Research Square

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

In Silico Directed Evolution of Humanized Peptide Transporters via Attention-Guided Epistatic Rescue

Alper Karagöl

alper.karagol@gmail.com

Istanbul University Istanbul Faculty of Medicine https://orcid.org/0009-0001-7864-0732

Taner Karagöl

taner.karagol@gmail.com

Istanbul University Istanbul Faculty of Medicine https://orcid.org/0009-0005-1011-7661

Research Article

Keywords: in silico evolution, Live Biotherapeutic Products, integral membrane proteins, docking, peptidomimetic therapeutics

Posted Date: March 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9198655/v1

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

Additional Declarations: The authors declare no competing interests.

Abstract

The human intestinal peptide transporter 1 (SLC15A1, PepT1) is a critical determinant of oral drug bioavailability, yet its thermodynamic characterization and structural tractability remain challenging. While bacterial orthologues like the Escherichia coli DtpA transporter offer highly stable structural surrogates, significant cross-species variations in the binding microenvironment limit their pharmacological fidelity. Traditional structure-based engineering aimed at engrafting the human pharmacophore onto bacterial scaffolds is frequently hindered by combinatorial explosion and severe thermodynamic frustration, as non-native side-chains disrupt co-evolved local packing. To bypass the empirical bottlenecks of in vitro directed evolution, we present a fully in silico co-evolutionary pipeline leveraging deep contextual protein language models (PLMs). By mapping 19 pharmacologically critical human PepT1 residues onto the DtpA sequence, we utilized the 650-million parameter ESM-2 model to conduct zero-shot mutational profiling. We identified 14 primary humanizing mutations that triggered profound local destabilization (fitness probabilities $< 10 - 5$). To resolve these energetic conflicts, we deployed an automated, attention-guided epistatic rescue algorithm to sweep the flanking microenvironmental topologies. This computationally efficient heuristic successfully identified localized, non-native compensatory mutations for all high-risk targets without relying on traditional molecular dynamics. Most notably, the highly deleterious Q41R and R305F substitutions were buffered by Y38D and I304K secondary mutations, yielding 5,215-fold and 11,614-fold improvements in contextual stability, respectively. Binding dynamics is confirmed via docking calculations and all-atom membranous molecular dynamics simulations. This deep learning-driven framework successfully rationalizes the engraftment of the human binding pocket, drastically reducing the required experimental screening space and yielding a thermodynamically robust, humanized surrogate primed for advanced drug discovery.

Introduction

The human intestinal peptide transporter 1 (SLC15A1, PepT1) is a ubiquitous proton-coupled symporter that dictates the oral bioavailability and pharmacokinetic profiles of a vast array of peptidomimetics and prodrugs [1, 2, 3]. Despite its outsized role in rational drug design, the structural and thermodynamic characterization of human integral membrane proteins remains a formidable biophysical challenge [5, 6, 7, 8, 9]. Consequently, bacterial structural orthologues, such as the Escherichia coli DtpA transporter, are frequently employed as highly stable, tractable scaffolds for structural biology and high-throughput screening [10, 11]. However, divergent evolutionary trajectories have yielded significant discrepancies in the binding pocket microenvironments, limiting the translational fidelity of bacterial surrogates in human pharmacological modeling [12, 13].

Beyond resolving the immediate biophysical bottlenecks of structural pharmacology, the successful in silico humanization of the DtpA scaffold establishes a critical foundation for advanced synthetic biology applications. By physically engrafting these rationally designed chimeras into living bacterial chassis, such as probiotic Escherichia coli strains, we can pioneer novel Live Biotherapeutic Products (LBPs) [14,

15]. Such engineered commensal organisms, surrogate within the gastrointestinal tract, offer unprecedented potential for active pharmacokinetic modulation: enabling the localized sequestration, modification, or targeted delivery of orally administered peptidomimetic therapeutics directly at the intestinal epithelium [14, 15, 16, 17]. Furthermore, these functionalized biological platforms could serve as active metabolic sinks [18] in various disease states, scavenging toxic luminal peptides before systemic absorption.

The ideal solution is the rational humanization of the bacterial scaffold: engrafting the human binding pharmacophore onto the stable orthologous backbone. Yet, traditional structure-based protein engineering is frequently derailed by the combinatorial explosion of the mutational search space and severe thermodynamic frustration. The introduction of non-native, humanizing side-chains often triggers localized structural collapse, steric clashing, and disruption of epistatic networks that have co-evolved over millions of years [19, 20, 21, 22]. Resolving these deleterious primary mutations through in vitro directed evolution requires screening massive, prohibitively expensive mutant libraries [23, 24, 25].

To bypass these experimental bottlenecks, we present a fully in silico directed evolution pipeline leveraging deep contextual protein language models (PLMs). Advanced transformer architectures, such as the 650-million parameter ESM-2 [26], have internalized the fundamental grammar of protein folding by undergoing unsupervised training on hundreds of millions of cross-species sequences. By mining the high-dimensional latent space of this model, we can map zero-shot fitness landscapes and accurately predict the thermodynamic tolerability of engineered mutations without relying on computationally expensive, physics-based molecular dynamics simulations.

In this study, we structurally align the binding interfaces of human PepT1 and E. coli DtpA, identifying discrete pharmacological targets for engraftment. Anticipating the structural destabilization inherent to non-native residue substitution, we deploy an automated, attention-guided epistatic rescue algorithm. By sweeping the microenvironmental topology flanking deleterious target sites, we systematically identify secondary compensatory mutations capable of resolving energetic conflicts. To evaluate the pharmacological fidelity and structural resilience of our engineered chimeras, we integrated a computational validation pipeline. Specifically, the binding affinity and microenvironmental compatibility of the engineered central cavity were assessed via blind molecular docking with valacyclovir. Furthermore, the thermodynamic viability and dynamic alternating-access mechanics of the ligand-bound humanized scaffolds were confirmed through 50 ns all-atom molecular dynamics (MD) simulations within a native-like mixed lipid bilayer.

Herein, we demonstrate that this deep learning-driven co-evolution framework successfully rationalizes the engraftment of the human PepT1 pocket, achieving profound fold-improvements in local stability and yielding a robust, humanized structural surrogate primed for pharmacological screening and evolutionary pharmacology.

Results and Discussions

Zero-Shot Mutational Profiling Highlights Severe Local Destabilization

To engineer a humanized model of the E. coli di/tripeptide transporter (DtpA) capable of mimicking the pharmacological binding profile of human PepT1 (SLC15A1), we first established a sequence-structure correspondence. Global sequence alignment utilizing the BLOSUM62 substitution matrix yielded a basal alignment score of 209.0. This degree of homology allowed for the accurate spatial mapping of the substrate-binding pocket. For humanized variant 1 we only utilized blind structural data to select targets, determining 19 residues (Table S1-S2). For the second variant, based on docking analysis for human PepT1, 26 active-site and pore-lining residues were successfully mapped to analogous spatial coordinates within the bacterial DtpA scaffold (Table S3).

