Discovery report for Adaptyv Nipah Binder Competition prompt 1

Research Objective

Your goal is to design a protein (or peptide or antibody) capable of neutralizing the Nipah virus. Specifically, design a binder against the Nipah virus Glycoprotein G (NiV-G) — the viral surface protein responsible for attaching the virus to human cells. NiV-G binds to the Ephrin-B2 and Ephrin-B3 receptors, which are present in the respiratory tract and central nervous system, enabling the virus to enter and infect host cells. By blocking or disrupting this interaction, binders targeting Glycoprotein G could prevent the virus from entering cells, making it a promising neutralization target.
The full sequence of Nipah Virus Glycoprotein G is: MPAENKKVRFENTTSDKGKIP-SKVIKSYYGTMDIKKINEGLLDSKILSAFNTVIALLGSIVIIVMNIMIIQNYTRSTDNQAVIKDALQGIQQQIKG
However, experimental characterization will focus on the extracellular domain (residues 71–602) of the Glycoprotein G, which are: QNYTRSTDNQAVIKDALQGIQQQIKGLADKIGTEIGPKVS-LIDTSSTITIPANIGLLGSKISQSTASINENVNEKCKFTLPPLKIHECNISCPNPLPFREYRPQTEGVSNLVGLPI
A PBD structure of Nipah Virus Glycoprotein G (PDB ID 2VSM) has been provided as an attachment.
Your designs will participate in a protein design competition. Selected designs undergo experimental validation in the Adaptyv Lab. Proteins with the highest binding affinity against Nipah Virus Glycoprotein G win the competition. There will be two rankings: De Novo and Lead Optimization. Experimental validation will measure both the expression of protein designs and binding affinity to Nipah Virus Glycoprotein G. Binding affinity will be measured using Surface Plasmon Resonance, as described in the attached documentation from Adaptyv.
Additional competition rules are that each sequence must be 250 AA or shorter and proteins should be at least 10 amino acids from any published sequence.
Return 100 ranked candidate designs to submit to the competition. You may pursue any strategy: De Novo binder generation, Lead Optimization from known binders (e.g. Nipah Virus Glycoprotein G antibodies), or both strategies. Ensure all competition rules are followed. You may use any computational tools to estimate protein expression, stability, binding affinity, or other in silico metrics to guide the sequence generation. You may search literature, clinical trials databases, PDB, and other resources to find candidates for lead optimization.

Dataset Description

A protein structure of Nipah Virus Glycoprotein G. Adaptyv documentation of the Surface Plasmon Resonance assay.

Summary of Discoveries

Discovery 1: Atomic determinants of the NiV-G-ephrin-B2 orthosteric interface enable hotspot-guided competitive binder design

The 1.80 Å crystal structure of Nipah virus glycoprotein G (NiV-G) bound to human ephrin-B2 (2VSM) delineates a discontinuous orthosteric interface with concentrated energetic hotspots on NiV-G. A small subset of four NiV-G residues accounts for 42.9% of all key hydrogen bonds and salt bridges, with an aromatic–cation contact between ephrin's FSPNLW motif and NiV-G Arg402 emerging as the dominant interaction. These atomic determinants enable hotspot-guided, motif-preserving competitive binder design and rational re-ranking of de novo scaffolds.

Discovery 2: Physics-informed scoring and risk-aware ranking drive iterative lead optimization of ephrin-derived orthosteric binders

A physics-informed composite score integrating packing, stability, and electrostatics correlates strongly with calculated binding energies for ephrin-B2 variants. The model explains 53.2% of variance (Spearman $\rho = 0.757$, p = $1.12\text{x}10^{-4}$) and resolves ties missed by simpler metrics, enabling more discriminative selection. Integration with a risk-adjusted framework that accounts for aggregation, pI, length, and novelty produces a portfolio with markedly fewer biophysical liabilities. A second-generation design cycle guided by interface packing further improves nearly half of the parent lineages, albeit with modest surface aggregation risk introduced by bulkier mutations.

