Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

In silico approach for the prospecting of molecules from Cereus jamacaru (Cereeae) in the treatment of obesity and cardiovascular diseases

Maria Izadora Oliveira Cardoso
Universidade Federal do ABC (UFABC)
Júlia Oliveira
Universidade Federal do ABC (UFABC)
V. Bassaneze
Universidade Federal do ABC (UFABC)
D. T. Amaral
danilo.trabuco@ufabc.edu.br
Universidade Federal do ABC (UFABC)

Research Article

Keywords: Bioactive proteins, bioinformatics, cardiovascular diseases, mandacaru, triacylglycerol

Posted Date: February 19th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-5961003/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

The Caatinga biome in Brazil is home to unique plant species with significant bioactive potential, such as *Cereus jamacaru* (Mandacaru). Historically used in traditional medicine for treating kidney issues, diabetes, and cardiovascular conditions, this plant contains macromolecules, which may play a role in therapeutic applications. Here, we integrate transcriptomics, molecular docking, and molecular dynamics approaches to explore the potential of *C. jamacaru* proteomics to binding to triacylglycerol molecules, particularly targeting triacylglycerol formed by lauric, myristic, and palmitic acids. Transcriptome analysis identified 128,942 transcripts, with 14,739 homologous proteins screened for binding affinities. Molecular docking highlighted an isoform of *Banyan Peroxidase* as a versatile candidate, exhibiting strong binding energies across all triacylglycerols, particularly palmitic acid (-7.632 kcal/mol). *Xyloglucan Endotransglycosylase* demonstrated specificity for myristic acid (-7.752 kcal/mol), while *Non-specific Lipid-Transfer Protein* showed exceptional structural stability in dynamic simulations. The molecular dynamics simulations, performed over 100 ns, revealed key insights into protein stability and ligand interactions. *Banyan Peroxidase* displayed moderate flexibility, enhancing its adaptability to diverse triacylglycerol substrates. Conversely, *Xyloglucan Endotransglycosylase* exhibited compact stability, making it a strong candidate for targeted new applications. These findings highlight the untapped potential of *C. jamacaru* as a source of bioactive proteins, bridging traditional knowledge with advanced computational methodologies.

1. Introduction

Brazil has a rich floral diversity, with the Caatinga biome being particularly notable for its variety of plant species adapted to water scarcity, such as cacti. This region provides valuable resources for local communities, especially in the food and herbal medicine sectors, due to the nutritional and bioactive potential of regional plants [1]. Among these species, the genus *Cereus* Mill. (Cereeae; Cactaceae), particularly *Cereus jamacaru* D.C., known as Mandacaru, stands out. Native to Brazil's semi-arid region, this plant is rich in macromolecules, with its spines and stems containing unsaturated fatty acids, such as oleic and linoleic acids, as well as saturated fatty acids like palmitic and stearic acids [2]. For decades, the root and stem of Mandacaru have been used in infusions and teas by the populations of northeastern Brazil to treat a broader range of conditions, including diuresis, kidney issues, diabetes, and worm infections. It has also been employed as a laxative, antihypertensive, and anti-rheumatic agent [3, 4].

With advancements in omics sciences, research on the medicinal properties and macromolecular composition of natural sources like Mandacaru has increased [5]. The bioactive compounds found in plants form the basis for discovering pharmaceuticals and industrial enzymes through innovative techniques. This aligns with traditional practices of plant prospecting, which have long been integral to cultures worldwide, observing their medicinal, cosmetic, and culinary properties. It is estimated that 80% of the global population relies on traditional medicine, with 85% of this dependence rooted in the use of plants and their bioactive compounds [6, 7]. Mandacaru, with its diverse applications and bioactive

potential, exemplifies the intersection of traditional knowledge and modern scientific exploration, reinforcing the importance of its study for both local and global health solutions.

Recently, approaches to facilitate the bioprospecting of biomolecules have relied on computer-assisted (*in silico*) strategies, such as molecular docking, due to their advantages of quick access and low cost, allowing for large-scale application [8, 9]. Inverse virtual screening (IVS), which is based on molecular docking methods, involves predicting the binding mode and affinity of a ligand with a receptor. During the IVS process, molecular docking is used as a search step within a protein database. In this context, computational methods for protein modeling and molecular docking are highlighted. Both methods aim to predict protein structures and energetically assess the affinity between the protein and its substrate [10].

Fatty acids play a critical role in cardiovascular health, particularly long-chain saturated fatty acids such as myristic and palmitic acids, which are associated with elevated Low-Density Lipoprotein cholesterol (LDL-c) levels and increased cardiovascular risk. This risk arises through mechanisms like cholesterol esterification [11]. A significant association exists between reducing LDL-c levels and lower rates of cardiovascular events, including cardiovascular death, acute myocardial infarction (MI), coronary revascularization, and stroke [12]. LDL-c is generated via the chylomicron pathway [13]. This pathway begins with the digestion of triacylglycerols by gastric lipase in the stomach and pancreatic lipase in the intestine, followed by the absorption of fatty acids and the subsequent production of chylomicrons by enterocytes [13]. Lipase, an enzyme central to this process, is targeted by pharmacological agents such as orlistat, which binds to the enzyme's active site [14]. This strategy promotes the elimination of triacylglycerols through the intestinal tract before digestion, thereby reducing cardiovascular risk. However, this approach indiscriminately eliminates all types of triacylglycerols, including beneficial unsaturated ones. To date, no strategy directly and specifically targets triacylglycerols for human health.

Recent epidemiological and genetic studies have established that triacylglycerols are among the primary contributors to atherosclerotic cardiovascular disease (ASCVD; [15]). Obesity is a key underlying risk factor for ASCVD, as it facilitates other risk factors such as hypertriglyceridemia and insulin resistance, both of which can trigger ASCVD. This constellation of factors constitutes metabolic syndrome, also referred to as metabolic risk factors [16]. The bioactive composition of *C. jamacaru*, positions this plant as a promising source for therapeutic proteins capable of modulating triacylglycerol digestion. This unique biochemical profile, coupled with its historical use in traditional medicine, underscores its potential for addressing contemporary health challenges.

