

Resurrection of *Anopheles darlingi* FREP1 Ancestor Reveals Adaptive Evolution Characterized by Changes in Protein Stability and *Plasmodium falciparum* Interaction

Krishnendu Sinha^{1*}

¹Department of Zoology, Jhargram Raj College, Jhargram-721507

***Corresponding author**

Dr. Krishnendu Sinha
Assistant Professor in Zoology
Jhargram Raj College
Jhargram-721507
dr.krishnendusinha@gmail.com

Abstract:

Fibrinogen-related protein 1 (FREP1), a midgut-localized fibrinogen-like lectin in *Anopheles* mosquitoes, mediates *Plasmodium* ookinete attachment by binding α -tubulin-1. To elucidate the evolutionary forces shaping this interaction, the study analysed FREP1 sequences from 29 *Anopheles* species using codon-based tests, ancestral sequence reconstruction, stability modelling, and docking. Both aBSREL and branch-site codeml identified the *Anopheles darlingi* lineage as the sole branch experiencing episodic diversifying selection, with MEME detecting a single positively selected site (codon 173/residue 218) within the FBG domain. Ancestral reconstruction revealed a Ser to Asn substitution at this position in the extant protein. Although Rosetta $\Delta\Delta G$ analyses indicated only modest local effects of this substitution, docking showed that extant *A. darlingi* FREP1 exhibits markedly reduced predicted binding affinity to *Plasmodium falciparum* α -tubulin-1 relative to its reconstructed ancestor. Complementary cophylogenetic analyses (PACo and ParaFit) identified significant global phylogenetic congruence between *Anopheles* and *Plasmodium* lineages, consistent with broad lineage-level structuring of vector-parasite compatibility rather than strict co-speciation. Within this evolutionary backdrop, the lineage-specific adaptive change at residue 218 suggests fine-scale molecular tuning of the FREP1 ligand-binding interface, potentially reflecting an arms-race-like dynamic in which *A. darlingi* FREP1 functionally diverged in response to historical parasite pressures despite the absence of tight tree-wide coevolution.

Keywords: *Anopheles darlingi*, FREP1, episodic selection, aBSREL, ancestral reconstruction, codeml, α -tubulin-1, *Plasmodium falciparum*, ookinete

1.INTRODUCTION

Malaria transmission relies on tightly coordinated interactions between *Plasmodium* parasites and their *Anopheles* vectors [1], particularly during ookinete traversal of the peritrophic matrix and midgut epithelium, a process governed by molecular compatibility between parasite ligands and vector midgut factors. Among these, fibrinogen-related protein 1 (FREP1) has emerged as a key determinant of vector competence, functioning as a midgut-expressed receptor that binds *Plasmodium* ookinetes and facilitates their invasion [2–8]. FREP1 belongs to the invertebrate fibrinogen-related (FReD) family and is defined by a conserved C-terminal fibrinogen-like (FBG) domain, preceded by an N-terminal signal peptide and a variable glycosylated region. The FBG domain forms a structured β -sheet core with exposed loops that mediate carbohydrate and protein interactions, enabling FREP1's dual roles in innate immune recognition and parasite binding. Recent structural and biochemical evidence indicates that *Plasmodium* ookinetes hijack specific FBG-domain surface motifs—most notably those interacting with α -tubulin-1, to secure attachment to the peritrophic matrix and penetrate the midgut barrier. Perturbation of this binding interface substantially reduces infection, underscoring the FBG domain as a critical molecular nexus in the establishment of malaria within mosquito vectors [3,4,6].

Although functional studies have demonstrated that FREP1 is essential for parasite binding[3,6], considerably less is known about how this gene has evolved across mosquito lineages, whether its protein-protein interaction properties have diverged, and whether individual amino acid changes reflect adaptive responses to parasite pressures. FREP1 is particularly interesting in this regard because the *Plasmodium* ligand responsible for midgut recognition, α -tubulin-1 exposed on the ookinete surface, is highly conserved within the parasite and interacts with specific regions of the FREP1 FBG domain[5–7]. The evolutionary history of FREP1 therefore has the potential to illuminate long-term coevolution between mosquitoes and the parasites they transmit.

Despite its functional importance, no systematic analysis has previously examined patterns of molecular evolution, lineage-specific adaptive changes, or the structural consequences of such changes across FREP1 orthologs in *Anopheles*. Questions remain as to whether FREP1 is evolving under purifying selection due to functional constraint, or whether certain lineages exhibit signatures of positive selection indicative of parasite-driven adaptation. Furthermore, any structural and biochemical outcomes of adaptive amino acid substitutions, particularly

those occurring within the FBG domain responsible for parasite binding, remain unresolved. Understanding how evolutionary forces shape FREP1 structure and binding properties may provide mechanistic insights into lineage-specific differences in parasite susceptibility.

This study has addressed these knowledge gaps by integrating comparative genomics, codon-based evolutionary modeling, ancestral sequence reconstruction, protein structural prediction, stability estimation, and protein–protein docking. This study first identifies 29 unique FREP1 orthologs across 29 *Anopheles* species and reconstruct a robust maximum likelihood phylogeny. Using multiple complementary molecular evolution frameworks, including aBSREL, branch-site codeml, and MEME, the study detects a single episode of diversifying selection along the *Anopheles darlingi* lineage, pinpointing a specific residue within the FBG domain. Then the ancestral FREP1 sequence for the lineage leading to *A. darlingi* was reconstructed and examined the functional consequences of the derived amino acid state using Rosetta-based stability analysis and HADDOCK–PRODIGY binding predictions with the *P. falciparum* α -tubulin-1 ligand.

Together, these complementary approaches provide a mechanistic view of how adaptive evolution has shaped FREP1 in *A. darlingi* lineage, revealing the structural and functional implications of a positively selected residue within a key parasite-interacting domain. This integrative framework establishes a foundation for understanding the evolutionary dynamics of mosquito midgut factors and their influence on malaria transmission.

