Journal of Medicinal Plants and By-products (2026) 03: 383– 393

Bioinformatic Analysis of the Hawthorn PAL gene's Role in OPC Biosynthesis and its Sstructural Comparison to the Cardiomyopathy Gene *MYBPC3*

Seyed Ali Forghani[1], Sara Rashki Ghalehnoo[2], Yeganeh Shafiei[3] and Bahman Fazeli-Nasab[4*]

[1] *Student Research Committee, Zabol University of Medical Sciences, Zabol, Iran*
[2] *Department of Cardiology, School of Medicine, Amir al momenin Hospital, Zabol University of Medical Sciences, Zabol, Iran*
[3] *Pharmaceutical Sciences Research Center, Health Institute, Kermanshah University of Medical Sciences, Kermanshah, Iran*
[4] *Department of Agronomy and Plant Breeding, Agriculture Institute, Research Institute of Zabol, Zabol, Iran*

Article Info

Article Type
Original Article

Article History
Received: 01 October 2025
Accepted: 19 November 2025

***Corresponding author**
bfazelinasab@gmail.com

ABSTRACT

This investigation conducted a computational examination of the phenylalanine ammonia-lyase (PAL) gene within *Crataegus* species, a pivotal enzyme governing the phenylpropanoid pathway and the biosynthesis of phenolic and flavonoid metabolites. The study also probed for potential structural homology between the plant PAL enzyme and the human *MYBPC3* protein, a sarcomeric constituent whose mutations are a predominant genetic etiology for familial hypertrophic cardiomyopathy. The methodological framework integrated a suite of bioinformatic applications. Protein sequence homology was assessed using BLASTp algorithms, while tertiary protein architecture was predicted via the Swiss-Model workspace. Molecular docking simulations, executed with AutoDock, characterized putative protein-ligand interfaces. Complementary analyses on dedicated computational servers facilitated active site prognostication, determination of pivotal residue constituents, and evaluations of conformational stability. Comparative sequence alignment established a lack of significant structural conservation between the PAL and *MYBPC3* polypeptides. Notwithstanding this divergence, molecular docking simulations demonstrated a capacity for oligomeric procyanidin (OPC) ligands to engage the PAL active site, a interaction characterized by favorable binding affinities and the formation of stable hydrogen bonds. Furthermore, structural profiling elucidated functional domains and critical catalytic residues intrinsic to PAL function. In conclusion, this research underscores the integral role of the PAL gene in the OPC biosynthetic machinery of hawthorn, as delineated by sophisticated *in silico* methodologies. The absence of structural mimicry with *MYBPC3* confirms the functional specificity of each protein within their respective biological contexts. These insights provide a foundational platform for subsequent exploration in metabolic engineering and the development of plant-based therapeutic agents.

Keywords: Plant Secondary Metabolism, Oligomeric Procyanidins, Cardiac Protein Structure, Hypertrophic Cardiomyopathy Mechanism, Computational Protein Modeling

How to cite this paper
Forghani, S.A., Rashki Ghalehnoo, S., Shafiei, Y., Fazeli-Nasab, B. Bioinformatic Analysis of the Hawthorn PAL gene's Role in OPC Biosynthesis and its Sstructural Comparison to the Cardiomyopathy Gene *MYBPC3*. Journal of Medicinal Plants and By-products, 2026; 15(3): 383-393. doi: 10.22034/jmpb.2025.370886.2064

INTRODUCTION

The phenylalanine ammonia-lyase (PAL) enzyme, encoded by the PAL gene, serves as a fundamental biocatalyst in the metabolic network responsible for phenolic biosynthesis in plants. It facilitates the non-oxidative deamination of L-phenylalanine, yielding cinnamic acid as the primary substrate for the phenylpropanoid cascade [1]. This pathway acts as a metabolic gateway to a diverse spectrum of specialized metabolites, such as flavonoids, lignins, and oligomeric proanthocyanidins (OPCs), which contribute to phytodefense mechanisms, radical quenching activities, and the augmentation of therapeutic attributes in medicinal plants [2].

OPCs products of phenylpropanoid metabolism, have drawn considerable scientific interest owing to their powerful antioxidative capacity and cardioprotective benefits [3, 4]. *Crataegus* spp. (hawthorn) represents a pharmacologically relevant species abundant in OPCs, with a longstanding history in traditional medicine for managing cardiovascular ailments and supporting vascular integrity. Concurrently, various organs of palm plants—such as fruits, foliage, and seeds—have been employed ethnobotanically across different cultures to address gastrointestinal, inflammatory, and cardiac conditions. These documented uses underscore the pharmaceutical promise inherent in palm-derived metabolites and reinforce the value of investigating bioactive constituents in *Crataegus* and taxonomically related species [5]. Nevertheless, elucidating the genetic control and expression dynamics of the PAL gene in hawthorn, particularly its contribution to OPC accumulation, remains an understudied area.

A study indicates that modulating PAL transcriptional activity can exert substantial effects on the yield and profile of phenolic compounds. Under abiotic or biotic stressors—including chilling, hydric deficit, or pathogenic assault—PAL expression is frequently induced, activating defense-related metabolic fluxes and amplifying the synthesis of protective antioxidants [6]. Consequently, the isolation and structural characterization of the PAL gene in medicinal species such as hawthorn represent a pivotal

research objective with implications for augmenting their pharmacologic value.

Accumulating evidence further points to a connection between dietary phenolic intake and cardiovascular wellness in humans. Epidemiological data correlate high consumption of flavonoids and proanthocyanidins with a decreased incidence of cardiovascular pathologies [7]. These phytochemicals are thought to promote cardiac function through mitigation of oxidative damage, inflammatory regulation, and enhancement of endothelial performance.

A central genetic determinant in cardiac pathology is the *MYBPC3* gene, which encodes myosin-binding protein C. This sarcomeric protein is instrumental in maintaining cardiac muscle architecture and contractility [8, 9]. Pathogenic variants in *MYBPC3* constitute the predominant genetic etiology of hypertrophic cardiomyopathy (HCM), a disorder marked by left ventricular hypertrophy and diastolic dysfunction. Emerging data propose that phenolic compounds may modulate the expression of cardiac genes such as *MYBPC3* through epigenetic modifications and signal transduction pathways [10-12].

In this context, a bioinformatics-driven examination of the PAL gene sequence in *Crataegus*, along with its comparative analysis with cardioprotective genes like *MYBPC3*, may reveal novel mechanistic intersections between plant secondary metabolism and human pathophysiology. Such an integrated strategy not only provides a foundation for the metabolic engineering of therapeutic plants but also fosters innovative approaches for preventing and managing cardiovascular disease. In the present investigation, the PAL gene homolog from *Crataegus* was identified and comprehensively analyzed using in silico tools. We characterized the secondary and tertiary conformation of the corresponding protein, delineated its functional domains, and predicted potential biomolecular interactions. Additionally, structural and sequential comparisons were conducted to evaluate possible biological parallels with the *MYBPC3* protein. This multifaceted methodology yields novel perspectives on the role of PAL in OPC biosynthesis and its putative influence on genes associated with cardiomyopathies.

MATERIALS AND METHODS

Acquisition and Curation of the Pal Gene Sequence for Investigation

Given the absence of a fully annotated *Phenylalanine Ammonia-Lyase (PAL)* gene sequence for *Crataegus* spp., the nucleotide sequence from the phylogenetically proximate species *Malus x domestica* (apple) was employed as a reference. The complementary DNA (cDNA) sequence corresponding to the PAL gene, designated under the accession identifier AF494403.1, was procured from the National Center for Biotechnology Information (NCBI) repository. To delineate evolutionarily conserved and functionally critical domains, homologous PAL sequences from additional taxa within the Rosaceae family were collated. A multiple sequence alignment (MSA) was executed utilizing the Clustal Omega algorithm, enabling the identification of conserved protein motifs that were subsequently leveraged for in-depth structural and functional characterization [13, 14].

Profiling of the PAL Protein's Physicochemical and Structural Properties

The deduced amino acid sequence of the PAL gene from *Malus x domestica*, serving as a proxy for the hawthorn ortholog, was subjected to computational analysis via the ExPASy ProtParam platform. A comprehensive suite of intrinsic physicochemical attributes was computed, encompassing molecular mass, theoretical isoelectric point (pI), instability index, molar extinction coefficient at 280 nm, quantitative amino acid profiling (differentiating polar and hydrophobic residues), aliphatic index, grand average of hydropathicity (GRAVY), and predicted half-life under various cellular expression systems [15].

