Black Sea Journal of Engineering and Science
doi: 10.34248/bsengineering.1779567

Open Access Journal
e-ISSN: 2619 – 8991

Research Article
Volume 9 - Issue 2: 742-758 / March 2026

A SYSTEMATIC *IN-SILICO* SCREENING OF BACTERIAL ALPHA GALACTOSIDASES: INTEGRATING STRUCTURAL, FUNCTIONAL, AND IMMUNOINFORMATIC ANALYSES TO IDENTIFY POTENTIAL THERAPEUTIC CANDIDATES

Ozan KILIÇKAYA[1]*, Yunus ENSARİ[2]

[1]*Afyonkarahisar Health Sciences University, Faculty of Pharmacy, Department of Pharmaceutical Biotechnology, 03030, Afyonkarahisar, Türkiye*
[2]*Kafkas University, Faculty of Engineering and Architecture, Department of Bioengineering, 36100, Kars, Türkiye*

Abstract: This study aims to identify novel enzyme candidates from bacterial sources to enhance the therapeutic potential of human alpha-galactosidase A (α-Gal). The limitations of current enzyme replacement therapies for Fabry disease, such as immunogenicity, necessitate the search for alternative homologs with superior properties. In this context, a BLASTp search using human α-Gal as a reference identified 100 potential bacterial homologs. The three-dimensional structural models of these homologs were subjected to a rigorous quality control process using the SAVES server, and candidates with inadequate structural integrity were eliminated. The immunogenic potential of the selected candidates was assessed by predicting B-cell epitopes via the ElliProt server. For functional analysis, molecular docking simulations were performed with the natural substrate, globotriaosylceramide (Gb3), and the artificial substrate, p-nitrophenyl α-galactopyranoside (*p*NP-Gal). The results highlighted proteins such as A0A1M5FVV3 and Q5LFG6, which showed the highest binding affinity for the Gb3 substrate, and proteins like R6DB23 and R5RG66, which exhibited the highest affinity for the *p*NP-Gal substrate. Furthermore, the interactions of conserved Aspartate residues, which play a key role in substrate binding and are critical for catalytic activity, were confirmed. This study identifies specific bacterial α-Gal homologs that combine high substrate affinity with low immunogenicity potential as promising candidates for further experimental validation as next-generation biotechnological and novel bacterial homologs for Fabry disease.

Keywords: Alpha-galactosidase, Fabry disease, Bioinformatics, Molecular docking, Immunogenicity, Enzyme replacement therapy

***Corresponding author:** Afyonkarahisar Health Sciences University, Faculty of Pharmacy, Department of Pharmaceutical Biotechnology, 03030, Afyonkarahisar, Türkiye
E mail: ozan.kilickaya@afsu.edu.tr (O. KILIÇKAYA)
Ozan KILIÇKAYA https://orcid.org/0000-0003-2710-2149
Yunus ENSARİ https://orcid.org/0000-0002-4757-4197

Received: September 07, 2025
Accepted: February 16, 2026
Published: March 15, 2026

Cite as: Kılıçkaya, O., Ensari, Y. (2026). A systematic *in silico* screening of bacterial alpha galactosidases: integrating structural, functional, and immunoinformatic analyses to identify novel therapeutic candidates. *Black Sea Journal of Engineering and Science*, *9*(2), 742-758.

1. Introduction

Alpha-galactosidase (α-Gal) is formally classified as an enzyme belonging to the glycoside hydrolase (GH) group, bearing the systematic name α-D-galactoside galactohydrolase and the Enzyme Commission (EC) number EC 3.2.1.22. It functions as an exoglycosidase, characterized by its specific catalytic activity of hydrolyzing terminal, non-reducing α-linked D-galactose residues from a variety of glycoconjugate substrates. This enzymatic action is crucial in numerous catabolic pathways involving glycoproteins, glycolipids, and polysaccharides (Garman, 2007). Fabry disease is a rare lysosomal storage disorder caused by mutations in the *GLA* gene that lead to a deficiency in α-galactosidase A activity, resulting in the consequent accumulation of globotriaosylceramide (Gb3) and other neutral glycosphingolipids in body fluids and lysosomes throughout the body (Azevedo et al., 2021).

Based on amino acid sequence similarities and structural features, α-galactosidases are distributed across several distinct GH families within the Carbohydrate-Active enZymes (CAZy) database, including GH4, GH27, GH36, GH57, GH97, and GH110 (Lafond et al., 2020). Bacterial α-galactosidases are predominantly found in families GH27 and GH36, while the clinically significant human lysosomal enzyme, α-galactosidase A, belongs to GH27 family (Cervera-Tison et al., 2012; Lafond et al., 2020). The classification of α-galactosidase activity across these multiple GH families underscores significant structural and evolutionary divergence despite a conserved catalytic function. This suggests that nature has developed different protein architectures, such as the common (β/α)8-barrel fold found in GH27 and GH36 enzymes (Modrego et al., 2021) versus potentially different scaffolds in other families, to achieve the hydrolysis of terminal α-galactosyl linkages. Such convergent evolution highlights the biological importance of this enzymatic activity across diverse life

forms and provides a rich source of enzymes with potentially varied substrate specificities, stability profiles, or regulatory features suitable for distinct biological roles and biotechnological exploitation.

α-Galactosidase functions as an exoglycosidase that hydrolyzes terminal α-galactosyl linkages from various glycoconjugates (Garman, 2007). Its broad substrate spectrum includes plant oligosaccharides (e.g., raffinose), polysaccharides, and glycoproteins, but its most clinically relevant target is the glycosphingolipid globotriaosylceramide (Gb3) (Lafond et al., 2020). In laboratory settings, enzyme activity is typically quantified using synthetic chromogenic or fluorogenic substrates such as p-nitrophenyl-α-D-galactopyranoside (*p*NP-Gal) (Shi et al., 2009). This biochemical versatility supports critical biological functions, including human lysosomal catabolism and microbial nutrient acquisition (Garman, 2007). Consequently, the enzyme holds significant value in diverse biotechnological applications, ranging from industrial food processing to therapeutic interventions for Fabry disease (Anisha, 2023; Ju et al., 2019; Katrolia et al., 2014).

α-Galactosidase activity is ubiquitous in nature, occurring widely across microorganisms, plants, and animals (Katrolia et al., 2014). While the human lysosomal enzyme is central to the pathophysiology of Fabry disease (Azevedo et al., 2021), microbial sources are of paramount importance for biotechnology due to their high production yields, cost-effectiveness, and amenability to genetic manipulation (Katrolia et al., 2014). Bacteria, including gut commensals such as *Bacteroides* and *Ruminococcus*, represent a particularly diverse reservoir of these enzymes (Katrolia et al., 2014; Lafond et al., 2020). This extensive biodiversity facilitates the identification of homologs with distinct biochemical profiles, such as enhanced thermostability or specific pH optima (Ju et al., 2019; Wang et al., 2021). Consequently, while recombinant human α-Gal A remains the standard for therapy, exploring bacterial diversity provides a strategic pathway to discover novel enzymes with superior physicochemical properties for next-generation treatments (Azevedo et al., 2021; Borzova and Varbanets, 2024). However, the systemic administration of non-human proteins carries an inherent risk of immunogenicity, regardless of their origin. Therefore, relying solely on the commensal nature of these bacteria is insufficient for therapeutic safety. Instead, a rigorous selection process targeting 'structural silence'—minimizing B-cell epitopes while maximizing stability—is essential to identify candidates that can evade immune surveillance.

This study aims to identify novel enzyme candidates from bacterial sources to enhance the therapeutic potential of Fabry disease treatment. We systematically screened 100 bacterial α-Gal A homologs using an integrated *in silico* pipeline. Candidates were prioritized based on physicochemical stability, structural integrity, and reduced immunogenicity via B-cell epitope prediction. Subsequent molecular docking with Gb3 and *p*NP-Gal validated substrate binding affinities, establishing a shortlist of stable, high-potency bacterial enzymes for future experimental validation.

2. Materials and Methods

All analyses in this study were performed in-silico using publicly available databases and web-based bioinformatic tools.

2.1. Homolog Screening and Physicochemical Analysis

The amino acid sequence of human α-Gal (UniProt ID: P06280) was used as a reference to perform a homolog search against the bacterial database using the NCBI BLASTp tool (Altschul et al., 1997). The selection focused on sequences from four major bacterial taxa prevalent in the human gut microbiome: *Bacteroides*, *Prevotella*, *Clostridium*, and *Ruminococcus*. The top 100 homologs with significant E-values were selected for further analysis. The fundamental physicochemical properties of these proteins, such as isoelectric point (pI), instability index, and half-life, were calculated using the ExPASy ProtParam server (Gasteiger et al, 2005). The solubility potential of the proteins was predicted using the SoluProt tool (Hon et al., 2021).

2.2. Phylogenetic Analysis

To elucidate the evolutionary relationships among the identified bacterial α-galactosidase homologs and the human α-Gal protein, a phylogenetic analysis was conducted. The amino acid sequences of the selected proteins were first aligned using the MUSCLE (Multiple Sequence Comparison by Log-Expectation) algorithm integrated within the MEGA 12 software.

The phylogenetic tree was constructed using the Maximum Likelihood (ML) method. The best-fit substitution model for the protein sequence alignment was determined by MEGA 12's model selection tool based on the Bayesian Information Criterion (BIC). The Jones-Taylor-Thornton (JTT) model with a gamma distribution (G) was identified as the most appropriate model and was subsequently used for the ML analysis. To assess the statistical reliability and robustness of the tree's branching patterns, a bootstrap analysis was performed with 500 replicates. The resulting phylogenetic tree was visualized and edited for publication using the tree editor tools available in MEGA 12 (Kumar et al., 2024).

2.3. Structural Modeling and Quality Assessment

Three-dimensional (3D) structural models for the 31 selected α-galactosidase sequences were generated using homology modeling. The specific modeling server used was YASARA Structure (Land and Humble, 2018). After modeling, the quality and validity of each 3D model were assessed using the SAVES v6.0 server, which integrates several validation tools:

- QMEAN: Provides a composite quality score for the entire model (Benkert et al., 2009, 2008).

- ProSA-web: Checks for potential errors in the 3D model by analyzing its Z-score (Wiederstein and Sippl, 2007).
- ERRAT: Analyzes the statistics of non-bonded atom interactions.
- Verify3D: Determines the compatibility of an atomic model (3D) with its own amino acid sequence (1D).
- PROCHECK: Evaluates the stereochemical quality of the protein structure, including Ramachandran plot analysis to assess disallowed residue conformations (Laskowski et al., 1993).

Models with high-quality scores and minimal stereochemical errors were considered suitable for further analysis.

2.4. Immunobioinformatic Analysis

To evaluate the potential immunogenicity of the candidate enzymes, B-cell epitope prediction was performed. B-cell epitopes are regions on the protein surface that are recognized by antibodies. Proteins with fewer or less prominent epitopes are generally less likely to elicit a strong immune response. Two complementary web-based tools were used for this analysis. ElliPro predicts linear and discontinuous B-cell epitopes based on the 3D structure of the protein. It calculates a Protrusion Index (PI) score for residues, with higher scores indicating a greater likelihood of being part of an epitope (Ponomarenko et al., 2008). DiscoTope 3.0 predicts discontinuous B-cell epitopes by combining surface accessibility data with amino acid properties derived from the protein's 3D structure. It provides a score for each residue, with higher scores indicating a higher probability of being part of an epitope (Høi et al., 2024).

The results from both servers were combined to create a comprehensive immunogenicity profile for each of the 31 candidate enzymes.

2.5. Molecular Docking

Molecular docking simulations were performed using YASARA Structure to investigate the binding interactions between the candidate α-galactosidases and their substrates. Two different ligands were used:

- Globotriaosylceramide (Gb3): The natural substrate that accumulates in Fabry disease.
- p-nitrophenyl-α-D-galactopyranoside (*p*NP-Gal): A synthetic chromogenic substrate commonly used to assay α-galactosidase activity.

For each of the 31 enzyme models, the catalytic active site residues were identified based on conserved motifs in the Glycoside Hydrolase (GH) families 27 and 36. Docking simulations were run using both AutoDock Vina and AutoDock LGA algorithms within YASARA (Gaillard, 2018; Land and Humble, 2018; Trott and Olson, 2010). The results were analyzed based on the binding energy (kcal/mol) and the specific amino acid residues involved in the interaction with the substrate. Lower binding energy values indicate more favorable and stable interaction.

