Received: 10 June 2024 | Revised: 10 October 2024 | Accepted: 22 January 2025 | Published online: 5 February 2025

RESEARCH ARTICLE

Medinformatics
2026, Vol. 3(1) 74–87
DOI: 10.47852/bonviewMEDIN52023586

BON VIEW PUBLISHING

Elucidating and Confirming the Substrate and Cofactor Binding Site of RetSat and Identifying the Optimal Cofactor for Retinol Saturase – A Computational Insight

Virangi Tharika Hewage[1,2], Ranga Srinath Jayakody[1,2,*], Anura Jayasooriya[3], Tharindu Senapathi[2] and Krishna Govendar[4]

[1]*Center for Scientific Computing and Advanced Drug Discovery, University of Sri Jayewardenepura, Sri Lanka*

[2]*Department of Chemistry, University of Sri Jayewardenepura, Sri Lanka*

[3]*Department of Basic Veterinary Sciences, University of Peradeniya, Sri Lanka*

[4]*Department of Chemical Sciences, University of Johannesburg, South Africa*

Abstract: Retinol saturase (RetSat) is an enzyme that inhibits the conversion of all-trans-retinol to all-trans-retinoic acid (ATRA), which is crucial for protecting cells against malignancy. Despite its significance, no effective inhibitors targeting RetSat currently exist, and limited knowledge about its structure, cofactor, and substrate binding has hindered inhibitor design. In this study, we utilized computational techniques to reconstruct and validate the 3D atomic structure of RetSat, identifying its cofactor and substrate binding sites. NADH was determined to be the optimal cofactor and successfully docked into its binding site, followed by docking of retinol into the substrate-specific site. Analysis of the equilibrated complexes revealed key amino acids involved in cofactor and substrate binding, enabling the identification of potential catalytic residues. This study provides validated insights into RetSat's structure and interactions, offering a foundation for designing RetSat inhibitors as potential anticancer agents and advancing the understanding of its role in cancer biology, thus paving the way for targeted therapeutic strategies.

Keywords: retinol saturase, NADH docking in RetSat, retinol-ATRA conversion inhibition, RetSat 3D atomic structure, RetSat inhibitors for cancer therapy

1. Introduction

Retinol saturase (RetSat), also known as all-trans-13, 14-dihydroretinol saturase, is an essential member of the oxido-reductase enzyme family, pivotal for retinoid metabolism [1–3]. While present throughout the body, RetSat exhibits prominent expression in the liver, kidney, intestine, and adipose tissues [1]. Its physiological functions encompass crucial roles in regulating lipid metabolism and adipogenesis, including the modulation of important transcriptional regulators such as PPARα and PPARγ, which govern lipid metabolism pathways [4–7]. Additionally, RetSat contributes to the generation of reactive oxygen species, serving as a cellular response to oxidative stress [8].

RetSat is recognized for its selective action on all-trans-retinol as its primary substrate [1, 9]. This enzymatic activity plays a crucial role in the conversion of all-trans-retinol to all-trans-retinoic acid, a bioactive compound [1, 10]. The conversion process involves two oxidation steps and has demonstrated efficacy in cancer treatment by inhibiting cell proliferation markers and inducing apoptosis in cancer cells [11]. This is achieved through the activation of retinoic acid receptors (RAR) or retinoid x receptors (RXR) bound to all-trans-retinoic acid [12–14]. The binding of these receptors to retinoic acid response elements (RARE) leads to downstream gene expression, which can result in various cellular outcomes such as proliferation, arrest, or apoptosis [12, 15].

As an oxidoreductase enzyme, RetSat plays a unique role in retinoid metabolism by saturating the double bond between carbons 13 and 14 in all-trans-retinol [1, 3, 5, 10]. This enzymatic action impedes the conversion of all-trans-retinol to all-trans-retinoic acid, resulting in the formation of a novel metabolite called all-trans-13,14-dihydroretinol [9, 16]. Despite its formation, the precise mechanism and function of this metabolite remain unknown [1, 14]. It has been hypothesized that modulating RetSat activity could potentially salvage all-trans-retinol and increase the pool of all-trans-retinoic acid in the body, which may have beneficial implications in inhibiting carcinogenesis [9, 10].

***Corresponding author:** Ranga Srinath Jayakody, Center for Scientific Computing and Advanced Drug Discovery, University of Sri Jayewardenepura and Department of Chemistry, University of Sri Jayewardenepura, Sri Lanka. Email: ranga@sci.sjp.ac.lk

Consequently, further investigations into RetSat are crucial to unravel its role in cellular metabolism and evaluate its potential as a target for cancer therapy [17].

Recent research indicates that RetSat exhibits sequence similarity with the carotenoid isomerase enzyme (CRTISO) found in plants and cyanobacteria, as well as a relationship with the γ-carotenoid desaturase [1]. The human and rat versions of RetSat share a remarkable 90% sequence identity, with the human sequence consisting of 610 amino acids [1]. Analysis of the RetSat sequence has revealed a putative dinucleotide-binding motif, represented as where U denotes a hydrophobic residue and X represents any other residue. $U_4G(G/A)GUXGLX_2(A/S)X_2L(X_{6\text{-}12})UX(L/V)UE(X_4)UGG(X_{9\text{-}13})(G/V)X_3(D/E)XG$ [1]. However, the specific cofactor responsible for providing the hydride during the retinol saturation reaction remains unknown, although existing literature proposes that either NADH or FADH2 could serve as suitable cofactors [5, 10].

Structural information regarding RetSat plays a vital role in comprehending the binding sites for cofactors and substrates, as well as the interactions between small molecule ligands and other proteins, which is crucial for the design of RetSat inhibitors. However, at present, these critical aspects of RetSat remain unknown. To address these gaps in knowledge, this study employed computational tools to provide insights into these questions. By leveraging computational approaches, we aim to shed light on the structural characteristics of RetSat and uncover the underlying mechanisms involved in its function, thereby facilitating the development of potential inhibitors for this enzyme.

2. Methodology

To obtain the 3D structure of human RetSat, the amino acid sequence comprising 610 residues [1] was retrieved from the UniProtKB database [18]. The modeling process was conducted utilizing the AlphaFold server [19], which employs artificial intelligence techniques to construct protein structures based on their amino acid sequences. The confidence of the generated model was assessed using the pLDDT measurement, which provides per residue confidence scores ranging from 0 to 100. This evaluation was carried out using the IDDT-Cα metric, a method that predicts the local accuracy of the model [19]. It is particularly useful for assessing the accuracy of multi-domain predictions, where individual domains may be accurate even if their relative positions are uncertain. By analyzing the confidence bands, the confidence levels of different regions within the model, along with their respective confidence scores, can be determined [20].

2.1. Molecular dynamics (MD) simulations

The RetSat structure was initially subjected to protonation at pH 7 using the ProPka server [21] to determine the correct protonation state of the amino acids. Subsequently, the relaxation of the protonated RetSat structure was achieved through MD simulations. These simulations were performed using the GROMACS [22] software packages employing the CHARMM36 force field [23].

To create an aqueous system for the protein, a cubic solvent box with a side length of 3.0 nm was generated, utilizing the TIP3P water model [24]. The solvated system was then neutralized and subjected to energy minimization using the steepest descent algorithm [25] for 50,000 steps, with a minimization step size of 0.01. This step aimed to eliminate steric hindrance and correct any improper geometry in the system.

Following the initial minimization, an equilibration process was carried out. First, the temperature was maintained at 300K in the NVT (ensemble, while the pressure was set to 1 bar in the NPT ensemble). This pre-equilibration phase in the NVT and the subsequent NPT ensemble ensured the system's stability before commencing the dynamic simulation.

The MD simulation for the protein was performed for a duration of 500 ns using the leap-frog integrator [25]. During the simulation, neighbor searching was handled using the Verlet scheme [26], while long-range electrostatic interactions were computed using the Particle Mesh Ewald (PME) method [27].

The equilibrated RetSat structure was subjected to various validation and quality assurance techniques. Geometric validation was performed using the Ramachandran plot [28]. The overall quality and energy deviation of the structure were assessed through the *z*-score obtained from ProSA [28]. ProQ [29] was utilized to predict the model's quality based on atom-atom contacts, residue-residue contacts, surface area exposure, and secondary structure agreement. Verify 3D [30] technique examined the compatibility of the 3D structure with the protein sequence. The ModFOLD server [31] provided global and local predictions of the structure's quality. Hydrophobicity analysis was conducted to differentiate non-polar and polar regions and calculate the Solvent Accessible Surface Area (SASA) using VADAR server [32].

Furthermore, the MD trajectory of the equilibrated RetSat structure was analyzed to generate RMSD (Root Mean Square Deviation) and RG (Radius of Gyration) time series plots. These plots were used to evaluate the dynamics stability of the predicted structure. The RMSD plot provided insights into the structural deviations from the starting conformation throughout the simulation, while the RG plot reflected the compactness and overall size of the protein over time. By examining these time series plots, the dynamics stability of the predicted RetSat structure could be assessed. Flexibility of the 500 ns long protein trajectory was investigated through principal component analysis (PCA) using the Galaxy server [33].