Computational Elucidation of Secondary and Tertiary Protein Architecture

The secondary structural topology of the PAL polypeptide was forecasted using the Self-Optimized Prediction Method with Alignment (SOPMA) server. This analysis quantified the proportional representation of α-helices, β-strands, β-turns, and random coils, providing foundational insights into the protein's folding propensity and potential functional regions.

For de novo prediction of the three-dimensional conformation, homology modeling was conducted using the SWISS-MODEL workspace. From the suite of generated models, two high-fidelity structural templates were identified and rigorously evaluated based on Global Model Quality Estimation (GMQE) and QMEAN scoring functions. The optimal model, derived from the template with UniProt ID Q8RVM4.1.A, demonstrated complete sequence identity with the query. This model achieved a superior GMQE score of 0.95 and a QMEANDisCo Global score of 0.86 ± 0.09, attesting to its high predictive accuracy and structural plausibility [16, 17].

Model validation was performed using the integrated structure assessment tools within SWISS-MODEL. MolProbity analysis yielded an exceptional MolProbity score of 0.50, with a clash score of 0.00 indicating no steric hindrances. Ramachandran plot analysis revealed that 98.98% of residues occupied the most favored regions, with no outliers detected. The model exhibited only two disallowed bond angles, which were deemed inconsequential to the global structural integrity. The finalized tertiary structure was rendered and interrogated using PyMOL for subsequent visual and functional analysis.

Sourcing, Preparation, and Conformational Optimization of OPC Ligands

To probe potential molecular recognition events between bioactive constituents of *Crataegus* and the PAL enzyme, three specific OPC -Procyanidin B1, B2, and C1-were selected as ligands. This selection was predicated on their documented cardioprotective efficacy and confirmed presence in hawthorn phytochemical profiles. Their canonical SMILES notations were acquired, and initial 3D structural data files in SDF format were retrieved from the PubChem chemical database [18, 19].

For ligands where a pre-optimized 3D conformation was unavailable (notably Procyanidin C1), the 2D structural representations were converted into three-dimensional atomic coordinates using Avogadro software. All ligand geometries subsequently underwent energy minimization to attain a stable low-energy conformation. This optimization was performed using the Merck Molecular Force Field 94 (MMFF94) within Avogadro, ensuring thermodynamic stability for subsequent computational analyses [20]. The resulting refined structures were archived in Protein Data Bank (PDB) format for deployment in molecular docking simulations (Table 1).

Table 1 Molecular properties and database identifiers of selected OPC compounds used in this study.

Compound	CID	Molecular formula	Molecular weight	Storage format
Procyanidin B1	11250133	$C_{30}H_{26}O_{12}$	578.50 g/mol	SDF / PDB
Procyanidin B2	130556	$C_{30}H_{26}O_{12}$	578.50 g/mol	SDF / PDB
Procyanidin C1	169853	$C_{45}H_{38}O_{18}$	866.80 g/mol	SDF / PDB

Structural and Functional Profiling of the Human *MYBPC3* Gene

To delineate the structural and functional attributes of the human cardiac Myosin-Binding Protein C (*MYBPC3*) gene, its canonical protein sequence was acquired from the UniProt knowledgebase (Accession: Q14896). A comprehensive survey of associated pathologies, mutational spectra, and molecular etiologies was conducted by mining data from OMIM, GeneCards, and UniProt.

Querying the OMIM repository with "*MYBPC3*" elucidated prevalent mutation classes—encompassing missense, nonsense, and frameshift variants—linked to disorders such as Familial Hypertrophic Cardiomyopathy (HCM), alongside proposed pathomechanisms. The Mendelian Inheritance in Man (MIM) identifier for the *MYBPC3* locus was documented as 600958.

For three-dimensional structural interrogation, the AlphaFold Protein Structure Database was employed. The predicted atomic coordinates for the human *MYBPC3* protein (Model ID: AF-Q14896-F1-v4), generated by the AlphaFold v4 deep learning system, were retrieved. This model was characterized by high per-residue confidence scores (pLDDT), endorsing its reliability for subsequent structural evaluations [17, 21, 22].

Computational Docking of OPC Ligands with the *MYBPC3* Protein

To probe potential molecular interactions between OPCs and the *MYBPC3* protein, in silico docking simulations were performed using AutoDock Tools v1.5.6 and AutoDock 4.2. The initial *MYBPC3* structure was sourced from the Protein Data Bank, followed by the removal of crystallographic water molecules and extraneous heteroatoms to optimize the receptor file. Both the protein and OPC ligands were subsequently prepared in PDBQT format, incorporating Gasteiger charges and rotatable bond definitions.

A docking grid parameterized to envelop all putative functional domains of the *MYBPC3* protein was established. A Lamarckian Genetic Algorithm (LGA) was implemented with standard parameters to exhaustively sample the conformational landscape and identify the most thermodynamically favorable ligand-binding poses.

Post-simulation analysis entailed the extraction of binding affinity (ΔG) values, ligand spatial coordinates, and detailed interaction profiles. The specific nature of intermolecular forces-including hydrogen bonding, van der Waals contacts, and hydrophobic effects—was scrutinized to pinpoint critical binding epitopes and evaluate the complex stability of OPCs within the *MYBPC3* binding pocket [23, 24].

Interaction Profiling and Structural Visualization

For granular analysis of the docking outcomes, the resulting protein-ligand complexes were imported into PyMOL. This enabled detailed examination of the ligand-binding loci, identification of participating amino acid residues, and assessment of ligand steric accommodation within the protein's surface topography.

A quantitative and structural dissection of the binding regions was further conducted leveraging data from the PDBsum database. Salient characteristics of the binding clefts—including volume, depth, solvent-excluded surface area, and the biochemical properties of the lining residues—were extracted and juxtaposed with the identified OPC docking sites. Additionally, PDBsum was utilized to analyze internal protein pores and putative transport tunnels, determining if ligand engagement sites coincided with these structural features. Metrics such as channel lumen radius, longitudinal extent, polarity profile, and inter-cavity connectivity were assessed.

Finally, the computed binding free energy (ΔG) and the enumeration of potential hydrogen bonds (inferred from interatomic distances in PyMOL) served as principal quantitative indicators for comparative evaluation of the OPC compounds' binding potential [25].

Interaction Network and Pathway Enrichment Analysis

To investigate the putative regulatory interplay between OPCs and *MYBPC3* gene expression, a systems biology approach was undertaken using the STITCH and STRING databases. These platforms facilitated the reconstruction of protein-protein interaction (PPI) networks centered on *MYBPC3*, with the aim of identifying direct interactors, associated signaling cascades, and potential modulatory roles for phytochemicals on *MYBPC3* function or expression.

The resulting network topology was characterized by metrics such as node cardinality, average node degree, clustering coefficient, and an interaction enrichment p-value, the latter signifying the statistical robustness of the observed interactions against a random background model [26, 27].

RESULTS

Physicochemical Profiling of the PAL Protein

The primary structure of the PAL protein was derived through in silico translation of the corresponding *Malus x domestica* mRNA sequence. The deduced polypeptide comprises a chain of 920 amino acid residues, corresponding to a predicted molecular mass of 94,235.71 Da. Computational determination of the theoretical isoelectric point (pI) yielded a value of 8.66, classifying the protein as fundamentally basic in character. Deconvolution of the amino acid constituency within the translated hawthorn PAL sequence identified a distinct compositional bias. Specifically, alanine (Ala, 15.1%), glycine (Gly, 12.3%), and threonine (Thr, 10.0%) constituted the most predominant residues within the primary structure, as delineated in Figure 1. The computed instability index was 50.76, a value which posits that the protein possesses inherent structural lability and is likely unstable under standard *in vitro* conditions. An aliphatic index of 63.90 was calculated, pointing toward a moderate degree of thermal resilience. Furthermore, the grand average of hydropathicity (GRAVY) index was determined to be 0.072, signifying that the protein, as a whole, exhibits a marginally hydrophilic character. The in vivo and in vitro half-life of the protein was projected to be context-dependent: approximately 1.1 hours in a mammalian reticulocyte lysate system (in vitro), a mere 3 minutes in a yeast cellular environment (in vivo), and exceeding 10 hours within *E. coli*. Analysis of the molar extinction coefficient at 280 nm predicted a value of 67,910 M/cm, indicating a substantial inherent capacity for ultraviolet light

absorption, largely attributable to its constituent aromatic amino acids.