2. MATERIALS AND METHODS

2.1. Retrieval of FREP1 sequence and ortholog identification

The protein sequence of FREP1 from *Anopheles gambiae* PEST (VectorBase gene ID: AGAP006914) was retrieved from VectorBase[9]. These 280 amino-acid sequences served as the query for ortholog identification across *Anopheles* clade and the outgroup, *Culex quinquefasciatus* (VectorBase gene ID: CPIJ000937). Complete proteomes for 29 *Anopheles* and one *Culex* species available in VectorBase (release 68, accessed November 2025) were downloaded. Orthologs were identified using a reciprocal BLAST hit (RBH) workflow implemented through an in-house Python script implementing soft masking and Smith–Waterman alignments[10] with an E-value threshold of 1×10^{-5} , and a minimum alignment coverage of 90%. Twenty-nine candidate sequences returning *A. gambiae* FREP1 as the top reciprocal match were accepted as true orthologs[11]. All sequences were further validated for

the presence of the fibrinogen-related (FBG) domain using InterProScan[12], and each ortholog was confirmed to contain a canonical FBG region. Corresponding coding sequences (CDS) for all confirmed orthologs were retrieved from VectorBase using gene-level identifiers.

2.2. *Multiple sequence alignment and trimming*

Protein sequences were aligned using PRANK v.170427 [13] with the codon-aware settings (default parameters) to preserve evolutionary signal and minimize gap misplacement. The resulting protein alignment was used to generate a codon-preserving nucleotide alignment via PAL2NAL v.14[14], producing an alignment of 1686 nucleotide positions. Both the protein and codon alignments were refined using ClipKIT v1.3 [15] with the kpic-smart-gap mode. For the codon alignment, ClipKIT produced a trimmed alignment of 747 positions, removing 55.69% of sites while preserving informative residues for evolutionary analysis.

2.3. *Phylogenetic inference*

A maximum likelihood phylogeny of FREP1 protein orthologs was reconstructed using IQ-TREE3 v3.0.1[16]. ModelFinder [17] identified LG+I+G4 as the best-fitting amino acid substitution model under the Bayesian information criterion. Tree reconstruction included 1000 ultrafast bootstrap replicates and 1000 SH-aLRT tests[18,19]. The final ML tree contained 29 taxa and 274 amino-acid sites, with 241 parsimony-informative sites. The tree was used as the fixed topology for all downstream molecular evolutionary analyses[19].

2.4. *Detection of positive selection*

Episodic diversifying selection across branches was assessed using the adaptive Branch-Site Random Effects Likelihood (aBSREL) model [20] implemented in HyPhy v2.5[21]. The trimmed codon alignment and the ML tree were provided as input. Branch-specific likelihood ratio tests identified the *A. darlingi* FREP1 lineage (ADAR2_011252) as the only branch with significant evidence of episodic diversification ($p = 0.0$ after correction). To further investigate codon-specific selective pressures, a branch-site test in codeml (PAML v4.10.9) [22,23] was performed using the *A. darlingi* branch as the foreground. Bayes Empirical Bayes (BEB) analysis identified several sites with elevated posterior probability, including codon positions 51 (PP = 0.969) and 226 (PP = 0.842). Site-level episodic selection was tested using MEME [21] holding *A. darlingi* FREP1 as foreground, which detected codon 173 (CDS alignment) as significantly evolving under episodic selection (LRT = 4.63; $p = 0.05$). Mapping this site using

in-house python script revealed that it corresponded to residue 218 in *A. darlingi* and residue 191 in the ancestral node.

2.5. Ancestral sequence reconstruction

Ancestral sequence reconstruction (ASR) was performed using IQ-TREE v3.0.1 to infer the historical amino-acid states of FREP1 across the *Anopheles* phylogeny. The analysis used the same maximum-likelihood (ML) protein phylogeny that was previously inferred from the 29-sequence, 274-amino-acid alignment, along with the best-fit substitution model (LG+I+G4) selected by ModelFinder. IQ-TREE's ASR procedure estimates, for every internal node and every alignment position, the most likely ancestral amino acid and its associated posterior probability, based on the fixed tree topology, branch lengths, and substitution model. The internal node representing the most recent common ancestor of *A. darlingi* and its sister taxon *A. aquasalis* was identified from the labeled ML tree and designated "Node 26" following the software's node indexing. For each node, IQ-TREE provides reconstructed amino-acid sequences that can be exported as standard FASTA files using in house python script.

To determine the ancestral state of the positively selected site, the codon identified by MEME (codon 173 in the CDS alignment) was mapped to its corresponding position in the ungapped protein sequence. After accounting for alignment gaps and restoring original residue numbering, this site corresponded to amino-acid position 218 in the *A. darlingi* FREP1 protein. Examination of the reconstructed Node 26 sequence showed that the corresponding position was occupied by a serine (S). In contrast, the extant *A. darlingi* sequence contains an asparagine (N) at the same position, indicating that the S→N substitution occurred along the *A. darlingi* lineage after divergence from *A. aquasalis*. The full ancestral sequence of Node 26 was used for all subsequent structural modeling, stability estimation, and protein–protein docking analyses, enabling direct comparison between the reconstructed ancestral state and the modern *A. darlingi* FREP1 protein.

2.6. Protein structure modeling

Three-dimensional structures of the extant *A. darlingi* FREP1, the ancestral Node 26 variant, and the engineered N218S back-mutation were generated using the AlphaFold 3 prediction server[24]. For the N218S variant, the amino acid substitution was introduced manually using AliView [25] followed by de novo structure prediction. All predicted structures were used as starting models for docking and stability simulations.

2.7. Rosetta-based stability estimation

Protein stability and the energetic effects of individual substitution were estimated using PyRosetta (Rosetta v2025)[26]. A standardized pipeline was applied to all variants using a Python script that performed 50 independent FastRelax replicates per sequence. Each replicate consisted of structure relaxation using the `fa_scorefxn` scoring function (full-atom score function) followed by calculation of Rosetta Energy Units (REU). $\Delta\Delta G$ values were computed as the difference between mutant and wild-type energies. This protocol produced stability profiles for the extant *A. darlingi* FREP1, the N218S mutant, and the Node 26 ancestor.

2.8. Protein–protein docking with α -tubulin-1

Protein–protein docking between FREP1 variants and *Plasmodium falciparum* α -tubulin-1 (UniProt Q6ZLZ9) was performed using HADDOCK 2.4[27]. Active residues within FREP1 were defined as those in the FBG domain (positions 92–302) based on InterProScan annotation. The experimentally mapped α -tubulin-1 linear epitope REDLAALEKD (residues 422–431) [28] as the core active site in HADDOCK docking, and expanded this region to residues 419–434 to allow for flanking contacts (passive residues 412–418 and 435–440 were auto-assigned/added). Passive residues were assigned automatically[27]. Docking was run using default parameters, and the resulting structures were clustered based on interface RMSD. Ten clusters were produced for each variant, each containing four water-refined models. Although the entire ensemble was used for binding energy assessment, the top-scoring HADDOCK cluster per variant was used for reporting docking statistics.