2.2. Tunnel and binding pocket analysis

The CAVER ANALYST 2.0 software [34] was employed to examine the tunnel pathways connecting the buried sites to an external solvent. This analysis encompassed all nine structures generated through cluster analysis after conducting MD simulations. The selection of structures with prominent and distinguishable binding sites was based on a comparison of the binding residues found in adenosine-specific sites of FAD/FADH2 and NAD/NADH proteins obtained from the Protein Data Bank [35]. The binding pockets of the cofactor and retinol were further validated using the GHECOM server [36]. This server employs a combination of grid representation of molecular shapes and mathematical morphology to detect the pockets [36].

2.3. Molecular docking calculations

Before conducting the docking process, the substrate (all-trans-retinol) and the cofactor underwent geometry optimization using the B3LYP/6-31G(d,p) level of density functional theory (DFT) in the Gaussian 09 software package [37]. The RetSat protein and ligands were prepared for the docking calculations using the MGL Tools-1.5.6 software package [38]. Gasteiger charges [39] were assigned to each atom of the protein and ligand during its preparation for docking. Prior to performing the docking

calculation, a 3-D grid representation was constructed to calculate non-covalent interaction energies with electrostatic and desolvation maps. The Autogrid module [40] was utilized for this purpose, with the grid box centered around the predicted catalytic residues for both the substrate and cofactor. The docking calculations were carried out using the AutoDock 4.2 software [38], employing the Lamarckian genetic algorithm [41]. Cofactors were docked into the predicted binding site for selected structures, focusing on the binding residues associated primarily with the adenosine group of each cofactor. Following the docking of the cofactor, a 50 ns MD simulation was conducted. The docking of retinol was performed on the last frame of the trajectory.

2.4. Molecular dynamics simulations of complexes

MD simulations were performed on both the RetSat:cofactor and RetSat:cofactor:substrate complexes. The parameterization of the cofactor and at-retinol was conducted using the CHARMM-GUI server [42] to enable simulation using the NAMD software [43] package with the CHARMM36 force field [23].

After obtaining the ligand parameters for the cofactor and at-retinol from CHARMM-GUI [42], a cubic water box was created using the TIP3P water [24] model, and the system was neutralized by adding counterions. Water molecules overlapping with the protein and ions were removed, followed by short minimization steps at each iteration. Finally, periodic boundary conditions were set up, and further minimization was carried out with 50 steps using the Steepest Descent [44] and 50 steps using the Adopted Basis Newton-Raphson (ABNR) algorithm [45] before the equilibration step. Non-bonded parameters were established and calculated through the CHARMM 44b1 software [46].

The NAMD software package [43] was employed to generate a 50 ns trajectory for the protein:cofactor complex and a 10 ns trajectory for the protein:cofactor:at-retinol complex, including the equilibration. Non-bonded parameters were scaled to 1–4 interactions, where electrostatics were scaled to 1–3 with a factor of 1.0, and van der Waals interactions were scaled to 1–4. Non-bonded interactions were truncated at a cutoff of 12.0. The integrator time step was set to 2 fs, and the non-bonded frequency was set to 1.0. During the equilibration phase, constraints were applied with a scaling factor of 1.0. Subsequently, the protein: ligand complex was simulated for 20 ns using Langevin dynamics [47] with a time step of 2 fs.

The PME method [47] was employed with a spacing of 1.0, and the temperature was set to 300 K. The protein:ligand trajectory was analyzed for further investigations.

3. Results and Discussion

The study conducted by Moise et al. [1] identified RetSat as a member of the oxidoreductase family, comprising 610 amino acid residues. Figure 1(A) provides the canonical sequence of human

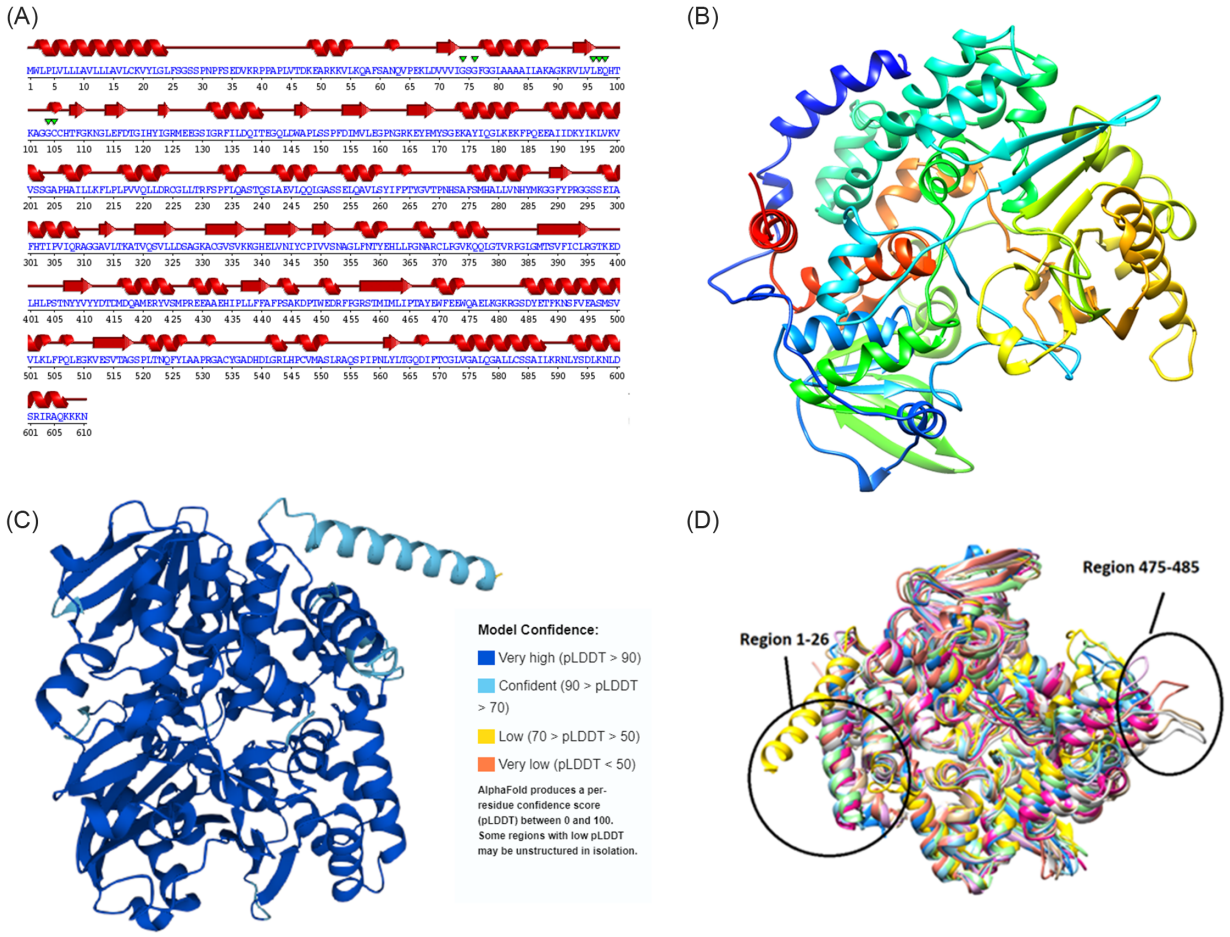

Figure 1. (A) Predicted secondary structure composition of the 3D model for human RetSat. (B) The 3D structure of RetSat presented in a cartoon representation. (C) Color-coded pLDDT score distribution for each amino acid in the predicted 3D model of RetSat. (D) Superimposition of structures from 9 clusters for identification of significant changes.

RetSat, while Figure 1(B) depicts a cartoon representation of its modeled 3D structure. The constructed 3D model shows various structural elements, including 5 beta sheets, 1 beta alpha beta unit, 6 beta hairpins, 6 beta bulges, 20 strands, 34 helices, 47 helix-helix interactions, 42 beta turns, and 3 gamma turns. These elements play crucial roles in maintaining the enzyme's stability.

The model's confidence is reinforced by the calculated confidence bands, with the predicted 3D model of human RetSat colored in blue, indicating a pLDDT score above 90. This high pLDDT score further strengthens the confidence in the accuracy of the model. Visual representation of the pLDDT score distribution for RetSat is given in Figure 1(C). By conducting cluster analysis on the MD trajectory of the predicted RetSat structure, nine distinct clusters or conformations were identified. From each cluster, a representative member was selected, and these representatives were superimposed to investigate the significant variations among the conformations. The analysis revealed that the relative motion of two regions, 484–473 and 590–610, played a pivotal role in driving the prominent changes observed. These findings are visually depicted in Figure 1(D).

The secondary structure composition of the predicted model indicates that it comprises 16.6% strands, 41.1% alpha-helix, 4.3% 3–10 helix, and 38% other structures, contributing to the overall stability of the enzyme. Further, Figure 2(A), (B), and (C) present contact maps for hydrogen bonding, Cα, and Cβ atoms, respectively. The observed patterns in these contact maps strongly suggest the existence of various secondary structures, notably alpha-helices, anti-parallel beta sheets, and parallel beta sheets.