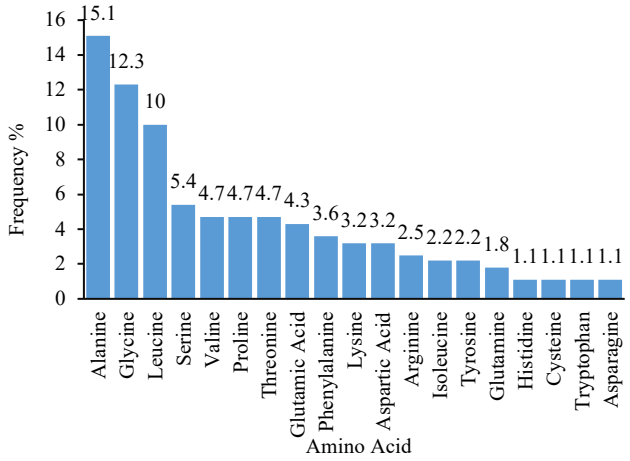

Fig. 1 Amino acid composition (%) in the PAL protein sequence

Prediction of Secondary and Tertiary Architecture for the PAL Protein

Computational prognostication of the PAL protein's secondary structure, executed via the SOPMA algorithm, indicated a pronounced predominance of α-helical conformations within its polypeptide backbone. This pronounced α-helical propensity is a structural motif congruous with the protein's catalytic function in mediating specific biochemical interactions within metabolic pathways. The quantitative distribution of secondary structural elements—encompassing α-helices, β-sheets, random coils, and β-turns—is itemized in the dedicated results summary (Table 2).

Table 2 Secondary Structure Prediction of PAL Protein (UNK_274810) Using GOR4

Secondary structure type	Number of residues	Percentage (%)
Alpha Helix (H)	25	25.00%
3-Helix (G)	0	0.00%
Pi Helix (I)	0	0.00%
Beta Bridge (B)	0	0.00%
Extended Strand (E)	17	17.00%
Beta Turn (T)	0	0.00%
Bend Region (S)	0	0.00%
Random Coil (C)	58	58.00%
Ambiguous States (?)	0	0.00%
Other States	0	0.00%

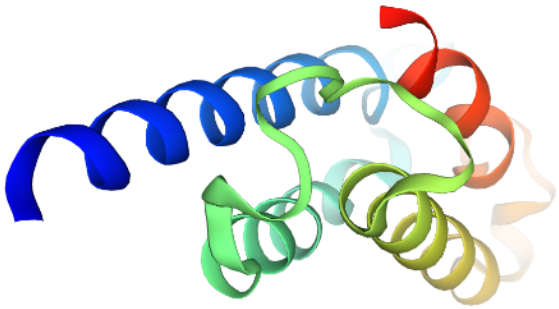

Fig. 2 Predicted 3D structure of the PAL protein from Malus domestica, generated using Swiss-Model (template Q8RVM4.1.A). GMQE = 0.95, QMEANDisCo = 0.86 ± 0.09, indicating high-quality structural prediction

The computationally modeled tertiary architecture of the PAL protein, presented in Figure 2, was generated through homology modeling on the SWISS-MODEL platform using the homologous structure Q8RVM4.1.A as a template. The predicted conformation demonstrates a significant preponderance of α-helical domains, a structural characteristic that aligns mechanistically with its recognized catalytic function. Quantitative evaluation of the model's fidelity produced a Global Model Quality Estimation (GMQE) score of 0.95 and a QMEANDisCo Global score of 0.86 ± 0.09, collectively affirming the high degree of confidence and stereochemical robustness of the predicted structure.

Structural validation of the PAL protein homology model via MolProbity revealed exceptional stereochemical quality. The Ramachandran plot analysis confirmed that 98.98% of all residues occupy energetically favored regions, with an absence of significant Ramachandran outliers, steric clashes, or aberrant torsion angles. This collective data underscores the model's high structural integrity and conformational reliability (Fig. 3). In summary, the refined model exhibits robust structural stability and is of sufficient resolution and quality to serve as a credible foundation for subsequent computational investigations.

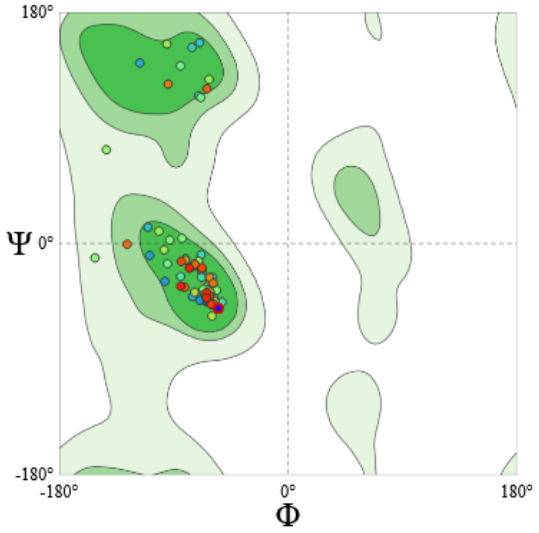

Fig. 3 Ramachandran plot of the PAL protein model evaluated using MolProbity. 98.98% of residues are located within allowed regions, confirming structural integrity

Results of Identification, Extraction, and Optimization of Procyanidin (OPCS) Structures

Molecular Docking Ligand Preparation

To facilitate the computational analysis of molecular interactions between bioactive constituents of *Crataegus* and the PAL enzyme, a panel of three representative OPCs—Procyanidin B1, Procyanidin B2, and Procyanidin C1—was curated. This selection was predicated on extensive pharmacological literature documenting their cardioprotective efficacy and their established abundance in hawthorn phytochemical profiles.

The three-dimensional structural coordinates for Procyanidin B1 (PubChem CID: 11250133) and Procyanidin B2 (CID: 130556) were acquired in SDF format from the PubChem database. In contrast, a pre-optimized 3D conformation for Procyanidin C1 (CID: 169853) was unavailable; consequently, its two-dimensional structural representation was retrieved and subsequently converted into a three-dimensional molecular model using the molecular editing suite Avogadro (v1.2.0) [18, 19].

To ensure physicochemical realism and conformational accuracy for subsequent docking simulations, all ligand structures underwent energy minimization. This optimization was performed employing the Merck Molecular Force Field 94 (MMFF94) within the Avogadro environment, which iteratively refined atomic

coordinates to arrive at low-energy, thermodynamically stable conformers [20].

Stereochemical analysis of the finalized models confirmed the conserved presence of reactive phenolic hydroxyl moieties and characteristic B-type interflavan linkages, structural hallmarks definitive of the procyanidin class found in *Crataegus*. The optimized structures were subsequently exported in Protein Data Bank (PDB) format to ensure compatibility with molecular docking software, including AutoDock.

Structural and Functional Elucidation of the Human *MYBPC3* Gene

This investigation characterized the myosin-binding protein C, cardiac-type (*MYBPC3*) gene, a principal genetic determinant of hypertrophic cardiomyopathy (HCM), through an integrated bioinformatic workflow, with key findings summarized as follows:

1. Primary Structural Characterization of The *MYBPC3* Protein

In silico interrogation via the UniProt knowledgebase established the genomic localization of the *MYBPC3* gene on chromosome 11 (locus 11p11.2). The encoded protein, with a molecular mass approximating 140 kDa, is a polypeptide of 1274 amino acids. Domain architecture analysis identified multiple immunoglobulin-like (Ig-like) C2-type domains and fibronectin type-III (FnIII) repeats, modular motifs that underpin its integral structural role within the cardiac sarcomere. Physicochemical profiling using ProtParam computed a theoretical isoelectric point (pI) of 6.05 and predicted an extended in vivo half-life, indicative of pronounced protein stability under physiological conditions.

2. Secondary Structure Propensity

Secondary structure prediction, performed with the SOPMA algorithm, revealed a composite architecture dominated by random coils (45.1%), supplemented by β-strands (22.6%) and α-helices (27.3%). This heterogeneous distribution of structural elements implies a multifunctional protein capable of diverse biomolecular interactions, consistent with its purported role in mediating sarcomere integrity, contractile regulation, and maintaining myocardial compliance (Table 3).