2.9. Binding affinity prediction

Binding free energies (ΔG) for all docked complexes were estimated using PRODIGY v2.1 [29,30] installed locally through the Conda Bioconda distribution. For each variant, all 40 structures (10 clusters $\times$ 4 models) were processed independently using default temperature (25°C) through inhouse python script. The resulting distributions of predicted ΔG values were compared using two-tailed t-tests, after confirming normality with the Shapiro–Wilk test, implemented in an in-house R script.

2.10. Anopheles-Plasmodium coevolution through PACo and ParaFit

Phylogenetic trees for 29 *Anopheles* species and 12 *Plasmodium* species were obtained in Newick format from published genomic resources and converted into patristic distance

matrices using `cophenetic()` in the `ape` package (R v4.5.2). A natural host–parasite association matrix was constructed using only documented field infections and confirmed vector–parasite pairings, yielding a sparse 28×11 binary matrix in which taxa with no associations were removed to produce a final working matrix containing 20 *Anopheles* hosts, 5 *Plasmodium* parasites, and 5 confirmed natural links. Cophylogenetic congruence was evaluated using the Procrustean Approach to Cophylogeny (PACo) implemented in the `paco` package, with Cailliez correction applied to ensure Euclidean distance matrices and significance assessed using 10,000 permutations. To complement PACo, we applied ParaFit using the `ade4` package, again using Cailliez-corrected host and parasite distance matrices and 9,999 permutations to obtain ParaFitGlobal and ParaFitLink statistics. Both analyses were performed on the trimmed natural matrix to ensure compatibility with the underlying algorithms.

3. Results

3.1. Ortholog identification and alignment

Using the *A. gambiae* FREP1 reference sequence (VectorBase ID AGAP006914; 280 amino acids), a total of 29 putative unique FREP1 orthologs were identified across available *Anopheles* species from VectorBase (release 68) and 1 FREP1 ortholog from the *Culex* outgroup. Domain validation with InterProScan indicated that each ortholog possessed an intact FBG domain characteristic of the FREP1 family. Protein sequences were aligned with PRANK, and the corresponding codon alignment was generated using PAL2NAL. After trimming ambiguous regions using ClipKit with the `kpic-smart-gap` model, the curated protein alignment contained 274 residues and the nucleotide alignment retained 747 codon sites (reduced from an initial 1,686 sites).

3.2. Phylogenetic reconstruction

Maximum likelihood phylogenetic inference using IQ-TREE3 selected the LG+I+G4 amino acid substitution model based on BIC. The resulting phylogeny ($\ln L = -9199.247$) exhibited strong support across major nodes, with both SH-aLRT and ultrafast bootstrap values generally exceeding 95%. The topology was largely congruent with established relationships among *Anopheles* species and placed *A. darlingi* in a New World clade alongside *A. aquasalis* (table 1; figure 1). IQ-TREE identified one very short internal branch (<0.0036), suggesting

uncertainty in a localized region of the tree, although the overall topology was stable and consistent with the consensus tree (Robinson–Foulds distance = 0).

3.3. Branch-level selection analysis

To detect episodic diversifying selection, aBSREL was applied to all branches of the phylogeny. Significant evidence of episodic positive selection was detected exclusively on the *A. darlingi* FREP1 branch ($p = 0$, reported as zero due to extremely small values). This branch exhibited heterogeneity in selective pressures, with approximately 76.4% of sites evolving under neutrality ($\omega = 1$) and 23.6% experiencing intense episodic positive selection ($\omega = 57.25$). The mean ω across sites on this branch was 14.26 (CoV = 1.67), reflecting substantial variability in selective constraints. No other branches showed statistically significant evidence of episodic selection.

3.4. Site-level selection analysis

MEME identified a single codon under episodic diversifying selection. Codon 173 (in the gapped CDS alignment) exhibited significant support for episodic selection (LRT = 4.63, $p = 0.05$), with a high estimated nonsynonymous rate ($\beta^+ \approx 2747.13$) in the positively selected rate class. In parallel, codeml branch-site analysis in PAML supported positive selection on the *A. darlingi* branch. The alternative model allowing $\omega_2 > 1$ produced a significantly higher likelihood ($\ln L = -14542.21619$) compared with the null model with ω_2 fixed at 1 ($\ln L = -14545.57582$). The resulting likelihood-ratio test (LRT = 6.72, $df = 1$) was significant ($p = 0.0095$), confirming the presence of diversifying selection on the focal lineage. Although codeml's branch-site BEB analysis identified codon 51 with high posterior probability (PP > 0.95), the codeml and MEME tests evaluate different aspects of selection even when applied to the same foreground branch. The codeml branch-site model infers site-level probabilities indirectly through its mixture of site classes, which favors sites that experienced more sustained shifts in ω on the foreground branch. In contrast, MEME tests each site independently and is sensitive to short episodic bursts of positive selection, allowing detection of sites where only a brief interval of accelerated nonsynonymous change occurred. Codon 173 was selected for downstream analysis because it was the site supported at both the branch level (aBSREL) and the site level (MEME) specifically on the *A. darlingi* lineage. Moreover, codon 173 lies within the FREP1 FBG binding domain, directly involved in parasite–vector interactions, making it the most functionally plausible target of adaptive change.

3.5. Ancestral sequence reconstruction

To place the selected residue change in an evolutionary context, ancestral sequence reconstruction was performed using IQ-TREE3 under the LG+I+G4 model and the inferred ML topology. The ancestral node shared by *A. darlingi* and its sister taxon *A. aquasalis*, hereafter referred to as Node 26, was reconstructed at high confidence across the FBG region. At the site corresponding to codon 173, the ancestral amino acid state was inferred to be serine (S), which corresponds to residue 191 in the ungapped Node 26 sequence. In the extant *A. darlingi* FREP1, the homologous position corresponds to residue 218 and carries an asparagine (N). This substitution (S → N) represents the derived state in *A. darlingi* and was the focus of subsequent structural modelling.