In an efficiently packed protein, all residues should have a fractional volume close to 1.0, indicating that they occupy their expected volume. However, if a residue is located in a cavity, its fractional volume exceeds 1.20, indicating extra space around it. On the other hand, if the residues are situated in a compressed

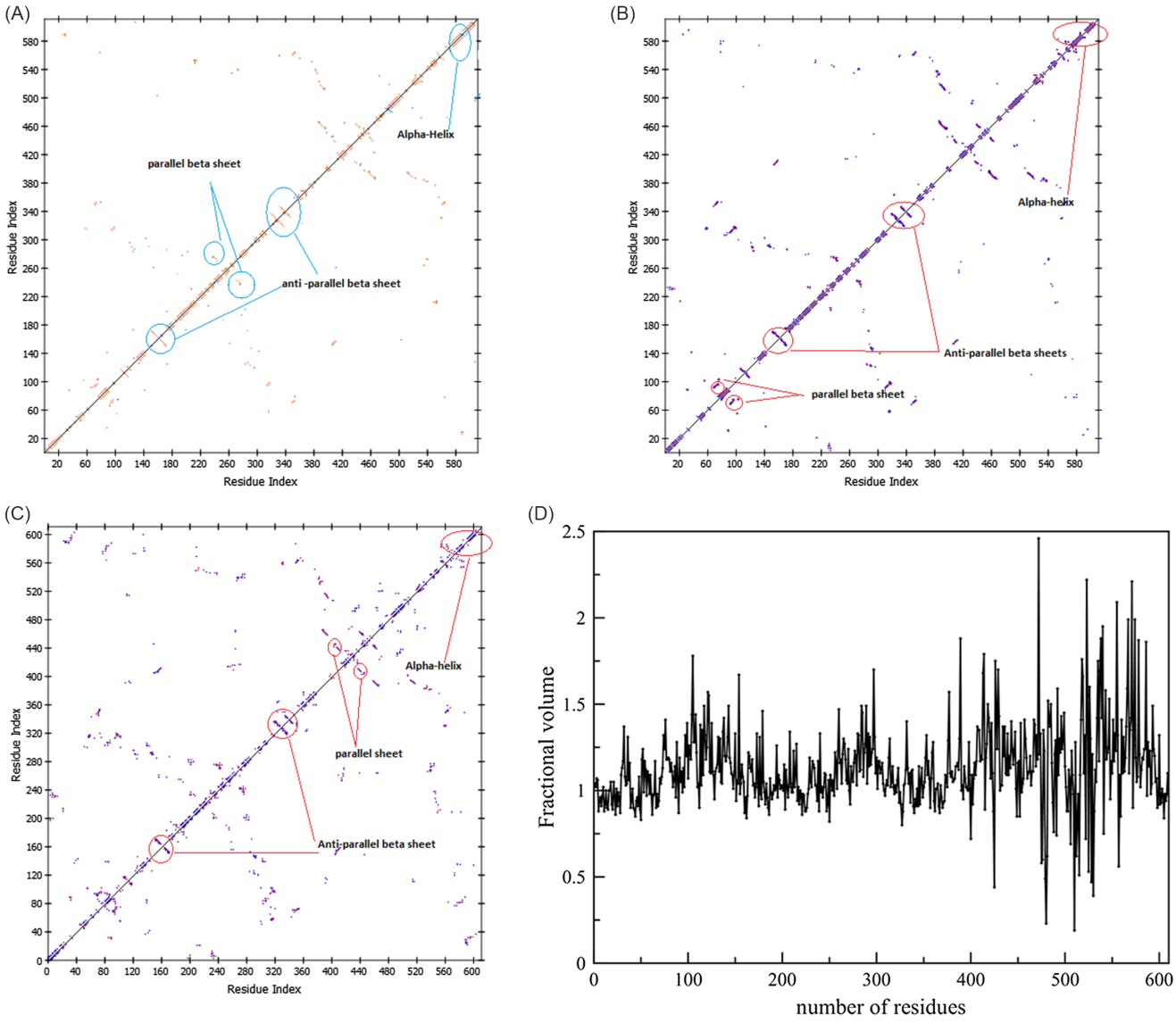

Figure 2. (A) Contact plot of hydrogen bond interactions. (B) Contact plot of Cα atoms. (C) Contact plot of Cβ atoms. (D) The plot of fractional accessible surface area vs residue number.

region, their fractional volume decreases and falls below 0.8, implying tighter packing. Based on the observations from Figure 2(D), where fractional volumes range from 0.24 to 2.4, it can be deduced that the RetSat model comprises a combination of both cavities and compressed areas.

The stability of the protein during the MD simulation was evaluated using RMSD and Radius of Gyration plots of the modeled protein. RMSD is a vital measure of the protein's stability in the aqueous medium throughout the simulation, while the radius of gyration reflects the compactness of the modeled protein.

In the RMSD plot (Figure 3(A)), it was observed that the structure reached a plateau after 350 ns, with an approximate RMSD value of 0.563. This plateau indicates the stability of the protein in the aqueous medium during the simulation. Moreover, the RetSat protein demonstrated a consistently steady *Rg* value of 2.7 in the radius of gyration plot (Figure 3(B)) throughout the 500 ns simulation, suggesting that it remains stably folded.

Both the RMSD and radius of gyration plots for the protein are shown in Figure 3(A) and (B), respectively. To further validate the quality and stereochemical properties of the modeled structure, the last frame of the 500 ns long MD trajectory was subjected to standard validation techniques as discussed below and subsequent docking calculations.

The Ramachandran plot [48] was generated using PROCHECK [28], and its representation is depicted in Figure 3(C). A significant percentage of ϕ-ψ angles, approximately 94.4%, is situated within the most allowed region, indicating the high quality of the predicted structure with favorable stereochemical properties. A very small proportion of residues, 5.4%, fall within the additional allowed region, while an even smaller fraction, 0.2%, occupy the generously allowed region. Notably, no residues are found in the disallowed region, further affirming the reliability of the model. With over 90% of residues falling within the most favored region, these results support the notion that the predicted structure is of good quality, exhibiting strong stereochemical characteristics.

The *z*-score obtained from PROSA [28] for experimentally determined structures typically falls within the range of +10 to −20. In the case of RetSat, the *z*-score was calculated to be −11.95 within the native conformations, suggesting the presence of reasonable side chain interactions in the protein. The majority of residues exhibit significantly negative energies, indicating a lack of errors in their positioning. However, a few residues may show deviations, but overall, the results reinforce that RetSat's structure is reliable and well-arranged.

The 3-D profile of the predicted protein structure was verified using Verify3D [30], which revealed that approximately 84.59% of residues achieved an average score equal to or greater than the threshold value of 3 (Figure 3(D)). This indicates that a significant majority of residues in the predicted structure are well-supported and fit well within the expected parameters.

The quality of the model was also assessed using the ModFOLD server [31] in two aspects. First, the local quality was evaluated,

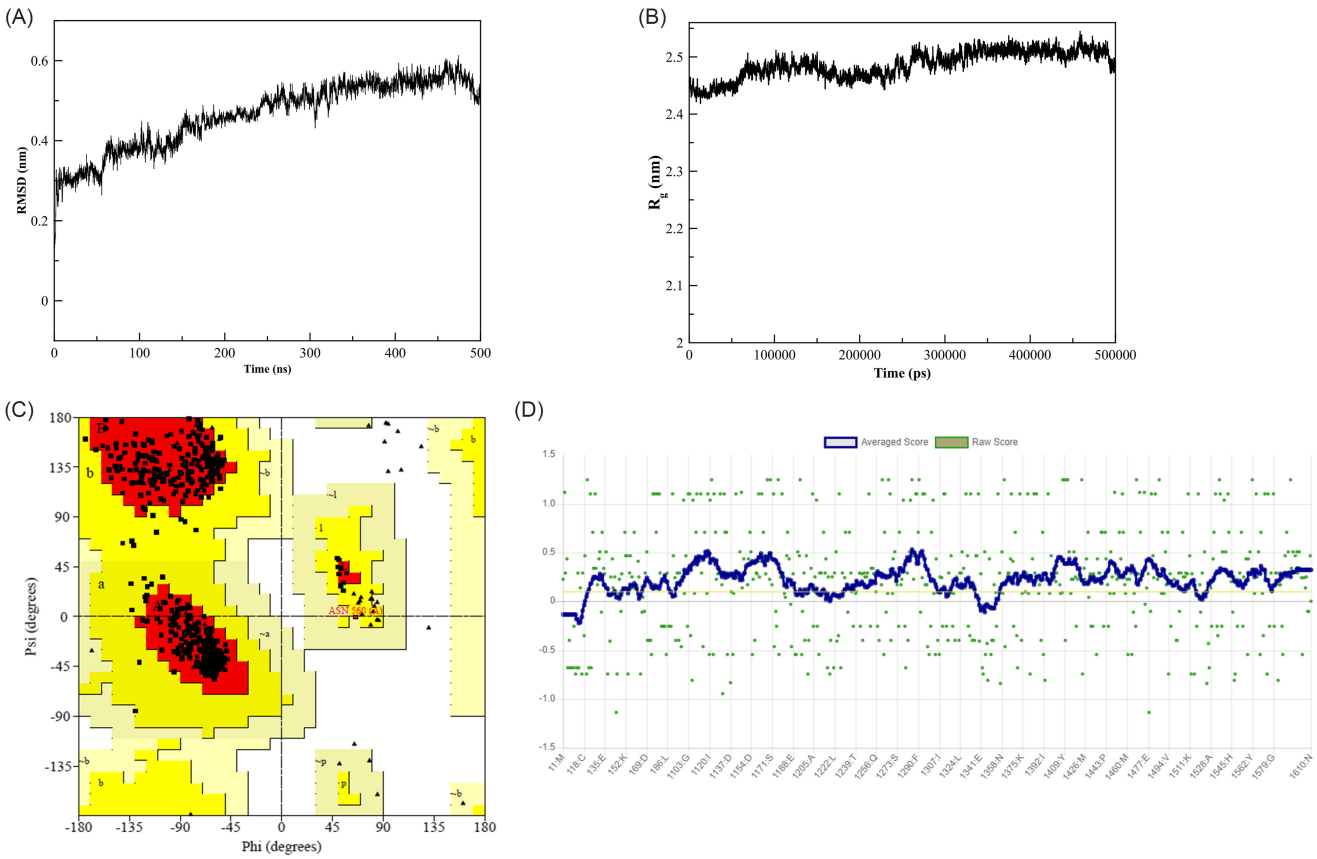

Figure 3. (A) RMSD plot. (B) RG plot. (C) Ramachandran plot. (D) verify 3D plot of RetSat protein.

yielding a high confidence score or *p*-value of 4.367E-3. This score indicates that there is less than a 1/100 chance that the model is incorrect at the local level, suggesting a high degree of accuracy in the model's specific regions. Second, the global model quality value, which is 0.4092, signifies that the overall model is complete and exhibits a high similarity to the native structure, indicating a confident and reliable prediction.