The viability of directly engrafting the human binding pocket was evaluated using the ESM-2 [26] (t33_650M_UR50D) protein language model. We conducted a zero-shot fitness assessment to quantify the thermodynamic and structural tolerability of each primary humanizing mutation. While certain positions demonstrated high evolutionary conservation and thermodynamic tolerability (e.g., E32E, R33R, and Y70Y maintained structural probabilities of ≈ 1.0), the unguided introduction of non-native human residues frequently resulted in severe fitness penalties. Specifically for humanized variant 1, six pharmacologically critical target sites exhibited profound structural destabilization, characterized by single-mutant marginal likelihoods plunging into the $10 - 5$ to $10 - 7$ range.

For our second variant, of the 26 target sites, several demonstrated absolute evolutionary conservation between DtpA and PepT1 (e.g., E33E, Y71Y, D82D, K130K), displaying near-perfect fitness probabilities. However, the substitution of structurally or electrostatically divergent human residues was highly destabilizing in the native bacterial context. We identified 14 high-risk humanizing point mutations that exhibited severe fitness penalties, falling below our predetermined structural viability threshold ($p < 0.001$). For instance, the introduction of bulky hydrophobic or altered-charge side chains, such as A151F and A337D, yielded catastrophic predicted stability scores of $7.23 \times 10 - 7$ and $3.45 \times 10 - 7$, respectively. These scores strongly indicate that naive point mutations induce severe local steric clashes or void volumes within the DtpA architecture. These precipitous drops in ESM-2 probability indicate pronounced thermodynamic frustration, likely stemming from severe local steric clashes, disruption of native side-chain packing, or the introduction of unfavorable electrostatic microenvironments.

Attention-Guided Epistatic Rescue Restores Thermodynamic Viability

To salvage the structurally deleterious humanizing mutations, we deployed an automated, spatially restricted epistatic rescue. By systematically mutating the local topology (± 3 residues) flanking each high-risk target site, the transformer's attention mechanism identified secondary, non-native substitutions capable of energetically buffering the primary humanizing mutations. The in silico directed

evolution successfully identified compensatory epistatic pairs for all six high-risk targets, significantly elevating their contextual fitness scores (Table S2). Notably, the most pronounced topological conflict, the substitution of serine with the bulky histidine at position 64 (S64H), was profoundly mitigated by the introduction of T61Y. This secondary mutation, yielded a 46.4-fold improvement in local thermodynamic stability. Similarly, the highly destabilizing R305F substitution was buffered by an adjacent I304K mutation, achieving a 22.1-fold recovery in structural probability. These results demonstrate that deep learning-guided co-evolution can successfully bypass the energetic minima associated with cross-species sequence engraftment, drastically reducing the required empirical screening space for the generation of humanized transmembrane scaffolds.

The epistatic sweep successfully identified compensatory secondary mutations for all 14 high-risk targets of variant 2, frequently yielding orders-of-magnitude improvements in predicted thermodynamic stability (Table 1). Notably, the deleterious Q41R mutation (initial score: $7.75\times10-6$) was remarkably rescued by a proximal Y38D mutation, resulting in a 5,215-fold improvement in structural viability ($p = 4.04\times10-2$). The most profound local epistatic interaction was observed for the R305F substitution; a contiguous I304K mutation recovered target stability by an exceptional 11,614-fold. Further analysis of the rescue networks revealed clear structural logic governing the compensations. For example, the severe penalty incurred by substituting a bulky tyrosine (F63Y, $p = 8.27\times10-4$) was robustly compensated by the introduction of localized backbone flexibility and reduction of local side-chain volume via F66G (878-fold improvement, resulting in a highly stable $p = 0.727$).While the majority of these epistatic pairings successfully restored viable fitness probabilities ($p > 10-2$), specific regions remained highly constrained. For example, the A285R mutation, when paired with the best local rescuer E284I, yielded an 11.4-fold improvement, yet its absolute fitness remained marginal ($3.08\times10-6$). This suggests that for hyper-constrained loci, local microenvironmental remodeling is insufficient, and broader, long-range allosteric compensatory networks are required to achieve full structural accommodation of the humanized pocket.

Table 1
Epistatic Rescue Results for Human2 Variant

Target_Site	Original_Score	Best_Rescue_Mutation	New_Score	Fold_Improvement
L31F	1.61e-04	R34Y	2.03e-01	1261.0x
Q41R	7.75e-06	Y38D	4.04e-02	5215.7x
F63Y	8.27e-04	F66G	7.27e-01	878.4x
S64H	3.10e-06	T61Y	7.02e-03	2265.5x
S67V	1.55e-04	A68P	4.54e-01	2940.0x
T87K	1.28e-05	K88P	1.78e-02	1395.2x
L98Y	3.15e-07	A95L	3.63e-05	115.3x
A151F	7.23e-07	M154R	8.22e-05	113.7x
A285R	2.71e-07	E284I	3.08e-06	11.4x
F302W	3.92e-05	A303G	6.07e-03	154.9x
R305F	3.46e-06	I304K	4.02e-02	11614.2x
N306D	1.38e-05	V307R	3.98e-02	2884.2x
L324V	2.29e-04	N325Y	2.96e-01	1290.3x
A337D	3.45e-07	A338R	1.29e-05	37.2x

To evaluate the global architectural integrity of these engineered chimeras, high-resolution structural models were generated using AlphaFold3 and superimposed against the native state (Fig. 1). Despite the extensive local microenvironmental remodeling required for epistatic rescue, the global transmembrane topology was remarkably conserved. Quantitative structural alignment revealed highly stable Root Mean Square Deviation (RMSD) values of 0.434 Å for Humanized Variant 1 and 0.697 Å for Humanized Variant 2 relative to the native DtpA scaffold (Table S4). The deviation between the two humanized variants was also minimal (RMSD = 0.615 Å). When cross-referenced against the native human PepT1 target, the baseline DtpA structure exhibited an RMSD of 1.111 Å. Notably, Humanized Variant 1 reduced this structural divergence (RMSD = 1.022 Å), indicating a spatial shift toward the human conformational state, whereas Variant 2 maintained a comparable deviation (RMSD = 1.127 Å). These metrics confirm that the computationally driven co-evolutionary pipeline successfully accommodated the humanized binding pocket without triggering the structural collapse characteristic of extensive localized mutagenesis.