3.6. Structural modelling and stability analyses

Three-dimensional structural models of the extant *A. darlingi* FREP1, the N218S mutant, and the reconstructed Node 26 ancestral sequence were generated using AlphaFold3, and all displayed high-confidence predictions across the FBG domain. To quantify changes in protein stability associated with lineage-specific or site-specific substitutions, Rosetta all-atom relax and $\Delta\Delta G$ calculations were performed using 50 independent replicate scoring cycles for each comparison. For the N218S point mutation, Rosetta $\Delta\Delta G$ estimates showed substantial replicate-to-replicate variability, with values spanning both stabilizing and destabilizing directions (table 2; figure 2). Consistent with this dispersion, the Wilcoxon rank-sum test comparing WT and N218S energies was significant ($p = 0.00136$). Mean energies for the WT and N218S models were -910.12 REU and -906.07 REU, respectively, indicating a moderate net destabilizing trend, but with considerable stochastic overlap among simulations. These results indicate that the N218S substitution perturbs local structural stability rather than inducing a large-scale change in the global fold of FREP1. In functional terms, this suggests that the positively selected substitution at residue 218 modulates localized stability within the FBG binding domain, consistent with a fine-tuning adaptive mechanism rather than a disruptive structural shift. In contrast, comparison of the WT structure with the Node 26 ancestral model revealed a dramatic and highly consistent stability shift. The ancestral models exhibited uniformly higher (less favourable) Rosetta energies, with the Wilcoxon rank-sum test yielding an extreme result ($p = 7.07 \times 10^{-18}$). The mean Rosetta energy for Node 26 was -672.90 REU compared with -909.88 REU for the WT, reflecting a massive loss of structural stability in the reconstructed ancestor. Given that this comparison involves many substitutions accumulated

along the *A. darlingi* lineage, not a single site, the magnitude of this Δ is best interpreted as the cumulative effect of broad sequence and backbone divergence, rather than the influence of any individual amino-acid replacement.

Together, these analyses indicate that the modern *A. darlingi* FREP1 protein is substantially more stable than its resurrected ancestor, whereas the derived N218S substitution itself causes a real but modest destabilizing effect. This pattern is consistent with the hypothesis that stabilization of FREP1 preceded or accompanied adaptive functional refinement in the *A. darlingi* lineage, with the N218S replacement contributing a comparatively fine-tuned adjustment rather than a major structural shift.

3.7. Docking with *Plasmodium falciparum* α -tubulin-1 and binding affinity prediction

To evaluate whether the S to N substitution influences interactions with the parasite ligand, docking simulations were performed between FREP1 variants and *P. falciparum* α -tubulin-1 using HADDOCK 2.4. Ten clusters containing four models each were generated for every FREP1 variant, and the most representative clusters were selected based on HADDOCK scoring. The best cluster for the wild-type *A. darlingi* FREP1 (Cluster 8) exhibited a HADDOCK score of 125.7 ± 22.8 and an RMSD of 1.3 ± 1.4 Å. The N218S mutant (Cluster 1) showed a similar HADDOCK score of 127.6 ± 4.0 but substantially higher RMSD (22.3 ± 0.4 Å), suggesting greater structural variability among the top-scoring complexes. In contrast, the Node 26 ancestral FREP1 (Cluster 4) displayed a markedly more favorable HADDOCK score of 62.5 ± 10.6 and a low RMSD of 3.2 ± 0.1 Å, indicating stronger and more stable predicted binding than either the WT or N218S complexes. The Z-score for the Node 26 cluster (-2.2) was also more favorable than those of WT (-1.7) or N218S (-1.9) (table 3; Figure 4).

Binding free energy predictions from PRODIGY were generated for all 40 models per variant (10 clusters $\times$ 4 models each). Mean predicted ΔG values were -12.16 kcal $\cdot$ mol $^{-1}$ for WT, -13.87 kcal $\cdot$ mol $^{-1}$ for Node 26, and approximately -12.5 kcal $\cdot$ mol $^{-1}$ for the N218S mutant. Statistical comparison of WT and Node 26 binding affinities showed a highly significant difference ($t = 7.136$, $df \approx 77.22$, $p = 4.52 \times 10^{-10}$), with Node 26 binding more strongly on average (table 4; figure 3). Differences between WT and N218S were not statistically significant, consistent with the similar HADDOCK scores and the greater structural variability observed for the mutant complexes.

3.8. Significant cophylogenetic congruence

PACo revealed significant global cophylogenetic congruence between the *Anopheles* and *Plasmodium* phylogenies based on natural host–parasite associations, with a PACoGlobal p-value of 0.0386 and a Procrustes sum of squares of 78.2 across 10,000 permutations, indicating that the observed mosquito–parasite associations deviate from random expectations. ParaFit similarly supported a significant global evolutionary relationship, with ParaFitGlobal yielding a p-value of 0.00537 after Cailliez correction. However, both methods failed to produce meaningful link-level statistics: PACo did not return usable residuals due to the extremely sparse structure of the association matrix, and ParaFitLink returned NULL because no parasite infected more than one host and no host carried multiple parasites within the natural dataset. Thus, while both global tests showed significant phylogenetic dependence, neither method could identify individual host–parasite pairs as primary contributors to the signal.

4. Discussion

Fibrinogen-related proteins (FREPs) are central mediators of mosquito–parasite interactions, yet their evolutionary dynamics and structural consequences remain poorly understood across *Anopheles* lineages. This study reveals that FREP1 in *A. darlingi* has undergone a pronounced episode of diversifying selection, highlighted by a single amino acid replacement (S → N) at a site identified independently by aBSREL, MEME, and branch-site codeml analyses. This substitution defines a derived state unique to *A. darlingi* relative to its reconstructed ancestor (Node 26), suggesting a lineage-specific adaptive event potentially linked to parasite recognition or midgut invasion.

The ancestral sequence reconstruction indicated that the ancestral FREP1 variant possessed serine at the positively selected codon, whereas extant *A. darlingi* carries an asparagine. Structural analyses revealed that the ancestral FREP1 was substantially less stable than the derived form. Rosetta-based $\Delta\Delta G$ calculations showed very large positive stability differences for the WT versus Node 26 comparison, reflecting global structural divergence and indicating

that acquisition of the derived state occurred on a background already undergoing broader conformational stabilization. Although single-site stability estimates must be interpreted with caution when structural models differ substantially, the pattern is clear: extant *A. darlingi* FREP1 is markedly more stable than its inferred ancestor, and this stabilization likely occurred through multiple coordinated substitutions, one of which was under episodic diversifying selection.

Conversely, the N218S back-mutation introduced into the extant protein produced only modest and variable destabilization, with mean $\Delta\Delta G$ values near +4–6 REU and substantial replicate variance. These results indicate that the selected residue contributes to fine-scale stability modulation but is not solely responsible for the profound stability difference between the ancestral and extant forms. The selected site therefore represents a potential adaptive refinement rather than the primary determinant of global structural stabilization.