This suggests that certain regions within the protein have excess space, while others exhibit tighter packing, contributing to its overall structural complexity.

3.1. Dynamics of RetSat: principle component analysis

The PCA of the RetSat protein incorporates three main components: PC1, PC2, and PC3, each capturing distinct aspects of the protein's conformational dynamics. The PCA scatter plots, as depicted in Figure 4(A), (B), and (C), utilize red and blue dots to represent the temporal progression of the protein's conformations. Notably, in the PC1 and PC3 planes (Figure 4(A) and (B)), the observed clustering patterns do not show clear

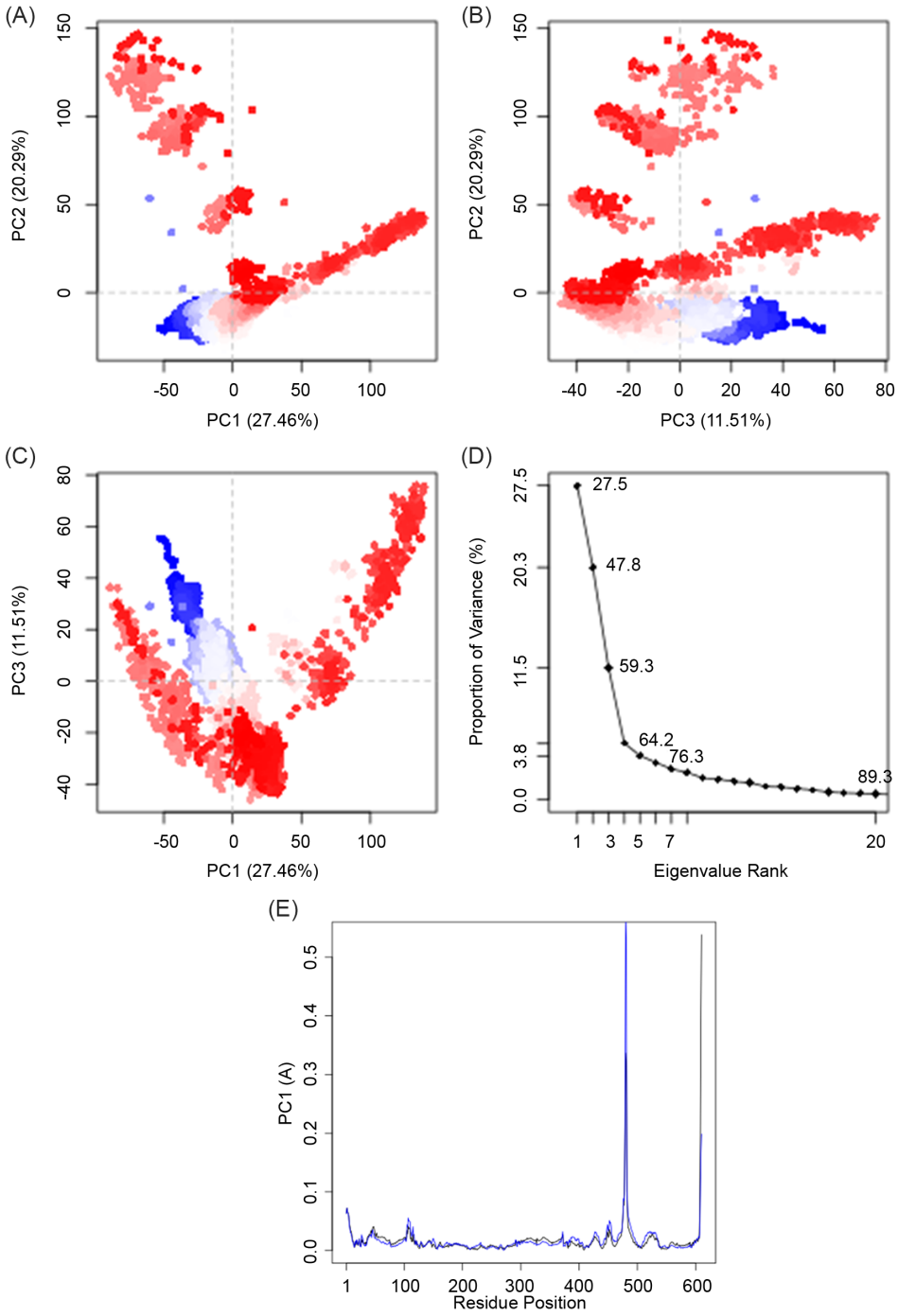

Figure 4. PCA graphs for (A) PC2 vs PC1, (B) PC2 vs PC1, (C) PC3 vs PC1, (D) eigenvalue rank, and (E) superimposed PC1 (blue) with RMSF (black) of RetSat protein

separation, indicating that the protein's global motions along these axes are periodic in nature. This periodicity suggests a certain regularity or repeating patterns in the protein's structural transformations along these dimensions.

In contrast, the PC1 and PC2 planes (Figure 4(C)) exhibit two distinct clusters, separated from each other, implying that the motions along these axes are non-periodic. This distinction points to a more diverse and possibly irregular range of conformational changes, which might reflect different functional states or significant structural transitions within the protein.

The eigen plot, presented in Figure 4(D), quantifies the variance contributions of the principal components. Collectively, the first three components account for 59.3% of the total variance in the protein's conformational data, with PC1 being the most dominant, contributing 27.5% of the variance. This dominance underscores the importance of the motions captured by PC1 in the overall structural dynamics of the protein.

Further analysis reveals that PC1 is closely linked to twisting motions in specific loop regions, located between amino acids 607 to 610 and 478 to 482. Interestingly, these regions are not part of the active site of the protein, suggesting that while these motions are significant and contribute greatly to the protein's overall variance, they do not directly influence the active site. This could imply that the major conformational changes represented by PC1 are more related to the protein's overall structural integrity, stability, or its interactions with other molecules, rather than its catalytic or binding functions.

Figure 4(E) illustrates the relationship between the atomic motion of the RetSat protein along Principal Component 1 (PC1) and the protein's root mean square fluctuation (RMSF). The PC1 motion is represented in blue, while the RMSF data is shown in black. The near overlap of the two graphs suggests a high degree of correspondence between the atomic motions captured by PC1 and the regions of the protein exhibiting the greatest fluctuations as measured by RMSF.

From this close correspondence, we can infer that the regions with the highest RMSF – which likely correspond to free loops or flexible linkers in the RetSat protein structure – contribute significantly to the conformational dynamics described by PC1. These areas of flexibility within the modeled protein are therefore implicated as the primary drivers of the major motions or structural changes that PC1 represents. This observation underscores the importance of these flexible regions in the overall dynamics and possibly the functional mechanisms of the RetSat protein.

3.2. Tunnel and binding site analysis of RetSat

Previous work by Moise et al. [1] suggested that a cofactor is involved in the saturation of the double bond between the 13th and 14th carbons in at-retinol. However, the literature does not provide detailed information on the binding pockets of the cofactor and at-retinol. Tunnel analysis can provide insights into the recognition of cofactor and substrate binding pockets in proteins.

Tunnel analysis was performed on nine average structures of RetSat obtained via cluster analysis using the Caver Analyst 1.0 software package [35]. The binding residues of the cofactor binding pocket in RetSat were initially identified by superimposing the last frame of a 500 ns simulation of RetSat with gamma carotenoid desaturase (PDB ID: 4REP), based on the study by Moise et al. [1]. This is because RetSat and gamma carotenoid desaturase have been shown to have related activities. Based on this superposition, tunnel I was postulated to be the cofactor binding tunnel in RetSat, as most of the amino acids in the tunnel are similar to the binding residues obtained by superimposing gamma carotenoid desaturase with the last frame of the 500 ns simulation of RetSat. Tunnel I in RetSat is 29.6Å long with a bottleneck radius of 1.9Å located 13.5Å away from the starting point, which is HIS121. It is composed of amino acid residues 74–77, 97–98, 103–106, 116–120, 122, 123, 264–266, 281, 285, 353–356, 358, 409, 439, 461, 463, 523, 526–530, 534–537, 567, and 571–573.