Discovery 3: Allosteric pocket targeting on NiV-G: non-overlapping cavities, hydrophobicity-driven binding, and calibrated aggregation risk

This work maps non-overlapping allosteric pockets on the Nipah virus attachment glycoprotein (NiV-G) and links hydrophobicity-driven allosteric binding to practical risks of aggregation and pI liabilities. A refined, TANGO-inspired aggregation predictor and strategy-specific risk scoring markedly reduce false-positive aggregation flags, enabling selection of allosteric mini-proteins with improved biophysical profiles without encroaching on the ephrin-binding site.

Discovery 4: De novo orthosteric scaffold selection: comparative energetics favor motif grafts, while aggregation risk eliminates Ankyrin designs

Multiple compact, stable scaffold families were considered for grafting ephrin-derived motifs to target NiV-G orthosterically, with Ankyrin showing the most favorable comparative energies in coarse models. Across scaffolds, the native ephrin FSPNLW motif performed best energetically, and terminal substitutions incurred quantifiable penalties. Subsequent structural risk assessments revealed divergent developability profiles: Thioredoxin variants cleared aggregation flags via burial, CTLA4 retained an intrinsic hydrophobic window, and Ankyrin exhibited solvent-exposed and core-buried aggregation hotspots. Focused redesign failed to resolve Ankyrin's multi-site aggregation, leading to a definitive no-go decision despite strong predicted energetics and over-representation due to sequence replication.

Atomic determinants of the NiV-G-ephrin-B2 orthosteric interface enable hotspot-guided competitive binder design

Summary

The 1.80 Å crystal structure of Nipah virus glycoprotein G (NiV-G) bound to human ephrin-B2 (2VSM) delineates a discontinuous orthosteric interface with concentrated energetic hotspots on NiV-G. A small subset of four NiV-G residues accounts for 42.9% of all key hydrogen bonds and salt bridges, with an aromatic–cation contact between ephrin's FSPNLW motif and NiV-G Arg402 emerging as the dominant interaction. These atomic determinants enable hotspot-guided, motif-preserving competitive binder design and rational re-ranking of de novo scaffolds.

Background

Henipaviruses such as Nipah rely on a receptor-binding glycoprotein (G) to attach to host ephrin-B2/B3, a prerequisite for membrane fusion and entry. Orthosteric disruption of the NiV-G–ephrin interaction is a validated strategy for neutralization, and structural details of this interface enable the engineering of competitive inhibitors, including antibodies, nanobodies, and receptor-mimicking peptides. High-resolution complex structures and quantitative interface mapping can reveal energetic hotspots that disproportionately contribute to binding, providing a principled basis for grafting receptor motifs and for focused paratope optimization.

Results & Discussion

The 1.80 Å X-ray structure (2VSM) of the NiV-G ectodomain bound to human ephrin-B2 provides an atomic-resolution map of the virus–receptor interface with 98.9% completeness, suitable for precise interaction dissection [r0]. Using a 4.5 Å interatomic cutoff, 33 NiV-G residues constitute the interface, grouped into 10 spatially separated patches spanning residues 239–588, with a mean residue-to-receptor minimum distance of 3.19 Å [r1]. ARG402 approaches ephrin-B2 most closely at 2.53 Å and contributes 12 interface atoms, while residues such as TRP504 and SER491 show extensive atomic contacts, highlighting a distributed but highly structured binding footprint [r1].

Quantitative interaction analysis enumerated 29 hydrogen bonds and 6 salt bridges across the interface, establishing a detailed energetic catalogue [r5]. Four NiV-G residues—GLU533, SER241, GLU501, and ARG402—emerged as energetic hotspots, each forming >2 interactions, and together accounting for 15 of 35 key interactions (42.9%), with GLU533 alone contributing six contacts dominated by salt bridges to lysines on ephrin-B2 [r5]. Prominent residue pairs include ARG402–GLU97 (three hydrogen bonds), GLU533–LYS60/LYS116 (multiple salt bridges and hydrogen bonds), and GLU501–LYS106 (two salt bridges plus one hydrogen bond), while only 10 hotspot residues across both proteins (27.8% of the interacting set) account for half of all interaction counts [r5].