The functional consequences of the substitution were examined through docking with *Plasmodium falciparum* α -tubulin-1, the principal known ookinete-exposed ligand for FREP1. Docking simulations demonstrated that the ancestral FREP1 sequence formed a more stable and tightly bound complex than the extant *A. darlingi* FREP1. Node 26 consistently exhibited lower (more favorable) HADDOCK scores, lower RMSD values, and greater buried surface area relative to the extant protein. PRODIGY binding affinity predictions across 40 models per variant confirmed that Node 26 binds α -tubulin-1 significantly more strongly than the WT, with a highly significant difference between the two distributions, whereas the N218S mutation did not significantly alter binding strength relative to WT. Taken together, these findings suggest that the evolutionary transition from the ancestral serine to the extant asparagine coincided with reduced ligand-binding affinity, despite overall protein stabilization.

The combined evolutionary, structural, and binding analyses point toward a subtle coevolutionary adjustment of FREP1 in *A. darlingi* within an ongoing mosquito–*Plasmodium* arms race. The reconstructed Node 26 ancestor exhibits markedly stronger predicted affinity for the ookinete ligand α -tubulin-1, consistent with a scenario in which *Plasmodium* more effectively hijacked ancestral FREP1 to anchor to the peritrophic matrix during midgut traversal. In contrast, the extant *A. darlingi* FREP1 shows significantly reduced binding affinity, suggesting that natural selection on the mosquito may have incrementally weakened this parasite-exploited interface. The derived S→N replacement at residue 218, detected as episodically selected and located within the functional FBG binding domain, introduces only

modest and localized destabilization, aligning with a model of fine-scale tuning rather than wholesale restructuring of the protein. Together, these results support the interpretation that FREP1 has undergone nuanced evolutionary refinement, wherein a single adaptive substitution contributes to a measurable shift in parasite binding potential, reflecting the molecular footprint of antagonistic coevolution between *A. darlingi* and *Plasmodium* ookinetes.

It is important to note that docking and binding energy predictions are *in silico* approximations and may not fully capture *in vivo* interaction dynamics influenced by glycosylation, midgut pH, or conformational flexibility. Nonetheless, the congruence between evolutionary signals, biophysical stability estimates, and binding predictions strengthens the interpretation that this residue represents a meaningful evolutionary modification. Experimental validation using recombinant FREP1 variants, midgut binding assays, or α -tubulin pull-down experiments would provide further mechanistic insights.

The significant PACoGlobal and ParaFitGlobal values demonstrate that natural Anopheles–Plasmodium associations exhibit non-random phylogenetic structure, suggesting that vector–parasite compatibility is influenced by evolutionary constraints, conserved physiological traits, and ecological filtering rather than occurring by chance. The inability of PACo and ParaFitLink to resolve individual coevolving pairs reflects the highly specialized and sparsely connected nature of natural malaria transmission networks, in which most Plasmodium species infect only a single or very few Anopheles vectors. Thus, the global phylogenetic signal indicates broad lineage-level structuring rather than strict co-speciation or strong pairwise coevolution. Importantly, this macroevolutionary pattern does not conflict with lineage-specific molecular adaptations, such as the observed strong binding between ancestral An. darlingi FREP1 and P. falciparum α -tubulin, which likely represents localized molecular arms-race dynamics rather than deep reciprocal coevolution across entire clades.

5. Conclusion

This study provides a comprehensive evolutionary and structural framework for understanding FREP1 adaptation in *Anopheles*. The positively selected N218S residue in *A. darlingi* FREP1 represents an adaptive modification that affects both protein stability and parasite-binding affinity. Because the cophylogenetic analyses show only broad lineage-level congruence, and molecular selection occurs exclusively on the *A. darlingi* branch, the FREP1– α -tubulin interaction evolved through a localized arms-race event driven by lineage-specific parasite pressures rather than tree-wide reciprocal coevolution. These findings enhance our

understanding of vector–parasite coevolution and may aid in identifying molecular determinants influencing malaria transmission.

6. Funding details

This research received no funding.

7. Conflicts of interest

The author report that there are no competing interests to declare.

8. Data availability statement

All data associated with this study are available on Figshare (DOI: 10.6084/m9.figshare.30618974). The materials provided include both raw and processed datasets, as well as all scripts required to reproduce the analyses described in the manuscript. Supplementary files contained within the repository and the manuscript ensure full transparency and reproducibility of the work.

9. Acknowledgments

I thank the global open-access community for their commitment to freely accessible science, without which this study would not have been possible. I gratefully acknowledge VectorBase for providing high-quality genomic resources essential to this work. I also recognize the contributions of openly available language models that assisted in refining the manuscript. Finally, I am deeply grateful to my wife, Dr. Nabanita Ghosh, Assistant Professor of Zoology at Maulana Azad College, Kolkata, whose insightful discussions helped shape the central idea of this study.

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

During the preparation of this work the author(s) used ChatGPT in order to refine this manuscript's readability. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