The substrate binding tunnel was selected based on the fact that the cofactor and substrate binding pockets must be in close proximity for the reaction mechanism to proceed, as a hydride ion is abstracted from the cofactor. Closer to the cofactor binding tunnel (tunnel I), two tunnels (tunnel II and tunnel III) interlock to form a long tunnel. All three tunnels are interconnected, forming a T-shaped structure. Tunnel II is connected to tunnel III to form a long hydrophobic tunnel for the substrate at-retinol, which is a hydrophobic molecule with two openings to the solvent exterior. Tunnel II is longer than tunnel III, with a length of 17.2Å and a bottleneck radius of 2.0Å. The amino acid composition of tunnel II is 109, 116–120, 122–123, 264–265, 285, 387–390, 407–409, 439, 461–463, and 517. Tunnel III is composed of amino acid residues 116–120, 123–126, 265, 280–286, 439, 570–573 and has a length of 14.3Å with a bottleneck radius of 2.1Å. Figure 5(A) highlights the orientation of tunnels for cofactor and at-retinol. Table 1 indicates the amino acid composition and relevant dimensions of both cofactor and substrate tunnel.

The existence of the tunnels was evaluated by the GHECOM 1.0 server [36] using the theory of mathematical morphology. The grid representation was used to locate the binding pockets of cofactor and substrate in the modeled RetSat protein and the contribution of each amino acid residue to the pocket. It was identified that both the cofactor and substrate binding pockets predicted via tunnel analysis fall into the same cluster of pockets indicated in red color. This suggests that the cofactor and substrate binding pockets are located in the same region of RetSat, and based on the grid of pocketness, the higher the pocket score for a considered residue, the deeper the residue in the pocket. Therefore, Figure 5(B) highlights that the substrate pocket is located deeper in the protein than the cofactor binding pocket, merely based on the pocket score value for respective binding residues.

3.3. The proposed cofactor and its binding to RESTSAT

Binding of FADH2 and NADH cofactors to the proposed binding pocket (tunnel I) of RetSat was studied using molecular docking simulations. The orientation of the cofactors in RetSat was predicted based on the superposition of the bioactive conformers of FADH2 and NADH in a number of PDBs available in the Protein Data Bank. Based on the estimated binding free energy, NADH was chosen as the cofactor that would bind in RetSat the most effectively. The stability of the NADH:RetSat complex and the cofactor binding tunnel was further evaluated by RMSD plot. The NADH complex was stable after 20 ns and

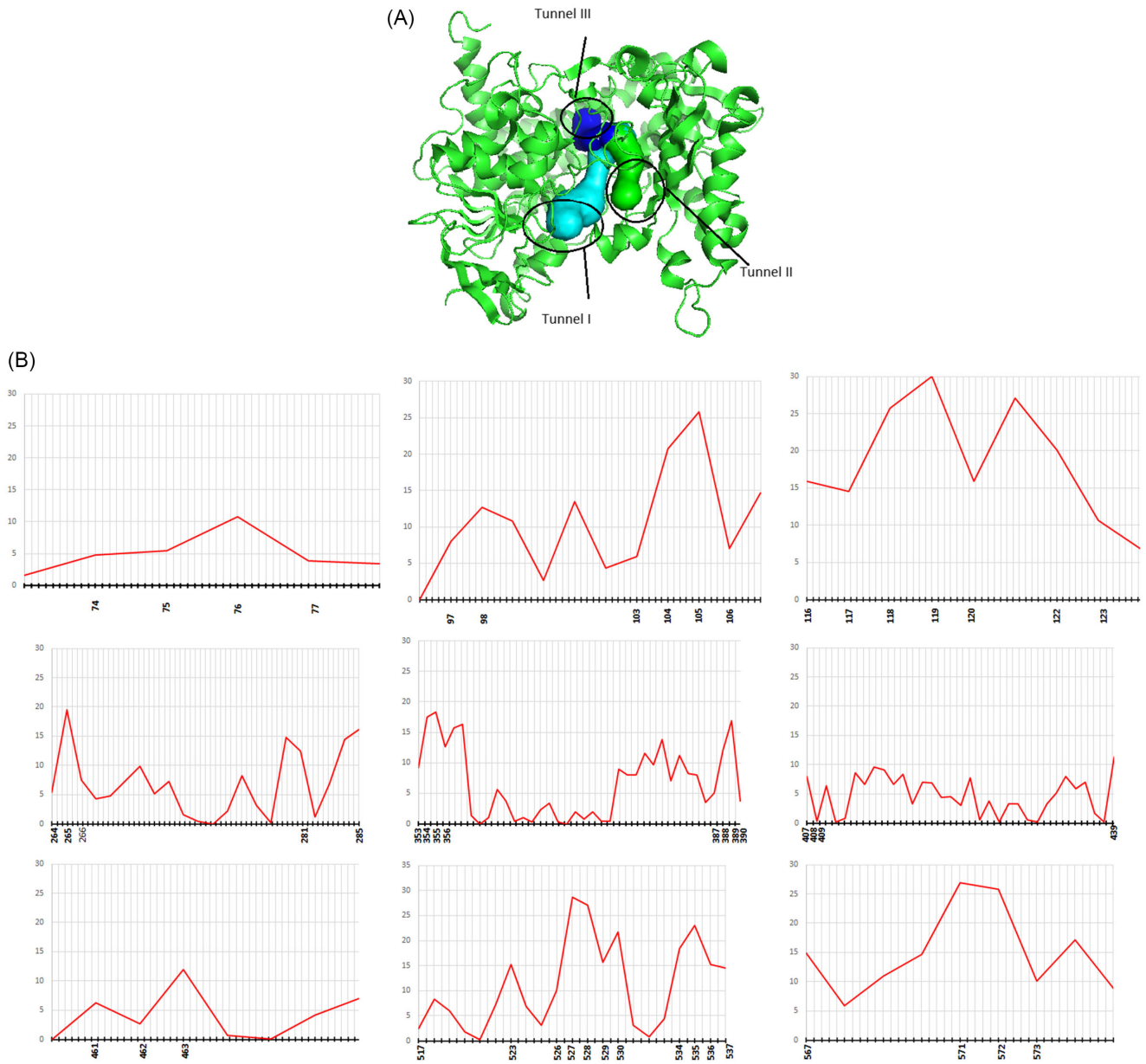

Figure 5. (A) The orientations of binding pockets for cofactor (tunnel I) and at-retinol (tunnel II and tunnel III). (B) grid of pocketness of cofactor and ligand binding tunnel in RetSat protein.

maintained an approximate value of 0.33. The binding pocket was also stable/rigid overall, with the RMSD difference being about 0.06. Figure 6(A) and (B) illustrate the stability of the protein backbone in RetSat and the binding pocket of NADH upon binding to RetSat with respectively. The protein-cofactor interactions were analyzed using 2D protein-ligand interaction plot.

In the binding pocket of RetSat, NADH exhibits remarkable stability, primarily due to robust hydrogen bonding. Key residues such as GLY78, GLN98, GLU97, CYS105, PHE77, CYS571, and GLY572 engage with the phosphate backbone, while VAL320 interacts with the adenosine base, and THR265, HIS121, and TYR122 associate with the nicotinamide ring. The adenosine nucleotide and ribose sugar ring of NADH are nestled within a hydrophobic enclave formed by residues ALA318, LEU96, ILE73, ALA354, ALA528, LEU527, SER75, and GLY74. Hydrophobic interactions are further supplemented by Pi-alkyl and alkyl bonds involving ILE120 and ALA354 with the nicotinamide ring and the ribose sugar, respectively. A comprehensive network of van der Waals interactions with LEU363, ILE73, THR319, LEU96, GLN566, CYS584, TYR585, GLY119, TYR285, LEU573, LEU527, HIS99, SER75, ALA528, and GLY104 also contributes to the stabilization of NADH within the binding site with respective to in Figure 7(A).

Table 1. Internal dimensions and substrate compositions of tunnels

Type of the tunnel	Tunnel ID	Length (Å)	Bottleneck radius (Å)	Residue composition
Cofactor binding tunnel	I	29.6	1.9	ILE74, SER75, GLY76, PHE77, GLU97, GLN98, GLY103, GLY104, CYS105, CYS106, PHE116, ASP117, THR118, GLY119, ILE120, TYR122, ILE123, PRO264, THR265, TYR266, LEU281, TYR285, ASN353, ALA354, GLY355, LEU356, ASN358, ASN407,TYR409, PHE439, ILE461, LEU463, ASN523, TYR526, LEU527, ALA528, ALA529, PRO530, ALA533, CYS534, TYR535, GLY536, ALA537, THR564, GLY565, GLN566, ASP567, CYS571, GLY572, LEU573
Substrate binding site	II	14.3	2.1	PHE116, ASP117, THR118, GLY119, ILE120, TYR122, ILE123, GLY124, ARG125, MET126, THR265, LEU280, LEU281, VAL282, ASN283, TYR285, MET286, PHE439, THR570, CYS571, GLY572, LEU573, VAL574, GLY575.
	III	17.2	2.0	PHE109, PHE116, ASP117, THR118, GLY119, ILE120, TYR122, ILE123, PRO264, THR265, TYR285, MET387, THR388, SER389, VAL390, ASN407, TYR409, PHE439, ILE461, MET462, LEU463, ALA517, GLY518, SER519, THR522, ASN523, TYR526, LEU527, ALA533, CYS534, TYR 535.