Molecular Docking within the Engineered Cavity

Following the structural validation of the humanized scaffolds, we investigated the ligand-binding capacity of the engineered central cavities. Blind molecular docking was executed via the SwissDock server to sample the entire protein surface before localizing to the central pore. The resulting binding poses (Fig. 2) highlight the successful generation of an accommodating pharmacological microenvironment. In the native E. coli DtpA transporter, the ligand assumes a basal binding configuration dictated by bacterial-specific volumetric and electrostatic constraints. In contrast, within both Humanized Variant 1 and Humanized Variant 2, the engineered side-chains and their corresponding epistatic rescue networks successfully recalculate the internal cavity dimensions and polar contact geometries. The coordinated positioning of the ligand within the humanized chimeras suggests that the rationally designed structural modifications correctly emulate the steric and electrostatic binding determinants required for targeted human PepT1 interactions.

Atomistic Stability and Membrane Dynamics of Humanized Chimeras

To validate the thermodynamic viability and dynamic stability of the humanized binding pockets under physiological conditions, we subjected the native and engineered structures to all-atom molecular dynamics (MD) simulations. The structural complexes were embedded within a symmetric mixed lipid bilayer comprising cholesterol and POPC, specifically designed to mimic the anisotropic dielectric properties of a representative plasma membrane environment. Following a stringent multi-step equilibration protocol, unrestrained production MD simulations were executed for 50 ns utilizing the CHARMM36m force field, following methodology to analyze initial binding dynamics established in our previous studies [27, 28, 29].

Visual inspection of the final simulation frames revealed distinct differences in the positioning of the substrate among the three models, suggesting altered binding or translocation dynamics (Fig. 3). The native E. coli DtpA transporter failed to stably retain the substrate within the deep central pocket. Over the course of the 50 ns simulation, the substrate dissociated from the primary binding site and migrated outward toward the extracellular vestibule. Humanized Variant 1 demonstrated intermediate retention. While the substrate did not fully exit the binding cleft like in the native model, it exhibited significant mobility and shifted closer to the pore opening. Notably, Humanized Variant 2 successfully anchored the substrate. The substrate remained stably bound deep within the central cavity throughout the entire 50 ns trajectory (Fig. 3). These structural dynamics indicate that the sequence modifications introduced in Humanized Variant 2 significantly enhance the stability of the protein-substrate complex. The engineered pocket in Variant 2 provides a more favorable energetic environment for the substrate, preventing the rapid dissociation observed in the wild-type DtpA.

System stability and global architectural integrity were r evaluated through continuous trajectory analysis, primarily by calculating the root mean square deviation (RMSD) of the protein backbone and the radius of gyration (Rg). The terminal trajectory snapshots highlight robust structural conservation across all simulated systems (Fig. 3). The native E. coli DtpA transporter maintained its canonical transmembrane packing and a highly stable lipid-protein interface throughout the simulated timeframe.

Critically, both Humanized Variant 1 and Humanized Variant 2 exhibited equivalent structural resilience. Following an initial membrane relaxation phase, the backbone RMSF trajectories for the engineered chimeras is below 2A (Figure S1-S6), confirming that the introduction of 14 high-risk primary humanizing substitutions did not trigger the global unravelling or helical dissociation typically associated with extensive cross-species engraftment.

Beyond global fold preservation, localized structural dynamics were assessed by computing the root mean square fluctuation (RMSF) of C-alpha atoms. Remarkably, the transmembrane helices harboring the densest clusters of non-native mutations did not display anomalous hyper-flexibility. This strongly implies that the attention-guided compensatory mutations (e.g., the I304K secondary mutation buffering the highly deleterious R305F substitution) successfully re-established tightly packed local microenvironments. Furthermore, calculations of the solvent accessible surface area (SASA) demonstrated that the hydrophobic core of the transporter remained effectively shielded (Figure S7-S9). The deep learning-directed substitutions did not introduce severe internal void volumes or energetic conflicts that would otherwise lead to aberrant water penetration and subsequent structural collapse.

Finally, the dynamic behavior of the bound ligand within the central cavity was monitored to evaluate continuous pharmacological engagement. Throughout the 50 ns production phase, the ligand remained stably coordinated within the engineered binding pockets of second variant, without uncoupling from the pore. The persistent association of the ligand within the highly remodeled cavities confirms that the in silico directed evolution pipeline effectively translated the static human PepT1 binding determinants into the dynamic, alternating-access mechanism of the bacterial scaffold, yielding a thermodynamically robust and pharmacologically faithful surrogate.

Translational Implications: From In Silico Design to In Vivo Chassis

The successful computational engineering of the DtpA scaffold establishes a critical foundation for in vivo synthetic biology applications. To physicalize these humanized chimeras, the optimized sequences of Variant 1 and Variant 2 can be synthesized and cloned into inducible expression vectors. The choice of bacterial chassis dictates the translational trajectory: high-yield expression strains, such as E. coli BL21(DE3) or C43(DE3), provide an optimized environment for mitigating membrane protein toxicity during large-scale production for structural biology. Conversely, deploying these plasmids within commensal or probiotic strains, such as E. coli Nissle 1917, bridges the gap between static structural models and dynamic, living therapeutic platforms.

Clinical Deployment and Pharmacokinetic Modulation

Introducing engineered bacteria that functionally express a humanized PepT1 surrogate into the human gastrointestinal tract represents a highly disruptive approach to Live Biotherapeutic Products (LBPs). Because native gut microbiota inherently influence systemic drug metabolism, an engineered commensal strain expressing a high-affinity human transport surrogate could actively modulate local

pharmacokinetics [14, 15, 16]. This platform could be leveraged to precisely sequester, enzymatically modify, or achieve targeted release of orally administered peptidomimetic drugs and prodrugs directly at the intestinal epithelium. Furthermore, these functionalized bacterial chassis offer novel avenues for biosensing and metabolic scavenging. In disease states characterized by the accumulation of toxic peptide byproducts or inflammatory biomarkers, the humanized DtpA could function as an active molecular sink, internalizing these substrates from the lumen before they breach the epithelial barrier.

Conclusion

The fundamental challenge in rational drug design for integral membrane proteins lies in the balance between structural tractability and pharmacological fidelity [30, 31, 32]. While bacterial orthologues offer exceptional stability for in vitro screening, their divergent binding pockets frequently yield false-negative or translationally irrelevant interaction profiles [12, 13]. To fully translate this in silico co-evolutionary pipeline into clinical utility, extensive empirical validation remains imperative. The remarkable thermodynamic stability imparted by the epistatic rescue networks presents an ideal opportunity for high-resolution structural validation. Advanced cryogenic electron microscopy (Cryo-EM) or co-crystallization studies of the humanized surrogates bound to novel prodrug candidates will be essential to definitively confirm the predictive accuracy of the ESM-2 and AlphaFold3 pipelines.