11. References

1. Daily, J. P. & Parikh, S. Malaria. *New England Journal of Medicine* **392**, 1320–1333 (2025).

2. Shaw, W. R., Marcenac, P. & Catteruccia, F. Plasmodium development in Anopheles: a tale of shared resources. *Trends Parasitol* **38**, 124 (2021).
3. Zhang, G. *et al.* Anopheles Midgut FREP1 Mediates Plasmodium Invasion. *J Biol Chem* **290**, 16490 (2015).
4. Niu, G. *et al.* The fibrinogen-like domain of FREP1 protein is a broad-spectrum malaria transmission-blocking vaccine antigen. *Journal of Biological Chemistry* **292**, 11960–11969 (2017).
5. Zhang, G. *et al.* Targeting plasmodium α -tubulin-1 to block malaria transmission to mosquitoes. *Front Cell Infect Microbiol* **13**, 1132647 (2023).
6. Zhang, G., Niu, G., Perez, L., Wang, X. & Li, J. Malaria transmission assisted by interaction between Plasmodium α -tubulin-1 and Anopheles FREP1 protein. *bioRxiv* 2019.12.16.878082 (2019) doi:10.1101/2019.12.16.878082.
7. Hanington, P. C. & Zhang, S. M. The Primary Role of Fibrinogen-Related Proteins in Invertebrates Is Defense, Not Coagulation. *J Innate Immun* **3**, 17 (2010).
8. Molina-Cruz, A. *et al.* Some strains of Plasmodium falciparum, a human malaria parasite, evade the complement-like system of Anopheles gambiae mosquitoes. *Proc Natl Acad Sci U S A* **109**, (2012).
9. Giraldo-Calderón, G. I. *et al.* VectorBase: an updated bioinformatics resource for invertebrate vectors and other organisms related with human diseases. *Nucleic Acids Res* **43**, D707 (2014).
10. Moreno-Hagelsieb, G. & Latimer, K. Choosing BLAST options for better detection of orthologs as reciprocal best hits. *Bioinformatics* **24**, 319–324 (2008).
11. Hernández-Salmerón, J. E. & Moreno-Hagelsieb, G. Progress in quickly finding orthologs as reciprocal best hits: comparing blast, last, diamond and MMseqs2. *BMC Genomics* **2020** 21:1 21, 741- (2020).
12. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
13. Löytynoja, A. Phylogeny-Aware Alignment with PRANK and PAGAN. *Methods Mol Biol* **2231**, 17–37 (2021).
14. Suyama, M., Torrents, D. & Bork, P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res* **34**, W609 (2006).
15. Steenwyk, J. L., Buida, T. J., Li, Y., Shen, X. X. & Rokas, A. ClipKIT: A multiple sequence alignment trimming software for accurate phylogenomic inference. *PLoS Biol* **18**, e3001007 (2020).
16. Wong, T. K. F. *et al.* IQ-TREE 3: Phylogenomic Inference Software using Complex Evolutionary Models. <https://doi.org/10.32942/X2P62N> (2025) doi:10.32942/X2P62N.
17. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A. & Jermini, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* **2017** 14:6 14, 587–589 (2017).

18. Sinha, K. Episodic Positive Selection Signatures in Arabidopsis CONSTANS-like Genes COL3 and COL5 Indicating Adaptive Evolution in Red-Light Signaling Pathways. *bioRxiv* 2025.04.03.646976 (2025) doi:10.1101/2025.04.03.646976.
19. Sinha, K., Jana, S., Pramanik, P. & Bera, B. Selection on synonymous codon usage in soybean (Glycine max) WRKY genes. *Scientific Reports* 2024 14:1 **14**, 26530- (2024).
20. Smith, M. D. *et al.* Less Is More: An Adaptive Branch-Site Random Effects Model for Efficient Detection of Episodic Diversifying Selection. *Mol Biol Evol* **32**, 1342 (2015).
21. Kosakovsky Pond, S. L. *et al.* HyPhy 2.5—A Customizable Platform for Evolutionary Hypothesis Testing Using Phylogenies. *Mol Biol Evol* **37**, 295 (2019).
22. Yang, Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol Biol Evol* **24**, 1586–1591 (2007).
23. Álvarez-Carretero, S., Kapli, P. & Yang, Z. Beginner’s Guide on the Use of PAML to Detect Positive Selection. *Mol Biol Evol* **40**, (2023).
24. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* 2021 596:7873 **596**, 583–589 (2021).
25. Larsson, A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* **30**, 3276–3278 (2014).
26. Sora, V. *et al.* RosettaDDGPrediction for high-throughput mutational scans: From stability to binding. *Protein Sci* **32**, e4527 (2023).
27. Honorato, R. V. *et al.* The HADDOCK2.4 web server for integrative modeling of biomolecular complexes. *Nature Protocols* 2024 19:11 **19**, 3219–3241 (2024).
28. Zhang, G. *et al.* Targeting plasmodium α -tubulin-1 to block malaria transmission to mosquitoes. *Front Cell Infect Microbiol* **13**, 1132647 (2023).
29. Xue, L. C., Rodrigues, J. P., Kastitis, P. L., Bonvin, A. M. & Vangone, A. PRODIGY: a web server for predicting the binding affinity of protein–protein complexes. *Bioinformatics* **32**, 3676–3678 (2016).
30. Vangone, A. & Bonvin, A. M. J. J. Contacts-based prediction of binding affinity in protein–protein complexes. *Elife* **4**, (2015).

12. Tables

Table 1. Summary of molecular evolution analyses identifying episodic and site-specific positive selection in *Anopheles* FREP1 orthologs. This table presents the results of aBSREL, the branch-site codeml model, and MEME analysis. The aBSREL test detected a single branch under episodic diversifying selection, corresponding to *A. darlingi* FREP1 (ADAR2_011252_R18153). The branch-site model similarly supported positive selection on this lineage, with one codon site (*) showing elevated posterior probabilities ($\text{BEB} \geq 0.95$). MEME analysis identified codon 173 (corresponding to residue 218 in *A. darlingi* and residue 191 in Node 26) as evolving under episodic positive selection with a significant LRT.

The combined results support the presence of a lineage-specific adaptive substitution along the *A. darlingi* branch.

Analysis	Key Result	Statistical Support	Notes
aBSREL	<i>A. darlingi</i> branch under episodic diversifying selection	$p = 0.0$; $\omega_2 = 57.25$ at 23.6% sites	Strong evidence for branch-specific selection
Branch-site codeml	Several sites under selection in <i>A. darlingi</i> foreground	LRT = 6.72; $p = 0.01$ BEB $\geq 0.969^*$ at codon 51; BEB ≥ 0.842 at codon 226; moderate support at others	Confirms aBSREL signal
MEME	Codon 173 under episodic selection	LRT = 4.63; $p = 0.05$	Selected residue maps to FREP1 FBG domain

Table 2. Rosetta $\Delta\Delta G$ stability results for extant *A. darlingi* FREP1, the N218S mutant, and the reconstructed Node 26 ancestor.

Rosetta FastRelax was performed with ten independent replicates to estimate the energetic consequences of the selected substitution. The extant *A. darlingi* FREP1 exhibited more favorable energies than the Node 26 ancestor in all replicates, whereas the N218S mutation introduced into the extant protein produced moderate and variable destabilization.

Variant	Mean $\Delta\Delta G$ (REU) $\pm$ SD	Interpretation	Notes
N218S (relative to WT)	3.93 ± 7.04	Mild destabilization	Highly variable between replicates
Node 26 ancestor (relative to WT)	237.62 ± 14.68	Strong destabilization	Indicates globally less stable ancestral fold

Table 3. HADDOCK docking statistics for FREP1 variants interacting with *Plasmodium falciparum* α -tubulin-1.

This table summarizes the best-scoring docked clusters for each FREP1 variant. The Node 26 ancestral sequence exhibited the most favorable docking energetics, lower RMSD, and greater buried surface area compared to the WT and N218S forms. These patterns are consistent with stronger and more stable binding in the ancestral FREP1-tubulin complex.