Collectively, these multifaceted interactions underpin the stable integration of NADH into the RetSat protein complex.

3.4. Binding of the natural substrate: retinol

The final frame of the 50 ns simulation of the NADH:RetSat complex was extracted for the docking of at-retinol. The nicotinamide ring of NADH was oriented towards tunnel III. Tunnel III was chosen as the binding tunnel for at-retinol because the reactive center of the 13th and 14th double-bonded carbons of at-retinol must be located orienting towards the nicotinamide ring of NADH. The docking process of at-retinol to tunnel III was carried out with the Autodock 4.2 software package [49].

In the docking study conducted, the interaction with at-retinol was focused on, especially around its 13th and 14th carbon atoms. Tyrosine 409 (TYR409) was identified as a key hydrogen donor in proximity to these carbon atoms. It was observed that the 13th carbon could form a more stable tertiary cation compared to the secondary cation at the 14th carbon. This led to the hypothesis of a natural alignment between the 14th carbon and TYR409. However, when this alignment was not observed initially, a manual adjustment was done to establish this alignment, while ensuring that the alignment between the 13th carbon and the reactive hydrogen of the NADH cofactor, identified as H63, was maintained, where NADH was presumed to be the secondary H donor in the next step of the saturation mechanism.

The distances in these molecular interactions were carefully measured. The distance between H63 of NADH and the 14th carbon was found to be 3.50Å, and the manually adjusted distance between TYR409 and the 14th carbon was 2.50Å. To further validate these alignments and gain additional insights, a short 20 nanoseconds of MD simulations were conducted, allowing for a deeper understanding of the molecular behavior and stability in this system.

The stability of the RetSat:NADH:at-retinol complex was further validated using MD simulations. The 20 ns MD trajectory of the complex revealed that at-retinol and NADH were stable and retained in their respective pockets, interacting strongly with the binding residues. The RMSD plot of the protein backbone is stable after 5 ns according to Figure 6(C). The NADH and retinol-binding pockets were stable after 5 ns (Figure 6(D) and (E)), possibly due to the relaxation of the respective binding pockets.

However, in the course of the MD simulation, a notable deviation was observed in the alignment, specifically concerning the distance between TYR409 and the 14th carbon of at-retinol. The respective distances between C13 of retinol and TYR409 and C14 of retinol and H63 are 4.5Å and 5.9Å. This deviation emerged as a significant finding, strongly suggesting that TYR409 is highly unlikely to be the initial hydrogen donor in the anticipated mechanism. This unexpected result indicates a need for further investigations to comprehensively understand the role of TYR409 and to explore alternative mechanisms or donor candidates within the molecular system.

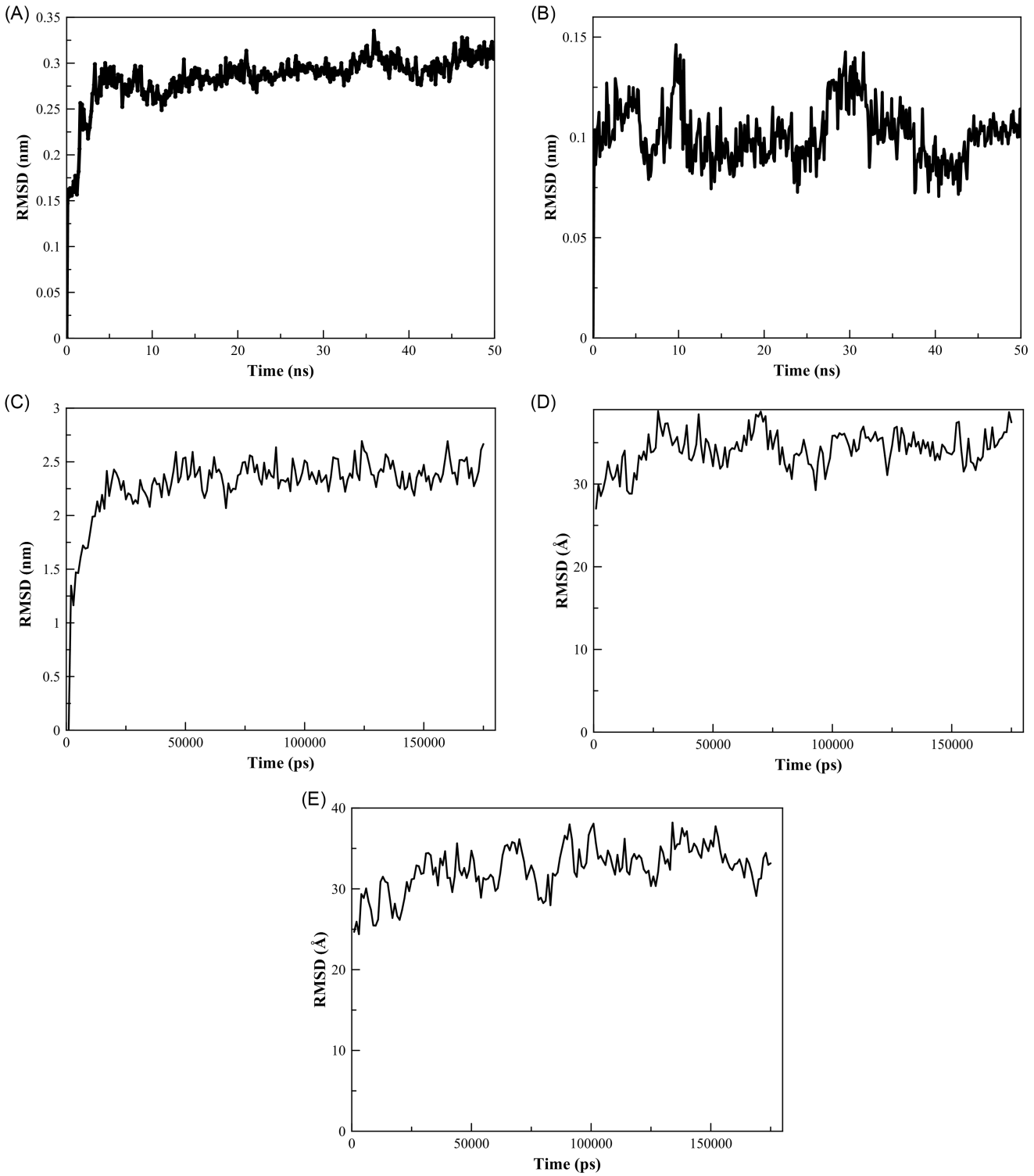

Figure 6. (A) RMSD plot of protein backbone after 50 ns simulation for RetSat: NADH complex. (B) RMSD plot for NADH binding pocket in 50 ns simulation of protein: NADH complex. (C) RMSD plot for protein backbone for RetSat: NADH: at-retinol complex in 20 ns long MD simulation. RMSD plot for binding pockets of (D) NADH and (E) at-retinol in 20 ns of MD simulation in RetSat: NADH: at-retinol.

In the final frame of the MD simulation involving at-retinol, a variety of interactions were observed between the molecule and the complex. Notably, a strong hydrogen bond formed between the LEU438 residue and the OH group of at-trans-retinol. However, when analyzing the 2D ligand-protein interaction plot, it was evident that hydrophobic interactions, including pi-alkyl and van der Waals forces, were more dominant than hydrogen bonds in this scenario with respective to Figure 7(B).

Specifically, with at-retinol, several residues were involved in distinct types of interactions. ILE120, PRO465, PRO264, LEU437, and PHE439 engaged in both alkyli and pi-alkyli interactions. Additionally, ILE123, THR265, MET387, TYR409, VAL410,