3. Results

3.1. Selection and Physicochemical Profiling of α-Galactosidases

Starting with 100 homologous sequences, a filtering process based on physicochemical properties led to the selection of 31 candidate enzymes for further analysis. The initial BLAST search confirmed the homology of these sequences to human α-Gal A. The ProtParam and SoluProt analyses provided key insights into the stability and solubility of these enzymes. The selected 31 candidates all showed a predicted half-life of over 10 hours in E. coli, an instability index below 40 (indicating stability), and high predicted solubility, making them promising candidates for recombinant production. The detailed results for all 100 initial sequences are provided in Table 1, with the 31 selected candidates highlighted Table 2.

Table 1. Protein parameters, SoluProt results and structure quality assessments

UniProt ID	Length (aa)	BLAST E-Value	pI	Instability Index	Solubility	QMEAN	VERIFY_3D	PROCHECK Disallowed (%)	Status
A0A069D0E7_9BACE	406	6.10E-85	6.16	25.95	0.417	0.70	Fail	1.2	Eliminated
A0A099BWT8_9BACT	406	7.80E-76	6.55	27.19	0.572	0.65	Fail	0.6	Eliminated
A0A174GRB4_9BACE	545	1.40E-44	5.42	32.75	0.467	-	-	-	Eliminated
A0A174UAB7_9BACE	538	1.20E-42	5.27	24.67	0.604	0.52	Pass	1.3	Eliminated
A0A1C7GX28_9BACE	660	1.50E-37	6.21	29.95	0.476	-	-	-	Eliminated
A0A1C7H5M1_9BACE	545	3.40E-42	5.24	30.37	0.424	-	-	-	Eliminated
A0A1H4DKF5_9BACT	410	6.40E-72	6.56	29.23	0.435	-	-	-	Selected
A0A1H8AZH4_9BACT	409	5.20E-74	5.93	26.74	0.499	-	-	-	Selected
A0A1I0MM87_9BACT	409	9.30E-75	6.51	32.92	0.499	-	-	-	Selected
A0A1I2M3X7_9BACT	408	4.70E-71	5.83	41.02	0.356	-	-	-	Eliminated
A0A1I2R6K0_9BACT	407	3.90E-69	6.12	31.57	0.495	-	-	-	Selected
A0A1I5K101_9BACT	412	1.60E-73	6.01	26.97	0.665	0.65	Pass	0.6	Selected
A0A1J0GFI3_9CLOT	387	4.50E-65	5.25	29.78	0.436	-	-	-	Selected
A0A1M4WPX1_9BACE	404	1.60E-84	6.20	29.31	0.512	0.66	Pass	0.3	Selected
A0A1M5CBM0_9BACE	543	3.30E-42	5.16	28.69	0.604	0.52	Fail	1.5	Eliminated
A0A1M5DWA2_9BACE	680	2.00E-41	6.12	21.67	0.395	-	-	-	Selected
A0A1M5E8B8_9BACE	525	7.40E-46	6.25	39.20	0.414	-	-	-	Selected
A0A1M5FVV3_9BACE	389	2.00E-78	5.53	26.98	0.618	0.70	Pass	0.7	Selected
A0A1T4X998_9CLOT	428	8.30E-25	5.37	32.54	0.620	0.78	Fail	0.6	Eliminated

Table 1. Protein parameters, SoluProt results and structure quality assessments (continued)

UniProt ID	Length (aa)	BLAST E-Value	pI	Instability Index	Solubility	QMEAN	VERIFY_3D	PROCHECK Disallowed (%)	Status
A0A1V4IHN6_9CLOT	387	9.60E-68	5.07	29.48	0.510	0.72	Pass	0.6	Selected
A0A255S3I7_9BACT	409	3.30E-75	7.08	30.12	0.464	-	-	-	Selected
A0A255SZI7_9BACT	564	2.50E-38	5.95	24.00	0.470	-	-	-	Selected
A0A255TC87_9BACT	406	1.80E-70	5.80	30.11	0.337	-	-	-	Selected
A0A264Y6Y3_9BACT	403	1.00E-72	5.55	30.97	0.524	0.72	Pass	0.3	Selected
A0A374VC22_9BACE	502	1.20E-39	6.95	24.27	0.505	0.65	Pass	0.7	Selected
A0A374W632_9BACE	520	2.00E-41	5.52	21.87	0.438	-	-	-	Selected
A0A374W9N3_9BACE	406	4.00E-78	5.47	31.74	0.378	-	-	-	Selected
A0A374WDC0_9BACE	406	1.20E-68	5.31	29.37	0.484	-	-	-	Selected
A0A410DZH3_9CLOT	392	1.00E-22	5.06	42.71	0.506	-	-	-	Eliminated
A0A416GE36_9BACE	406	1.20E-68	6.15	31.92	0.358	-	-	-	Selected
A0A5D3F0L8_9BACE	406	1.20E-68	5.77	39.19	0.398	-	-	-	Selected
A0A6I1MJ60_9CLOT	406	1.20E-68	5.65	28.02	0.493	-	-	-	Selected
A0A6I1MN87_9CLOT	752	1.20E-42	5.32	26.90	0.571	0.43	Fail	0.4	Eliminated
A0A6I1MNI8_9CLOT	660	1.20E-42	4.89	21.70	0.505	0.54	Fail	0.7	Eliminated
A0A7X9XG25_9CLOT	431	1.20E-42	5.02	29.94	0.781	0.79	Pass	0.3	Selected
A0A934HWH4_9CLOT	406	1.20E-42	5.09	25.74	0.410	-	-	-	Selected
C9MTB0_9BACT	371	1.20E-42	5.95	36.02	0.624	0.73	Pass	0.3	Selected
C9MTQ6_9BACT	406	2.00E-79	7.12	31.09	0.268	-	-	-	Selected
D3IEJ2_9BACT	406	1.20E-42	8.23	27.06	0.480	-	-	-	Selected
D3IF54_9BACT	406	1.20E-42	8.51	30.87	0.383	-	-	-	Selected
D3IKF5_9BACT	406	2.00E-75	7.52	30.44	0.272	-	-	-	Selected
D4LCA1_RUMC1	419	1.20E-42	5.17	34.39	0.662	0.79	Pass	0.4	Selected
D5EVX6_PRER2	406	1.20E-42	6.61	28.67	0.231	-	-	-	Selected
E5CAS9_9BACE	406	1.20E-42	6.18	31.99	0.385	-	-	-	Selected
E5CHE4_9BACE	406	1.20E-42	5.32	27.81	0.460	-	-	-	Selected
E6SNX6_BACT6	406	1.20E-42	6.31	35.65	0.265	-	-	-	Selected
F0F449_9BACT	406	1.20E-42	8.97	40.91	0.272	-	-	-	Eliminated
F3PSS3_9BACE	406	1.20E-42	6.10	35.61	0.335	-	-	-	Selected
G6AJI3_9BACT	406	1.20E-42	8.52	40.81	0.413	-	-	-	Eliminated
H1HL25_9BACT	406	1.20E-42	7.94	34.06	0.273	-	-	-	Selected
I9GBD4_9BACE	406	1.20E-42	5.91	36.06	0.287	-	-	-	Selected
K6U4Z0_9CLOT	406	9.00E-72	4.95	40.33	0.506	-	-	-	Eliminated
Q5L7T7_BACFN	480	3.00E-75	5.40	33.18	0.787	0.89	Pass	0.2	Selected
Q5LFG6_BACFN	484	4.00E-73	5.42	31.51	0.797	0.90	Pass	0.2	Selected
Q5LIM3_BACFN	406	1.20E-42	8.20	29.72	0.318	-	-	-	Selected
Q8A064_BACTN	406	1.20E-42	8.42	35.45	0.284	-	-	-	Selected
Q8A1R7_BACTN	406	1.20E-42	6.50	30.87	0.456	-	-	-	Selected
Q8A389_BACTN	406	4.00E-74	6.23	32.58	0.493	-	-	-	Selected
Q8A4D7_BACTN	507	1.40E-44	5.09	27.42	0.628	0.53	Pass	0.9	Selected
Q8ABP5_BACTN	406	1.20E-42	4.86	30.59	0.497	-	-	-	Selected
R5ACL5_9CLOT	406	1.00E-22	5.02	42.15	0.558	-	-	-	Eliminated
R5ATE1_9BACE	406	1.20E-42	5.15	31.82	0.402	-	-	-	Selected
R5FC01_9BACT	406	1.20E-42	7.78	28.10	0.351	-	-	-	Selected
R5GC83_9BACT	406	1.20E-42	6.26	31.17	0.317	-	-	-	Selected
R5KIG7_9CLOT	406	1.20E-42	7.57	32.25	0.314	-	-	-	Selected
R5PB58_9BACT	406	1.20E-42	7.06	26.72	0.409	-	-	-	Selected
R5PTT6_9FIRM	406	1.20E-42	5.25	30.95	0.486	-	-	-	Selected
R5RG66_9CLOT	427	4.00E-22	4.77	36.36	0.647	0.77	Pass	0.3	Selected
R5RYG8_9BACE	406	1.20E-42	6.15	25.97	0.332	-	-	-	Selected
R5SAE2_9BACE	406	1.20E-42	5.25	43.58	0.338	-	-	-	Eliminated
R5ZTF6_9BACT	967	1.20E-42	5.05	24.49	0.424	-	-	-	Selected
R6AFM2_9BACT	404	1.20E-42	6.51	37.64	0.555	0.70	Pass	0.3	Selected
R6C3E9_9BACE	406	1.20E-42	5.47	36.12	0.299	-	-	-	Selected
R6D626_9BACE	406	1.20E-42	5.17	27.51	0.492	-	-	-	Selected
R6DB23_9BACE	550	1.20E-42	5.34	27.78	0.725	0.60	Fail	1.3	Eliminated
R6EFG5_9FIRM	391	1.20E-42	5.06	21.05	0.559	0.73	Pass	0.6	Selected
R6FEQ0_9BACE	406	1.20E-42	7.12	23.18	0.279	-	-	-	Selected
R6FRK8_9BACT	670	1.20E-42	4.98	23.88	0.683	0.61	Pass	0.5	Selected
R6FUL4_9BACT	406	1.20E-42	6.10	35.89	0.469	-	-	-	Selected
R6HUK9_9FIRM	406	1.20E-42	4.97	38.34	0.439	-	-	-	Selected
R6I6P9_9FIRM	406	1.20E-42	4.81	23.19	0.472	-	-	-	Selected
R6J1B0_9CLOT	406	1.20E-42	6.26	29.85	0.317	-	-	-	Selected
R6K6D9_9BACE	406	1.20E-42	5.07	22.02	0.474	-	-	-	Selected
R6KIE4_9BACE	406	1.20E-42	5.74	22.16	0.284	-	-	-	Selected
R6NBZ8_9CLOT	348	1.20E-42	4.93	28.90	0.552	0.75	Fail	0.0	Eliminated
R6NMU9_9CLOT	406	1.20E-42	6.64	23.04	0.423	-	-	-	Selected
R6S872_9BACE	406	1.20E-42	5.78	40.19	0.300	-	-	-	Eliminated
R6SS45_9CLOT	406	1.20E-42	5.17	31.57	0.334	-	-	-	Selected
R6U293_9BACE	406	1.20E-42	5.31	41.79	0.326	-	-	-	Eliminated

Table 1. Protein parameters, SoluProt results and structure quality assessments (continued)

UniProt ID	Length (aa)	BLAST E-Value	pI	Instability Index	Solubility	QMEAN	VERIFY_3D	PROCHECK Disallowed (%)	Status
R6VVG3_9BACT	406	1.20E-42	5.54	28.88	0.392	-	-	-	Selected
R6W0L0_9FIRM	406	1.20E-42	4.95	35.18	0.435	-	-	-	Selected
R6WC51_9FIRM	377	2.00E-76	5.39	30.39	0.673	0.73	Pass	0.3	Selected
R6WM48_9CLOT	538	1.20E-42	8.95	25.87	0.519	0.51	Fail	0.6	Eliminated
R6XHN1_9BACT	405	1.00E-69	6.42	28.89	0.526	0.72	Pass	0.3	Selected
R7C4R7_9CLOT	391	1.20E-42	5.19	34.81	0.552	0.72	Pass	0.0	Selected
R7C9F1_9CLOT	412	1.80E-22	4.66	35.99	0.711	0.76	Fail	0.4	Eliminated
R7KWP3_9FIRM	406	8.00E-73	5.13	40.22	0.394	-	-	-	Eliminated
R7P5T8_9BACT	406	1.20E-42	6.54	29.61	0.337	-	-	-	Selected
R9I1D0_BACUN	480	1.20E-42	5.10	30.71	0.605	0.69	Pass	1.0	Eliminated
R9I1J1_BACUN	397	8.00E-83	6.27	27.47	0.599	0.70	Fail	1.2	Eliminated

3.2. Phylogenetic Analysis

To place the identified bacterial homologs in an evolutionary context relative to the human enzyme, a phylogenetic analysis was performed. The amino acid sequences of the selected candidates and human α-Gal A were aligned, and a Maximum Likelihood (ML) phylogenetic tree was constructed using the JTT substitution model, with the statistical robustness of the branching patterns confirmed via 500 bootstrap replicates. The resulting tree visually maps the evolutionary distances and relationships among these proteins (Figure 1).