Variant	Best Cluster	HADDOCK Score	Cluster Size	RMSD (Å)	Buried Surface Area (Å ²)	Z-score
WT	Cluster 8	125.7 ± 22.8	5	1.3 ± 1.4	2202.0 ± 98.0	-1.7
Node 26 ancestor	Cluster 4	62.5 ± 10.6	22	3.2 ± 0.1	2834.4 ± 38.4	-2.2
N218S mutant	Cluster 1	127.6 ± 4.0	23	22.3 ± 0.4	2194.8 ± 29.6	-1.9

Table 4. PRODIGY binding free energy predictions (ΔG) for the interaction between FREP1 variants and *P. falciparum* α -tubulin-1.

Binding energies were computed for all clusters (40 structures per variant). Node 26 exhibited significantly stronger binding than WT, whereas N218S did not differ significantly from WT.

Comparison	Mean ΔG (kcal·mol ⁻¹)	Difference	Statistical Test	p-value	Interpretation
WT vs Node 26	-12.16 vs -13.87	+1.71 kcal·mol ⁻¹	Mann-Whitney U	4.52×10^{-10}	Node 26 binds significantly more strongly
WT vs N218S	-12.16 vs -12.50	+0.34 kcal·mol ⁻¹	Mann-Whitney U	ns	No significant effect

13. Figures

Figure 1. Maximum likelihood phylogeny of FREP1 across 29 *Anopheles* species. The phylogeny was inferred using IQ-TREE3 under the LG+I+G4 substitution model, with 1000 ultrafast bootstrap replicates and 1000 SH-aLRT tests. Branch lengths represent amino-acid substitutions per site. The tree is rooted using the *Culex quinquefasciatus* FREP1 ortholog (VectorBase gene ID: CPIJ000937), which served as the designated outgroup. The *A. darlingi* lineage (ADAR2_011252_R18153) is highlighted as the only branch exhibiting significant evidence of episodic diversifying selection in the aBSREL analysis.

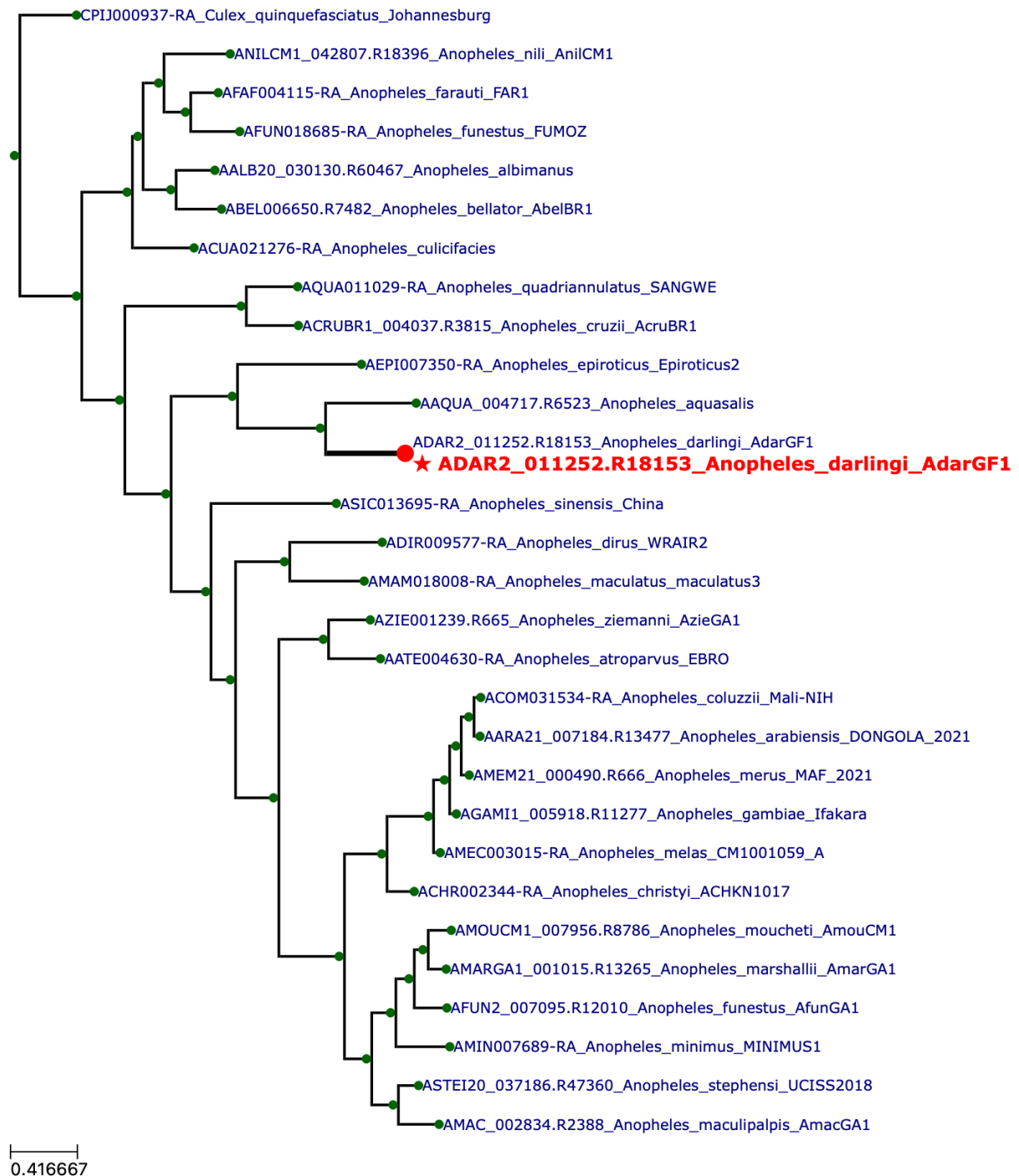


Figure 2. Rosetta stability estimates for WT, N218S, and Node 26 FREP1. Replicate $\Delta\Delta G$ values are shown for each comparison. The Node 26 ancestor exhibits uniformly large positive $\Delta\Delta G$ values relative to WT, consistent with global destabilization (left panel), whereas the N218S mutation shows modest and variable effects (right panel).