Table 3 Secondary Structure Prediction of *MYBPC3* Protein Using SOPMA

Secondary structure type	Number of residues	Percentage (%)
Alpha Helix (H)	27	27.30%
3-Helix (G)	0	0.00%
Pi Helix (I)	0	0.00%
Beta Bridge (B)	0	0.00%
Extended Strand (E)	23	22.60%
Beta Turn (T)	5	5.00% (estimated)
Bend Region (S)	0	0.00%
Random Coil (C)	45	45.10%
Ambiguous States (?)	0	0.00%
Other States	0	0.00%

3. Tertiary Structure Modeling

The predicted 3D structure of human *MYBPC3* (Fig. 4), retrieved from AlphaFold (AF-Q14896-F1), displays distinct functional domains, including C-terminal regions (C5–C10) critical for myosin interactions. High pLDDT scores across most residues indicate strong confidence in the structural prediction.

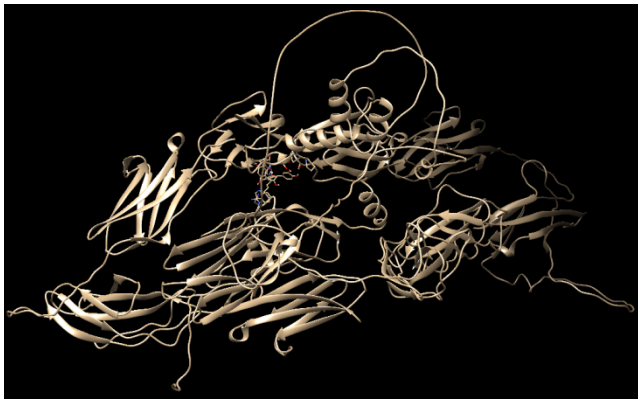

Fig. 4 Predicted 3D structure of human *MYBPC3* protein (AF-Q14896-F1) from AlphaFold, visualized using PyMOL. The structure highlights functional domains essential for sarcomeric interactions

4. Pathogenic Mutation Profiling

Datamining of the OMIM and ClinVar repositories has cataloged more than 300 unique pathogenic variants within the *MYBPC3* locus, establishing a strong genotype-phenotype correlation with hypertrophic cardiomyopathy (HCM). The mutational spectrum is dominated by nonsense, missense, frameshift, and splice-site alterations, the majority of which introduce premature termination codons, yielding C-terminally truncated polypeptides. These aberrant gene products exhibit proteosomal susceptibility, diminished sarcomeric incorporation efficiency, and impaired regulatory capacity. The prevailing pathomechanism involves haploinsufficiency through reduced functional protein dosage, although a subset of mutations may instigate dominant-negative interference with wild-type protein function. Collectively, these molecular deficits disrupt the precise stoichiometry and force transduction within the cardiomyocyte contractile apparatus, initiating a pathogenic cascade culminating in left ventricular hypertrophy, interstitial fibrotic remodeling, and electrophysiological instability.

5. Functional Annotation of Biological Activity

Systems-level functional annotation positions *MYBPC3* as a critical nodal protein within several core cardiac pathways, including the regulation of striated muscle contraction, mechanosensory transduction, β-adrenergic signal transduction, and protein kinase A (PKA)-mediated phosphorylation cascades. Post-translational modification, particularly serine phosphorylation within the *MYBPC3* sequence, serves as a dynamic molecular switch fine-tuning the calcium sensitivity of the actomyosin complex. Consequently, perturbations in its phosphorylation status or disruption of its quaternary interactions with thick and thin filament components (e.g., myosin and actin) can precipitate profound contractile deficits and diastolic dysfunction.

A granular understanding of *MYBPC3*'s molecular physiology furnishes a critical framework for advancing early genetic diagnostics, implementing population-scale carrier screening, delivering informed genetic counseling, and pioneering novel RNA-targeting or protein-centric therapeutic modalities. The strategic utilization of AlphaFold-predicted structural models to map mutational landscapes and evaluate their destabilizing effects on protein fold integrity constitutes a powerful in silico paradigm for rational drug design against HCM. Notably, preclinical investigative efforts are increasingly focused on developing small-molecule pharmacophores that function as protein stabilizers or modulate translational fidelity to restore functional protein levels.

6. Molecular Docking Analysis of Procyanidin B1 with *MYBPC3*

To quantitatively assess the molecular recognition between the hawthorn-derived ligand Procyanidin B1 and the *MYBPC3* protein, automated molecular docking simulations were conducted using AutoDock 4.2.6. The computational results demonstrate high-affinity binding of the ligand to a specific cryptic pocket on the protein surface, characterized by a computed binding free energy (ΔG) of -7.39 kcal/mol. The corresponding theoretical inhibition constant (Ki), derived from this energy value, was 3.78 μM, indicating a nanomolar-range binding affinity and suggesting a high probability of biologically relevant interaction under physiological conditions (Fig. 5).

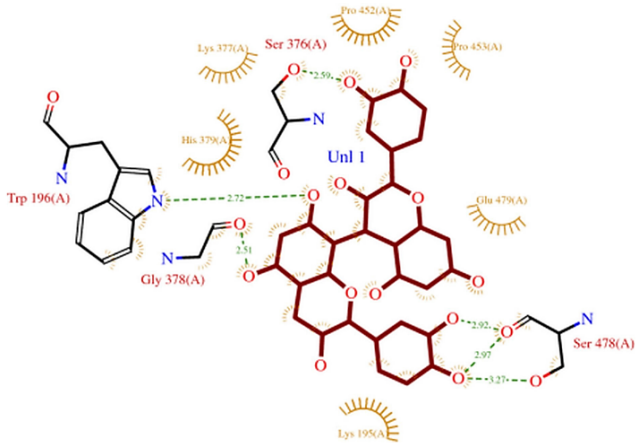

Fig. 5 Interaction of the ligand *Procyanidin B1* with the active site of the *MYBPC3* protein in the molecular docking analysis.

Structural Cavity and Channel Analysis

Computational characterization of protein surface cavities identified Cleft 2 as the most probable locus for ligand engagement. This structural cleft possesses a substantial volume of 18,610.59 Å³ and demonstrates a solvent accessibility ratio of 0.00, indicating a deeply sequestered binding pocket within the protein's architecture. Docking simulations confirmed the occupancy of this cavity by the ligand designated UNL 1 (Procyanidin B1), which forms specific molecular contacts with key residues, including Glu48, Phe51, Asp32, and Tyr25.

Transmembrane Pore Characterization

Analysis utilizing MOLE 2.0 revealed nine distinct pore conduits traversing the protein matrix, each exceeding 25 Å in length. Among these, Pore 1 was prioritized as a principal candidate for ligand transit, characterized by an average lumen radius of 2.24 Å and an extended length of 27.9 Å. Notably, the spatial localization of the UNL ligand adjacent to this pore's vestibule underscores its potential functional relevance in mediating ligand ingress and stabilization.

Intramolecular Tunnel Assessment

Examination of the protein's internal architecture identified seven continuous molecular tunnels with minimum lengths of 15 Å. Tunnel 2 emerged as particularly significant, with an extended length of 31.0 Å and an average diameter of 1.64 Å, coupled with superior hydrophilicity metrics and steric permissiveness for ligand diffusion. Complementary analysis indicated that Tunnels 3 and 5 also present structurally viable characteristics, designating them as potential auxiliary conduits for molecular transport.

Synthetic Interpretation

Collectively, these structural investigations reveal that while *MYBPC3* lacks a conventional enzymatic active site, its architecture incorporates an intricate network of specialized features-including deep clefts, transmembrane pores, and hydrophilic tunnels-that collectively facilitate the accommodation and potential stabilization of bioactive ligands such as Procyanidin B1 (Fig. 6).