This study demonstrates that deep contextual protein language models can systematically resolve the thermodynamic frustration associated with cross-species pocket engraftment. By employing an attention-guided epistatic rescue algorithm, we bypassed the combinatorial explosion that strictly limits traditional directed evolution, identifying highly non-intuitive secondary mutations that rescued catastrophically deleterious primary substitutions (e.g., the 11,614-fold stabilization of R305F via I304K).

The resulting chimeras not only preserve the highly stable bacterial structural backbone but dynamically recapitulate the human PepT1 binding microenvironment. This deep learning-driven framework establishes a highly tractable, thermodynamically robust surrogate primed for advanced structural biology and high-throughput screening of novel prodrugs. Furthermore, the in silico evolutionary pipeline detailed herein presents a universal, scalable strategy for the rational humanization of other pharmacologically critical but structurally intractable transport proteins and receptors.

Methods

Global Sequence Alignment and Orthologous Pharmacophore Mapping

To identify structural orthologues of the human SLC15A1 (PepT1) binding pocket within the Escherichia coli DtpA scaffold, global pairwise sequence alignment was conducted using a dynamic programming approach. The Needleman-Wunsch algorithm was executed via the Biopython pairwise alignment module [33], utilizing the BLOSUM62 substitution matrix to account for evolutionary divergence. Gap open and gap extension penalties were stringently set to -10 and -0.5, respectively, to maintain strict structural register. Possible pharmacologically critical residues defining the human PepT1 binding

interface were subsequently mapped onto the DtpA coordinate space, establishing a precise residue-by-residue correspondence for targeted humanization. These 19 defined positional coordinates were subsequently mapped onto the DtpA coordinate space. Residues possibly responsible for anchoring the polar dipoles of canonical substrates were identified. These included the primary acidic anchor (Glu594) and its co-anchor (Asn328) for N-terminal amine coordination, alongside basic residues (Arg27, Arg34, and Lys140). To target the transporter's intrinsic peptide-bond recognition machinery, the central tyrosine triad (Tyr64, Tyr91, Tyr167) and anamide-binding residue (Asn171) were isolated. Residues defining the volumetric and lipophilic constraints of the binding cavity were prioritized for altering substrate specificity. This targeted subset comprised the aromatic/hydrophobic core elements: Phe24, Tyr56, Phe162, Phe166, Trp294, and Phe297. Residues that cpould be essential for the proton-coupled alternating-access mechanism (the rocker-switch dynamic) were mapped to ensure that structural modifications would not uncouple transport activity. This included proton sensor (His57), the intra-helical salt-bridge gate (Arg282, Asp341), and the extracellular vestibular gate (Glu26). The finalized index array, corresponding strictly to the P46059 sequence, is as follows: 24, 26, 27, 34, 56, 57, 64, 91, 140, 162, 166, 167, 171, 282, 294, 297, 328, 341, and 594.

For humanized variant 2, docking data (with Valacyclovir as ligand) were incorporated to the target selection. The SMILES representation is used: CC(C)[C@@H](N)C(=O)OCCOCN1C=NC2=C1N=C(N)NC2=O. Native structure was generated using AlphaFold3 [34]. Molecular docking was executed via the SwissDock web server [35,36]. We performed a blind docking procedure to sample the entire protein surface and the central cavity. The selection of the optimal starting configuration was based on a SwissDock scoreing and three canditaes were selected. The residues within 4 Angstrom of the ligand were detected using Pymol 3. Accordingly; 31, 60, 63, 75, 80, 144, and 298 were included to the finalized array.

Zero-Shot Mutational Inference via Deep Contextual Protein Language Models

To evaluate the thermodynamic and structural tolerability of the proposed humanizing mutations, we deployed ESM-2 (t33_650M_UR50D), a state-of-the-art, transformer-based protein language model (PLM) comprising 33 attention layers and 650 million parameters. Computational analyses utilized a Google Colab instances, equipped with an NVIDIA g4 GPU, 96 GB of VRAM, 176GB of RAM, and Intel® Xeon® CPUs. Unlike traditional physics-based scoring functions, ESM-2 captures complex biophysical constraints and deep evolutionary grammar entirely unsupervised from billions of cross-species protein sequences [26]. For each single-point humanizing substitution, the zero-shot mutational fitness was quantified by performing a forward pass of the mutated sequence through the model. The raw output logits were normalized via a softmax activation function. This generated a probability distribution over the amino acid vocabulary at the target site. The marginal likelihood of the introduced human residue was extracted as a predictive proxy for folding stability and local structural compatibility.

In Silico Directed Evolution via Attention-Guided Epistatic Rescue

Primary humanizing mutations that exhibited sub-optimal zero-shot fitness probabilities (indicative of local structural destabilization, steric clashes, or thermodynamic frustration) were subjected to a highly targeted in silico directed evolution protocol. To circumvent the computationally prohibitive landscape of global combinatorial mutagenesis, we designed a spatially restricted epistatic rescue sweep leveraging the bidirectional context of the ESM-2 architecture [26].

This heuristic search operated on the hypothesis that localized secondary mutations can resolve folding conflicts introduced by primary non-native substitutions via compensatory epistasis. We defined a microenvironmental search space using a sliding window of ±3 residues flanking each structurally deleterious primary mutation. Within this defined local topology, we executed exhaustive computational saturation mutagenesis; every position within the window was systematically mutated to all 19 alternative amino acids in the persistent background of the primary humanizing mutation.

The ESM-2 attention mechanism continuously re-evaluated these transient double-mutant states, implicitly calculating the non-linear side-chain couplings and backbone shifts required for stabilization [26]. The objective function for this evolutionary sweep was the optimization of the compensatory fitness differential, calculated as the fold-improvement of the primary mutation's probability relative to its single-mutant baseline. Secondary substitutions that successfully maximized this quotient and bypassed the local energy minima were classified as epistatic rescuers. This computationally guided co-evolution yields a rationally optimized, mutationally robust scaffold, drastically reducing the empirical screening burden required for in vitro synthesis.

Molecular Docking and Ligand Architecture

To rigorously evaluate the pharmacological fidelity and structural resilience of our engineered chimeras, we integrated a robust computational validation pipeline. Specifically, the binding affinity and microenvironmental compatibility of the engineered central cavity were assessed via blind molecular docking with valacyclovir—a canonical, high-affinity human PepT1 prodrug substrate—utilizing its exact SMILES string for precise conformational ligand generation. Molecular docking was executed via the SwissDock web server [35,36]. We performed a blind docking procedure to sample the entire protein surface and the central cavity. The selection of the optimal starting configuration was based on a combination of the AC score.