3.3. Structural Modeling and Quality Assessment

High-quality 3D models were successfully generated for all 31 selected α-galactosidase candidates by YASARA Structure. The validation results from the SAVES server confirmed the reliability of these models. The majority of the models exhibited favorable scores across all validation metrics, including QMEAN, ProSA Z-scores, ERRAT, and Verify3D. The PROCHECK analysis showed that most residues were in the allowed regions of the Ramachandran plot, with a very low percentage in disallowed regions, indicating good stereochemical quality. A summary of the structural quality assessment for the 31 final candidates is presented in Table 2.

3.4. Immunogenicity Profile and Epitope Prediction

The prediction of B-cell epitopes is crucial for assessing the immunogenic potential of therapeutic proteins. The results from ElliPro identified both linear and discontinuous epitopes for each of the 31 candidates. The number and scores of predicted epitopes varied among the sequences. Similarly, DiscoTope 3.0 provided a list of residues predicted to form discontinuous epitopes. By combining these results, we identified several enzymes with a lower number of predicted high-scoring epitopes, suggesting a potentially reduced immunogenic risk. The complete ElliPro and DiscoTope epitope prediction results are summarized in Tables 3 and 4, respectively.

3.5. Molecular Docking Analyses of Substrate Interactions

Molecular docking simulations were performed with the natural substrate Gb3 and the artificial substrate *p*NP-Gal to evaluate the functional potential of the filtered candidates. According to the programs used, a higher positive binding energy score indicates a more stable interaction.

3.6. Interactions with Globotriaosylceramide (Gb3)

Docking with the natural substrate Gb3 is critical for assessing potential efficacy. The results showed that several bacterial homologs have strong binding potential. A0A1M5FVV3 achieved the highest binding energy score of +5.639 kcal/mol, followed by Q5LFG6 and R5RG66 with scores of +5.480 kcal/mol. Analysis of the binding poses revealed that conserved active site residues, such as Asp140 and Asp195 in A0A1M5FVV3, directly contact the substrate, supporting the biological relevance of the poses (Figure 2 and Figure 3). The top results are summarized in Table 5.

3.7. Interactions with p-Nitrophenyl α-galactopyranoside (*p*NP-Gal)

Docking with the artificial substrate *p*NP-Gal bridges in-silico predictions with future experimental assays. The analyses showed that different proteins excelled with this substrate. R6DB23 and R5RG66 displayed very high binding energies of +8.510 and +8.500 kcal/mol, respectively, suggesting that some enzymes may have a higher affinity for smaller, less complex artificial substrates. The best *p*NP-Gal binding proteins are detailed in Table 6.

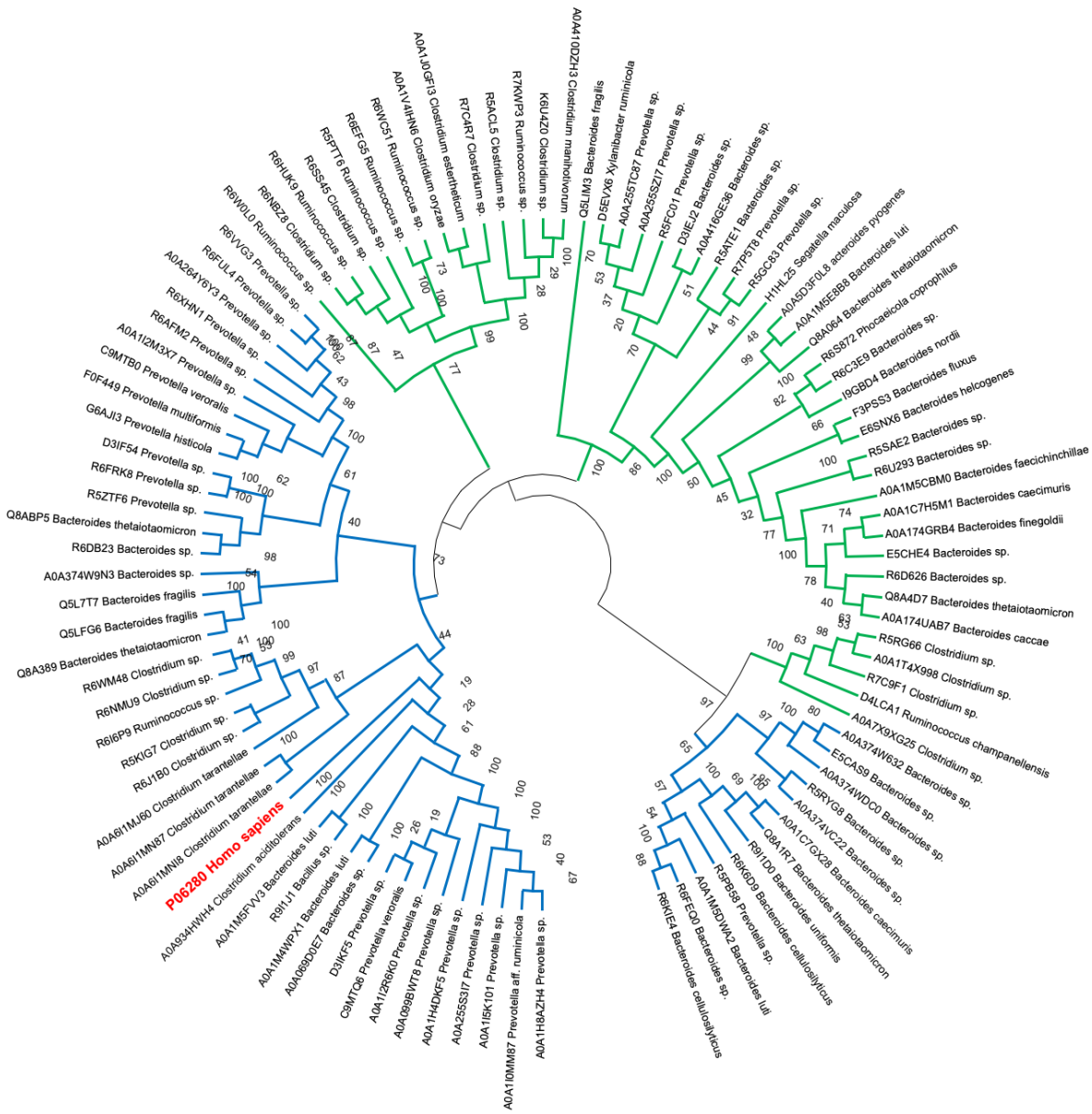

Figure 1. Phylogenetic analysis of the selected bacterial α-galactosidase homologs and human α-Gal A. The Maximum Likelihood tree illustrates two distinct evolutionary clusters corresponding to Glycoside Hydrolase (GH) families. Human α-Gal A reference sequence (UniProt: P06280) has been shown as red. Blue Branches (GH27 Family): This clade contains the human enzyme and closely related homologs from the Bacteroidetes phylum indicating a conserved 'human-like' structural architecture. Green Branches (GH36 Family): This clade represents evolutionarily distant homologs from the Firmicutes phylum selected for their distinct protein fold to potentially evade pre-existing anti-drug antibodies.

Table2. Detailed structural quality assessment for the selected alpha-galactosidase models

UniProt ID	ProSA	QMEAN Score	Errat Score	VERIFY_3D Status	PROCHECK*
A0A1I5K101		0.65	91.8159	Pass	2 - 0.6
A0A1M4WPX1		0.66	93.6387	Pass	1 - 0.3
A0A1M5CBM0	*	0.52	94.2056	Fail	7 - 1.5
A0A1M5FVV3		0.70	94.5695	Pass	9 - 0.7
A0A1T4X998		0.78	94.6301	Fail	9 - 0.6
A0A1V4IHN6		0.72	94.1953	Pass	2 - 0.6
A0A6I1MN87	*	0.43	91.1846	Fail	3 - 0.4
A0A6I1MNI8	*	0.54	96.0123	Fail	4 - 0.7
A0A7X9XG25		0.79	96.8620	Pass	4 - 0.3
A0A099BWT8		0.65	89.9240	Fail	9 - 0.6
A0A174UAB7	*	0.52	91.6828	Pass	6 - 1.3
A0A264Y6Y3		0.72	97.1354	Pass	1 - 0.3
A0A374VC22		0.65	91.8864	Pass	3 - 0.7
C9MTB0		0.73	96.9697	Pass	1 - 0.3
D4LCA1		0.79	94.5155	Pass	6 - 0.4
Q5L7T7		0.89	95.9227	Pass	1 - 0.2
Q5LFG6		0.90	93.1624	Pass	1 - 0.2
Q8A4D7		0.53	90.1606	Pass	4 - 0.9
R5RG66		0.77	95.9687	Pass	5 - 0.3
R6AFM2		0.70	95.1407	Pass	1 - 0.3
R6DB23	*	0.60	93.4334	Fail	6 - 1.3
R6EFG5		0.73	96.7828	Pass	2 - 0.6
R6FRK8		0.61	89.1239	Pass	3 - 0.5
R6NBZ8		0.75	93.7313	Fail	0 - 0.0
R6WC51		0.73	96.2008	Pass	2 - 0.3
R6WM48	*	0.51	96.2264	Fail	3 - 0.6
R6XHN1		0.72	94.2065	Pass	1 - 0.3
R7C4R7		0.72	96.3351	Pass	0 - 0.0
R7C9F1		0.76	95.5168	Fail	6 - 0.4
R9I1D0		0.69	95.9660	Pass	4 - 1.0
R9I1J1		0.70	96.8421	Fail	4 - 1.2

*Disallowed Regions (number of amino acids and %)

Table 3. Predicted B-cell epitope characteristics by Ellipro

UniProt ID	Total Residue Count	Predicted Epitope Residue Count	Epitope Percentage (%)
A0A6I1MNI8	660	48	7.27
A0A174UAB7	538	39	7.25
A0A1I5K101	412	30	7.28
R9I1D0	480	36	7.50
Q8A4D7	507	38	7.50
R6FRK8	670	51	7.61
A0A1M5CBM0	543	42	7.73
A0A7X9XG25	431	34	7.89
R6WC51	377	30	7.96
R7C9F1	412	34	8.25
D4LCA1	419	35	8.35
R5RG66	427	36	8.43
C9MTB0	371	32	8.63
A0A6I1MN87	752	65	8.64
Q5LFG6	484	44	9.09
A0A1T4X998	428	39	9.11
Q5L7T7	480	44	9.17
R6NBZ8	348	32	9.20
A0A1V4IHN6	387	37	9.56
R6EFG5	391	38	9.72
R6DB23	550	54	9.82
A0A1M5FVV3	389	39	10.03
A0A374VC22	502	60	11.95
A0A069D0E7	-	-	-
A0A099BWT8	24	406	5.91
A0A1M4WPX1	24	404	5.94
A0A264Y6Y3	28	403	6.95
R6AFM2	28	404	6.93
R6XHN1	27	405	6.67
R7C4R7	26	391	6.65
R9I1J1	25	397	6.30

Table 4. Summary of conformational B-cell epitope predictions by Discotope 3.0

UniProt ID	Predicted Epitope Residues
A0A069D0E7	23
A0A099BWT8	11
A0A174UAB7	13
A0A1I5K101	20
A0A1M4WPX1	12
A0A1M5CBM0	24
A0A1M5FVV3	39
A0A1T4X998	21
A0A1V4IHN6	11
A0A264Y6Y3	28
A0A374VC22	22
A0A6I1MN87	65
A0A6I1MNI8	48
A0A7X9XG25	17
C9MTB0	13
D4LCA1	13
Q5L7T7	44
Q5LFG6	14
Q8A4D7	17
R5RG66	17
R6AFM2	28
R6DB23	20
R6EFG5	12
R6FRK8	51
R6NBZ8	32
R6WC51	12
R6WM48	17
R6XHN1	27
R7C4R7	26
R7C9F1	17
R9I1D0	34

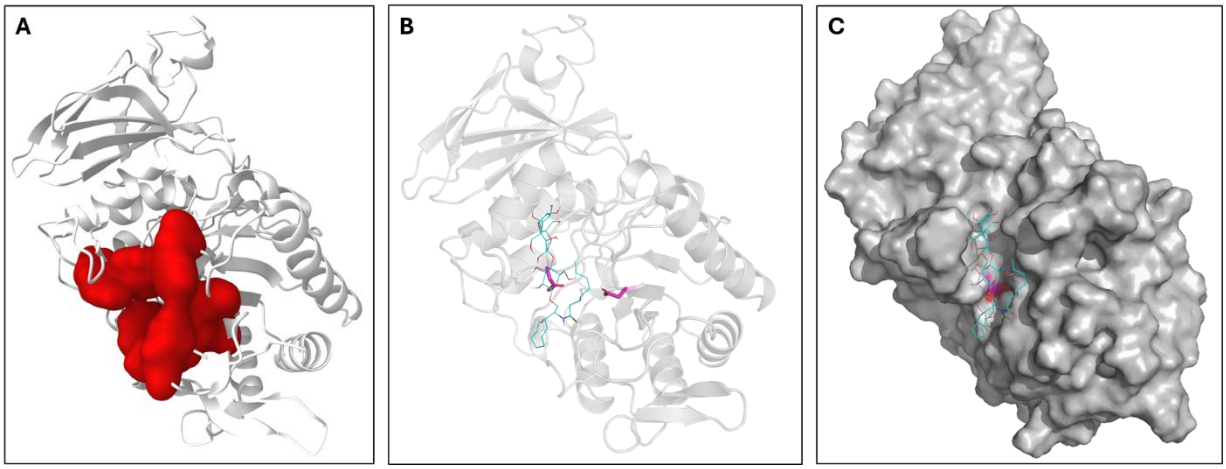

Figure 2. Binding site prediction and docking visualization of A0A1M5FVV3 with globotriaosylceramide (Gb3) as substrate. A: Binding site of A0A1M5FVV3 was predicted using P2RANK (www.prankweb.cz). Identified binding site was shown as red. B: Docking pose of A0A1M5FVV3 with globotriaosylceramide (Gb3) was shown as cartoon. C: Docking pose of A0A1M5FVV3 with globotriaosylceramide (Gb3) was shown as surface. Globotriaosylceramide (Gb3) substrate was shown as stick. Catalytic residues (Aspartate 140 and Aspartate 195) were shown as magenta color.