Thus, this study integrates transcriptomics and in *silico* methodologies to investigate the therapeutic the hypothesis of triacylglycerol binding capability to *C. jamacaru* proteins, which might be used to affected triacylglycerol availability and metabolism and, consequently, impact cardiovascular health. Here, we explored molecular docking and molecular dynamics to uncover the therapeutic potential of *C. jamacaru* proteins in mitigating triacylglycerol-related health risks. By identifying candidate proteins with high binding affinities to triacylglycerol, it bridges traditional knowledge with modern computational

techniques. The findings lay the groundwork for cost-effective therapeutic strategies aimed at addressing metabolic disorders and cardiovascular diseases. Moreover, this research highlights the untapped potential of *C. jamacaru* as a source of bioactive proteins, leveraging advanced in *silico* methods to target triglyceride metabolism and cardiovascular health.

2. Materials and Methods

2.1. Data collection and transcriptome assembly

The transcriptome data from the roots and epidermis of *C. jamacaru* were previously generated [17]. Sequencing was performed using the DNB-Seq platform (BGI Americas) in paired-end mode. *De novo* transcriptome assembly was carried out using the Trinity software (version 2.5.11; [18]) and its associated packages. After assembly, transcripts were translated into amino acid sequences using the TransDecoder program.

2.2. Inverse Virtual Screening

The inverse virtual screening (IVS) analysis was performed using an automated workflow, BioProtIS pipeline ([19]; available at https://github.com/BBMDO/BioProtIS), developed in our laboratory. This pipeline employs a set of Python programs to integrates "omics" data and stands out in the analysis of plant transcriptomics, detailed above:

2.2.1. Three-dimensional protein structure modeling

The modeling process of the three-dimensional protein structures was carried out using the Modeller v.10.4 [20]. Subsequently, the identification of known three-dimensional structures in the PDB to be used as templates was performed. For this purpose, the predicted protein sequences from the transcriptome were compared to the protein sequences available in the PDB using the locally Blastp program. Proteins from *Cereus* with less than 50% homology to the template were promptly excluded, as low identity can also result in low-quality modeled protein structures. After identifying the model proteins in the PDB, the *batch_download.sh* script (available at: https://www.rcsb.org/scripts/batch_download.sh), developed by the PDB itself, was applied for the automated download of tertiary structures, resulting in 3,847 proteins. Using the protein sequences of *C. jamacaru* and the tertiary structure of the model sequence, the Modeller program and its Python scripts were employed to perform sequence alignment between the target protein and the selected homologous protein. Based on this alignment, the initial model of the target protein was generated using the known structure of the homologous protein as a reference.

2.2.2. Preparation of inputs for molecular docking

After obtaining the models we used the Openbabel 3.1.1 software [21] to prepare the protein structure. This included removing hydrogens, adding atomic types, establishing interactions and bonds, and centralizing the molecule to define the search grid (grid box). These operations were performed using the *pdb2gmx* and *editconf* functionalities, as proposed by Senra & Fonseca [9]. Additionally, the

Openbabel 3.1.1 software [21] was applied to prepare the structures of triacylglycerol formed by lauric acid, myristic acid, and palmitic acid (CIDs: 10851, 11148, and 11147, respectively).

2.2.3. Molecular docking

With the protein structures and their ligands prepared, we proceeded to the docking analysis, conducted using the blind docking strategy. In blind docking the three triacylglycerol target substrates were virtually positioned within the docking grid without prior information about their binding sites, utilizing the AutoDock Vina program [22] with default parameters. This method enabled the exploration of different conformations and orientations of the compounds within potential binding cavities of the target protein. This approach is particularly relevant as it allows the investigation of flexible conformations, the identification of allosteric binding sites, and functional implications in a more comprehensive manner [23]. As a result, proteins were selected based on their lowest binding energy (kcal/mol), indicating strong interactions between the protein and the substrate. This process aimed to identify potential targets for the functional and metabolic annotation of the pathways of interest.

2.3. Functional annotation

To obtain information about the function and metabolic pathways of the target proteins of *C. jamacaru*, functional annotation was performed using a combination of tools and public databases, including the *blastp* program and the KEGG, NCBI-nr, and UniProt databases. Protein sequences were submitted to Prosite and InterProScan online tools to identify functional domains, protein families, and catalytic or binding residues. The data were analyzed to correlate structural properties with potential biological functions.

2.4. Molecular dynamics

The molecular dynamics simulation aimed to understand the binding potential and explore the structural and dynamic properties of the complex. This analysis was conducted using GROMACS v.2023 [24] in a multi-parallel setup to leverage computational efficiency. Initially, the protein and ligand files, commonly outputted from docking software in PDBQT format, were converted to PDB format using Open Babel. The ligand topology was then prepared using ACPYPE v.2023.10.27 [25] with the GAFF2 force field, which assigns parameters tailored for small molecules, ensuring an accurate representation of the ligand's geometry and electrostatic properties for reliable simulation results.

For the protein topology, the OPLS-AA force field and TIP3P water model were employed to describe system interactions, both selected for their robust performance in simulating biomolecular systems. A dodecahedron simulation box was generated around the protein with a buffer of 1.0 nm from the protein to the box edges, providing adequate space for solvation and interactions [24]. The system was solvated using the SPC216 water model, which balances computational efficiency and accuracy, and ions were added to neutralize the system and maintain a physiological ionic concentration of 0.1 M.

Following solvation and neutralization, the system underwent energy minimization to resolve steric clashes and bad atomic contacts. Equilibration was then performed in two phases. First, the system was

equilibrated under constant volume and temperature (NVT ensemble) at 300K, using a V-rescale thermostat. Next, equilibration was conducted under constant pressure and temperature (NPT ensemble) using the Parrinello-Rahman barostat to maintain thermal and pressure equilibrium at 1 bar. These steps were critical to achieve thermodynamic stability and prepare the system for production dynamics.

