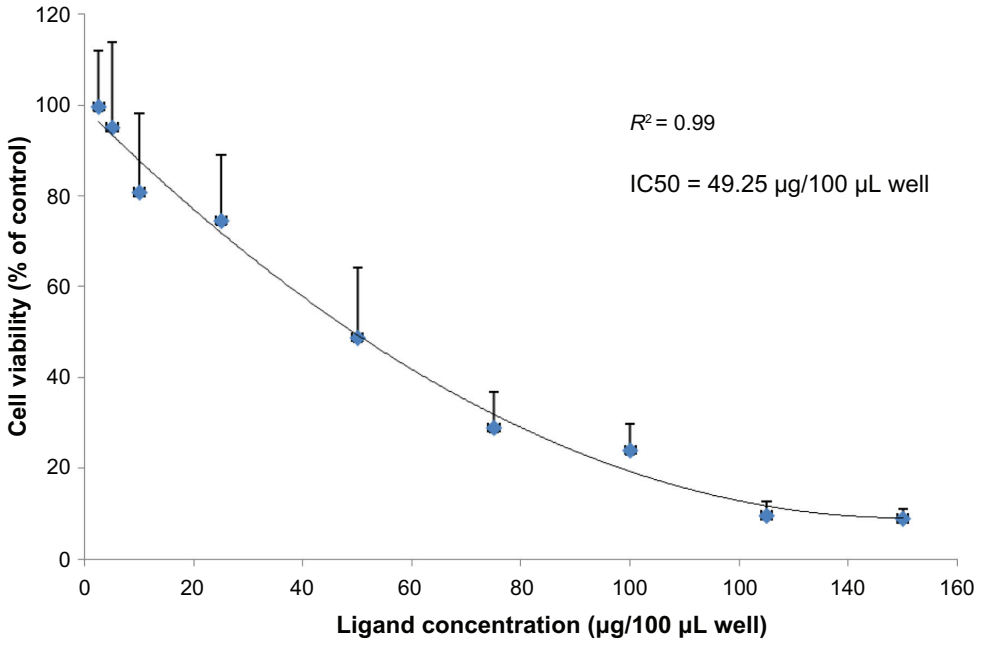

Figure 11. Dose-dependent cytotoxicity of ligand NSC71039 in RB cancer cells. Cytotoxic effect was evaluated by MTT assay. Values are expressed as mean ± SD of at least three separate experiments performed in triplicate.

quencher.[9] Additionally, many residues along the helix tunnel accommodate the growing fatty acid chain during the enzymatic process. Overall, the ligand NSC71039 satisfied most of the necessary attributes of a good inhibitor and, therefore, is suggested as a potential hFASN–DH domain inhibitor.

Anti-cancer effect of the top hit ligand. This inhibitor ligand NSC71039 was tested for its anti-cancer effect in retinoblastoma cancer cells *in vitro*. Dose-dependent cytotoxicity was observed when the cancer cells were treated with the increasing concentrations of ligand NSC71039 (Fig. 11). The IC_{50} was calculated to be 49.25 µg/µL well for NSC71039 in WERI-RB-1 cancer cells.

Conclusion

In the absence of a crystal structure for hFASN–DH enzyme domain, a comparative modeling approach was used here to prepare its structural model *in silico*. The highly plausible model generated in this study was validated for its stability using MD studies. Docking analysis of DH domain with its biochemical substrate yielded insights on its active site interactions with its substrate. Furthermore, chemical ligands listed in the NCI database were screened (by HTVS) for docking efficacy and tested for ADME properties. Among the best hit inhibitors, the ligand NSC71039 (3-deoxyhexonic acid) showed higher docking score of –7.18 kcal/mol and showed very good interactions with the GLN^{1035}, which involves in anchoring the substrate abstracting residue ASP^{1031}, ionic interaction with the important substrate proton quencher His^{878} showing contacts over the simulation period of 20 ns. This inhibitory ligand showed anticancer cytotoxicity in FASN-expressing cancer cells *in vitro*. Thus, the *in silico* structure prepared for hFASN–DH will be useful for identifying novel anti-FASN inhibitors that can be further tested for biologic and pharmacologic properties.

Author Contributions

Conceived and designed the experiments: PRD, VU, SK. Analyzed the data: AJ, AS, VU, MS, PRD, SK. Wrote the first draft of the manuscript: AJ, AS, VU, MS, PRD, SK. Contributed to the writing of the manuscript: AJ, AS, VU, MS, PRD, SK. Agree with manuscript results and conclusions: AJ, AS, VU, MS, PRD, SK. Jointly developed the structure and arguments for the paper: AJ, AS, VU, MS, PRD, SK. Made critical revisions and approved final version: AJ, AS, VU, MS, PRD, SK. All authors reviewed and approved of the final manuscript.