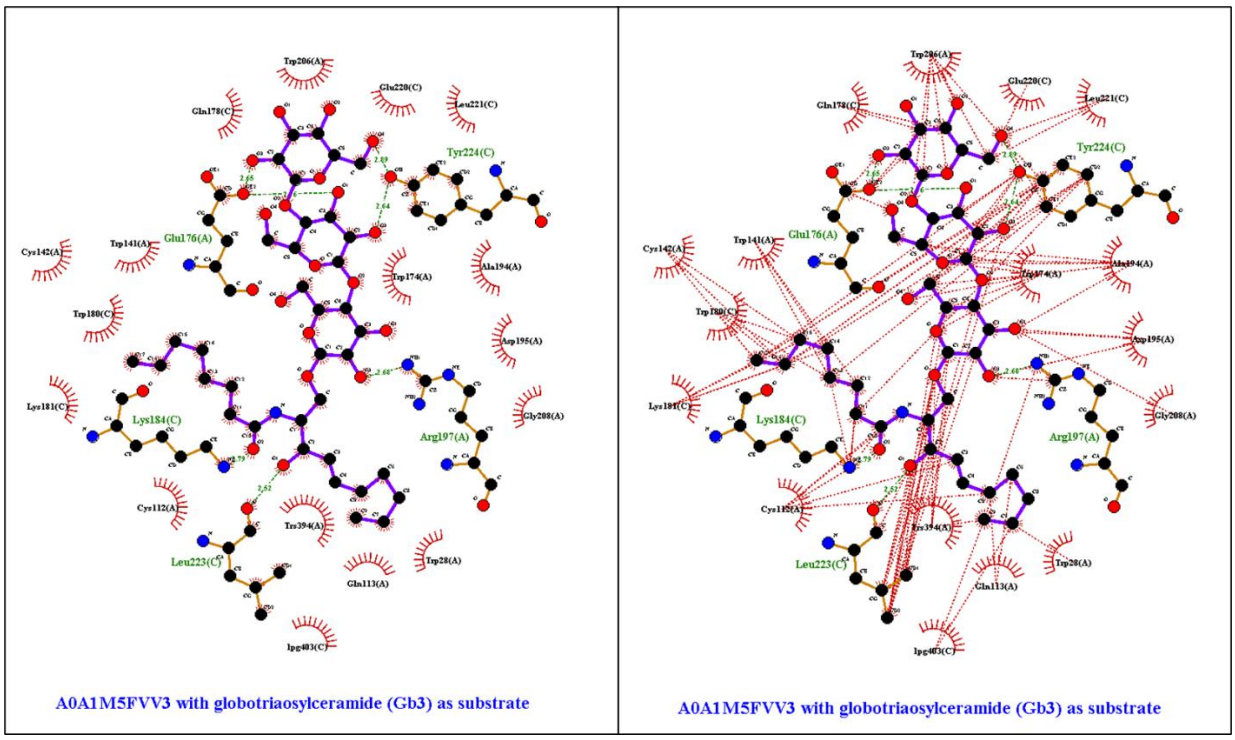

Figure 3. 2D interaction diagram obtained using LigPlot for Docking pose of A0A1M5FVV3 with globotriaosylceramide (Gb3). Left diagram shows only hydrogen bonds, while right diagram shows both hydrogen bonds and hydrophobic interactions.

Table 5. Molecular docking results with the natural substrate globotriaosylceramide (Gb3)

Uniprot ID	Docking	HBE (- kcal/mol)	Interacting Amino Acids
R6DB23	AutoDockLGA	8.510	A TRP 424 A LYS 428 A ASP 459 A ASP 460 A TYR 506 A SER 507 A ASP 508 A CYS 509 A THR 513 A CYS 514 A GLU 515 A LYS 516 A LYS 540 A GLU 541 A ASP 542 A TRP 543 A TYR 544 A PHE 545 A TYR 578 A CYS 595 A TRP 597 A ILE 599 A ARG 615 A TYR 618 A ASP 619 A HIS 620 A ARG 621 A TYR 628 A ASP 639 A GLY 640 A GLY 641 A ILE 642 A ASP 667 A SER 680
R5RG66	AutoDockLGA	8.500	A TYR 17 A ILE 52 A HIS 115 A ILE 116 A MET 117 A ARG 118 A CYS 148 A TRP 150 A ASN 151 A CYS 185 A ASP 186 A ASP 187 A ILE 188 A ARG 190 A SER 219 A PRO 220 A ARG 239 A ASP 242 A ASP 243 A TRP 245
R6WM48	AutoDockLGA	6.330	A TRP 46 A ARG 50 A ASP 81 A ASP 82 A ASN 83 A TRP 84 A VAL 85 A TYR 123 A SER 124 A SER 125 A THR 130 A CYS 131 A GLU 132 A LEU 134 A LYS 157 A TYR 158 A ASP 159 A PHE 160 A CYS 161 A HIS 162 A LEU 289 A LYS 290 A ASN 293 A TYR 306 A CYS 336 A TRP 338 A ARG 355 A PRO 358 A ASP 359 A ILE 360 A ARG 361 A TRP 366 A ASP 390
R6EFG5	AutoDockLGA	5.640	A ASN 11 A TRP 13 A ASP 48 A ASP 49 A TYR 90 A SER 91 A CYS 92 A ALA 93 A THR 97 A CYS 98 A ALA 99 A LYS 124 A TYR 125 A ASP 126 A TYR 127 A CYS 128 A TYR 129 A HIS 130 A TYR 140 A CYS 159 A TRP 161 A SER 163 A ARG 179 A THR 181 A GLY 182 A ASP 183 A ILE 184 A PHE 185 A LEU 194 A ASP 214 A MET 215
A0A1M5FVV3	VINA	5.639	A TRP 28 A TYR 104 A SER 105 A CYS 106 A ALA 107 A CYS 112 A GLN 113 A LYS 138 A TYR 139 A ASP 140 A TRP 141 A CYS 142 A CYS 172 A TRP 174 A GLU 176 A ARG 191 A THR 193 A ALA 194 A ASP 195 A ARG 197 A PHE 204 A TRP 206 A GLY 207 A GLY 208 A VAL 209 C GLN 178 C TRP 180 C LYS 181 C LYS 184 C GLU 220 C LEU 221 C LEU 223 C TYR 224 A TRS 394 C 1PG 403
Q5LFG6	AutoDockLGA	-9.010	A ASN 128 A TRP 130 A ASP 165 A ASP 166 A PHE 167 A TYR 207 A SER 208 A ASP 209 A CYS 215 A GLY 216 A LYS 241 A TYR 242 A ASP 243 A TYR 244 A CYS 245 A TYR 258 A CYS 277 A TRP 279 A GLN 281 A ARG 297 A GLY 300 A ASP 301 A ILE 302 A GLY 310 A GLY 314 A GLY 315 A LEU 316 A HIS 317 A ASP 342 A MET 343 A SER 354 A ILE 355
R9I1J1	AutoDockLGA	5.250	A ASN 35 A TRP 37 A ASP 72 A ASP 73 A TYR 113 A CYS 115 A ALA 116 A LEU 119 A CYS 121 A ALA 122 A GLY 123 A LYS 147 A TYR 148 A ASP 149 A TRP 150 A CYS 151 A PHE 152 A ASP 153 A CYS 181 A TRP 183 A ARG 200 A GLY 203 A ASP 204 A ILE 205 A ASN 207 A PHE 209 A ASP 243 A MET 244
R9I1D0	AutoDockLGA	4.930	A TRP 281 A ASP 315 A ASP 316 A TYR 355 A SER 357 A PRO 358 A CYS 363 A GLY 364 A LYS 389 A TYR 390 A ASP 391 A TYR 392 A CYS 393 A GLY 394 A TYR 395 A LEU 396 A GLU 397 A ILE 398 A TYR 411 A GLY 430 A TYR 431 A GLY 432 A ALA 433 A ARG 449 A THR 450 A THR 451 A ARG 452 A ASP 453 A ILE 454 A ASP 484 A MET 485 A SER 497 A HIS 500
Q5L7T7	AutoDockLGA	4.640	A TRP 130 A ARG 135 A ASP 165 A ASP 166 A PHE 167 A TYR 207 A THR 214 A CYS 215 A GLY 216 A LYS 241 A ASP 243 A TYR 244 A CYS 245 A ASN 246 A CYS 277 A TRP 279 A GLN 281 A ARG 297 A GLY 300 A ASP 301 A ILE 302 A THR 310 A GLY 314 A GLY 315 A LEU 316 A ASP 342 A MET 343 A ILE 355
A0A6I1MN87	AutoDockLGA	4.210	A TRP 343 A ASN 344 A GLY 347 A TYR 420 A ASN 421 A SER 422 A ALA 423 A CYS 428 A MET 429 A TYR 455 A ASP 456 A PHE 457 A CYS 458 A GLU 464 A TYR 600 A LYS 603 A CYS 634 A TRP 636 A ARG 653 A GLY 656 A ASP 657 A ILE 658 A PHE 660 A ASN 661 A HIS 662 A ASP 693 A GLA 853
R6AFM2	AutoDockLGA	3.940	A TRP 40 A ASN 41 A LYS 44 A TYR 116 A SER 117 A ASP 118 A ALA 119 A THR 123 A CYS 124 A ALA 125 A LYS 150 A TYR 151 A ASP 152 A TYR 153 A CYS 154 A ASN 155 A ALA 156 A TYR 167 A CYS 186 A TRP 188 A GLN 190 A ARG 206 A TYR 209 A ASP 210 A VAL 211 A ARG 212 A GLY 223 A ASP 249 A PRO 261 A SER 262 A SER 263
A0A264Y6Y3	AutoDockLGA	3.550	A TRP 38 A GLN 42 A ASP 73 A ASP 74 A LEU 75 A TYR 114 A SER 115 A ASP 116 A ALA 117 A CYS 122 A GLY 123 A LYS 148 A TYR 149 A ASP 150 A TYR 151 A CYS 152 A CYS 184 A TRP 186 A GLN 188 A ARG 204 A TYR 207 A ASP 208 A VAL 209 A ARG 210 A GLU 219 A GLY 220 A GLY 221 A MET 222 A SER 260 A SER 261
R6XHN1	AutoDockLGA	2.930	A TRP 42 A LYS 46 A GLY 47 A ASP 78 A LEU 79 A TYR 118 A SER 119 A ASP 120 A LEU 124 A THR 125 A CYS 126 A GLY 127 A GLY 128 A LYS 152 A TYR 153 A ASP 154 A TYR 155 A CYS 156 A ASN 157 A ALA 158 A CYS 188 A GLU 189 A TRP 190 A GLN 192 A ARG 193 A ARG 208 A TYR 211 A ASP 212 A VAL 213 A ARG 214 A GLY 223 A ASP 250
A0A374VC22	AutoDockLGA	1.320	A TRP 159 A ASP 195 A TYR 234 A PRO 237 A TYR 308 A ASP 309 A TRP 310 A SER 337 A ASN 338 A ARG 356 A GLY 359 A ASP 360 A ILE 361 A ARG 362 A ASP 363 A THR 364 A ARG 367 A ASP 391 A TRP 400 A ARG 408
R6FRK8	AutoDockLGA	1.230	A TRP 433 A ASN 434 A GLN 437 A ASP 468 A ASP 469 A LEU 470 A TYR 509 A THR 510 A ASP 511 A ALA 512 A ARG 515 A THR 516 A CYS 517 A ALA 518 A LYS 543 A CYS 544 A ASP 545 A TYR 546 A CYS 547 A TYR 548 A TYR 560 A CYS 579 A TRP 581 A ARG 599 A GLN 602 A ASP 603 A VAL 604 A ARG 605 A SER 613 A GLY 614 A ASP 640 A ASN 654 A ASP 655 A LEU 656 A CYS 657
R6WC51	AutoDockLGA	0.960	A TRP 13 A GLY 18 A CYS 50 A TYR 90 A SER 91 A CYS 92 A ALA 93 A CYS 98 A ALA 99 A VAL 100 A TYR 101 A LYS 124 A TYR 125 A ASP 126 A TYR 127 A CYS 128 A PHE 129 A LYS 130 A PRO 131 A HIS 133 A TYR 140 A CYS 159 A TRP 161 A ASP 183 A VAL 185 B ASP 315 A GLA 378
R7C4R7	AutoDockLGA	0.860	A TRP 23 A ASN 24 A THR 27 A TRP 28 A TYR 100 A CYS 102 A ALA 103 A CYS 108 A LYS 134 A TYR 135 A ASP 136 A PHE 137 A CYS 138 A PHE 139 A LYS 140 A TYR 150 A CYS 169 A ASN 170 A TRP 171 A GLU 173 A ARG 189 A PRO 192 A ASP 193 A ILE 194 A GLN 195 A ASP 224
A0A174UAB7	AutoDockLGA	-0.160	A TRP 37 A ASN 38 A ARG 41 A TYR 113 A THR 114 A ASP 115 A CYS 121 A TRP 125 A LYS 157 A ILE 158 A ASP 159 A TYR 160 A CYS 161 A CYS 193 A ARG 194 A TRP 195 A ARG 210 A GLY 213 A ASP 214 A ILE 215 A ASN 216 A ALA 217 A HIS 218 A SER 221 A ASP 245 A VAL 248 A PHE 251 A ASP 253 A LYS 256 A VAL 257 A GLY 258 A GLY 259
A0A1I5K101	AutoDockLGA	-2.040	A TRP 40 A GLN 44 A TYR 116 A SER 117 A ASP 118 A ALA 119 A CYS 124 A ALA 125 A LYS 150 A TYR 151 A ASP 152 A TRP 153 A CYS 154 A CYS 184 A GLU 185 A TRP 186 A ASN 188 A ARG 203 A PRO 206 A ASP 207 A ILE 208 A TRP 209 A GLN
A0A1T4X998	AutoDockLGA	-3.580	A TYR 18 A ILE 53 A GLN 68 A TYR 69 A HIS 116 A MET 118 A ARG 119 A SER 149 A LYS 150 A TRP 151 A ASN 152 A LYS 185 A VAL 186 A ASP 187 A ASP 188 A ILE 189 A ASN 191 A THR 192 A SER 225 A PRO 226 A ARG 245 A ASP 248 A ASP 249 A TRP 251 A ASP 252 A ASP 253 A LYS 255 A LEU 256 A PHE 291 A ARG 295 A ASN 298