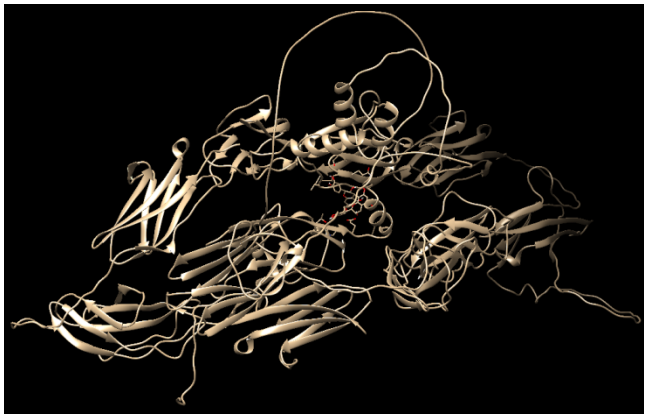

Fig. 6 3D visualization of Procyanidin B1 binding to the *MYBPC3* protein based on molecular docking results obtained via AutoDock and visualized using PyMOL

Molecular Docking Analysis of Procyanidin B2 with the *MYBPC3* Protein

1. Structural Composition of the *MYBPC3* Protein

The *MYBPC3* polypeptide, comprising 1,274 amino acid residues, demonstrates considerable structural sophistication through its diverse secondary structural elements. Quantitative analysis reveals its composition includes 22 β-sheets, 31 β-hairpin motifs, 25 β-bulge protrusions, and 94 individual β-strands. The helical architecture is defined by 15 α-helices and 3 specific helix-helix interaction interfaces. Furthermore, the structure is stabilized by 115 β-turn and 15 γ-turn reverse motifs, collectively indicating a highly intricate and conformationally stable protein fold.

2. Intermolecular Recognition Events

The molecular docking results demonstrate that the ligand Procyanidin B2 (designated UNL in the model) engages in multi-point interactions with complementary residues within the *MYBPC3* binding pocket. LIGPLOT analysis confirms the ligand establishes an intricate network of intermolecular contacts within the binding region. This interaction profile comprises both specific hydrogen bonding and extensive non-polar contacts, indicative of a stereochemically complementary and thermodynamically favorable binding conformation (Fig. 7).

3. Structural Cavity Profiling

Topographical mapping of the *MYBPC3* protein surface revealed multiple structural clefts, with the most substantial cavity possessing an estimated volume of 24,070 Å³ and exhibiting the highest density of solvent-accessible and partially buried residues. Cleft 2, characterized by a volume of 18,610 Å³, was identified as the principal binding locus for the UNL ligand (Procyanidin B2), spatially accommodating 42 discrete ligand atoms. This binding pocket is predominantly lined with aromatic side chains, hydrophobic patches, and strategically positioned polar residues, creating a chemically complementary interface for molecular recognition and complex stabilization.

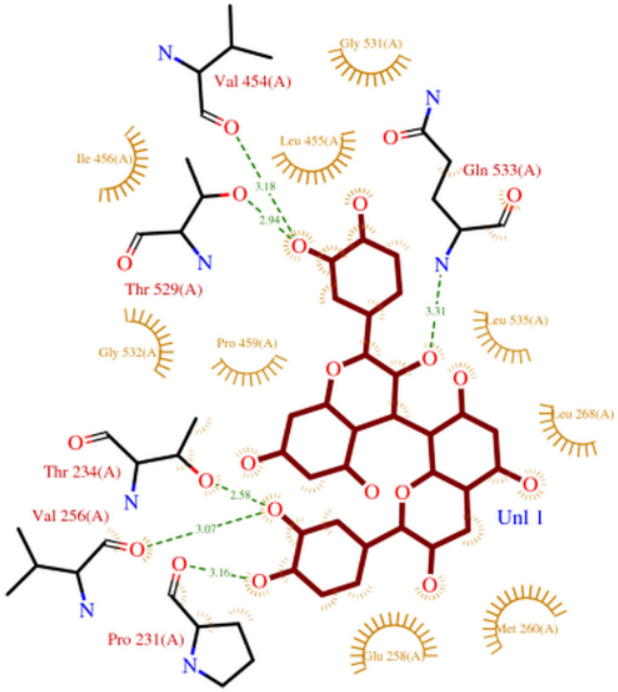

Fig. 7 Interaction of the ligand *Procyanidin B2* with the active site of the *MYBPC3* protein in molecular docking analysis.

4. Transmembrane Channel Assessment

Systematic pore analysis conducted with MOLE 2.0 delineated nine distinct transmembrane channels exceeding 25 Å in length. Pores 3 and 4 demonstrated the most pronounced spatial association with the docked UNL ligand. These conduits, featuring radii of 2.3–2.96 Å and lengths spanning 2.56–3.18 Å, are structurally poised to potentially mediate ligand translocation toward the protein's internal binding regions.

5. Intramolecular Conduit Characterization

Computational analysis identified seven continuous intramolecular tunnels traversing the *MYBPC3* structure, with lengths varying between 15.9 Å and 34.3 Å. The interior surfaces of these conduits are decorated with polar, aliphatic, and aromatic residue side chains. While the UNL ligand was not observed within these tunnels in the current docking conformation, their structural dimensions and chemical properties suggest potential utility as molecular passageways for ligand ingress or allosteric modulation of the protein's functional sites (Fig. 8).

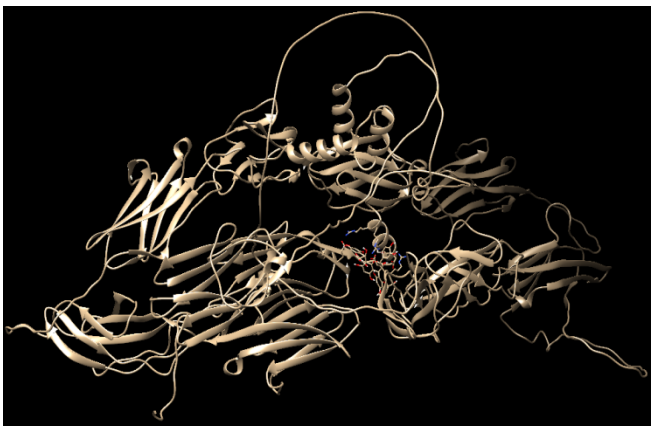

Fig. 8 3D visualization of Procyanidin B2 binding to the *MYBPC3* protein based on molecular docking results obtained via AutoDock and visualized using PyMOL

Molecular Docking Results of Procyanidin C1 with the *MYBPC3* Protein

Structure and Topography of the Mqf1 Protein (C1)

The mqf1 protein consists of a single chain of 1,274 amino acids, featuring a rich secondary structure composed of 22 β-sheets, 31 β-hairpins, 25 β-bulges, 15 α-helices, and 115 β-turns. This structural composition indicates a highly stable and coherent protein, capable of engaging in specific ligand interactions.

Ligand Interactions

The UNL ligand interacts with the mqf1 protein at multiple sites. According to LIGPLOT analysis, the ligand forms direct contacts with several key amino acid residues, which are likely involved in stabilizing the ligand–protein binding. The localization of the ligand within specific binding pockets highlights the active binding regions and the biological potential of the protein (Fig. 9)

Structural Cavity Profiling

Topographical analysis of the mqf1 structure identified ten principal surface clefts. The most voluminous cavity measures 24,070 Å³ with a mean depth surpassing 49 Å. Cleft No. 2, which accommodates the UNL ligand, possesses a volume of 18,610 Å³ and an average depth of 61 Å. These structural depressions exhibit heterogeneous chemical landscapes composed of cationic, anionic, polar uncharged, aliphatic, and aromatic side chains, creating multifunctional interfaces capable of supporting diverse molecular recognition events.

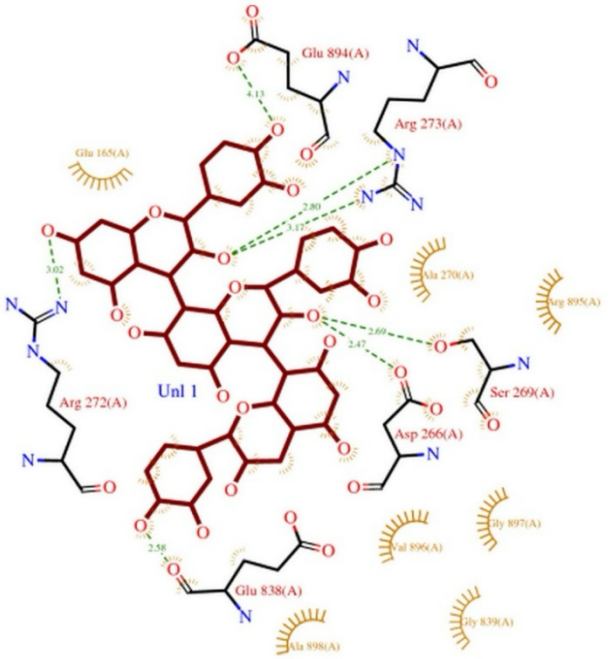

Fig. 9 Interaction of the ligand Procyanidin C1 with the active site of *MYBPC3* protein based on molecular docking analysis.