The production molecular dynamics simulation was run for 100 ns (50,000,000 steps) with a 2 fs time step, recording energy, coordinates, and velocities every 50,000 steps. Long-range electrostatic interactions were calculated using the Particle Mesh Ewald (PME) method. Periodic boundary conditions (PBC) were applied in all directions to simulate an infinite system and prevent edge effects. Bond lengths involving hydrogen atoms were constrained using the LINCS algorithm, and the Verlet scheme was utilized for neighbor searching with a 1.0 nm cutoff for Coulomb and Van der Waals interactions. In the final stages of analysis, the Root Mean Square Deviation (RMSD) was calculated to assess the structural deviation of the protein over time.

3. Results

3.1. Transcriptome assembly and modelling

The transcriptome assembly of *C. jamacaru* yielded a total of 128,942 transcripts, which were translated into protein sequences. The IVS analysis of the *C. jamacaru* proteomic data began with ~ 120,000 translated aminoacids sequences. Using the BioProtIS pipeline, the protein primary structures were screened for homologous sequences in the PDB, resulting in 14,739 proteins with identified similarities. This extensive dataset served as a foundational step for subsequent structural and functional analyses, enabling a robust exploration of the *C. jamacaru* proteome.

Next, the identified homologous sequences were subjected to three-dimensional structure modeling, generating 3,847 protein structures, which were used for molecular docking analysis. The execution time for each script was estimated based on the volume of information generated and processed. On average, the processing pipeline completed all phases, producing results for each protein in approximately 280 hours. Considering the amount of data used and the methodologies employed, the cost-benefit ratio is favorable, potentially reducing experimental time in the laboratory.

3.2. Molecular Docking

The binding affinities (kcal/mol) for lauric acid, myristic acid, and palmitic acid substrates indicate the interaction between the ligands and their respective protein targets. More negative binding energy values correspond to stronger ligand-protein interactions, while less negative values suggest weaker binding. The results consistently showed strong binding affinities across all ligands, reflecting robust data quality and validation according to standard binding score metrics.

A total of 3,847 proteins derived from *C. jamacaru* were subjected to molecular docking against triacylglycerol substrates composed of lauric acid, myristic acid, and palmitic acid (Table S1). Table 1 highlights the top twenty candidate proteins based on their binding energy scores (kcal/mol). These proteins represent the most promising targets for the identified ligands (Fig. 1). For lauric acid, notable candidates include *Cytokinin Dehydrogenase* (*T89423*), which exhibited the strongest binding energy of -7.782 kcal/mol, followed by *Glycolate Oxidase* (*T57692*) and *Cinnamic Acid 4-Hydroxylase* (*T113801*). The presence of *Banyan Peroxidase* (*T97361*) among the top candidates underscores its potential as a versatile protein for triacylglycerol interactions, a feature also observed for other substrates.

Table 1

C. jamacaru transcripts, their respective binding energies, and annotations for the top twenty candidate protein targets for triacylglycerol formed by lauric acid, myristic acid, and palmitic acid (triglycerides), based on blind docking.

Transcripts	PDB ID	Binding energy (Kcal/mol)	Target protein
Lauric acid triglyceride			
T89423	2q4w	-7,782	*Cytokinin dehydrogenase*
T57692	1gyl	-7,721	*Glycolate Oxidase*
T113801	6vby	-7,673	*Cinnamic acid 4-hydroxylase*
T97361	4cuo	-7,603	*Banyan Peroxidase*
T64835	5nzv	-7,579	*Coatomer subunit beta'*
Myristic acid triglyceride			
T91586	1un1	-7,752	*Xyloglucan Endotransglycosylase*
T32032	4i94	-7,563	*Probable serine/threonine-protein kinase*
T38514	4c44	-7,524	*2-on-2 Hemoglobin*
T97366	4cuo	-7,436	*Banyan Peroxidase*
T6778	7ksc	-7,394	*Non-specific lipid-transfer protein*
Palmitic acid triglyceride			
T97361	4cuo	-7,632	*Banyan Peroxidase*
T19750	7vka	-7,587	*Indole-3-acetic acid-amido synthetase*
T62011	7aaq	-7,57	*Sugar transport protein 10*
T97368	4cuo	-7,561	*Banyan Peroxidase*
T97353	4cuo	-7,484	*Banyan Peroxidase*
* RMSD values for all analysis were 0.0			

For myristic acid, the protein *Xyloglucan Endotransglycosylase* (*T91586*) displayed the strongest binding energy (-7.752 kcal/mol), followed closely by *Non-specific Lipid-Transfer Protein* (*T6778*). These results suggest their suitability for interactions with triacylglycerols containing this fatty acid. Palmitic acid, on the other hand, revealed significant involvement of *Banyan Peroxidase*, represented by three distinct transcripts (*T97361, T97368, and T97353*), with binding energies ranging from − 7.632 to -7.484 kcal/mol. This consistent representation of *Banyan Peroxidase* across all three triacylglycerol types reinforces its potential as a key player in ligand binding, making it a prime target for further functional annotation and experimental validation. These findings provide a strong foundation for subsequent project phases, which will focus on elucidating the functional roles and therapeutic applications of these proteins.

To refine the analyses of the binding site potential (allosteric or not) of *C. jamacaru* proteins, transcripts that appeared repeatedly in the blind docking results for the three different substrates (lauric acid, myristic acid, and palmitic acid) were selected. For this purpose, variations in binding energies and the proteins most frequently observed among the three ligands simultaneously were analyzed (see Table 2 for the top five ligands; for an overview of the table, refer to Table S1). The results revealed a total of 59 proteins that appeared across all three substrates, representing a significant sample of proteins with potential for exploring therapeutic applications.

Table 2
C. jamacaru transcripts; binding energies for triacylglycerol formed by lauric acid, myristic acid, and palmitic acid; and annotation of the twenty best candidate protein targets for the three substrates mentioned according to blind docking.