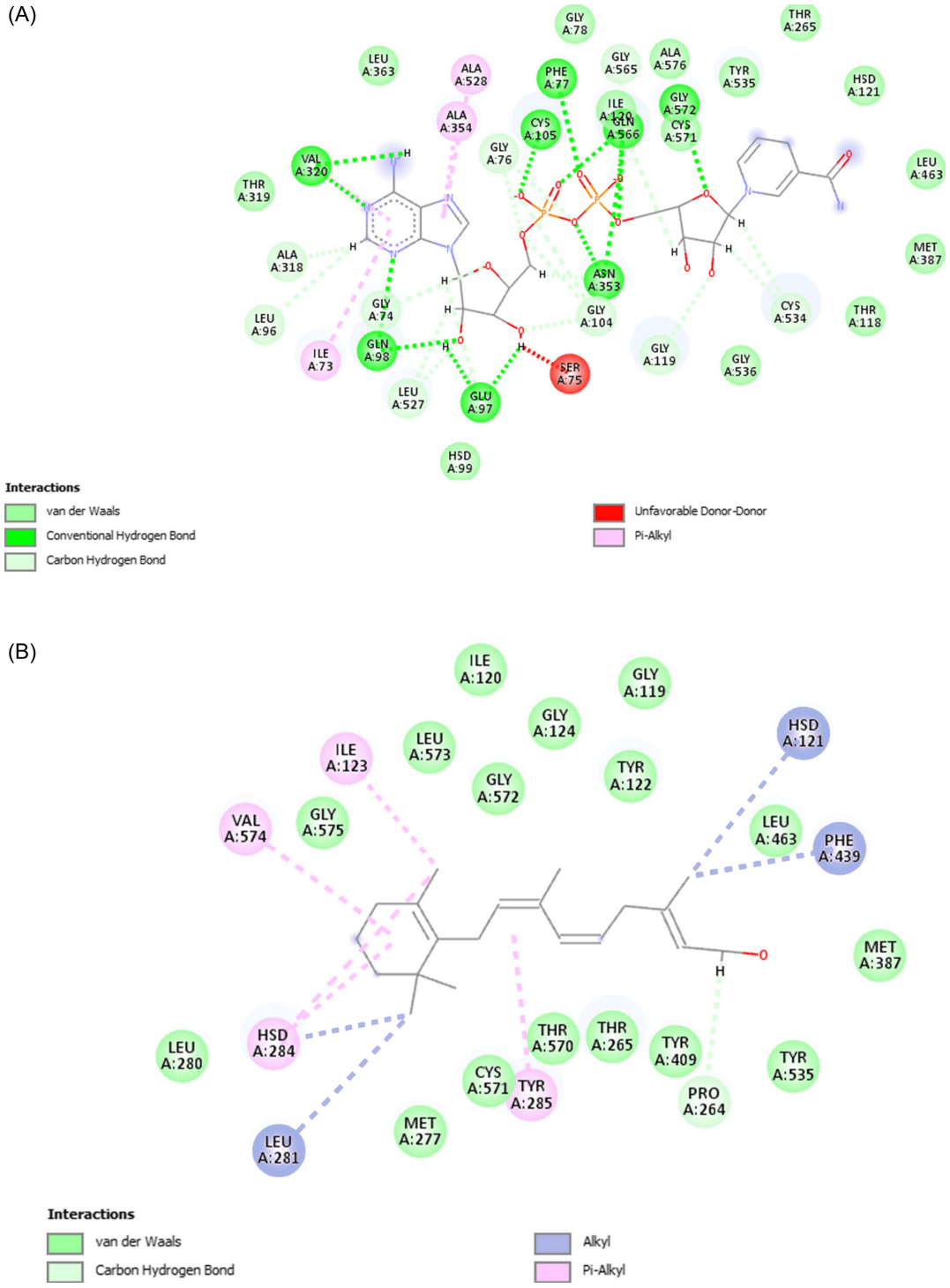

Figure 7. 2D interaction plot for (A) NADH cofactor and (B) 2D interaction plot for at-retinol after MD simulation

TYR411, LEU463, ILE464, TYR535, GLY572, and LEU573 were found to interact with NADH predominantly through van der Waals forces.

4. Conclusions

In this comprehensive study, we successfully remodeled and revalidated the complete 3-D atomic structure of RetSat using homology techniques, marking a significant advancement in the understanding of this vital enzyme. The model was rigorously validated through standard techniques, including dynamic stability assessments, ensuring its reliability. This study is pioneering in presenting the RetSat model structure, which includes the prediction of both cofactor and ligand binding sites – details previously unexplored in existing literature.

Our findings indicate that RetSat possesses a complex and stable structure. The enzyme is characterized by a diverse array of structural elements, such as beta sheets, alpha-helices, and various turns, all contributing to its stability. The high pLDDT score of the model and the validation through the Ramachandran

plot, *z*-score, Verify3D, and ModFOLD server assessments further reinforce the accuracy and quality of the model.

MD simulations and PCA revealed the dynamic nature of RetSat, with significant conformational variability. This aspect of the study emphasized the importance of certain loop regions, suggesting their role in the protein's overall structural integrity and functionality.

One of the study's key achievements is the comprehensive analysis of cofactor and substrate binding sites. Tunnel analysis and molecular docking simulations provided deep insights into these interactions. NADH was identified as the most suitable cofactor for RetSat through molecular docking calculations, a critical step forward in understanding the enzyme's mechanism. The docking process revealed strong hydrogen bonding and hydrophobic interactions, suggesting a stable integration of NADH and at-retinol into the RetSat protein complex.

Despite the detailed analysis, our study revealed unexpected findings, particularly regarding the role of TYR409. The observed deviation in the alignment during the MD simulation suggests that TYR409 is unlikely to be the initial hydrogen donor, calling for further investigations into alternative mechanisms.

Overall, this study not only provides a detailed structural and dynamic profile of RetSat but also opens avenues for future research. Given the diverse physiological roles and importance of RetSat in various biological processes, our findings serve as a foundational platform for deeper exploration into its functions and mechanisms. This research enriches our understanding of RetSat and lays the groundwork for potential therapeutic and biotechnological applications, underscoring the enzyme's significance in the biochemical field.

Acknowledgment

We gratefully acknowledge the Center for Scientific Computing and Advanced Drug Discovery, University of Sri Jayewardenepura, for providing essential computing facilities for our research. We also extend our thanks to the Research Council of the University of Sri Jayewardenepura for their financial support under grant ASP/01/RE/SCI/2018/19, which was pivotal in facilitating our work.

Funding Support

This work was supported by University of Sri Jayewardenepura Research Grant of Sri Jayewardenepura Research Grant – ASP/01/RE/SCI/2018/19.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support this work are available upon reasonable request to the corresponding author.

Author Contribution Statement

Virangi Tharika Hewage: Software, Validation, Investigation, Data curation, Writing – original draft, Visualization. **Ranga Srinath Jayakody:** Conceptualization, Methodology, Validation, Formal analysis, Resources, Data curation, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Anura Jayasooriya:** Formal analysis, Resources, Data curation, Writing – review & editing, Visualization, Supervision. **Tharindu Senapathi:** Software, Formal analysis, Resources, Data curation, Writing – review & editing, Visualization, Supervision. **Krishna Govendar:** Software, Formal analysis, Resources, Data curation, Writing – review & editing, Visualization, Supervision.

References

[1] Moise, A. R., Kuksa, V., Imanishi, Y., & Palczewski, K. (2004). Identification of all-*trans*-retinol: All-*trans*-13, 14-dihydroretinol saturase. *Journal of Biological Chemistry*, *279*(48), 50230–50242. https://doi.org/10.1074/jbc.M409130200

[2] Tarban, N., Halász, H., Gogolák, P., Garabuczi, É., Moise, A. R., Palczewski, K., . . . , & Szondy, Z. (2022). Regenerating skeletal muscle compensates for the impaired macrophage functions leading to normal muscle repair in retinol saturase null mice. *Cells*, *11*(8), 1333. https://doi.org/10.3390/cells11081333

[3] Weber, P., Flores, R. E., Kiefer, M. F., & Schupp, M. (2020). Retinol saturase: More than the name suggests. *Trends in Pharmacological Sciences*, *41*(6), 418–427. https://doi.org/10.1016/j.tips.2020.03.007

[4] Heidenreich, S., Witte, N., Weber, P., Goehring, I., Tolkachov, A., von Loeffelholz, C., . . . , & Schupp, M. (2017). Retinol saturase coordinates liver metabolism by regulating ChREBP activity. *Nature Communications*, *8*(1), 384. https://doi.org/10.1038/s41467-017-00430-w

[5] Jiang, X., He, Y., Shen, Q., Duan, L., Yuan, Y., Tang, L., . . . , & Chen, Y. (2021). *RETSAT* mutation selected for hypoxia adaptation inhibits tumor growth. *Frontiers in Cell and Developmental Biology*, *9*, 744992. https://doi.org/10.3389/fcell.2021.744992

[6] Li, C., Kiefer, M. F., Dittrich, S., Flores, R. E., Meng, Y., Yang, N., . . . , & Schupp, M. (2024). Adipose retinol saturase is regulated by β-adrenergic signaling and its deletion impairs lipolysis in adipocytes and acute cold tolerance in mice. *Molecular Metabolism*, *79*, 101855. https://doi.org/10.1016/j.molmet.2023.101855

[7] Schupp, M., Lefterova, M. I., Janke, J., Leitner, K., Cristancho, A. G., Mullican, S. E., . . . , & Lazar, M. A. (2009). Retinol saturase promotes adipogenesis and is downregulated in obesity. *Proceedings of the National Academy of Sciences*, *106*(4), 1105–1110. https://doi.org/10.1073/pnas.0812065106

[8] Pang, X.-Y., Wang, S., Jurczak, M. J., Shulman, G. I., & Moise, A. R. (2017). Retinol saturase modulates lipid metabolism and the production of reactive oxygen species. *Archives of Biochemistry and Biophysics*, *633*, 93–102. https://doi.org/10.1016/j.abb.2017.09.009

[9] O'Connor, C., Varshosaz, P., & Moise, A. R. (2022). Mechanisms of feedback regulation of vitamin A metabolism. *Nutrients*, *14*(6), 1312. https://doi.org/10.3390/nu14061312

[10] Jayasooriya, A. P. (2017). Cancer busting super fats: A quest for real cancer killer. *Wayamba Journal of Animal Science*, *9*, 1507–1516.