HBE= highest binding energy

Table 5. Molecular docking results with the natural substrate globotriaosylceramide (Gb3) (continued)

Uniprot ID	Docking	HBE (- kcal/mol)	Interacting Amino Acids
R7C9F1	AutoDockLGA	-3.750	A ILE 48 A GLU 60 A ARG 61 A PHE 62 A HIS 111 A MET 113 A ILE 143 A CYS 144 A GLY 145 A TRP 146 A ASN 147 A LYS 180 A CYS 181 A ASP 182 A ASP 183 A ILE 184 A ASP 188 A ALA 189 A ASP 190 A SER 215 A PRO 216 A GLY 217 A PRO 218 A ARG 235 A ASP 238 A ASP 239 A TRP 241 C TYR 141
A0A099BWT8	AutoDockLGA	-12.610	A TRP 37 A ASN 38 A GLN 41 A TYR 113 A SER 114 A ASP 115 A CYS 121 A GLY 122 A LYS 147 A TYR 148 A ASP 149 A TRP 150 A CYS 151 A CYS 181 A GLU 182 A TRP 183 A PRO 203 A ASP 204 A ILE 205 A ASP 208 A PHE 209 A ASP 246 C LYS 190 C ARG 193 C THR 232 C ALA 235 C TYR 236 C ALA 237 C GLY 238 C ARG 359
A0A1V4IHN6	AutoDockLGA	-35.310	A TRP 17 A TRP 22 A ASP 52 A ASP 53 A TYR 94 A SER 95 A CYS 96 A ALA 97 A THR 101 A CYS 102 A LYS 128 A TYR 129 A ASP 130 A TYR 131 A CYS 132 A TYR 144 A CYS 163 A ASN 164 A TRP 165 A GLY 166 A ALA 167 A ARG 183 A THR 185 A GLY 186 A ASP 187 A ILE 188 A GLN 189 A ASP 218 A MET 219 A PHE 232 A ILE 233 A SER 235 A LYS 236
R6NBZ8	AutoDockLGA	-35.360	A ASN 11 A THR 12 A TRP 13 A GLY 17 A GLU 18 A ILE 20 A ASP 48 A ASP 49 A CYS 50 A TYR 91 A CYS 93 A THR 98 A CYS 99 A ALA 100 A LYS 125 A TYR 126 A ASP 127 A PHE 128 A CYS 129 A TYR 130 A LYS 131 A TYR 141 A SER 158 A CYS 160 A TRP 162 A ASN 164 A MET 178 A ARG 180 A GLY 183 A ASP 184 A ILE 185 A ASN 186 A ASP 213 A ASP 215 A MET 216 A ASN 226 A MET 252
C9MTB0	AutoDockLGA	-79.060	A MET 4 A TRP 6 A ASN 7 A GLY 10 A GLU 11 A ASP 41 A ASP 42 A TYR 82 A CYS 90 A ALA 91 A LYS 116 A ASP 118 A CYS 120 A HIS 121 A CYS 152 A TRP 154 A ARG 172 A TYR 175 A ASP 176 A VAL 177 A ARG 178 A MET 190 A ASP 213 A ASP 215 A MET 216 A PRO 227 A SER 228 A SER 229 A LEU 231
A0A6I1MNI8	AutoDockLGA	-82.510	A TRP 56 A ASN 57 A PHE 59 A GLY 60 A GLY 61 A ASN 62 A ASP 91 A ASP 92 A GLY 93 A TYR 133 A SER 135 A ALA 136 A CYS 141 A MET 142 A LYS 167 A TYR 168 A ASP 169 A PHE 170 A CYS 171 A CYS 347 A TRP 349 A ASN 351 A ARG 366 A GLY 369 A ASP 370 A ILE 371 A THR 372 A PHE 373 A ASN 374 A HIS 375 A GLU 382 A ASP 406 A MET 407
A0A1M5CBM0	AutoDockLGA	-108.410	A TRP 43 A ASN 44 A ARG 47 A VAL 48 A ILE 50 A ASP 78 A ASP 79 A TYR 119 A THR 120 A ASP 121 A ALA 122 A THR 126 A CYS 127 A GLY 128 A ALA 131 A LYS 163 A ILE 164 A ASP 165 A TYR 166 A CYS 167 A TYR 180 A CYS 199 A TRP 201 A ARG 216 A GLY 219 A ASP 220 A ILE 221 A ASN 222 A ASP 251 A MET 252 A VAL 254
Q8A4D7	AutoDockLGA	-113.250	A TRP 6 A ASN 7 A ARG 10 A VAL 11 A ASP 41 A ASP 42 A GLY 43 A TYR 82 A THR 83 A ASP 84 A ALA 85 A THR 89 A CYS 90 A GLY 91 A TRP 94 A LYS 126 A ILE 127 A ASP 128 A TYR 129 A CYS 130 A TYR 143 A CYS 162 A ARG 163 A TRP 164 A ARG 179 A SER 181 A GLY 182 A ASP 183 A ILE 184 A ASP 214 A MET 215 A VAL 217
A0A1M4WPX1	AutoDockLGA	-165.630	A SER 37 A TRP 38 A ASN 39 A LYS 40 A PHE 41 A ALA 42 A CYS 43 A ASP 44 A ASP 73 A ASP 74 A CYS 75 A HIS 77 A TYR 114 A SER 115 A ASP 116 A ALA 117 A THR 121 A CYS 122 A GLY 123 A GLY 124 A ARG 125 A LYS 148 A TYR 149 A ASP 150 A TRP 151 A CYS 152 A CYS 182 A TRP 184 A GLY 204 A ASP 205 A ILE 206 A PHE 207 A ASN 208 A CYS 209 A PHE 210 A ASP 247
D4LCA1	AutoDockLGA	-331.660	A TRP 13 A ASP 47 A ILE 48 A GLN 49 A TYR 51 A TYR 60 A HIS 107 A ILE 108 A MET 109 A ARG 110 A CYS 140 A TRP 142 A ASN 143 A THR 144 A ASP 145 A MET 146 A LYS 176 A CYS 177 A ASP 178 A ASP 179 A ILE 180 A ARG 182 A SER 210 A PRO 211 A ARG 230 A ASP 233 A ASP 234 A TRP 236
A0A7X9XG25	AutoDockLGA	-1140.000	A ASN 15 A TRP 17 A ASP 18 A ASP 51 A ILE 52 A HIS 111 A ILE 112 A MET 113 A ARG 114 A TRP 146 A LYS 180 A MET 181 A ASP 182 A ASP 183 A SER 184 A SER 218 A PRO 219 A GLY 220 A ARG 238 A THR 240 A GLY 241 A ASP 242 A TYR 243 A TRP 244 A ALA 272 A ASP 273 A MET 274 A LEU 275 A MET 318

HBE= highest binding energy

Table 6. Molecular docking results with the artificial substrate p-Nitrophenyl α-D-galactopyranoside (*p*NP-Gal)

Uniprot ID	Docking	HBE (- kcal/mol)	Interacting Amino Acids
R6DB23	AutoDockLGA	8.200	A TRP 424 A ASP 459 A TYR 506 A SER 507 A ASP 508 A CYS 509 A CYS 514 A LYS 540 A GLU 541 A ASP 542 A TRP 543 A PHE 545 A TYR 578 A CYS 595 A TRP 597 A ARG 615 A TYR 618 A ASP 619 A HIS 620 A ARG 621 A TYR 628 A ASP 639 A GLY 640 A GLY 641 A ASP 667 A MET 668
R5RG66	AutoDockLGA	8.500	A TYR 17 A ILE 52 A HIS 115 A ILE 116 A MET 117 A ARG 118 A CYS 148 A TRP 150 A ASN 151 A CYS 185 A ASP 186 A ASP 187 A ILE 188 A ARG 190 A SER 219 A PRO 220 A ARG 239 A ASP 242 A ASP 243 A TRP 245
R6WM48	AutoDockLGA	9.690	A TRP 46 A TYR 123 A SER 124 A SER 125 A CYS 131 A LYS 157 A TYR 158 A ASP 159 A PHE 160 A CYS 161 A HIS 162 A ASN 163 A LEU 289 A LYS 290 A ASN 293 A TYR 306 A CYS 336 A GLU 337 A TRP 338 A ARG 355 A PRO 358 A ASP 359 A ILE 360 A ARG 361 A TRP 366 A ASP 390
R6EFG5	AutoDockLGA	9.100	A TRP 13 A ALA 17 A GLU 18 A ASP 48 A ASP 49 A CYS 50 A TYR 90 A CYS 92 A ALA 93 A CYS 98 A ALA 99 A LYS 124 A TYR 125 A ASP 126 A TYR 127 A CYS 128 A TYR 140 A CYS 159 A TRP 161 A ARG 179 A GLY 182 A ASP 183 A ILE 184 A PHE 185 A ASP 214 A MET 215
A0A1M5FVV3	VINA	6.970	A TYR 104 A SER 105 A CYS 106 A ALA 107 A CYS 112 A LYS 138 A TYR 139 A ASP 140 A TRP 141 A CYS 142 A CYS 172 A GLU 173 A TRP 174 A GLU 176 A ARG 191 A ALA 194 A ASP 195 C TRP 180 C LYS 181 C LYS 184 C LEU 223 C TYR 224 A TRS 394
R9I1J1	AutoDockLGA	8.560	A TRP 37 A ASP 72 A ASP 73 A TYR 113 A CYS 115 A ALA 116 A CYS 121 A LYS 147 A TYR 148 A ASP 149 A TRP 150 A CYS 151 A CYS 181 A TRP 183 A ASN 185 A ARG 200 A GLY 203 A ASP 204 A ILE 205 A ASN 207 A PHE 209 A ASP 243 A MET 244
R9I1D0	AutoDockLGA	7.400	A TRP 281 A TYR 355 A SER 357 A PRO 358 A CYS 363 A LYS 389 A TYR 390 A ASP 391 A TYR 392 A CYS 393 A LEU 396 A GLU 397 A TYR 411 A GLY 430 A TYR 431 A ARG 449 A ARG 452 A ASP 453 A ILE 454 A ASP 484 A MET 485
Q5L7T7	AutoDockLGA	7.240	A TRP 130 A ARG 135 A ASP 165 A ASP 166 A PHE 167 A TYR 207 A CYS 215 A GLY 216 A LYS 241 A ASP 243 A CYS 277 A TRP 279 A ARG 297 A GLY 300 A ASP 301 A ILE 302 A LEU 316 A ASP 342 A MET 343