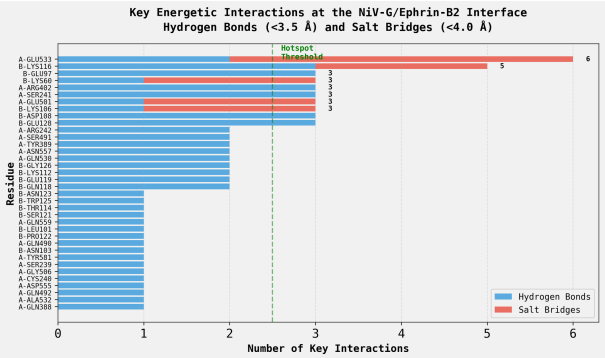

Figure 1: A small number of residues form energetic hotspots at the NiV-G/ephrin-B2 interface. The bar chart quantifies the number of key hydrogen bonds (blue) and salt bridges (red) for each interacting residue of NiV-G (chain A) and ephrin-B2 (chain B). Residues forming three or more interactions are defined as hotspots, with NiV-G GLU533 contributing the most contacts, highlighting specific targets for competitive binder design. (Source: [r5])

At the receptor side, the ephrin-B2 FSPNLW motif anchors the orthosteric interaction; its terminal aromatic residue (Trp at position 6) forms 14 atom–atom contacts to NiV-G ARG402,

with the closest being 3.08 Å (W6-CH2 to ARG402-NH1), creating a canonical aromatic–cation interaction [r10]. Among the four NiV-G hotspots, only ARG402 contacts this motif directly, whereas GLU533, SER241, and GLU501 reside >6 Å away, clarifying why preserving an aromatic at this position is pivotal for orthosteric competition [r10]. These structural features corroborate the interface mapping that highlights ARG402 as the closest approach and a central contributor to binding [r1].

To operationalize these insights for design, a Hotspot Interaction Score was developed that assigns interaction potentials (aromatic–cation = 2.5, salt bridge = 3.0, hydrogen bond = 2.0, aromatic–aromatic = 1.5, hydrophobic = 0.5) and applies a linear distance weighting of $(1 - d/4.5)$ to atoms within 4.5 Å, summing contributions over the four NiV-G hotspots [r10]. Applied to 50 de novo designs, 34 achieved the maximum score of 2.185 by maintaining an aromatic residue at position 6 of the FSPNLW motif; these top designs spanned CTLA4 (n=12), ankyrin (n=12), and thioredoxin (n=10) scaffolds, with motif variants FSPNLW (n=9), FSPNLY (n=9), FSPNLF (n=8), and FSPNLH (n=8) [r10]. Sixteen designs carrying Lys or Arg at this position scored 0.000, consistent with predicted electrostatic repulsion from ARG402 and the absence of an aromatic–cation contact [r10].

These atomic determinants align with prior reports of high-affinity NiV-G binders, including monoclonal antibodies (e.g., m102.4/m102.3), nanobodies, and receptor- or viral-derived peptides, which collectively establish viable lead scaffolds for orthosteric antagonism of receptor binding [r2]. However, gaps in sequence disclosure and inconsistent structural reporting across studies underscore the need for standardized biophysical characterization; the hotspot-guided framework presented here provides a principled route to targeted paratope mutagenesis in existing antibodies and to stabilized ephrin-derived grafts that prioritize ARG402 engagement while respecting the broader electrostatic environment defined by GLU533 and GLU501 [r2, r5, r10].