Transcripts	PDB ID	Binding energy Lauric acid (Kcal/mol)	Binding energy Myristic acid (Kcal/mol)	Binding energy Palmitic acid (Kcal/mol)	Target protein
T97361	4cuo	-7,603	-7,178	-7,632	*Banyan Peroxidase*
T6778	7ksc	-7,475	-7,394	-7,479	*Non-specific lipid-transfer protein*
T91586	1un1	-7,102	-7,752	-7,304	*Xyloglucan Endotransglycosylase*
T97353	4cuo	-7,425	-7,063	-7,484	*Banyan Peroxidase*
T97368	4cuo	-7,386	-6,933	-7,561	*Banyan Peroxidase*

3.3. Functional annotation

The analyses performed with Prosite and InterProScan, along with the molecular docking results, reveal that the three identified and studied proteins possess structural and functional characteristics that make them promising candidates for interacting with triacylglycerols. The *nsLTP* belongs to a family of

proteins recognized for their ability to bind and transport lipids such as triacylglycerols. The presence of the *nsLTP1* domain and a signal peptide suggests that this protein is exported to extracellular environments, where it may play a role in lipid transport and modulation, as corroborated by its affinity for triacylglycerols form by lauric, myristic, and palmitic acids indicated by docking analyses. The *Banyan Peroxidase*, on the other hand, is a plant peroxidase that features conserved domains associated with calcium and heme binding, essential for its role in redox processes. Interestingly, its ability to interact with triacylglycerols suggests a possible additional role, potentially modulating or oxidizing triacylglycerols, which could expand its applicability in metabolic contexts. The *Xyloglucan Endotransglycosylase Protein*, known for its role in carbohydrate metabolism and cell wall modification in plants, exhibited an unexpected capacity to interact with triacylglycerols, which may be related to the flexibility of its catalytic domain and its predisposition for interactions with nonpolar molecules. These findings indicate that all three proteins possess functional affinity for triacylglycerols, broadening their biological roles.

3.4. Molecular dynamics

The molecular dynamics results of the seven analyzed proteins revealed significant differences in structural and energetic stability, which may influence their potential as therapies for triacylglycerol binding and removal (Fig. 2). The RMSD analysis demonstrated that T91586 (*Xyloglucan Endotransglycosylase*) exhibited the highest structural stability, with an average value of 6.69 nm and low variation (0.0656 nm), indicating good consistency over time. In contrast, T89423 (*Cytokinin Dehydrogenase*) showed higher variability in its RMSD, suggesting structural instability under dynamic conditions.

The RMSF analysis highlighted that *T6778* (*Non-specific Lipid-Transfer Protein*) exhibited the lowest residue fluctuations, indicating a compact and stable structure. Meanwhile, *T97361* (*Banyan Peroxidase*) showed greater flexibility in specific regions, such as loops and termini, which may facilitate dynamic interactions with ligands. The radius of gyration (Rg) for *T6778* was also the most consistent, reinforcing its global stability. In contrast, *T32032* (*Probable Serine/Threonine-Protein Kinase*) displayed more pronounced variations, possibly linked to conformational transitions.

Regarding total energy, *T6778* (*Non-specific Lipid-Transfer Protein*) stood out with the lowest average energy (-2382.10 kJ/mol) and reduced variation, indicating an energetically stable conformation. On the other hand, *T97368* (*Banyan Peroxidase*) showed the highest total energy (-7359.22 kJ/mol), reflecting a greater energetic cost to maintain its equilibrium under dynamic conditions. Overall, the molecular dynamics data indicate that *T91586*, *T6778*, and *T97361* possess favorable characteristics for therapeutic applications, either due to global stability or the moderate flexibility needed for specific interactions.

4. Discussion

The therapeutic potential of natural products has been widely recognized, but transitioning traditional knowledge into modern applications poses challenges [26]. While current therapies for hypertriglyceridemia and cardiovascular diseases are often limited by accessibility and affordability, computational approaches like molecular docking present a cost-effective avenue for exploring plant-derived proteins as potential therapeutic agents. Despite extensive studies on well-known medicinal plants, there is a notable gap in the proteomic exploration of native species from underrepresented ecosystems such as the Caatinga. The unique adaptations of *C. jamacaru* to its environment likely yield specialized bioactive proteins, which remain largely untapped.

The molecular docking was employed in this study to evaluate, for the first time, the binding interactions between triacylglycerols composed of myristic, lauric, and palmitic acids and proteins from *Cereus jamacaru*. Notably, significant binding energies were observed, suggesting strong interactions between these triacylglycerols and the identified proteins. Functional annotation through the Prosite and InterProScan databases further supported the molecular docking findings, indicating a potential interaction of these proteins with lipids. This integrated approach demonstrated that the three main identified proteins exhibit structural and functional properties consistent with interactions with triacylglycerols—a feature not previously attributed to these candidates.

Banyan Peroxidase was first purified from the latex of *Ficus benghalensis* and underwent crystallographic analysis in 2012 [27]. While its crystallographic structure has been resolved, its specific biological functions remain largely unexplored, including any potential lipid-binding capabilities. This protein is generally recognized for its role in oxidative stress responses and the detoxification of reactive oxygen species in plants. *Xyloglucan Endotransglycosylase* has been intensely studied since its crystallization in 1992 [28] across several plant species. Its primary known function is the cutting and rejoining of interfibrillar xyloglucan chains, facilitating wall loosening required for plant cell expansion [28]. Despite its extensive study, no lipid-binding capacity has been demonstrated for this enzyme. This protein is primarily associated with plant growth and structural dynamics of the cell wall. The *Non-specific Lipid-Transfer Protein* was crystallized in 1995 from maize seedlings [29], and homologs have been identified in numerous plant species. In tobacco, this protein has been proposed to transfer phospholipids, contributing to plant defense against bacterial and fungal pathogens, and promoting cell wall extension [30]. It is widely recognized for its involvement in lipid transport and signaling within plant tissues, as well as its role in stress responses.