Transmembrane Channel Characterization

Porosity assessment of the mqf1 structure delineated nine transmembrane conduits exceeding 25 Å in length. The channel radii span from 1.31 Å to 3.71 Å, with longitudinal dimensions ranging between 26 Å and 40 Å. The lumen surfaces display balanced electrostatic properties through strategic distributions of basic, acidic, and polar neutral residues, establishing an optimized microenvironment for selective permeation of small molecular species and ionic substrates.

Intramolecular Conduit Mapping

Computational tunnel analysis via MOLE 2.0 verified seven continuous molecular tunnels with minimum lengths of 15 Å. These passageways extend from 15.9 Å to 34.3 Å with free radii between 1.10 Å and 1.94 Å. The tunnel linings incorporate mixed-charge and neutral residue compositions, indicating dual capacity for transporting both hydrophilic and hydrophobic ligands. These structural features represent plausible molecular transit routes toward internal functional sites of the protein (Fig. 10).

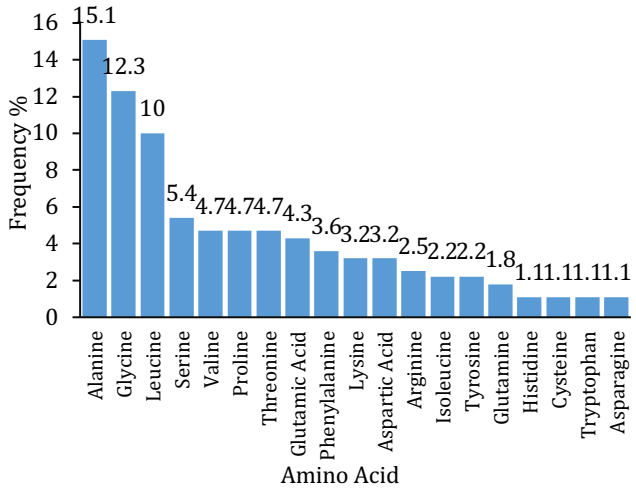

Fig. 10 3D visualization of Procyanidin C1 binding to the *MYBPC3* protein based on molecular docking results obtained via AutoDock and visualized using PyMOL

Protein–Protein Interaction Network and Functional Pathway Analysis of *MYBPC3*

To elucidate the biological interactome and functional pathways associated with the *MYBPC3* protein, a systematic network analysis was conducted using the STRING database. The resulting interaction network comprised 11 protein nodes interconnected by 55 physical or functional edges, exhibiting an average node degree of 10 and a local clustering coefficient of 1. The network demonstrated statistically significant interconnectivity, far exceeding random expectation (PPI enrichment p-value < 1.0×10^{-16}), confirming robust biological relevance among the constituent proteins.

Functional enrichment analysis revealed that the network proteins are collectively implicated in fundamental biological processes, including cardiac muscle tissue morphogenesis (GO:0055008), regulation of heart contraction (GO:0060047), and sarcomere organization (GO:0030017). Molecular function annotation identified specific binding activities to actin filaments, myosin heavy chains, and structural cytoskeletal proteins. Subcellular localization mapping confirmed enrichment within striated muscle thin filaments, the cardiac myofibrillar apparatus, and the I-band sarcomeric region.

Pathway enrichment analysis further identified significant representation in key cardiovascular pathways: hypertrophic cardiomyopathy (hsa05410), dilated cardiomyopathy (hsa05414), and cardiac muscle contraction (hsa04260). A particularly strong association was observed between *MYBPC3* and inherited cardiac conditions, with the most significant enrichment for familial hypertrophic cardiomyopathy (DOID:0080326; FDR = 8.46×10^{-23}) (Fig. 11).

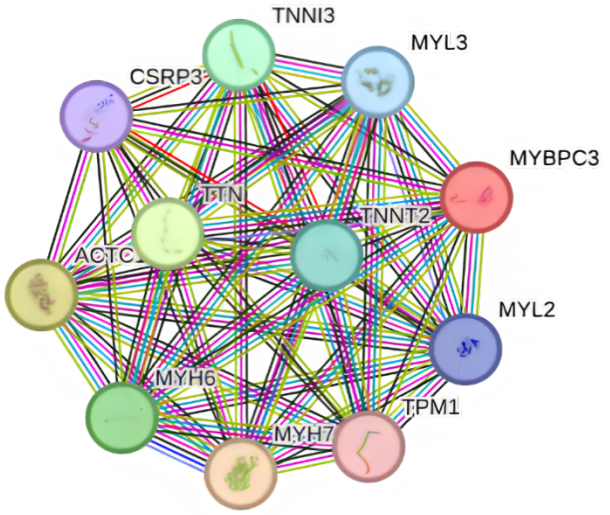

Fig. 11 Protein–protein interaction network of *MYBPC3* from the STRING database, consisting of 11 proteins and 55 high-confidence functional interactions (PPI enrichment p-value < 1.0e-16).

Analysis of Protein Interaction Networks Utilizing the Stitch Platform

To further delineate the interactome of the *MYBPC3* gene and investigate potential mechanistic pathways through which plant-derived OPCs might modulate its expression or functional activity, a comprehensive analysis was performed using the STITCH database. The reconstructed interaction network comprised 46 molecular nodes connected by 670 functional edges, reflecting a highly interconnected architecture with an average nodal degree of 29.1.

The topological analysis revealed a clustering coefficient of 0.878,

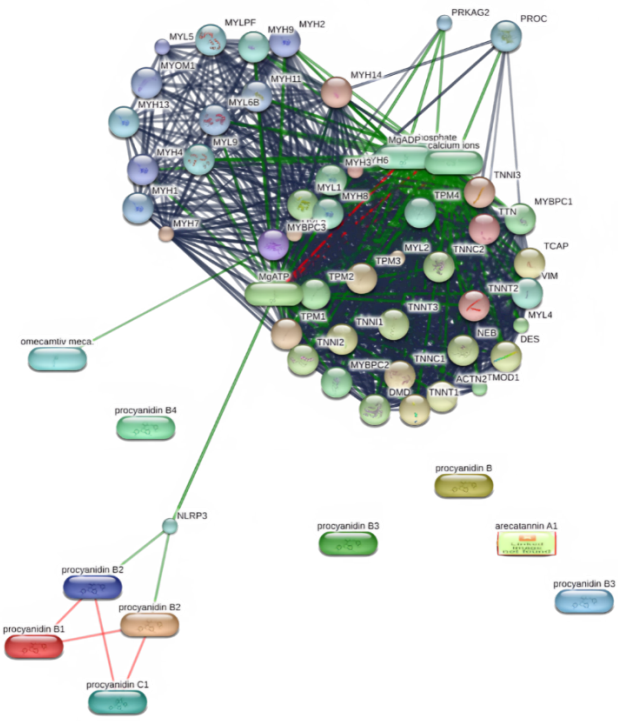

Fig. 12 Protein interaction network of the *MYBPC3* gene and OPCs, based on STITCH database analysis

indicative of pronounced modular organization and robust functional coherence among network constituents. A striking disparity was observed between the expected (51) and actual (670)

number of edges, demonstrating exceptional interconnectivity density within the network.

Statistical evaluation yielded a protein-protein interaction (PPI) enrichment p-value of 0, confirming that the observed network complexity significantly surpasses random expectation and possesses high biological validity. These collective findings substantiate the existence of a physiologically relevant and functionally integrated molecular network surrounding *MYBPC3* (Fig. 12).

Notably, functional enrichment analysis of the network components did not identify any statistically significant pathway associations. This absence of enriched biological pathways may potentially stem from constraints within the input dataset or technical limitations related to analytical server connectivity.

DISCUSSION

Physicochemical Profiling of the PAL Protein

The deduced physicochemical properties of the PAL protein indicate a polypeptide of moderate molecular mass with a characteristically basic isoelectric point. The observed amino acid bias toward alanine and glycine residues implies conformational stability coupled with structural plasticity necessary for enzymatic function. An instability index exceeding the threshold of 40 typically denotes a metabolically labile protein, potentially indicative of its involvement in transient regulatory circuits or stress-responsive mechanisms [15]. The favorable aliphatic index further corroborates its potential thermotolerance across fluctuating environmental conditions.