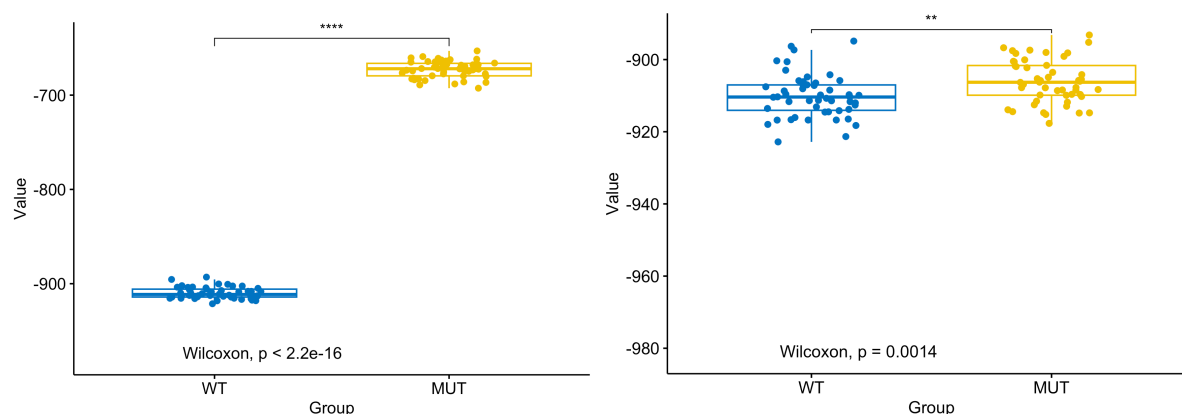


Figure 3. Distribution of PRODIGY-predicted binding energies for WT, N218S, and Node 26 FREP1. Boxplots display ΔG ($\text{kcal} \cdot \text{mol}^{-1}$) across 40 docked structures per variant. The Node 26 distribution is significantly shifted toward more negative binding energies, whereas the N218S distribution overlaps the WT (not shown).

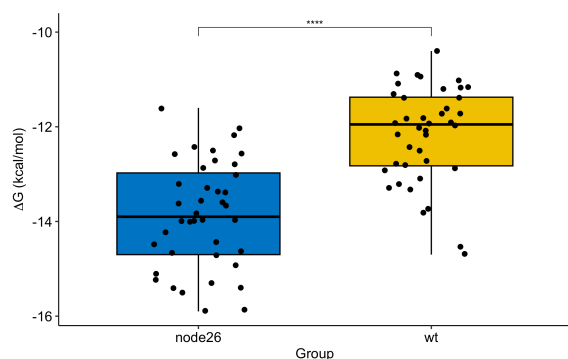
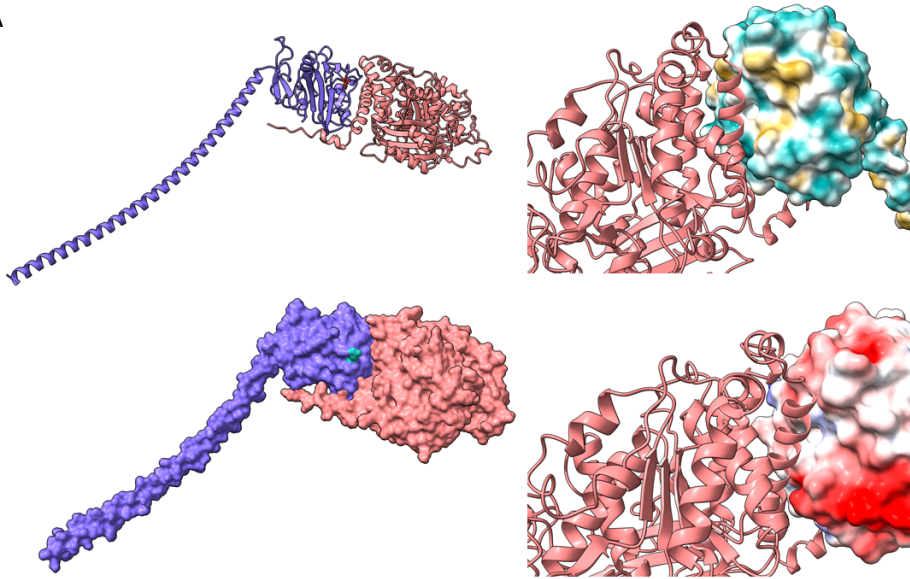
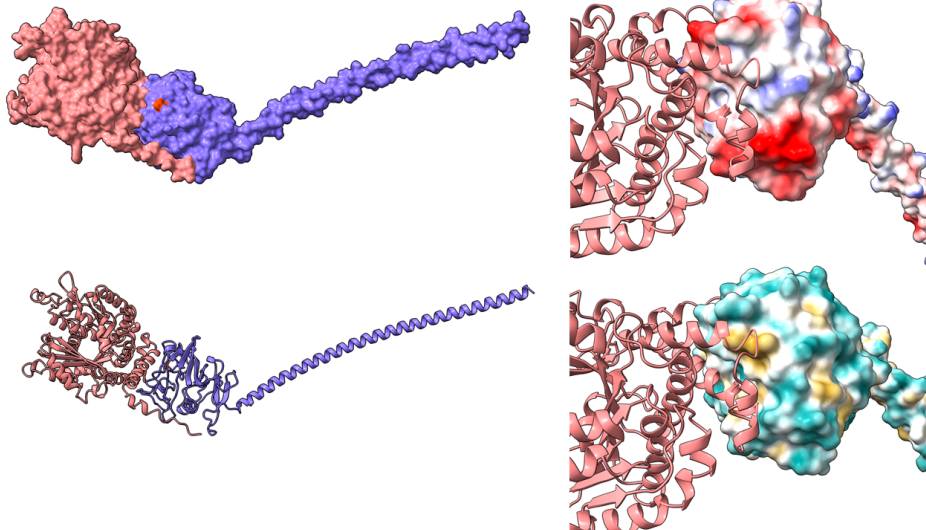


Figure 4. Comparative structural analysis of WT, N218S, and Node 26 FREP1 variants docked to *P. falciparum* α -tubulin-1. HADDOCK best-scoring complexes are shown for the three FREP1 variants: WT (A), N218S mutant (B), and Node 26 ancestral variant (C). For each variant, the overall docked assembly is displayed in ribbon representations, followed by a space-filling dimer view highlighting the variant-specific residue (WT: 218; N218S: Ser218; Node 26: position 191) in distinct color. Corresponding close-up views of the binding interface are shown as electrostatic surfaces and hydrophobicity-mapped surfaces, illustrating differences in charge complementarity, nonpolar contacts, and interface geometry among the three variants. Collectively, these comparisons highlight the expanded buried surface area and altered interaction landscape of the Node 26 complex relative to WT and N218S.

A



B



C