HBE= highest binding energy

Table 6. Molecular docking results with the artificial substrate p-Nitrophenyl α-D-galactopyranoside (*p*NP-Gal) (continued)

Uniprot ID	Docking	HBE (- kcal/mol)	Interacting Amino Acids
A0A6I1MN87	AutoDockLGA	7.330	A TYR 420 A ASN 421 A SER 422 A ALA 423 A CYS 428 A TYR 455 A ASP 456 A PHE 457 A CYS 458 A ASN 459 A ASN 460 A GLU 464 A LYS 603 A CYS 634 A TRP 636 A ARG 653 A GLY 656 A ASP 657 A ILE 658 A PHE 660 A ASP 693 A GLA 853
R6AFM2	AutoDockLGA	7.970	A TRP 40 A TYR 116 A SER 117 A ASP 118 A ALA 119 A CYS 124 A LYS 150 A TYR 151 A ASP 152 A TYR 153 A CYS 154 A ASN 155 A ALA 156 A TYR 167 A CYS 186 A TRP 188 A GLN 190 A TYR 209 A ASP 210 A VAL 211 A ARG 212 A GLN 221 A GLY 222 A GLY 223 A ASP 249
A0A264Y6Y3	AutoDockLGA	6.900	A TRP 38 A GLN 42 A ASP 73 A ASP 74 A TYR 114 A SER 115 A ASP 116 A ALA 117 A CYS 122 A LYS 148 A TYR 149 A ASP 150 A TYR 151 A CYS 152 A CYS 184 A TRP 186 A ARG 204 A ASP 208 A VAL 209 A ARG 210 A GLY 221 A SER 260 A SER 261
R6XHN1	AutoDockLGA	9.400	A TRP 42 A LYS 46 A GLY 47 A ASP 77 A ASP 78 A LEU 79 A TYR 118 A SER 119 A ASP 120 A CYS 126 A GLY 127 A LYS 152 A TYR 153 A ASP 154 A TYR 155 A CYS 156 A CYS 188 A TRP 190 A ARG 208 A TYR 211 A ASP 212 A VAL 213 A ARG 214 A GLY 223 A ASP 250
A0A374VC22	AutoDockLGA	7.790	A TRP 159 A ASP 195 A TYR 234 A SER 235 A GLY 236 A PRO 237 A TYR 308 A ASP 309 A TRP 310 A SER 337 A ASN 338 A ARG 356 A GLY 359 A ASP 360 A ILE 361 A ARG 362 A ARG 367 A ASP 391
R6FRK8	AutoDockLGA	7.100	A TRP 433 A GLN 437 A ASP 469 A LEU 470 A TYR 509 A THR 510 A ASP 511 A ALA 512 A CYS 517 A ALA 518 A LYS 543 A CYS 544 A ASP 545 A TYR 546 A CYS 547 A TYR 560 A CYS 579 A TRP 581 A ARG 599 A GLN 602 A ASP 603 A VAL 604 A ARG 605 A ASP 640 A ASN 654 A ASP 655 A LEU 656 A CYS 657
R6WC51	AutoDockLGA	6.210	A TRP 13 A TYR 90 A SER 91 A CYS 92 A ALA 93 A CYS 98 A ALA 99 A LYS 124 A TYR 125 A ASP 126 A TYR 127 A CYS 128 A PHE 129 A TYR 140 A CYS 159 A TRP 161 A ASP 183 A GLA 378
R7C4R7	AutoDockLGA	8.260	A TYR 100 A CYS 102 A ALA 103 A CYS 108 A LYS 134 A TYR 135 A ASP 136 A PHE 137 A CYS 138 A PHE 139 A LYS 140 A TYR 150 A CYS 169 A TRP 171 A GLU 173 A ARG 189 A PRO 192 A ASP 193 A ILE 194 A GLN 195 A ASP 224
A0A174UAB7	AutoDockLGA	7.090	A TRP 37 A ARG 41 A TYR 113 A THR 114 A ASP 115 A CYS 121 A TRP 125 A LYS 157 A ILE 158 A ASP 159 A TYR 160 A CYS 161 A CYS 193 A ARG 194 A TRP 195 A ARG 210 A GLY 213 A ASP 214 A ILE 215 A ASN 216 A ALA 217 A HIS 218 A ASP 245 A PHE 251 A LYS 256 A GLY 258 A GLY 259
A0A1I5K101	AutoDockLGA	5.310	A TRP 40 A GLN 44 A TYR 116 A SER 117 A ASP 118 A ALA 119 A CYS 124 A LYS 150 A TYR 151 A ASP 152 A TRP 153 A CYS 154 A CYS 184 A GLU 185 A TRP 186 A ASN 188 A ARG 203 A PRO 206 A ASP 207 A ILE 208 A TRP 209 A PHE 225 A ASP 250
A0A1T4X998	AutoDockLGA	9.290	A TYR 18 A ASP 52 A ILE 53 A GLN 68 A TYR 69 A HIS 116 A MET 118 A ARG 119 A SER 149 A LYS 150 A TRP 151 A ASN 152 A LYS 185 A VAL 186 A ASP 187 A ASP 188 A ILE 189 A SER 225 A PRO 226 A ARG 245 A ASP 249
R7C9F1	AutoDockLGA	6.460	A ARG 61 A HIS 111 A MET 113 A ARG 114 A ILE 143 A CYS 144 A GLY 145 A TRP 146 A LYS 180 A CYS 181 A ASP 182 A ASP 183 A ILE 184 A CYS 185 A ASP 188 A ASP 190 A SER 215 A PRO 216 A GLY 217 A PRO 218 A ARG 235 A ASP 238 C TYR 141
Q5LFG6	AutoDockLGA	7.970	A TRP 130 A TYR 207 A SER 208 A ASP 209 A CYS 215 A LYS 241 A TYR 242 A ASP 243 A TYR 244 A CYS 245 A TYR 258 A CYS 277 A GLU 278 A TRP 279 A ARG 297 A GLY 300 A ASP 301 A ILE 302 A GLY 310 A GLY 314 A GLY 315 A LEU 316 A HIS 317 A ASP 342 A MET 343
A0A099BWT8	AutoDockLGA	7.370	A TRP 37 A GLN 41 A GLY 42 A ASP 73 A TYR 113 A SER 114 A ASP 115 A CYS 121 A GLY 122 A LYS 147 A TYR 148 A ASP 149 A TRP 150 A CYS 151 A CYS 181 A TRP 183 A PRO 203 A ASP 204 C LYS 190 C ARG 193 C ALA 235 C TYR 236
A0A1V4IHN6	AutoDockLGA	5.180	A TRP 17 A TYR 94 A CYS 96 A ALA 97 A CYS 102 A LYS 128 A TYR 129 A ASP 130 A TYR 131 A CYS 132 A TYR 144 A CYS 163 A TRP 165 A ARG 183 A THR 185 A GLY 186 A ASP 187 A ILE 188 A GLN 189 A SER 194 A LEU 198 A ASP 218 A PHE 232 A LYS 236
R6NBZ8	AutoDockLGA	6.710	A TRP 13 A ASP 49 A TYR 91 A CYS 93 A THR 98 A CYS 99 A ALA 100 A LYS 125 A TYR 126 A ASP 127 A PHE 128 A CYS 129 A TYR 141 A CYS 160 A TRP 162 A ARG 180 A GLY 183 A ASP 184 A ILE 185 A ASP 215 A VAL 227
C9MTB0	AutoDockLGA	5.140	A TRP 6 A GLU 11 A ASP 41 A ASP 42 A TYR 82 A CYS 90 A ALA 91 A LYS 116 A ASP 118 A CYS 152 A TRP 154 A ARG 172 A ASP 176 A ARG 178 A SER 228 A LEU 231
A0A6I1MNI8	AutoDockLGA	5.840	A TRP 56 A ASN 57 A THR 58 A PHE 59 A GLY 60 A GLY 61 A ASP 92 A TYR 133 A SER 135 A ALA 136 A CYS 141 A MET 142 A LYS 167 A TYR 168 A ASP 169 A PHE 170 A CYS 171 A CYS 347 A TRP 349 A ARG 366 A GLY 369 A ASP 370 A ILE 371 A PHE 373 A ASN 374 A HIS 375 A ASP 406 A MET 407
A0A1M5CBM0	AutoDockLGA	8.030	A TRP 43 A TYR 119 A THR 120 A ASP 121 A CYS 127 A ALA 131 A LYS 163 A ILE 164 A ASP 165 A TYR 166 A CYS 167 A TYR 180 A CYS 199 A TRP 201 A ARG 216 A GLY 219 A ASP 220 A ILE 221 A ASP 251
Q8A4D7	AutoDockLGA	6.960	A TRP 6 A ASP 41 A TYR 82 A THR 83 A ASP 84 A ALA 85 A CYS 90 A TRP 94 A LYS 126 A ILE 127 A ASP 128 A TYR 129 A CYS 130 A TYR 143 A CYS 162 A TRP 164 A ARG 179 A GLY 182 A ASP 183 A ILE 184 A ASP 214 A MET 215
A0A1M4WPX1	AutoDockLGA	5.790	A TRP 38 A ASN 39 A LYS 40 A ALA 42 A TYR 114 A SER 115 A ASP 116 A ALA 117 A CYS 122 A LYS 148 A TYR 149 A ASP 150 A TRP 151 A CYS 182 A TRP 184 A ARG 201 A GLY 204 A ASP 205 A ILE 206 A PHE 207 A CYS 209 A PHE 210 A ASP 247
D4LCA1	AutoDockLGA	5.770	A ILE 48 A TYR 60 A HIS 107 A ILE 108 A MET 109 A ARG 110 A CYS 140 A TRP 142 A ASN 143 A LYS 176 A CYS 177 A ASP 178 A ASP 179 A ILE 180 A ARG 182 A SER 210 A PRO 211 A ARG 230 A ASP 233 A ASP 234 A TRP 236
A0A7X9XG25	AutoDockLGA	-4.320	A TRP 17 A ILE 52 A HIS 111 A ILE 112 A MET 113 A TRP 146 A LYS 180 A MET 181 A ASP 182 A ASP 183 A SER 184 A SER 218 A PRO 219 A ARG 238 A GLY 241 A ASP 242 A TRP 244 A ARG 246

4. Discussion

This study has identified novel bacterial α-galactosidases with therapeutic potential through a comprehensive bioinformatic pipeline. The findings shed light on the intricate relationships between structural quality, substrate specificity, immunogenicity, and physicochemical properties.

4.1. An Integrated Strategy for Identifying Superior α-Galactosidase Homologs

This study presents a comprehensive *in silico* screening pipeline designed to identify novel bacterial α-

galactosidases with therapeutic potential for Fabry disease. The work is predicated on the persistent clinical need for improved enzyme replacement therapies (ERTs) that can overcome the limitations of currently approved treatments (Lidove et al., 2010). While ERT with recombinant human α-galactosidase A (α-Gal A) has been the standard of care for two decades, its efficacy is often constrained by significant challenges, most notably the development of anti-drug antibodies (ADAs) that can neutralize the enzyme and trigger infusion-associated reactions (Lenders et al., 2025). Furthermore, the burdensome bi-weekly intravenous infusion schedule and evidence of incomplete substrate clearance in key tissues highlight the necessity for next-generation enzymes with superior biochemical and immunological properties (Lenders and Brand, 2021). In response, this study employs a rational, high-throughput computational strategy to systematically mine the vast enzymatic biodiversity of the bacterial kingdom, aiming to discover homologs of human α-Gal A that are not merely functional equivalents but are intrinsically superior in terms of substrate affinity, stability, and immunogenic profile.