Atomistic all-atom membrane systems

According to the methodology explained in our previous studies [27,28,29,37,38], all-atom membrane systems were constructed using the CHARMM-GUI Membrane Builder [39,40,41]. Protein spatial alignment within the lipid bilayers was optimized via the PPM 2.0 algorithm, which incorporates hydrogen bonding and anisotropic dielectric properties at the membrane-water interface [42]. The lipid composition was designed to mimic a representative plasma membrane, consistent with our previous models [27,28,29,37]. The membrane contained cholesterol and POPC distributed symmetrically across both leaflets. System neutrality was established by adding K^+ and Cl^- ions to an ionic strength of 0.15 M,

determined through 2,000 steps of Monte Carlo simulation using a primitive ion model. The CHARMM36m all-atom force field was applied throughout [43].

Molecular dynamics simulations

Molecular dynamics simulations were executed on AlphaFold-predicted membrane structures using GROMACS 2024.3 [44], adhering to protocols established in our prior studies on glutamate transporter variants and neuronal proteins [27,28,29]. Computational analyses utilized a cluster of three Google Colab instances, equipped with an NVIDIA L4 GPU, 126 GB of VRAM, 318 GB of RAM, and Intel® Xeon® CPUs [38]. To maximize efficiency, the software was recompiled with CUDA support, enabling parallelization across multiple cores [38]. Energy minimization was performed using the steepest descent algorithm for a maximum of 5,000 steps. Convergence was defined by a maximum force tolerance of 1000 kJ/mol/nm. Equilibration followed the standard six-step CHARMM-GUI protocol, characterized by the gradual release of restraints. The final equilibration step consisted of a 500 ps run utilizing the v-rescale thermostat to maintain a temperature of 303.15K (τ t =1.0ps) and the C-rescale barostat to maintain a pressure of 1bar via semi-isotropic coupling (τp=5.0ps). Production MD simulations were carried out for 50 ns with frames saved every 0.5 ns, consistent with our earlier studies [28,29]. Electrostatic interactions were computed via the Particle Mesh Ewald (PME) method, applying a 1.2 nm cutoff for both Coulomb and van der Waals terms. The system temperature (303.15 K) and pressure (1 bar) were maintained using the Nose-Hoover thermostat and the Parrinello-Rahman barostat under semi-isotropic coupling. Hydrogen bonds were constrained using the LINCS algorithm, allowing for a 1.2 nm cutoff for short-range van der Waals and electrostatic interactions.

System stability was evaluated through trajectory analysis. The radius of gyration (Rg) were calculated using the GROMACS rms and gyrate modules, respectively. Residue flexibility was assessed by computing the root mean square fluctuation (RMSF) of C-alpha atoms. Solvent accessible surface area (SASA) for protein side chains was determined using the gmx sasa tool with a probe radius of 1.4 Å [45]. For binding free energy estimations, the final 20 ns (31-50 ns) of the equilibrated trajectories were analyzed. Structures were visualized usiong UCSF ChimeraX [46].

Declarations

Ethics Approval

Ethics approval was not required for this computational study as it did not involve animal subjects, human participants, and identifiable data.

Consent to participate

Not applicable. This computational study did not involve human participants.

Consent for publication

Not applicable. This computational study did not involve human participants.

Availability of data and materials

The AlphaFold DB (https://alphafold.ebi.ac.uk), a database developed by DeepMind and the European Bioinformatics Institute (EMBL-EBI) at EMBL, is a repository for AlphaFold2 predictions, with over 200 million protein structures. Each statistical and computational analysis of this study, included with step-by-step instructions where possible, are publicly available to ensure repeatability. For more detailed information on the statistical analyses, input files and detailed outputs, including the AlphaFold calculations and codes to regenerate analyses, please visit the website: https://github.com/karagol-alper/Dtpa-evolution

Competing financial interests

None.

Funding

The author(s) received no specific funding for this work.

Author Contributions

A.K and T.K contributed equally to this work. All authors have read and agreed to the published version of the manuscript.

Acknowledgements

None.

References

1. Roy T, Nath M, Halder N, Saxena R, Velpandian T (2025) Functional and Physiological Implications of Oligopeptide Transporters: Potential Targets for Pharmacological Interventions. J Membr Biol 258(5):375–390. https://doi.org/10.1007/s00232-025-00348-1
2. Jordan JM, Ersoy BA (2025) Drug Transporters. The Pathophysiologic Basis of Drug Therapy, Principles of Pharmacology
3. Smith DE, Clémençon B, Hediger MA (2013) Proton-coupled oligopeptide transporter family SLC15: physiological, pharmacological and pathological implications. Mol Aspects Med 34(2–3):323–336. https://doi.org/10.1016/j.mam.2012.11.003
4. Ye J, Chen H, Wang K, Wang Y, Ammerman A, Awasthi S, Xu J, Liu B, Li W (2024) Structural insights into vesicular monoamine storage and drug interactions. Nature 629(8010):235–243. https://doi.org/10.1038/s41586-024-07290-7