The strategy allowed for the evaluation of the affinity of the modeled proteins for these substrates, providing a deeper understanding of their molecular interaction and functions. The results from this automated approach demonstrated effectiveness in identifying and modeling the proteins with the best binding energy models for the substrates. The discovered proteins may serve as potential targets for the development of future biotechnological production of pharmacological drugs for the treatment of obesity and cardiovascular diseases.

Fatty acids are classified according to the length of their carbon chain and the number and configuration of their bonds. These chemical characteristics are associated with the plasma concentration of cholesterol and its distribution in lipoproteins. Fatty acids can be divided into saturated and unsaturated [11]. Long-chain fatty acids are found in solid form at room temperature, with the main ones being lauric (12:0), myristic (14:0), palmitic (16:0), and stearic (18:0). Different saturated fatty acids may have diversified effects on the lipid profile and cardiovascular risk factors. Nonetheless, the definitive effect of a diet enriched for a specific fatty acid is still under investigation and debate.

The primary sources of dietary fat include butter fat and animal fat, which are rich in palmitic acid, as well as vegetable oils such as soybean oil and palm oil. The latter contains a low concentration of myristic acid but is high in palmitic acid [31, 32]. Due to the stronger cholesterol-raising effect of myristic acid, attributed to the length of its carbon chain, earlier conclusions suggested that palmitic acid could serve as a favorable dietary substitute in some cases [33]. However, more recent studies indicate that palmitic acid derived from palm oil may not be entirely beneficial, as it has been associated with an elevation in LDL-c levels [34]. A proposed mechanism for this effect involves a decrease in hepatic LDL receptor activity combined with an increased production rate of LDL cholesterol [35]. Nevertheless, a systematic review reported that the overall effect of palm oil on lipid profiles remains inconclusive [36], potentially due to its additional effect of raising HDL-c level. This HDL-c-raising effect may counterbalance the LDL-c-raising effect, resulting in a neutral overall impact on cardiovascular health [37].

4.1. Integration of molecular docking and molecular dynamics

The integration of molecular docking and molecular dynamics results provides a comprehensive perspective on the structural and functional behavior of the proteins analyzed. Docking studies revealed the binding affinities of lauric acid, myristic acid, and palmitic acid to key proteins derived from *C. jamacaru*. Among these, the *Banyan Peroxidase* (*T97361*) consistently demonstrated strong binding energies across all ligands, with particularly high affinities for palmitic acid (-7.632 kcal/mol) and lauric acid (-7.603 kcal/mol). This versatility suggests its potential as a primary candidate for interacting with diverse triacylglycerol substrates (Fig. 1). Complementing these findings, the molecular dynamics analysis showed that the *Banyan Peroxidase* maintains moderate structural stability during simulations, indicating that it can retain functional integrity under physiological conditions.

The *Xyloglucan Endotransglycosylase* (*T91586*) also emerged as a significant candidate (Fig. 1), particularly for myristic acid binding, with a binding energy of -7.752 kcal/mol. Molecular dynamics further supported its candidacy by highlighting its compact and stable structure, as evidenced by low RMSD values and consistent radius of gyration. This dual evidence of strong ligand affinity and structural robustness underscores its potential role in therapeutic applications targeting specific triacylglycerol types. Meanwhile, the *Non-specific Lipid-Transfer Protein* (T6778), while showing slightly weaker docking results (Fig. 1), stood out in dynamics analyses due to its superior energetic stability and minimal

structural fluctuations, suggesting its suitability for scenarios requiring high stability under diverse conditions.

The integration of these datasets reveals nuanced roles for each protein in triacylglycerol interaction. *Banyan Peroxidase* emerges as the most versatile and robust candidate, suitable for general applications across multiple triacylglycerol types. *Xyloglucan Endotransglycosylase* excels in specific interactions with myristic acid, while *Non-specific Lipid-Transfer Protein* offers remarkable stability for experimental or therapeutic conditions that demand structural resilience. Together, these results provide a strong framework for prioritizing proteins for further experimental validation and functional annotation, bridging in *silico* findings with practical applications in the reduction of triacylglycerols and cholesterol.

5. Conclusion

The *Banyan Peroxidase* (*T97361*) stands out as a versatile candidate for therapeutic applications, showing consistent binding to all three triacylglycerols analyzed (lauric, myristic, and palmitic acids) and moderate structural stability, making it a promising option for addressing triacylglycerol-related metabolic disorders. The *Xyloglucan Endotransglycosylase* (*T91586*) exhibited strong specificity for myristic acid, while the *Non-specific Lipid-Transfer Protein* (*T6778*) demonstrated superior energetic stability, highlighting their potential in targeted therapeutic applications. Possibly, these candidates might act as scavengers of triacylglycerols if appropriate conditions are offered.

From a global perspective, the increasing prevalence of cardiovascular diseases and obesity underscores the urgent need for innovative and cost-effective therapeutic solutions. The findings presented here could contribute to developing affordable treatments that are accessible to broader populations, particularly in regions with limited resources. Furthermore, leveraging bioactive compounds from native plants such as *C. jamacaru* aligns with sustainable and ethical practices, promoting biodiversity conservation while addressing pressing public health challenges.Experimental validation remains essential to confirm these findings and advance the therapeutic potential of these proteins for clinical applications.

Declarations

Acknowledgments

We acknowledge the support of São Paulo Research Foundation (FAPESP 2022/09910-9 to D.T.A.and 2023/06831-3 to M.I.O.C) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq 132515/2024-5 to M.I.O.C).

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Author contributions

The idea for this study was conceived by MIOC and DTA. Data collection and analyses were performed by MIOC, JO, and DTA. MIOC led the writing of the paper, while JO, VB, and DTA contributed with numerous conceptions and writing. All authors contributed to the intellectual development of the paper, made multiple revisions, and approved the final draft.