The methodology can be conceptualized as a stepwise filtration process, designed to progressively enrich a large pool of initial sequences for candidates possessing a combination of therapeutically desirable traits. The process began with a broad BLASTp search to identify 100 bacterial homologs, which were then subjected to a series of stringent filtering steps. The initial filter assessed fundamental physicochemical properties, selecting for enzymes predicted to be stable and highly soluble—critical prerequisites for successful recombinant production and formulation. This was followed by the generation of three-dimensional structural models, which underwent rigorous quality assessment to ensure their stereochemical and energetic validity. Only high-quality models were advanced to the functional and safety evaluation stages, which involved molecular docking simulations to predict substrate binding affinity and immunoinformatic analyses to predict B-cell epitopes and thus potential immunogenicity. This integrated approach represents a significant methodological strength, as it avoids the common pitfall of optimizing for a single parameter, such as catalytic activity, at the expense of other equally critical attributes like manufacturability or patient safety. By simultaneously evaluating efficacy, producibility, and safety *in silico*, the pipeline efficiently narrows the field to a small cohort of high-potential candidates for resource-intensive experimental validation.

A particularly noteworthy aspect of the screening strategy is the deliberate focus on sourcing enzyme candidates from bacteria prevalent in the human gut microbiome, specifically from the genera *Bacteroides*, *Prevotella, Clostridium*, and *Ruminococcus* (Forster et al., 2019; Gordo, 2019). This choice is not merely one of convenience but represents a sophisticated bio-inspired design principle aimed at mitigating the primary clinical barrier of immunogenicity. Unlike the assumption that gut-derived proteins are inherently tolerated, recent evidence suggests that the systemic immune system may recognize commensal antigens if the mucosal barrier is breached. Therefore, in this study, the selection of candidates such as A0A1M5FVV3 and R6EFG5 was driven not by their biological origin, but by their specific immunoinformatic profiles. By prioritizing sequences with low B-cell epitope density (predicted via ElliPro and DiscoTope), we aimed to identify 'stealth' enzyme variants that offer reduced surface visibility to the host immune system, independent of evolutionary familiarity.

In contrast, current ERTs, such as agalsidase beta produced in Chinese hamster ovary (CHO) cells, are recognized as foreign proteins by the patient's immune system, leading to the formation of ADAs in a substantial proportion of treated individuals, particularly males with the classic phenotype (Lidove et al., 2010). By sourcing homologous enzymes from bacteria that are a constant and integral part of the human biological landscape, there is a compelling theoretical basis to hypothesize that these proteins may possess structural motifs that are inherently less immunogenic or to which the human immune system is already partially tolerized. This strategic decision to leverage the principles of evolutionary biocompatibility represents a novel approach to addressing the challenge of immunogenicity in ERT, moving beyond post-translational modifications or immunosuppressive co-therapies to identify enzymes that are intrinsically better suited for human therapeutic use.

4.2. Delineating Substrate Specificity and Validating Catalytic Function through Molecular Docking

A central component of this study's functional analysis is the use of molecular docking to predict the binding affinity of the candidate enzymes for both the natural substrate, globotriaosylceramide (Gb3), and a common artificial substrate, p-nitrophenyl α-galactopyranoside (*p*NP-Gal). These simulations provide crucial, albeit predictive, insights into the potential catalytic efficacy and substrate specificity of the identified homologs. The study specifies that for the docking programs used, a higher positive binding energy score, reported in kcal/mol, indicates a stronger and more stable binding interaction. This convention will be used for the interpretation of the docking results throughout this discussion. Following this convention, the docking results reveal several candidates with exceptionally high predicted affinity for the natural substrate Gb3. Notably, the enzyme R6EFG5 shows a very strong predicted binding energy of +5.640 kcal/mol, and A0A1M5FVV3 exhibits a similarly high affinity with a score of +5.639 kcal/mol. Another strong candidate, R5RG66, was also highlighted with a binding score of +5.480 kcal/mol. These high-affinity binders are of primary therapeutic interest, as strong binding is often a prerequisite for efficient catalysis and effective substrate clearance at the low physiological concentrations found within lysosomes.

The comparative docking analysis with Gb3 and the smaller artificial substrate *p*NP-Gal reveals interesting distinctions in substrate preference, likely stemming from differences in the architecture of the enzyme active sites. For example, the enzyme R5RG66 was highlighted for its very high binding score with *p*NP-Gal (+8.500 kcal/mol) and also showed a strong interaction with Gb3 (+5.480 kcal/mol). This suggests a versatile active site. In contrast, some enzymes may have a compact, well-defined pocket ideally suited to accommodate the small, rigid *p*NP-Gal molecule but less capable of optimally binding the larger, more flexible Gb3 molecule with its long ceramide tail. Conversely, enzymes like A0A1M5FVV3, which show a strong preference for Gb3, likely possess a more expansive active site groove that can make favorable contacts with both the terminal galactose moiety and the hydrophobic lipid portion of the natural substrate. This distinction is of practical importance: enzymes with high affinity for Gb3 are the most promising candidates for developing new ERTs, whereas those that excel at binding *p*NP-Gal could be valuable as diagnostic reagents or tools for high-throughput screening assays.

Perhaps the most compelling finding from the molecular docking analysis is the consistent observation that the top-scoring, lowest-energy binding poses for both substrates feature direct interactions with the two conserved Aspartate residues that form the catalytic heart of GH27 and GH36 family α-galactosidases. The catalytic mechanism for these enzymes proceeds via a double displacement reaction involving a covalent glycosyl-enzyme intermediate, which requires two critical carboxylate residues: one acting as a catalytic nucleophile and the other as a general acid/base catalyst. In this study, the docking simulations, which function as unbiased search algorithms to find the most energetically favorable ligand position, independently and repeatedly placed the substrates within the correct active site cleft, oriented precisely for interaction with these key catalytic Aspartates (e.g., Asp140 and Asp195 in A0A1M5FVV3). This result serves as a powerful *in-silico* validation of the entire preceding workflow. It strongly implies that the homology models are accurate in their most functionally critical region—the active site—and that the predicted binding modes are not random computational artifacts but are biologically plausible representations of the enzyme-substrate complex at the initiation of the catalytic cycle. This internal consistency significantly bolsters confidence in the functional predictions and provides a solid rationale for prioritizing these specific candidates for experimental characterization.

The phylogenetic analysis reveals that the top-performing candidates, A0A1M5FVV3 and R6EFG5, cluster in distinct lineages that are evolutionarily distant from human α-Gal A. This significant divergence suggests these enzymes utilize unique protein architectures to achieve catalysis, potentially offering reduced immunogenicity while maintaining the high stability and Gb3 binding affinity identified in this study.

Strategic Divergence in Candidate Selection: The identification of candidates from two distinct phylogenetic clusters (GH27 and GH36) represents a dual strategy to overcome current therapeutic limitations:

1. The High-Affinity Homolog (GH27): Candidate A0A1M5FVV3 (Bacteroides luti) clusters closely with the human enzyme (GH27 family). This structural similarity preserves the active site architecture, explaining its superior binding energy to Gb3 (+5.639 kcal/mol). It represents an 'optimized human-like' candidate where catalytic efficiency is maximized.
2. The Non-Cross-Reactive Alternative (GH36): In contrast, R6EFG5 (Firmicutes sp.) belongs to the evolutionarily distant GH36 family. This distance is a strategic advantage for patients with high titers of neutralizing antibodies against the standard human enzyme (agalsidase beta). Due to its divergent structural scaffold, R6EFG5 is less likely to be recognized by pre-existing anti-drug antibodies (ADAs), offering a potential 'rescue therapy' option

Evolutionary Basis of Divergence: The phylogenetic split between the GH27 (Blue clade) and GH36 (Green clade) families reflects distinct ecological niches driven by substrate specificity.

The GH27 clade (containing *Bacteroides spp*. and Human α-Gal A) has evolved primarily to degrade complex host-derived glycoconjugates and glycolipids. This evolutionary path explains the structural similarity of candidate A0A1M5FVV3 to the human enzyme and its superior affinity for the lipid-tailed substrate Gb3.

Conversely, the GH36 clade (containing *Firmicutes spp*.) has evolved to metabolize dietary plant oligosaccharides (e.g., raffinose). These enzymes, such as R6EFG5, often adopt robust tetrameric structures for stability in the harsh gut environment. While evolutionarily distant from the human enzyme, their catalytic core is conserved enough to hydrolyze galactose, offering a stable and immunologically distinct 'scaffold' for therapy.

4.3. An Integrated Assessment of Therapeutic Potential: Synthesizing Efficacy, Safety, and Producibility

The development of a successful biotherapeutic is a multi-objective optimization problem. An ideal candidate for enzyme replacement therapy must not only exhibit high catalytic activity but also possess a low risk of immunogenicity and favorable physicochemical properties that permit robust, large-scale manufacturing. Therefore, a holistic assessment that integrates these disparate parameters is essential for rationally prioritizing the candidates identified in this study. By synthesizing the data on predicted Gb3 binding affinity (a proxy for efficacy), B-cell epitope content (a proxy for immunogenic risk), and biophysical characteristics like stability and solubility (proxies for manufacturability), a clearer picture of the most promising therapeutic leads emerges.

This integrated analysis highlights several bacterial homologs that exhibit a superior overall profile. The

protein A0A1M5FVV3 stands out as a particularly compelling candidate. It combines a strong predicted binding affinity for the natural substrate Gb3 (+5.639 kcal/mol) with exceptional stability (instability index of 26.98) and high predicted solubility (0.618). Although its predicted epitope burden is moderate (10.03%), its excellent biophysical properties and high predicted efficacy mark it as a protein of significant interest. Another strong candidate is R5RG66, which pairs a high Gb3 binding score (+5.480 kcal/mol) with a low predicted epitope percentage (8.43%), good stability (36.36), and high solubility (0.647). This powerful combination of high predicted efficacy, favorable immunogenicity profile, and good manufacturability characteristics makes R5RG66 a prime candidate for advancement. Similarly, R6EFG5 presents a promising profile with the top Gb3 binding score (+5.640 kcal/mol), exceptional stability (21.05), and good solubility (0.559), making it another key candidate for experimental validation.

To facilitate a direct and rational comparison of these leading candidates, the key performance metrics are summarized in Table 7. This matrix distills the complex, multi-source data into a single, accessible format, enabling an evidence-based prioritization of candidates for the next, more resource-intensive phases of drug development. This act of synthesis and prioritization is a core outcome of the *in-silico* screening process.

Table 7. Integrated therapeutic profile of top α-galactosidase candidates

UniProt ID	Predicted Gb3 Binding Energy (-kcal/mol)	Predicted B-Cell Epitope Percentage (%) [Ellipro]	Instability Index	Predicted Solubility Score	Structural Quality (PROCHECK Disallowed %)	Overall Therapeutic Potential
R6EFG5	5.640	9.72	21.05	0.559	0.6	Very High
A0A1M5FVV3	5.639	10.03	26.98	0.618	0.7	Very High
R5RG66	5.480	8.43	36.36	0.647	0.3	High
Q5L7T7	4.640	9.17	33.18	0.787	0.2	High
R6AFM2	3.940	6.93	37.64	0.555	0.3	Moderate-High

This integrated assessment transforms the study from a simple catalog of homologous enzymes into a targeted discovery engine. It successfully identifies a shortlist of bacterial α-galactosidases, led by R6EFG5 and A0A1M5FVV3, that not only show promise for high catalytic activity but are also predicted to be stable, soluble, and potentially less immunogenic than current therapies. These candidates represent the most valuable outputs of the computational pipeline and are the logical focus of all future experimental work (Table 7).

4.4. Contextualizing Novel Candidates within the Evolving Landscape of Fabry Disease Therapy

The identification of these novel bacterial enzymes is not an academic exercise; it directly addresses the significant unmet medical needs that persist in the clinical management of Fabry disease. Although ERT is the established standard of care, it is a treatment, not a cure, and its long-term utility is often compromised (Lenders et al., 2025; Lenders and Brand, 2021). A primary limitation is immunogenicity. The development of neutralizing ADAs, particularly in male patients with the classic phenotype who have little to no native α-Gal A protein, can significantly reduce the efficacy of therapy by accelerating enzyme clearance and blocking its catalytic activity (Lidove et al., 2010). This immune response can also lead to infusion-associated reactions, adding to the treatment burden. The candidates identified in this study, with their predicted lower B-cell epitope content, offer a potential pathway to a less immunogenic ERT, which could lead to more consistent therapeutic outcomes and improved patient safety. There is a risk that a protein tolerated in the gut may elicit an immune response when administered intravenously. T-cell epitopes were not assessed in this study, which constitutes a limitation. However, the scarcity of B-cell epitopes serves as a positive starting point for further studies and developments.