5. Lv X, Liu J, Shi Q, Tan Q, Wu D, Skinner JJ, Walker AL, Zhao L, Gu X, Chen N, Xue L, Si P, Zhang L, Wang Z, Katritch V, Liu ZJ, Stevens RC (2016) In vitro expression and analysis of the 826 human G protein-coupled receptors. Protein Cell 7(5):325–337. https://doi.org/10.1007/s13238-016-0263-8
6. Carpenter EP, Beis K, Cameron AD, Iwata S (2008) Overcoming the challenges of membrane protein crystallography. Curr Opin Struct Biol 18(5):581–586. https://doi.org/10.1016/j.sbi.2008.07.001
7. Zhang S, Tao F, Qing R, Tang H, Skuhersky M, Corin K, Maertens B (2018) QTY code enables design of detergent-free chemokine receptors that retain ligand-binding activities. Proceedings of the National Academy of Sciences, 115(37), E8652-E8659. https://doi.org/10.1073/pnas.1811031115
8. Karagöl T, Karagöl A, Zhang S (2024) Structural bioinformatics studies of serotonin, dopamine and norepinephrine transporters and their AlphaFold2 predicted water-soluble QTY variants and uncovering the natural mutations of L-> Q, I-> T, F-> Y and Q-> L, T-> I and Y-> F. PLoS ONE 19(3):e0300340. https://doi.org/10.1371/journal.pone.0300340
9. Pan E, Tao F, Smorodina E, Zhang S (2024) Structural bioinformatics studies of six human ABC transporters and their AlphaFold2-predicted water-soluble QTY variants. QRB discovery 5:e1. https://doi.org/10.1017/qrd.2024.2.pr1
10. Killer M, Finocchio G, Mertens HD, Svergun DI, Pardon E, Steyaert J, Löw C (2022) Cryo-EM structure of an atypical proton-coupled peptide transporter: di-and tripeptide permease C. Front Mol Biosci 9:917725. https://doi.org/10.3389/fmolb.2022.917725
11. Stauffer M, Jeckelmann JM, Ilgü H, Ucurum Z, Boggavarapu R, Fotiadis D (2022) Peptide transporter structure reveals binding and action mechanism of a potent PEPT1 and PEPT2 inhibitor. Commun Chem 5(1):23. https://doi.org/10.1038/s42004-022-00636-0
12. Bulitta JB, Hope WW, Eakin AE, Guina T, Tam VH, Louie A, Hoover JL (2019) Generating robust and informative nonclinical in vitro and in vivo bacterial infection model efficacy data to support translation to humans. Antimicrob Agents Chemother 63(5):10–1128. https://doi.org/10.1128/aac.02307-18
13. Ayyildiz M, Celiker S, Ozhelvaci F, Akten ED (2020) Identification of alternative allosteric sites in glycolytic enzymes for potential use as species-specific drug targets. Front Mol Biosci 7:88. https://doi.org/10.3389/fmolb.2020.00088
14. Heavey MK, Durmusoglu D, Crook N, Anselmo AC (2022) Discovery and delivery strategies for engineered live biotherapeutic products. Trends Biotechnol 40(3):354–369. https://doi.org/10.1016/j.tibtech.2021.08.002
15. Sadhu S, Paul T, Yadav N (2025) Therapeutic engineering of the gut microbiome using synthetic biology and metabolic tools: a comprehensive review with E. coli Nissle 1917 as a model case study. Arch Microbiol 207(9):213. https://doi.org/10.1007/s00203-025-04417-w
16. Ou B, Yang Y, Tham WL, Chen L, Guo J, Zhu G (2016) Genetic engineering of probiotic Escherichia coli Nissle 1917 for clinical application. Appl Microbiol Biotechnol 100(20):8693–8699. https://doi.org/10.1007/s00253-016-7829-5

17. Lynch JP, Goers L, Lesser CF (2022) Emerging strategies for engineering Escherichia coli Nissle 1917-based therapeutics. Trends Pharmacol Sci 43(9):772–786. https://doi.org/10.1016/j.tips.2022.02.002
18. Clare C, Rutter JW, Fedorec AJ, Frank S, Barnes CP (2024) Bacterial microcompartment utilization in the human commensal Escherichia coli Nissle 1917. J Bacteriol 206(12):e00269–e00224. https://doi.org/10.1128/jb.00269-24
19. Ose NJ, Campitelli P, Patel RP, Ozkan SB, Kumar S (2022) Protein dynamics provide mechanistic insights about the epistatic relationships among highly observed potentially adaptive missense variants. Biophys J 121(3):456a. https://doi.org/10.1016/j.bpj.2021.11.488
20. Buda K, Miton CM, Fan XC, Tokuriki N (2023) Molecular determinants of protein evolvability. Trends Biochem Sci 48(9):751–760. https://doi.org/10.1016/j.tibs.2023.05.009
21. Tharp CR, Catalano C, Khalifeh A, Ghaffari-Kashani S, Huang R, Kang G, Phillips AM (2026) Biophysical trade-offs in antibody evolution are resolved by conformation-mediated epistasis. https://doi.org/10.64898/2026.03.12.711465. bioRxiv, 2026-03
22. Karagöl T, Karagöl A, Zhang S (2025) Co-evolution of alpha-helical transmembrane protein residues: large-scale variant profiling and complete mutational landscape of 2277 known PDB entries representing 504 unique human protein sequences. J Mol Evol 93(5):581–599. https://doi.org/10.1007/s00239-025-10262-8
23. Chen S, Yang Z, Zhong Z, Yu S, Zhou J, Li J, Zhang G (2024) Ultrahigh-throughput screening-assisted in vivo directed evolution for enzyme engineering. Biotechnol Biofuels Bioprod 17(1):9. https://doi.org/10.1186/s13068-024-02457-w
24. McLure RJ, Radford SE, Brockwell DJ (2022) High-throughput directed evolution: a golden era for protein science. Trends Chem 4(5):378–391. https://doi.org/10.1016/j.trechm.2022.02.004
25. Wang Y, Xue P, Cao M, Yu T, Lane ST, Zhao H (2021) Directed evolution: methodologies and applications. Chem Rev 121(20):12384–12444. https://doi.org/10.1021/acs.chemrev.1c00260
26. Lin Z, Akin H, Rao R, Hie B, Zhu Z, Lu W, Rives A (2023) Evolutionary-scale prediction of atomic-level protein structure with a language model. Science 379(6637):1123–1130. https://doi.org/10.1126/science.ade2574
27. Johnsson F, Karagöl T, Karagöl A, Zhang S (2025) Structural bioinformatic study of six human olfactory receptors and their AlphaFold3 predicted water-soluble QTY variants and OR1A2 with an odorant octanoate and TAAR9 with spermidine. QRB discovery 6:e2. https://doi.org/10.1017/qrd.2024.18
28. Karagöl A, Karagöl T, Zhang S (2024) Molecular dynamic simulations reveal that water-soluble QTY-Variants of glutamate transporters EAA1, EAA2 and EAA3 retain the conformational characteristics of native transporters. Pharm Res 41(10):1965–1977. https://doi.org/10.1007/s11095-024-03769-0
29. Karagöl A, Karagöl T, Li M, Zhang S (2024) Inhibitory Potential of the Truncated Isoforms on Glutamate Transporter Oligomerization Identified by Computational Analysis of Gene-Centric Isoform Maps. Pharm Res 41(11):2173–2187. https://doi.org/10.1007/s11095-024-03786-z