Data Availability

The datasets reused in the present study are publicly available on NCBI under the project accession PRJNA1192998.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used chatGPT 4.0 in order to improve English grammar. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

References

1. Cartaxo SL, de Almeida Souza MM, de Albuquerque UP (2010) Medicinal plants with bioprospecting potential used in semi-arid northeastern Brazil. Journal of Ethnopharmacology, 131(2), 326–342. https://doi.org/10.1016/j.jep.2010.07.003
2. Meiado MV, Albuquerque LSC, Rocha EA, Rojas-Aréchiga M, Leal IR. (2010). Seed germination responses of *Cereus jamacaru* D.C. ssp. *Jamacaru* (Cactaceae) to environmental factors. Plant Species Biology 25: 120-128.
3. Sales M. de S. L., Martins LV, Souza I, Deus MSM de, Peron AP (2014) Cereus jamacaru DE CANDOLLE (CACTACEAE), O MANDACARU DO NORDESTE BRASILEIRO. Publicatio UEPG: Ciencias Biologicas e Da Saude, 20(2), 135–142. https://doi.org/10.5212/publ.biologicas.v.20i2.0006
4. Silva LFCR, Valle L de S, Nascimento ARC, Medeiros MFT (2018) Cereus jamacaru DC. (Cactaceae): From 17th century naturalists to modern day scientific and technological prospecting. Acta Botanica Brasilica, 33(2), 191–197. https://doi.org/10.1590/0102-33062018abb0352
5. Jorquera MA, Graether SP, Maruyama F (2019) Editorial: Bioprospecting and biotechnology of xtremophiles In: Frontiers in Bioengineering and Biotechnology (Vol. 7, Issue AUG). Frontiers Media S.A. https://doi.org/10.3389/fbioe.2019.00204
6. Arora J, Goyal S, Ramawat KG (2010) Biodiversity, biology and conservation of medicinal plants of the Thar Desert In: Ramawat KG, editor. Desert Plants Biology and Biotechnology. Berlin: Springer-Verlag, p. 3-36.
7. Yadav G, Meena M (2021) Bioprospecting of endophytes in medicinal plants of Thar Desert: An attractive resource for biopharmaceuticals. Biotechnology Reports, 30, e00629.
8. Xu X, Huang M, Zou X (2018) Docking-based inverse virtual screening: methods, applications, and challenges. Biophysics Reports, 4(1), 1–16. https://doi.org/10.1007/s41048-017-0045-8.

9. Senra MV, Fonseca AL (2023) Toxicological impacts and likely protein targets of bisphenol A in Paramecium caudatum. European Journal of Protistology, 88, 125958.
10. Xu M, Lill MA (2013) Induced fit docking, and the use of QM/MM methods in docking. Drug Discov Today Technol 10:e411– e418
11. Santos RD, Gagliardi ACM, Xavier HT, Magnoni CD, Cassani R, Lottenberg AMP, Casella Filho A, Araújo DB, Cesena FY, Alves RJ, Fenelon G, Nishioka SAD, Faludi AA, Geloneze B, Scherr C, Kovacs C, Tomazzela C, Carla C, Barrera-Arellano D, Ramos S (2013) I Diretriz sobre o consumo de gorduras e saúde cardiovascular. Arquivos Brasileiros De Cardiologia, 100(1), 1–40. https://doi.org/10.1590/S0066-782X2013000900001
12. Silverman MG, Ference BA, Im K, Wiviott SD, Giugliano RP, Grundy SM, Sabatine MS (2016) Association between lowering LDL-C and cardiovascular risk reduction among different therapeutic interventions: a systematic review and meta-analysis. Jama, 316(12), 1289-1297.
13. Hussain MM (2014) Intestinal lipid absorption and lipoprotein formation. Current opinion in lipidology, 25(3), 200-206.
14. Carrière F, Renou C, Ransac S, Lopez V, De Caro J, Ferrato F, Laugier R. (2001) Inhibition of gastrointestinal lipolysis by Orlistat during digestion of test meals in healthy volunteers. American Journal of Physiology-Gastrointestinal and Liver Physiology, 281(1), G16-G28.
15. Laufs U, Parhofer KG, Ginsberg HN, Hegele RA (2020) Clinical review on triglycerides. European Heart Journal, 41(1), 99–109. https://doi.org/10.1093/eurheartj/ehz785
16. Grundy SM (2004) Obesity, metabolic syndrome, and cardiovascular disease. Journal of Clinical Endocrinology and Metabolism, 89(6), 2595–2600. https://doi.org/10.1210/jc.2004-0372
17. Amaral DT, Bonatelli IA, Romeiro-Brito M, Telhe MC, Moraes EM, Zappi DC, Franco FF (2024) Comparative transcriptome analysis reveals lineage-and environment-specific adaptations in cacti from the Brazilian Atlantic Forest. Planta, 260(1), 4.
18. Grabherr, M., Haas, B., Yassour, M. et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29, 644–652. https://doi.org/10.1038/nbt.1883
19. Virgens GS, Oliveira J, Cardoso MIO, Teodoro JA, Amaral DT. BioProtIS: Streamlining protein-ligand interaction pipeline for analysis in genomic and transcriptomic exploration. J Mol Graph Model. 2024 May;128:108721. doi: 10.1016/j.jmgm.2024.108721. Epub 2024 Jan 30. PMID: 38308972.
20. Webb B, Sali A (2016) Comparative Protein Structure Modeling Using Modeller. Current Protocols in Bioinformatics 54, John Wiley & Sons, Inc., 5.6.1-5.6.37.
21. O'Boyle, N.M., Banck, M., James, C.A. et al. (2011) Open Babel: An open chemical toolbox. J Cheminform 3, 33. https://doi.org/10.1186/1758-2946-3-33.
22. Trott O, Olson AJ (2009) Autodock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. Journal of Computational Chemistry, 31, 455-461. https://doi.org/10.1002/jcc.21334.
23. Iorga B, Herlem D, Barré E, Guillou C (2006) Acetylcholine nicotinic receptors: finding the putative binding site of allosteric modulators using the “blind docking” approach. Journal of Molecular