REFERENCES

1. Wakil SJ. Fatty acid synthase, a proficient multifunctional enzyme. *Biochemistry*. 1989;28(11):4523–30.
2. Menendez JA, Lupu R. Fatty acid synthase and the lipogenic phenotype in cancer pathogenesis. *Nat Rev Cancer*. 2007;7(10):763–77.
3. Pasta S, Witkowski A, Joshi AK, Smith S. Catalytic residues are shared between two pseudosubunits of the dehydratase domain of the animal fatty acid synthase. *Chem Biol*. 2007;14(12):1377–85.
4. Helmkamp GM, Bloch K. β-Hydroxydecanoyl Thioester Dehydrase studies on molecular structure and active site. *J Biol Chem*. 1969;244(21):6014–22.
5. Akey DL, Razelun JR, Tehranisa J, Sherman DH, Gerwick WH, Smith JL. Crystal structures of dehydratase domains from the curacin polyketide biosynthetic pathway. *Structure*. 2010;18(1):94–105.
6. Kimber MS, Martin F, Lu Y, et al. The structure of (3R)-hydroxyacyl-acyl carrier protein dehydratase (FabZ) from *Pseudomonas aeruginosa*. *J Biol Chem*. 2004;279(50):52593–602.
7. Leesong M, Henderson BS, Gillig JR, Schwab JM, Smith JL. Structure of a dehydratase–isomerase from the bacterial pathway for biosynthesis of unsaturated fatty acids: two catalytic activities in one active site. *Structure*. 1996;4(3):253–64.
8. Joshi AK, Smith S. Construction, expression, and characterization of a mutated animal fatty acid synthase deficient in the dehydrase function. *J Biol Chem*. 1993;268(30):22508–13.
9. Maier T, Leibundgut M, Ban N. The crystal structure of a mammalian fatty acid synthase. *Science*. 2008;321(5894):1315–22.
10. Anupriya G, Roopa K, Basappa S, Chong YS, Annamalai L. Homology modeling and in silico screening of inhibitors for the substrate binding domain of human Siah2: implications for hypoxia-induced cancers. *J Mol Model*. 2011;17(12):3325–32.
11. Semighini EP, Taft CA, Silva CH. Homology modelling, virtual screening and molecular dynamics of the MARK3 KA1 domain for cancer drug design. *Mol Simul*. 2011;37(14):1186–206.
12. John A, Umashankar V, Krishnakumar S, Deepa PR. Comparative modeling and molecular dynamics simulation of substrate binding in human fatty acid synthase: enoyl reductase and β-ketoacyl reductase catalytic domains. *Genomics Inform*. 2015;13(1):15–24.
13. Martí-Renom MA, Stuart AC, Fiser A, Sánchez R, Melo F, Šali A. Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct*. 2000;29(1):291–325.
14. Webb B, Sali A. Comparative protein structure modeling using MODELLER. *Curr Protoc Bioinformatics*. 2014;47:5.6.1–32.
15. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK – a program to check the stereochemical quality of protein structures. *J App Cryst*. 1993;26:283–91.
16. Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci*. 1993;2(9):1511–9.
17. Xu D, Zhang Y. Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. *Biophys J*. 2011;101(10):2525–34.
18. Hekkelman ML, Te Beek TA, Pettifer SR, Thorne D, Attwood TK, Vriend G. WIWS: a protein structure bioinformatics web service collection. *Nucleic Acids Res*. 2010;38:W719–23.
19. Wallner B, Elofsson A. Can correct protein models be identified? *Protein Sci*. 2003;12(5):1073–86.
20. de Beer TAP, Berka K, Thornton JM, Laskowski RA. PDBsum additions. *Nucleic Acids Res*. 2014;42:D292–6.
21. Reddy SVG, Reddy KT, Kumari VV, Basha SH. Molecular docking and dynamic simulation studies evidenced plausible immunotherapeutic anticancer property by Withaferin A targeting indoleamine 2, 3-dioxygenase. *J Biomol Struct Dyn*. 2015;33(12):2695–709.
22. Sarvagalla S, Cheung CHA, Tsai JY, Hsieh HP, Coumar MS. Disruption of protein–protein interactions: hot spot detection, structure-based virtual screening and in vitro testing for the anti-cancer drug target–survivin. *RSC Adv*. 2016;6(38):31947–59.
23. Bowers KJ, Chow E, Xu H, et al. Scalable algorithms for molecular dynamics simulations on commodity clusters. *Proceedings of the ACM/IEEE Conference on Supercomputing (SC06), 2006 Nov 11–17; Tampa, FL, USA*. New York: Association for Computing Machinery; 2006:43.
24. Darden T, Perera L, Li L, Pedersen L. New tricks for modelers from the crystallography toolkit: the particle mesh Ewald algorithm and its use in nucleic acid simulations. *Structure*. 1999;7(3):R55–60.