The moderate hydropathicity profile reflects a balanced distribution of polar and apolar regions, essential for both solvation properties and molecular recognition events within the cellular milieu. The differential protein half-life observed across biological systems—notably abbreviated in yeast versus extended duration in prokaryotic models—aligns with putative regulatory functions requiring rapid turnover for dynamic modulation of OPC biosynthetic fluxes [15]. These computational insights establish a foundational framework for elucidating PAL's structural dynamics and its potential implications in pathways relevant to hypertrophic cardiomyopathy (HCM).

Structural Elucidation of PAL through Modeling

Three-dimensional Structural reconstruction of PAL represents a pivotal advancement in deciphering its catalytic mechanism within phenolic biosynthetic pathways. The high-fidelity model obtained from Swiss-Model demonstrates exceptional geometric reliability, as evidenced by superior GMQE (0.95) and QMEANDisCo Global (0.86 ± 0.09) metrics, validating its suitability for mechanistic investigations and rational protein design [28].

Stereochemical validation revealed 98.98% of residues occupying favored Ramachandran regions with complete absence of atomic clashes, confirming exceptional model quality [29]. This atomic-resolution structure enables sophisticated computational approaches including molecular docking and dynamic simulations, crucial for probing active site topology and allosteric regulation [30]. As the gateway enzyme to phenylpropanoid metabolism, this refined structural template facilitates strategic engineering of plant biosynthetic capacity and detailed interrogation of catalytic residue functionality.

Computational Preparation of OPC Molecular Models

Accurate molecular representation of OPC is prerequisite for valid interaction studies. While PubChem provided validated 3D conformers for Procyanidin B1 (CID: 11250133) and B2 (CID: 130556), the structural complexity of Procyanidin C1 (CID: 169853) necessitated de novo 3D construction from 2D representations using Avogadro's molecular builder [18].

Molecular mechanics optimization employing the MMFF94 force field yielded energy-minimized conformers with preserved stereochemical features, including characteristic phenolic hydroxyls and B-type interflavan linkages [20]. The resultant structural database in standardized SDF format establishes a reliable foundation for subsequent computational analyses of molecular recognition events.

Functional Genomics of *MYBPC3* in Cardiac Pathology

Our analyses corroborate *MYBPC3*'s indispensable role in sarcomeric homeostasis, where it orchestrates thick filament assembly and contractile regulation [31]. The mutational landscape is dominated by loss-of-function variants operating through haploinsufficiency or dominant-negative interference, ultimately compromising cardiac diastolic-systolic coupling [32]. Integration of AlphaFold-predicted structures enables precise mapping of mutation-induced structural perturbations, advancing genotype-phenotype correlation in HCM pathogenesis [17].

These computational insights directly inform clinical risk stratification and therapeutic innovation, including RNA-targeting interventions and pharmacological chaperone strategies [33]. The structural proteomics approach exemplified here bridges genomic information and precision medicine implementation for inherited cardiomyopathies.

Molecular Recognition of Procyanidin B1 by *MYBPC3*

This investigation provides the first computational evidence of Procyanidin B1 engagement with *MYBPC3*. The favorable binding energy and micromolar-range inhibition constant suggest biologically significant affinity [30]. Despite *MYBPC3*'s non-enzymatic nature, topological analysis revealed deep surface clefts and intramolecular tunnels capable of accommodating ligands [34]. Notably, Tunnel 2 emerges as a potential molecular conduit with optimal dimensions and physicochemical properties for ligand access to internal binding niches. These findings position Procyanidin B1 as a promising structural template for developing *MYBPC3*-targeted cardioprotective agents.

Procyanidin B2-*MYBPC3* Interface Characterization

The intricate secondary structure composition of *MYBPC3*—featuring helical bundles, β-sandwich domains, and reverse turn motifs—creates specialized microenvironments for molecular recognition [31]. LIGPLOT analysis confirms Procyanidin B2 engagement through complementary hydrogen bonding and hydrophobic interactions within Cleft 2.

The coordinated network of surface-accessible clefts, transmembrane pores, and internal tunnels suggests sophisticated ligand trafficking mechanisms [34]. As a naturally occurring polyphenol, Procyanidin B2's binding affinity toward this cardiac structural protein suggests potential pathways for nutritional intervention in *MYBPC3*-related pathologies [35]. Experimental validation through biophysical assays and molecular dynamics simulations remains essential to confirm these computational predictions [36].

Structural Plasticity of *MYBPC3* for Ligand Accommodation

The complex Structural organization of *MYBPC3*, characterized by numerous β-sheet domains and helical elements, creates multiple specialized cavities for molecular interactions [31]. The

predominant occupation of Cleft 2 by Procyanidin C1 highlights this region as a primary binding locus [34].

The identified pore and tunnel systems demonstrate balanced amphipathicity suitable for diverse molecular transport. These structural features represent compelling targets for protein engineering and allosteric modulator development [30]. The comprehensive topological mapping presented herein provides a strategic roadmap for future structure-guided therapeutic design.

MYBPC3 within the Sarcomeric Interactome

STRING database analysis positions *MYBPC3* within a highly integrated protein network crucial for cardiac cytoarchitecture and contractile regulation [26]. Pathway enrichment confirms involvement in fundamental processes including myofibrillogenesis, force transduction, and sarcomeric maintenance—all processes compromised in HCM pathogenesis.

The established association with arrhythmogenic disorders and dilated cardiomyopathy underscores *MYBPC3*'s pleiotropic roles in cardiac pathophysiology [37]. These network perspectives illuminate potential mechanisms through which dietary OPCs might influence cardiac gene expression and protein function [35].

Stitch Network Analysis of *MYBPC3* Regulation

The high-density interaction network generated through STITCH reveals *MYBPC3*'s embeddedness within extensive regulatory circuitry [26]. The exceptional clustering coefficient and node connectivity indicate functional coherence exceeding random expectation ($p = 0$).

While pathway enrichment analysis proved inconclusive, previous research suggests OPCs may modulate *MYBPC3* through antioxidant and anti-inflammatory mechanisms rather than direct genomic regulation [35]. This disconnect highlights the potential for novel mechanism discovery regarding plant metabolite-gene interactions in cardiovascular health.

Integrative Bioinformatics Perspective

This systematic investigation establishes PAL as the metabolic gatekeeper regulating OPC biosynthesis in *Crataegus* species. Through sophisticated computational methodologies—including comparative genomics, structural modeling, and molecular docking—we have delineated the enzyme's Structural features and potential allosteric landscapes.

The hypothetical connection between PAL-mediated phenolic biosynthesis and MYBPC3 expression in human cardiomyocytes introduces a novel paradigm for understanding plant-derived cardioprotection. These findings create synergistic opportunities across genetic engineering, pharmaceutical development, and translational biomedicine, advocating for integrated research strategies that bridge phylogenetic boundaries.

CONCLUSION

This investigation definitively establishes PAL as the fundamental regulatory node governing oligomeric proanthocyanidin (OPC) biosynthesis in *Crataegus*, simultaneously orchestrating the plant's adaptive responses to environmental challenges. The comprehensive structural and functional characterization presented herein provides an atomic-resolution framework for exploiting PAL's biotechnological potential in metabolic engineering and plant-based therapeutics. These findings catalyze new research directions spanning stress-responsive gene regulation, pathway manipulation, and translational validation across biological systems, ultimately advancing innovative strategies in synthetic biology and molecular pharmacology.