Modeling, 12 (3), pp.366-372. ff10.1007/s00894-005-0057-zff. ffhal-03161516f

24. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E (2015) "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers," SoftwareX, 1–2 19–25.
25. Sousa da Silva AW, Vranken WF (2012) ACPYPE-Antechamber python parser interface. *BMC research notes*, *5*, 1-8.
26. Porras G, Chassagne F, Lyles JT, Marquez L, Dettweiler M, Salam AM, Quave CL (2020) Ethnobotany and the role of plant natural products in antibiotic drug discovery. Chemical reviews, 121(6), 3495-3560.
27. Sharma A, Palm GJ, Kumari M, Panjikar S, Jagannadham MV, Hinrichs W (2012) Purification, crystallization and preliminary crystallographic analysis of banyan peroxidase. Acta Crystallographica Section F: Structural Biology and Crystallization Communications, 68(8), 931-934.
28. Fry SC, Smith RC, Renwick KF, Martin DJ, Hodge SK, Matthews KJ (1992) Xyloglucan endotransglycosylase, a new wall-loosening enzyme activity from plants. Biochemical Journal, 282(3), 821-828.
29. Shin DH, Lee JY, Hwang KY, Kim KK, Suh SW (1995) High-resolution crystal structure of the non-specific lipid-transfer protein from maize seedlings. Structure, 3(2), 189-199.
30. Wang NJ, Lee CC, Cheng CS, Lo WC, Yang YF, Chen MN, Lyu PC (2012) Construction and analysis of a plant non-specific lipid transfer protein database (nsLTPDB). In BMC genomics (Vol. 13, pp. 1-9). BioMed Central.
31. Fattore E, Bosetti C, Brighenti F, Agostoni C, Fattore G (2014) Palm oil and blood lipid–related markers of cardiovascular disease: a systematic review and meta-analysis of dietary intervention trials. The American journal of clinical nutrition, 99(6), 1331-1350.
32. Mancini A, Imperlini E, Nigro E, Montagnese C, Daniele A, Orrù S, Buono P (2015) Biological and nutritional properties of palm oil and palmitic acid: effects on health. Molecules, 20(9), 17339-17361.
33. Zock PL, de Vries JHM, Katan MB (1994) Impact of myristic acid versus palmitic acid on serum lipid and lipoprotein levels in healthy women and men. Arterioscler Thromb;14:567–75.
34. Sun Y, Neelakantan N, Wu Y, Lote-Oke R, Pan A, van Dam RM (2015) Palm oil consumption increases LDL cholesterol compared with vegetable oils low in saturated fat in a meta-analysis of clinical trials. The Journal of nutrition, 145(7), 1549-1558.
35. Nicolosi RJ (1997) Dietary fat saturation effects on low-density-lipoprotein concentrations and metabolism in various animal models. The American journal of clinical nutrition, 65(5), 1617S-1627S.
36. Ismail SR, Maarof SK, Siedar Ali S, Ali A (2018) Systematic review of palm oil consumption and the risk of cardiovascular disease. PLoS One, 13(2), e0193533.
37. Unhapipatpong C, Shantavasinkul PC, Kasemsup V, Siriyotha S, Warodomwichit D, Maneesuwannarat S, Thakkinstian A (2021) Tropical oil consumption and cardiovascular disease:

An umbrella review of systematic reviews and meta analyses. Nutrients, 13(5), 1549.