[11] Lidén, M., & Eriksson, U. (2006). Understanding retinol metabolism: Structure and function of retinol dehydrogenases. *Journal of Biological Chemistry*, *281*(19), 13001–13004. https://doi.org/10.1074/jbc.R500027200

[12] di Masi, A., Leboffe, L., de Marinis, E., Pagano, F., Cicconi, L., Rochette-Egly, C., . . . , & Nervi, C. (2015). Retinoic acid receptors: From molecular mechanisms to cancer therapy. *Molecular Aspects of Medicine*, *41*, 1–115. https://doi.org/10.1016/j.mam.2014.12.003

[13] Miller, A. P., Coronel, J., & Amengual, J. (2020). The role of β-carotene and vitamin A in atherogenesis: Evidences from preclinical and clinical studies. *Molecular and Cell Biology of Lipids*, *1865*(11), 158635. https://doi.org/10.1016/j.bbalip.2020.158635

[14] Moise, A. R., Kuksa, V., Blaner, W. S., Baehr, W., & Palczewski, K. (2005). Metabolism and transactivation activity of 13, 14-dihydroretinoic acid. *Journal of Biological Chemistry*, *280*(30), 27815–27825. https://doi.org/10.1074/jbc.M503520200

[15] Balmer, J. E., & Blomhoff, R. (2002). Gene expression regulation by retinoic acid. *Journal of Lipid Research*, *43*(11), 1773–1808. https://doi.org/10.1194/jlr.R100015-JLR200

[16] Xu, D., Yang, C., Shen, Q., Pan, S., Liu, Z., Zhang, T., . . . , & Shi, P. (2021). A single mutation underlying phenotypic convergence for hypoxia adaptation on the Qinghai-Tibetan Plateau. *Cell Research*, *31*(9), 1032–1035. https://doi.org/10.1038/s41422-021-00517-6

[17] Jayasooriya, A. P. (2018). How to safeguard an appropriate "all *trans* retinoic acid" concentration to keep cell division on track: Exploring therapeutic hotspots from metabolomics. *Medical Hypotheses*, *121*, 56. https://doi.org/10.1016/j.mehy.2018.09.020

[18] The UniProt Consortium. (2017). UniProt: The universal protein knowledgebase. *Nucleic Acids Research*, *45*(D1), D158–D169. https://doi.org/10.1093/nar/gkw1099

[19] Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., . . . , & Velankar, S. (2022). AlphaFold protein structure database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research*, *50*(D1), D439–D444. https://doi.org/10.1093/nar/gkab1061

[20] David, A., Islam, S., Tankhilevich, E., & Sternberg, M. J. E. (2022). The AlphaFold database of protein structures: A biologist's guide. *Journal of Molecular Biology*, *434*(2), 167336. https://doi.org/10.1016/j.jmb.2021.167336

[21] Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., & Baker, N. A. (2004). PDB2PQR: An automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. *Nucleic Acids Research*, *32*(suppl_2), W665–W667. https://doi.org/10.1093/nar/gkh381

[22] Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, *1–2*, 19–25. https://doi.org/10.1016/j.softx.2015.06.001

[23] Huang, J., & MacKerell, A. D. (2013). CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *Journal of Computational Chemistry*, *34*(25), 2135–2145. https://doi.org/10.1002/jcc.23354

[24] van Gunsteren, W. F., & Berendsen, H. J. C. (1988). A leap-frog algorithm for stochastic dynamics. *Molecular Simulation*, *1*(3), 173–185. https://doi.org/10.1080/08927028808080941

[25] Mark, P., & Nilsson, L. (2001). Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *The Journal of Physical Chemistry A*, *105*(43), 9954–9960. https://doi.org/10.1021/jp003020w

[26] Grubmüller, H., Heller, H., Windemuth, A., & Schulten, K. (1991). Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions. *Molecular Simulation*, *6*(1–3), 121–142. https://doi.org/10.1080/08927029108022142

[27] Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., & Pedersen, L. G. (1995). A smooth particle mesh Ewald method. *The Journal of Chemical Physics*, *103*(19), 8577–8593. https://doi.org/10.1063/1.470117

[28] Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). *PROCHECK*: A program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, *26*(2), 283–291. https://doi.org/10.1107/S0021889892009944

[29] Larsson, P., Skwark, M. J., Wallner, B., & Elofsson, A. (2009). Assessment of global and local model quality in CASP8 using Pcons and ProQ. *Proteins: Structure, Function, and Bioinformatics*, *77*(S9), 167–172. https://doi.org/10.1002/prot.22476

[30] Eisenberg, D., Lüthy, R., & Bowie, J. U. (1997). VERIFY3D: Assessment of protein models with three-dimensional profiles. *Methods in Enzymology*, *277*, 396–404. https://doi.org/10.1016/s0076-6879(97)77022-8

[31] McGuffin, L. J. (2008). The ModFOLD server for the quality assessment of protein structural models. *Bioinformatics*, *24*(4), 586–587. https://doi.org/10.1093/bioinformatics/btn014

[32] Willard, L., Ranjan, A., Zhang, H., Monzavi, H., Boyko, R. F., Sykes, B. D., & Wishart, D. S. (2003). VADAR: A web server for quantitative evaluation of protein structure quality. *Nucleic Acids Research*, *31*(13), 3316–3319. https://doi.org/10.1093/nar/gkg565

[33] Tyryshkina, A., Coraor, N., & Nekrutenko, A. (2019). Predicting runtimes of bioinformatics tools based on historical data: Five years of Galaxy usage. *Bioinformatics*, *35*(18), 3453–3460. https://doi.org/10.1093/bioinformatics/btz054

[34] Jurcik, A., Bednar, D., Byska, J., Marques, S. M., Furmanova, K., Daniel, L., . . . , & Kozlikova, B. (2018). CAVER Analyst 2.0: Analysis and visualization of channels and tunnels in protein structures and molecular dynamics trajectories. *Bioinformatics*, *34*(20), 3586–3588. https://doi.org/10.1093/bioinformatics/bty386

[35] Westbrook, J., Feng, Z., Jain, S., Bhat, T. N., Thanki, N., Ravichandran, V., . . . , & Berman, H. M. (2002). The protein data bank: Unifying the archive. *Nucleic Acids Research*, *30*(1), 245–248. https://doi.org/10.1093/nar/30.1.245

[36] Kawabata, T. (2010). Detection of multiscale pockets on protein surfaces using mathematical morphology. *Proteins: Structure, Function, and Bioinformatics*, *78*(5), 1195–1211. https://doi.org/10.1002/prot.22639

[37] Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., . . . , & Fox, D. J. (2009). *Gaussian 09 Revision D. 01* [Computer software]. Gaussian.

[38] Rizvi, S. M. D., Shakil, S., & Haneef, M. (2013). A simple click by click protocol to perform docking: AutoDock 4.2 made easy for non-bioinformaticians. *EXCLI Journal*, *12*, 831–857.

[39] Gasteiger, J., & Marsili, M. (1980). Iterative partial equalization of orbital electronegativity—A rapid access to atomic charges. *Tetrahedron*, *36*(22), 3219–3228. https://doi.org/10.1016/0040-4020(80)80168-2

[40] Augusto, M., Sallem, S., de Sousa, S. A., & da Silva e Silva, F. J. (2007). AutoGrid: Towards an autonomic grid middleware. In *16th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises*, 223–228. https://doi.org/10.1109/WETICE.2007.4407158

[41] Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., & Olson, A. J. (1998). Automated docking using a Lamarckian genetic algorithm and an

empirical binding free energy function. *Journal of Computational Chemistry*, *19*(14), 1639–1662. https://doi.org/10.1002/(SICI)1096-987X(19981115)19:14%3C1639::AID-JCC10%3E3.0.CO;2-B

[42] Lee, J., Cheng, X., Jo, S., MacKerell, A. D., Klauda, J. B., & Im, W. (2016). CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. *Biophysical Journal*, *110*(3), 641A. https://doi.org/10.1016/j.bpj.2015.11.3431

[43] Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., . . . , & Schulten, K. (2005). Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, *26*(16), 1781–1802. https://doi.org/10.1002/jcc.20289

[44] Meza, J. C. (2010). Steepest descent. *WIREs Computational Statistics*, *2*(6), 719–722. https://doi.org/10.1002/wics.117

[45] Chu, J.-W., Trout, B. L., & Brooks, B. R. (2003). A super-linear minimization scheme for the nudged elastic band method. *The Journal of Chemical Physics*, *119*(24), 12708–12717. https://doi.org/10.1063/1.1627754

[46] Brooks, B. R., Brooks, C. L., Mackerell, A. D., Nilsson, L., Petrella, R. J., Roux, B., . . . , & Karplus, M. (2009). CHARMM: The biomolecular simulation program. *Journal of Computational Chemistry*, *30*(10), 1545–1614. https://doi.org/10.1002/jcc.21287

[47] Bussi, G., & Parrinello, M. (2007). Accurate sampling using Langevin dynamics. *Physical Review E*, *75*(5), 056707. https://doi.org/10.1103/PhysRevE.75.056707

[48] Hollingsworth, S. A., & Karplus, P. A. (2010). A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins. *Biomolecular Concepts*, *1*(3–4), 271–283. https://doi.org/10.1515/bmc.2010.022

[49] Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry*, *30*(16), 2785–2791. https://doi.org/10.1002/jcc.21256

How to Cite: Hewage, V. T., Jayakody, R. S., Jayasooriya, A., Senapathi, T., & Govendar, K. (2026). Elucidating and Confirming the Substrate and Cofactor Binding Site of RetSat and Identifying the Optimal Cofactor for Retinol Saturase – A Computational Insight. *Medinformatics*, *3*(1), 74–87. https://doi.org/10.47852/bonviewMEDIN52023586