30. Garibsingh RAA, Schlessinger A (2019) Advances and challenges in rational drug design for SLCs. Trends Pharmacol Sci 40(10):790–800. https://doi.org/10.1016/j.tips.2019.08.006
31. Li H, Sun X, Cui W, Xu M, Dong J, Ekundayo BE, Vogel H (2024) Computational drug development for membrane protein targets. Nat Biotechnol 42(2):229–242. https://doi.org/10.1038/s41587-023-01987-2
32. Das N, Singh S, Swaminathan P (2025) Rational Drug Designing for Antimicrobial Resistance: New Strategies and Targets. Curr Pharmacol Rep 11(1):44. https://doi.org/10.1007/s40495-025-00424-z
33. Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, De Hoon MJ (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics 25(11):1422. https://doi.org/10.1093/bioinformatics/btp163
34. Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore L, Ballard AJ, Bambrick J, Bodenstein SW, Evans DA, Hung CC, O'Neill M, Reiman D, Tunyasuvunakool K, Wu Z, Žemgulytė A, Arvaniti E, Beattie C, Jumper JM (2024) Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature 630(8016):493–500. https://doi.org/10.1038/s41586-024-07487-w
35. Grosdidier A, Zoete V, Michielin O (2011) SwissDock, a protein-small molecule docking web service based on EADock DSS. Nucleic acids research, 39(Web Server issue). W270–W277. https://doi.org/10.1093/nar/gkr366
36. Bugnon M, Röhrig UF, Goullieux M, Perez MAS, Daina A, Michielin O, Zoete V (2024) SwissDock 2024: major enhancements for small-molecule docking with Attracting Cavities and AutoDock Vina. Nucleic Acids Res 52:W324–W332. https://doi.org/10.1093/nar/gkae300
37. Sajeev-Sheeja A, Karagöl A, Karagöl T, Zhang S (2025) Molecular dynamics simulations and structural bioinformatics of bacterial integral alpha-helical membrane enzymes and their AlphaFold2-predicted water-soluble QTY analogues. Mol Simul 51(15):984–998. https://doi.org/10.1080/08927022.2025.2562932 4
38. Karagöl T, Karagöl A (2024) Benchmarking GROMACS on Optimized Colab Processors and the Flexibility of Cloud Computing for Molecular Dynamics. bioRxiv. 2024–2011. https://doi.org/10.1101/2024.11.14.623563
39. Wu EL, Cheng X, Jo S, Rui H, Song KC, Dávila-Contreras EM, Qi Y, Lee J, Monje-Galvan V, Venable RM, Klauda JB, Im W (2014) CHARMM-GUI Membrane Builder toward realistic biological membrane simulations. J Comput Chem 35(27):1997–2004. https://doi.org/10.1002/jcc.23702
40. Jo S, Kim T, Iyer VG, Im W (2008) CHARMM-GUI: a web-based graphical user interface for CHARMM. J Comput Chem 29(11):1859–1865. https://doi.org/10.1002/jcc.20945
41. Jo S, Lim JB, Klauda JB, Im W (2009) CHARMM-GUI Membrane Builder for mixed bilayers and its application to yeast membranes. Biophys J 97(1):50–58. https://doi.org/10.1016/j.bpj.2009.04.013
42. Lomize MA, Pogozheva ID, Joo H, Mosberg HI, Lomize AL (2012) OPM database and PPM web server: resources for positioning of proteins in membranes. Nucleic Acids Res 40. https://doi.org/10.1093/nar/gkr703. D370-D376

43. Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, De Groot BL, Grubmüller H, MacKerell AD Jr (2017) CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nat Methods 14(1):71–73. https://doi.org/10.1038/nmeth.4067
44. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1:19–25. https://doi.org/10.1016/j.softx.2015.06.001
45. Eisenhaber F, Lijnzaad P, Argos P, Sander C, Scharf M (1995) The double cubic lattice method: Efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies. J Comput Chem 16(3):273–284. https://doi.org/10.1002/jcc.540160303
46. Meng EC, Goddard TD, Pettersen EF, Couch GS, Pearson ZJ, Morris JH, Ferrin TE (2023) UCSF ChimeraX: Tools for structure building and analysis. Protein Sci 32(11):e4792. https://doi.org/10.1002/pro.4792

Figures

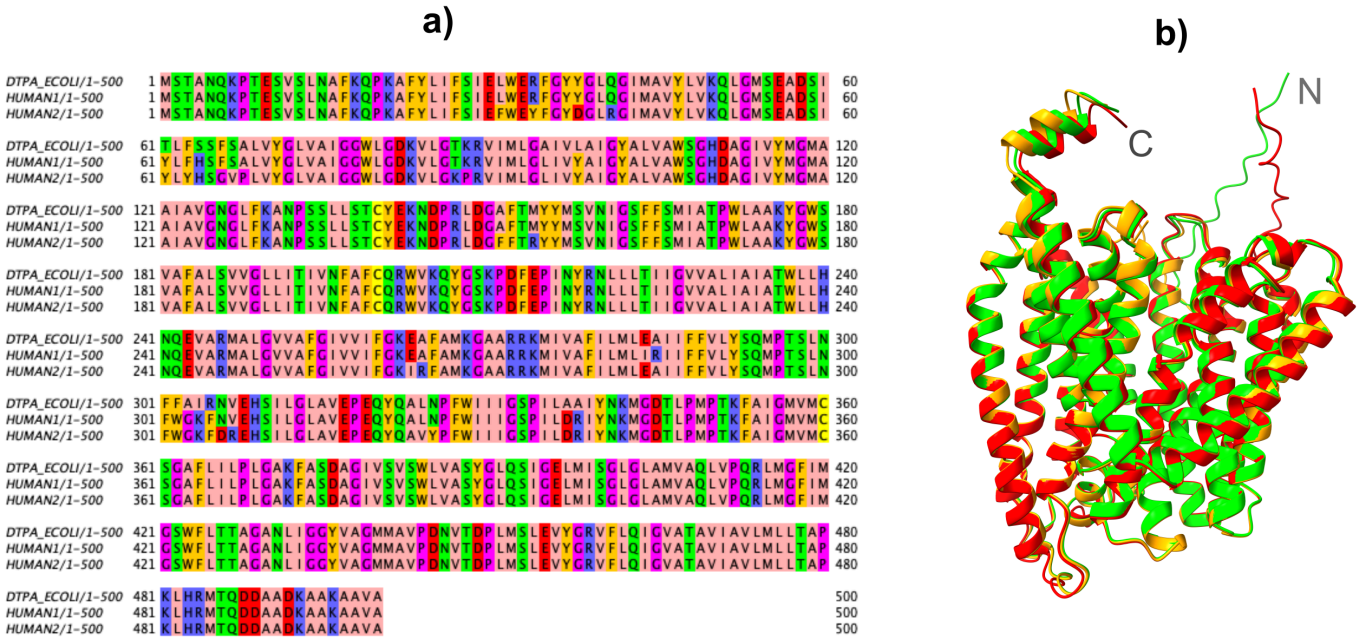


Figure 1

Sequence and structural comparison of native E. coli DtpA and engineered humanized variants. a) Global sequence alignment of the native E. coli DtpA transporter and the engineered humanized models, Humanized Variant 1 and Humanized Variant 2. The alignment highlights the targeted sequence modifications designed to engraft the human PepT1 binding pharmacophore onto the stable orthologous bacterial backbone. =**b)** Structural superimposition of the native DtpA and the engineered humanized chimeras, generated using AlphaFold3. The superimposed 3D models demonstrate the

preservation of the overall transmembrane architecture following the introduction of non-native primary humanizing substitutions and their corresponding attention-guided epistatic rescue mutations. Quantitative structural comparison confirms this scaffold stability, with Humanized Variant 1 and Humanized Variant 2 displaying low Root Mean Square Deviation (RMSD) values of 0.434 Å and 0.697 Å, respectively, relative to the native DtpA structure. The structural deviation between the two humanized variants is also minimal (RMSD = 0.615 Å).