The potential advantages of these novel enzymes are further illuminated by the long-standing clinical debate surrounding the two first-generation ERTs: agalsidase alfa (Replagal®) and agalsidase beta (Fabrazyme®). These enzymes are biochemically very similar, yet they are approved at a five-fold dose difference (0.2 mg/kg for alfa vs. 1.0 mg/kg for beta). Multiple comparative studies have demonstrated that the higher dose of agalsidase beta results in a more robust and sustained reduction of the key disease biomarker, globotriaosylsphingosine (lyso-Gb3), and may have a more favorable impact on cardiac hypertrophy (Arends et al., 2018). This suggests a dose-response relationship for clinical efficacy. However, administering a higher protein dose inherently increases the cost of therapy and potentially elevates the risk of immunogenicity (Lenders et al., 2025; Lenders and Brand, 2021). This creates a difficult clinical trade-off between maximizing efficacy and minimizing risk and cost. The novel bacterial candidates identified here present an opportunity to break this paradigm. An enzyme with an intrinsically higher catalytic efficiency, as suggested by the strong Gb3 binding affinity predicted for candidates like R6EFG5 and A0A1M5FVV3, could achieve the superior substrate clearance associated with high dose agalsidase beta but at a significantly lower administered protein dose. A more potent enzyme would be more efficient at degrading Gb3, especially at the low substrate concentrations present in tissues. This could allow for a therapeutic regimen with a lower dose, which would simultaneously reduce the immunogenic burden on the patient and lower the cost of manufacturing, directly addressing the two most significant drawbacks of the current market-leading therapy.

The therapeutic landscape for Fabry disease is rapidly evolving, with the recent approval of next-generation therapies and a robust pipeline of investigational agents (van der Veen et al., 2020). Pegunigalsidase alfa (Elfabrio®), a recently approved ERT, utilizes PEGylation—a chemical modification—to extend the enzyme's plasma half-life and potentially reduce its immunogenicity (Lidove et al., 2010). This represents a strategy of improving an existing enzyme through protein engineering. The current study offers a complementary approach: discovering an intrinsically superior enzyme from a natural source. A highly potent, low-immunogenicity bacterial α-galactosidase could potentially be used on its own or in conjunction with technologies like PEGylation to create a third-generation ERT with unparalleled efficacy and safety.

Furthermore, the relevance of an improved ERT remains high even in the era of potentially curative gene therapies and convenient oral treatments. Gene therapies, such as those being developed by uniQure and Sangamo Therapeutics, offer the profound promise of a one-time treatment but still face significant hurdles related to long-term safety, durability of expression, and stringent patient eligibility criteria (Lenders et al., 2025; Lenders and Brand, 2021). Oral substrate reduction therapies (SRTs), like venglustat and lucerastat, offer the advantage of broad applicability across mutation types and ease of administration but primarily act to prevent new substrate accumulation rather than clearing existing deposits. Therefore, for the foreseeable future, a large segment of the Fabry patient population will continue to depend on ERT. The development of a 'biobetter' enzyme, as proposed by this study, is not merely an incremental improvement but a vital therapeutic goal that will continue to have a major clinical impact, serving patients who are not candidates for other modalities or potentially being used in combination strategies.

Additionally, bacterial enzymes do not carry the signal for lysosomal localization. Consequently, obtained candidates cannot be used directly. However, these enzymes should be considered highly stable 'biological parts' suitable for genetic engineering to incorporate signals or for encapsulation within liposomes/nanoparticles.

5. Conclusion

This study has successfully implemented a comprehensive *in-silico* pipeline to identify and characterize novel bacterial homologs of human α-galactosidase A. The process involved a combinatorial assessment, including rigorous structural quality control, immunogenicity prediction, and molecular docking analyses. The results highlighted specific candidates, such as A0A1M5FVV3 and R6EFG5, which exhibit high potential for binding the natural substrate Gb3, making them prime candidates for further development as treatments for Fabry disease. This work demonstrates the power of computational methods to accelerate the discovery of new enzyme candidates and represents a significant step toward developing safer and more effective next-generation enzyme replacement therapies. It is important to acknowledge that the therapeutic potential identified in this study relies on predictive models and awaits biological verification. Future work will focus on the recombinant expression of the top-ranked bacterial homologs to validate their kinetic properties and stability profiles in a wet-lab setting.

Finally, while these bacterial enzymes exhibit promising *in-silico* properties, they lack mammalian post-translational modifications and lysosomal targeting signals (M6P). Therefore, they should be viewed as valuable 'biological parts' for future bioengineering efforts, likely requiring encapsulation (e.g., liposomes) or fusion with targeting peptides for effective intracellular delivery.

Author Contributions

The percentages of the authors' contributions are presented below. All authors reviewed and approved the final version of the manuscript.

	O.K.	Y.E.
C	50	50
D	50	50
DCP	50	50
DAI	50	50
L	50	50
W	50	50
CR	50	50
SR	50	50
PM	50	50
FA	50	50

C= concept, D= design, S= supervision, DCP= data collection and/or processing, DAI= data analysis and/or interpretation, L= literature search, W= writing, CR= critical review, SR= submission and revision, PM= project management, FA= funding acquisition.

Conflict of Interest

The authors declared that there is no conflict of interest.

Ethical Consideration

Ethics committee approval was not required for this study because there was no study on animals or humans.

References

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research*, *25*(17), 3389–3402.

Anisha, G. S. (2023). Biopharmaceutical applications of α-galactosidases. *Biotechnology and Applied Biochemistry*, *70*(1), 257–267.

Arends, M., Biegstraaten, M., Wanner, C., Sirrs, S., Mehta, A., Elliott, P. M., Oder, D., Watkinson, O. T., Bichet, D. G., Khan, A.,

Iwanochko, M., Vaz, F. M., van Kuilenburg, A. B. P., West, M. L., Hughes, D. A., & Hollak, C. E. M. (2018). Agalsidase alfa versus agalsidase beta for the treatment of Fabry disease: An international cohort study. *Journal of Medical Genetics*, *55*(5), 351–358.

Azevedo, O., Cordeiro, F., Gago, M. F., Miltenberger-Miltenyi, G., Ferreira, C., Sousa, N., & Cunha, D. (2021). Fabry disease and the heart: A comprehensive review. *International Journal of Molecular Sciences*, *22*(9), 4434.

Benkert, P., Künzli, M., & Schwede, T. (2009). QMEAN server for protein model quality estimation. *Nucleic Acids Research*, *37*(suppl_2), W510–W514.

Benkert, P., Tosatto, S. C. E., & Schomburg, D. (2008). QMEAN: A comprehensive scoring function for model quality assessment. *Proteins: Structure, Function, and Bioinformatics*, *71*(1), 261–277.

Borzova, N. V., & Varbanets, L. D. (2024). Distribution, properties, and practical significance of α-galactosidase. *Mikrobiolohichnyi Zhurnal*, *86*(1), 90–113.

Cervera-Tison, M., Tailford, L. E., Fuell, C., Bruel, L., Sulzenbacher, G., Henrissat, B., Berrin, J. G., Fons, M., Giardina, T., & Juge, N. (2012). Functional analysis of family GH36 α-galactosidases from *Ruminococcus gnavus* E1: Insights into the metabolism of a plant oligosaccharide by a human gut symbiont. *Applied and Environmental Microbiology*, *78*(21), 7720–7732.

Forster, S. C., Kumar, N., Anonye, B. O., Almeida, A., Viciani, E., Stares, M. D., Dunn, M., Mkandawire, T. T., Zhu, A., Shao, Y., Pike, L. J., Louie, T., Browne, H. P., Mitchell, A. L., Neville, B. A., Finn, R. D., & Lawley, T. D. (2019). A human gut bacterial genome and culture collection for improved metagenomic analyses. *Nature Biotechnology*, *37*(2), 186–192.

Gaillard, T. (2018). Evaluation of AutoDock and AutoDock Vina on the CASF-2013 benchmark. *Journal of Chemical Information and Modeling*, *58*(8), 1697–1706.

Garman, S. C. (2007). Structure–function relationships in α-galactosidase A. *Acta Paediatrica*, *96*, 6–16.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S. E., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. In *The proteomics protocols handbook* (pp. 571–607). Humana Press.

Gordo, I. (2019). Evolutionary change in the human gut microbiome: From a static to a dynamic view. *PLoS Biology*, *17*(1), e3000155.

Høi, M. H., Gade, F. S., Johansen, J. M., Würtzen, C., Winther, O., Nielsen, M., & Marcatili, P. (2024). DiscoTope-3.0: Improved B-cell epitope prediction using inverse folding latent representations. *Frontiers in Immunology*, *15*, 1322712.

Hon, J., Marusiak, M., Martinek, T., Kunka, A., Zendulka, J., Bednar, D., & Damborsky, J. (2021). SoluProt: Prediction of soluble protein expression in *Escherichia coli*. *Bioinformatics*, *37*(1), 23–28.

Ju, L. K., Loman, A. A., & Mahfuzul Islam, S. M. (2019). α-Galactosidase and its applications in food processing. In L. Melton, F. Shahidi, & P. Varelis (Eds.), *Encyclopedia of food chemistry* (pp. 124–128). Academic Press.

Katrolia, P., Rajashekhara, E., Yan, Q., & Jiang, Z. (2014). Biotechnological potential of microbial α-galactosidases. *Critical Reviews in Biotechnology*, *34*(4), 307–317.

Kumart, S., Stecher, G., Suleski, M., Sanderford, M., Sharma, S., & Tamura, K. (2024). MEGA12: Molecular Evolutionary Genetic Analysis Version 12 for Adaptive and Green Computing. *Molecular Biology and Evolution*, *41*(5), msae086.

Lafond, M., Tauzin, A. S., Bruel, L., Laville, E., Lombard, V., Esque, J., André, I., Vidal, N., Pompeo, F., Quinson, N., Perrier, J., Fons, M., Potocki-Veronese, G., & Giardina, T. (2020). α-Galactosidase and sucrose-kinase relationships in a bi-functional AgaSK enzyme produced by the human gut symbiont *Ruminococcus gnavus* E1. *Frontiers in Microbiology*, *11*, 579521.

Land, H., & Humble, M. S. (2018). YASARA: A tool to obtain structural guidance in biocatalytic investigations. *Methods in Molecular Biology*, *1685*, 43–67.

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

Lenders, M., & Brand, E. (2021). Fabry disease: The current treatment landscape. *Drugs*, *81*(6), 635–645.

Lenders, M., Menke, E. R., & Brand, E. (2025). Progress and challenges in the treatment of Fabry disease. *BioDrugs*, *39*(4), 517–535.

Lidove, O., West, M. L., Pintos-Morell, G., Reisin, R., Nicholls, K., Figuera, L. E., Parini, R., Carvalho, L. R., Kampmann, C., Pastores, G. M., & Mehta, A. (2010). Effects of enzyme replacement therapy in Fabry disease—A comprehensive review of the medical literature. *Genetics in Medicine*, *12*(11), 668–679.

Modrego, A., Amaranto, M., Godino, A., Mendoza, R., Barra, J. L., & Corchero, J. L. (2021). Human α-Galactosidase A mutants: Priceless tools to develop novel therapies for Fabry disease. *International Journal of Molecular Sciences*, *22*(12), 6518.

Ponomarenko, J., Bui, H. H., Li, W., Fusseder, N., Bourne, P. E., Sette, A., & Peters, B. (2008). ElliPro: A new structure-based tool for the prediction of antibody epitopes. *BMC Bioinformatics*, *9*, 514.

Shi, Z. D., Motabar, O., Goldin, E., Liu, K., Southall, N., Sidransky, E., Austin, C. P., Griffiths, G. L., & Zheng, W. (2009). Synthesis and characterization of a new fluorogenic substrate for alpha-galactosidase. *Analytical and Bioanalytical Chemistry*, *394*(7), 1903–1909.

Trott, O., & Olson, A. J. (2010). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, *31*(2), 455–461.

van der Veen, S. J., Hollak, C. E. M., van Kuilenburg, A. B. P., & Langeveld, M. (2020). Developments in the treatment of Fabry disease. *Journal of Inherited Metabolic Disease*, *43*(5), 908–921.

Wang, J., Cao, X., Chen, W., Xu, J., & Wu, B. (2021). Identification and characterization of a thermostable GH36 α-galactosidase from *Anoxybacillus vitaminiphilus* WMF1 and its application in synthesizing isofloridoside by reverse hydrolysis. *International Journal of Molecular Sciences*, *22*(19), 10778.

Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research*, *35*(suppl_2), W407–W410.