Figures

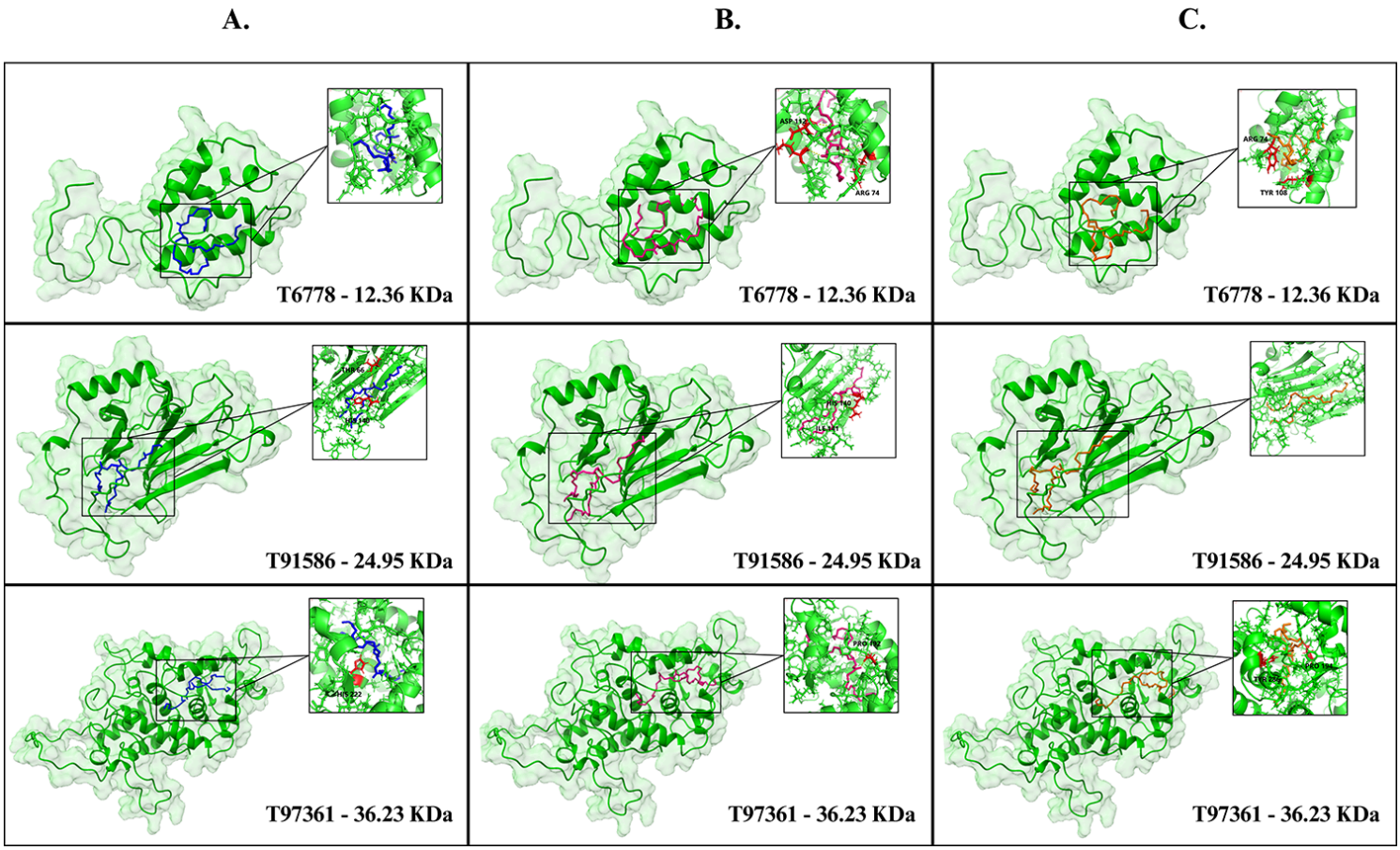

Figure 1

Three-dimensional structure of the three main proteins with the best binding energy for the three substrates: T97361 - b*anyan peroxidase* (BP), T6778 - *nonspecific lipid transfer protein* (NSLT), and T91586 - *xyloglucan endotransglycosylase* (XE). A) Docking for the anchoring of proteins with the lauric acid ligand. B) Docking for the anchoring of proteins with the palmitic acid ligand. C) Docking for the anchoring of proteins with the myristic acid ligand.

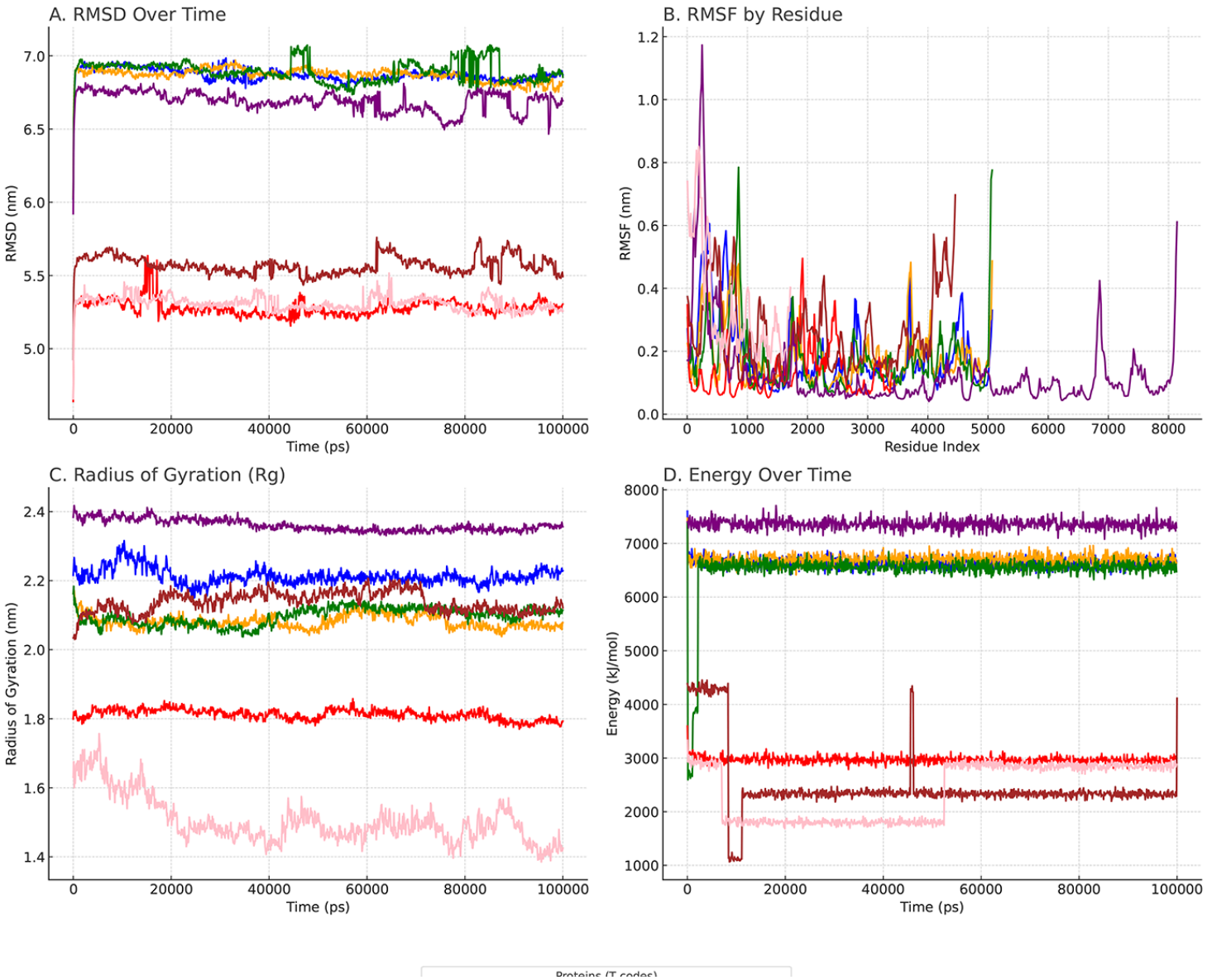

Figure 2

Analysis of molecular dynamics parameters for candidate proteins, presenting the results for seven proteins derived from *C. jamacaru*. The panel include, (A) RMSD over time, showing the structural stability of each protein; (B) RMSF per residue; (C) radius of gyration (Rg); and (D) total energy, reflecting the stability and energetic requirements of the protein conformations during the simulation. Each line corresponds to a protein identified by its transcript code (T), with colors matching the legend provided.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS1.xlsx