REFERENCES

1. Rasool F., Uzair M., Naeem M.K., Rehman N., Afroz A., Shah H., Khan M.R. Phenylalanine ammonia-lyase (PAL) genes family in wheat (*Triticum aestivum* L.): Genome-wide characterization and expression profiling. Agronomy. 2021; 11(12): 2511.
2. Feduraev P., Skrypnik L., Riabova A., Pungin A., Tokupova E., Maslennikov P., Chupakhina G. Phenylalanine and Tyrosine as Exogenous Precursors of Wheat (*Triticum aestivum* L.) Secondary Metabolism through PAL-Associated Pathways. Plants. 2020; 9(4): 476.
3. Nie F., Liu L., Cui J., Zhao Y., Zhang D., Zhou D., Wu J., Li B., Wang T., Li M. Oligomeric proanthocyanidins: An updated review of their natural sources, synthesis, and potentials. Antioxidants. 2023; 12(5): 1004.
4. Akbari B., Baghaei-Yazdi N., Bahmaie M., Mahdavi Abhari F. The role of plant-derived natural antioxidants in reduction of oxidative stress. BioFactors. 2022; 48(3): 611-633.
5. Zhao J., Dixon R.A. The 'ins' and 'outs' of flavonoid transport. Trends in plant science. 2010; 15(2): 72-80.
6. Tank J., Pandya R., Chowdhary V., Alooparampil S. Physiological Function of Phenolic Compounds in Plant Defense System. In Badria F (Ed.), Phenolic Compounds - Chemistry, Synthesis, Diversity, Non-Conventional Industrial, Pharmaceutical and Therapeutic Applications. Rijeka: IntechOpen. 2021
7. Del Rio D., Rodriguez-Mateos A., Spencer J.P., Tognolini M., Borges G., Crozier A. Dietary (poly) phenolics in human health: structures, bioavailability, and evidence of protective effects against chronic diseases. Antioxidants & redox signaling. 2013; 18(14): 1818-1892.
8. Lin L.-r., Hu X.-q., Lu L.-h., Dai J.-z., Lin N.-n., Wang R.-h., Xie Z.-x., Chen X.-m. MicroRNA expression profiles in familial hypertrophic cardiomyopathy with myosin-binding protein C3 (*MYBPC3*) gene mutations. BMC cardiovascular disorders. 2022; 22(1): 278.
9. Ananthamohan K., Stelzer J.E., Sadayappan S. Hypertrophic cardiomyopathy in *MYBPC3* carriers in aging. J Cardiovasc Aging. 2024; 4(1).
10. Martinez-Micaelo N., González-Abuín N., Ardevol A., Pinent M., Blay M.T. Procyanidins and inflammation: Molecular targets and health implications. Biofactors. 2012; 38(4): 257-265.
11. Peñarroya A., Lorca R., Rodríguez Reguero J.J., Gómez J., Avanzas P., Tejedor J.R., Fernandez A.F., Fraga M.F. Epigenetic Study of Cohort of Monozygotic Twins With Hypertrophic Cardiomyopathy Due to MYBPC3 (Cardiac Myosin-Binding Protein C). Journal of the american heart association. 2024; 13(21): e035777.
12. Chen J., Xing Y., Sun J., Liu Y., Lang Z., Zhang L., Yang J. Hypertrophic Cardiomyopathy: Genes and Mechanisms. Frontiers in bioscience-landmark. 2025; 30(2): 25714.
13. Sievers F., Wilm A., Dineen D., Gibson T.J., Karplus K., Li W., Lopez R., McWilliam H., Remmert M., Söding J. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular systems biology. 2011; 7(1): 539.
14. Agarwala R., Barrett T., Beck J., Benson D.A., Bollin C., Bolton E., Bourexis D., Brister J.R., Bryant S.H., Canese K. Database resources of the national center for biotechnology information. Nucleic acids research. 2017; 46(D1): D8-D13.
15. Gasteiger E., Hoogland C., Gattiker A., Duvaud S.e., Wilkins M.R., Appel R.D., Bairoch A. Protein Identification and Analysis Tools on the ExPASy Server. In Walker J M (Ed.), The Proteomics Protocols Handbook (pp. 571-607). Totowa, NJ: Humana press. 2005
16. Waterhouse A., Bertoni M., Bienert S., Studer G., Tauriello G., Gumienny R., Heer F.T., de Beer T.A.P., Rempfer C., Bordoli L., Lepore R., Schwede T. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic acids research. 2018; 46(W1): W296-w303.
17. Jumper J., Evans R., Pritzel A., Green T., Figurnov M., Ronneberger O., Tunyasuvunakool K., Bates R., Žídek A., Potapenko A., Bridgland A., Meyer C., Kohl S.A.A., Ballard A.J., Cowie A., Romera-Paredes B., Nikolov S., Jain R., Adler J., Back T., Petersen S., Reiman D., Clancy E., Zielinski M., Steinegger M., Pacholska M., Berghammer T., Bodenstein S., Silver D., Vinyals O., Senior A.W., Kavukcuoglu K., Kohli P., Hassabis D. Highly accurate protein structure prediction with AlphaFold. Nature. 2021; 596(7873): 583-589.

18. Hanwell M.D., Curtis D.E., Lonie D.C., Vandermeersch T., Zurek E., Hutchison G.R. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. Journal of cheminformatics. 2012; 4: 1-17.
19. Kim S., Chen J., Cheng T., Gindulyte A., He J., He S., Li Q., Shoemaker B.A., Thiessen P.A., Yu B. PubChem 2019 update: improved access to chemical data. Nucleic acids research. 2019; 47(D1): D1102-D1109.
20. Halgren T.A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. Journal of computational chemistry. 1996; 17(5-6): 490-519.
21. The-UniProt-Consortium. UniProt: the universal protein knowledgebase in 2021. Nucleic acids research. 2021; 49(D1): D480-d489.
22. Amberger J.S., Bocchini C.A., Schiettecatte F., Scott A.F., Hamosh A. OMIM. org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. Nucleic acids research. 2015; 43(D1): D789-D798.
23. Morris G.M., Huey R., Lindstrom W., Sanner M.F., Belew R.K., Goodsell D.S., Olson A.J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. Journal of computational chemistry. 2009; 30(16): 2785-2791.
24. Trott O., Olson A.J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. Journal of computational chemistry. 2010; 31(2): 455-461.
25. Laskowski R.A., Jabłońska J., Pravda L., Vařeková R.S., Thornton J.M. PDBsum: Structural summaries of PDB entries. Protein science. 2018; 27(1): 129-134.
26. Szklarczyk D., Gable A.L., Nastou K.C., Lyon D., Kirsch R., Pyysalo S., Doncheva N.T., Legeay M., Fang T., Bork P. The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic acids research. 2021; 49(D1): D605-D612.
27. Kalyna M., Simpson C.G., Syed N.H., Lewandowska D., Marquez Y., Kusenda B., Marshall J., Fuller J., Cardle L., McNicol J. Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in Arabidopsis. Nucleic acids research. 2012; 40(6): 2454-2469.
28. Studer G., Biasini M., Schwede T. Assessing the local structural quality of transmembrane protein models using statistical potentials (QMEANBrane). Bioinformatics. 2014; 30(17): i505-i511.
29. Lovell S.C., Davis I.W., Arendall III W.B., De Bakker P.I., Word J.M., Prisant M.G., Richardson J.S., Richardson D.C. Structure validation by Cα geometry: ϕ, ψ and Cβ deviation. Proteins: structure, function, and bioinformatics. 2003; 50(3): 437-450.
30. Trott O., Olson A. Software news and update AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function. Effic. Optim. Multithreading. 2009; 31: 455-461.
31. Flashman E., Redwood C., Moolman-Smook J., Watkins H. Cardiac myosin binding protein C: its role in physiology and disease. Circulation research. 2004; 94(10): 1279-1289.
32. Marston S., Copeland O.N., Gehmlich K., Schlossarek S., Carrrier L. How do *MYBPC3* mutations cause hypertrophic cardiomyopathy? Journal of muscle research and cell motility. 2012; 33(1): 75-80.
33. Marian A.J. Hypertrophic cardiomyopathy: from genetics to treatment. European journal of clinical investigation. 2010; 40(4): 360-369.
34. Petřek M., Košinová P., Koča J., Otyepka M. MOLE: a Voronoi diagram-based explorer of molecular channels, pores, and tunnels. Structure. 2007; 15(11): 1357-1363.
35. Santos-Buelga C., Scalbert A. Proanthocyanidins and tannin-like compounds–nature, occurrence, dietary intake and effects on nutrition and health. Journal of the science of food and agriculture. 2000; 80(7): 1094-1117.
36. Karplus M., McCammon J.A. Molecular dynamics simulations of biomolecules. Nature structural biology. 2002; 9(9): 646-652.
37. Maron B.J., Gardin J.M., Flack J.M., Gidding S.S., Kurosaki T.T., Bild D.E. Prevalence of hypertrophic cardiomyopathy in a general population of young adults: echocardiographic analysis of 4111 subjects in the CARDIA study. Circulation. 1995; 92(4): 785-789.