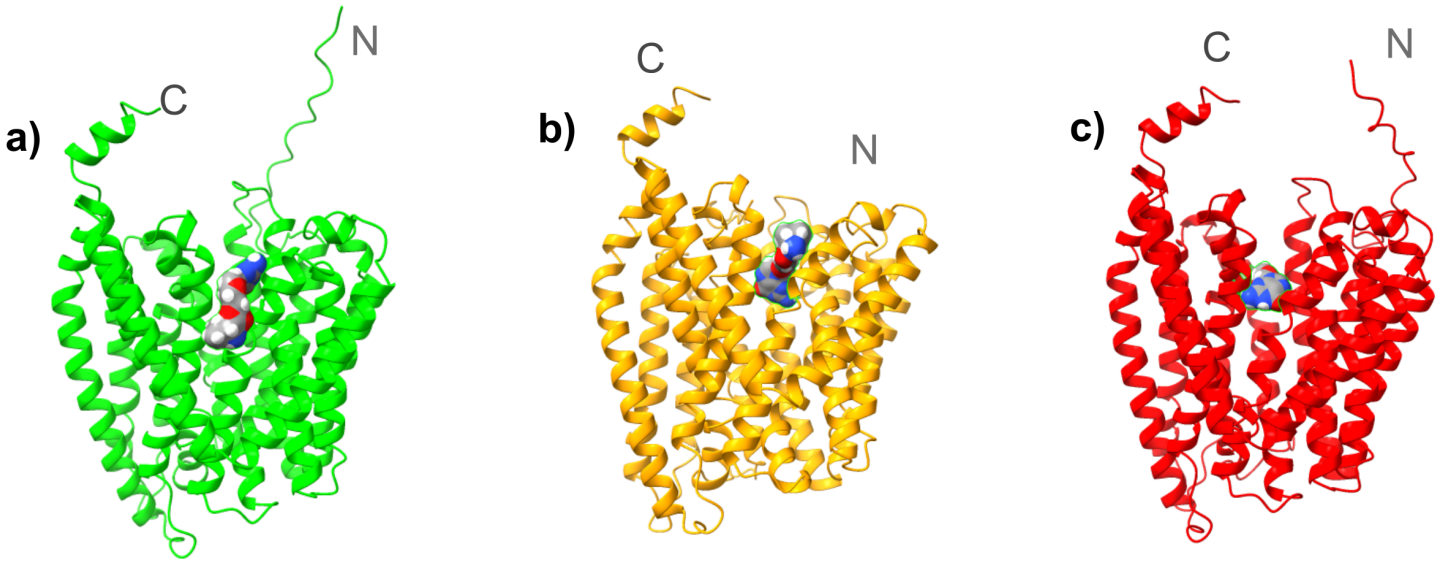


Figure 2

Molecular docking predictions within the central cavities of native and humanized transporter models. **a)** Blind docking configuration of the ligand within the central cavity of the native E. coli DtpA transporter (green). **b**) Docked ligand configuration within the engineered Humanized Variant 1 (yellow). **c)** Docked ligand configuration within the engineered Humanized Variant 2 (red). Molecular docking for all structural models was executed via the SwissDock web server.

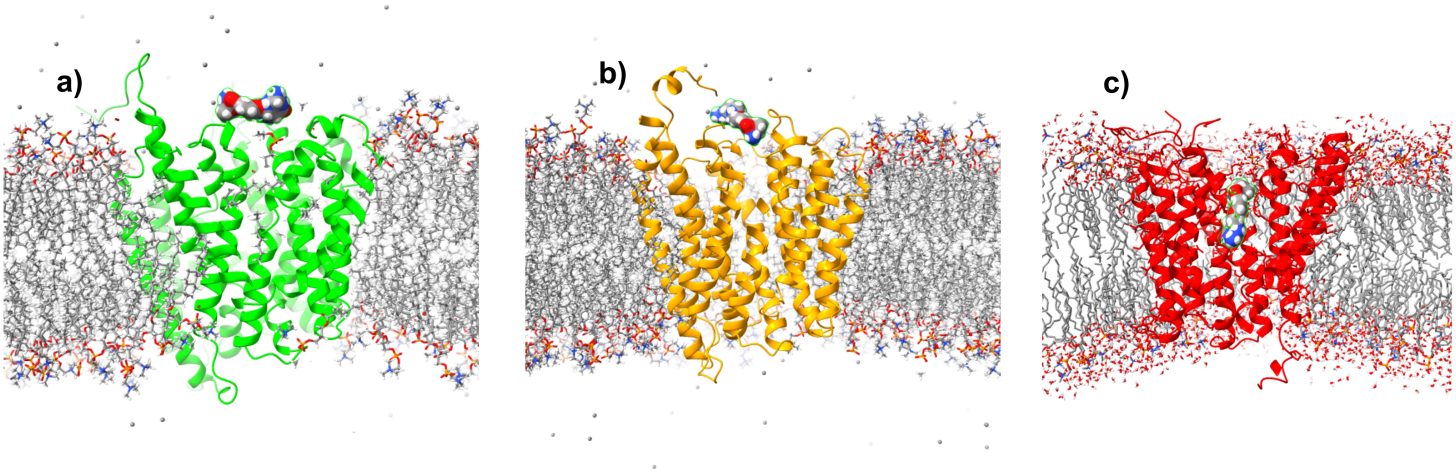


Figure 3

Final structural snapshots following 50 ns of all-atom molecular dynamics simulations. All-atom membrane systems representing the transporter models embedded within a lipid bilayer , shown after 50 ns of equilibrated production MD. **a)** The native E. coli DtpA transporter (green) within the simulated membrane environment. **b)** The engineered Humanized Variant 1 (yellow) after 50 ns of simulation. **c)** The engineered Humanized Variant 2 (red) after 50 ns of simulation.

Supplementary Files

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

- SupplementaryFile.pdf