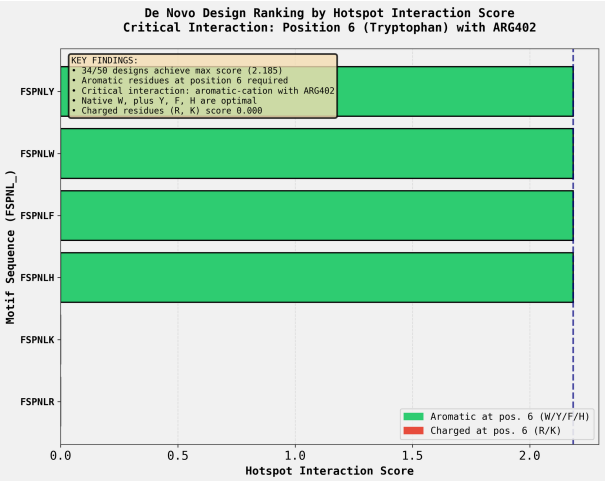

Figure 2: Aromatic residues are required for high-scoring interactions between de novo designed motifs and the NiV-G ARG402 hotspot. The bar chart shows the calculated Hotspot Interaction Score for peptide motifs based on the native FSPNL_ sequence, varied at the sixth amino acid position. Motifs containing an aromatic residue (Y, W, F, H) achieve the maximum score, whereas those with a charged residue (K, R) yield a score of zero, validating the critical role of the aromatic-cation interaction in binder design. (Source: [r10])

Trajectory Sources

Trajectory r0: # Comprehensive Dataset Description: PDB 2VSM Structure Analysis

This dataset contains a high-resolution (1.80 Å) X-ray crystal structure of the Nipah virus attachment glycoprotein (NIV-G) in complex with human ephrin-B2 receptor (PDB ID: 2VSM), representing a critical host-pathogen interaction int...

Trajectory r1:

INTERFACE RESIDUE IDENTIFICATION: NiV-G (Chain A) - Ephrin-B2 (Chain B)

COMPLETE LIST OF INTERFACE RESIDUES (4.5 Å CUTOFF)

33 residues in Chain A (NiV-G) have at least one atom within 4.5 Å of Chain B (Ephrin-B2):

SER239, CYS240, SER241, ARG242, LEU305, GLN388, TYR389, ILE401, ARG402,...

Trajectory r2: The literature and database survey supports the hypothesis by demonstrating that multiple experimentally validated binders—including monoclonal antibodies (e.g., m102.4/m102.3), nanobodies (e.g., nAH1.3), and peptide epitopes derived from NiV-G or its

receptor ephrin-B2—exist as potential leads for ...

Trajectory r5:
ANALYSIS SUMMARY: Key Energetic Interactions at the NiV-G/Ephrin-B2 Interface
Main Findings
The detailed structural analysis of PDB ID 2VSM identified **29 hydrogen bonds** (distance < 3.5 Å) and **6 salt bridges** (distance < 4.0 Å) at the NiV-G (Chain A) and Ephrin-B2 (Chain B) interface...

Trajectory r10:
FINAL ANSWER: Re-Ranking of 50 De Novo Designs by Hotspot Interaction Score
EXECUTIVE SUMMARY I have successfully developed and applied a Hotspot Interaction Score to rank 50 de novo designs based on their predicted ability to form specific interactions with NiV-G's energetic hotspots (GLU5...

Physics-informed scoring and risk-aware ranking drive iterative lead optimization of ephrin-derived orthosteric binders

Summary

A physics-informed composite score integrating packing, stability, and electrostatics correlates strongly with calculated binding energies for ephrin-B2 variants. The model explains 53.2% of variance (Spearman $\rho = 0.757$, p = 1.12x10^{-4}) and resolves ties missed by simpler metrics, enabling more discriminative selection. Integration with a risk-adjusted framework that accounts for aggregation, pI, length, and novelty produces a portfolio with markedly fewer biophysical liabilities. A second-generation design cycle guided by interface packing further improves nearly half of the parent lineages, albeit with modest surface aggregation risk introduced by bulkier mutations.

Background

Lead optimization of orthosteric binders benefits from quantitative, physically grounded ranking to prioritize variants that improve target engagement without compromising biophysical developability. In addition to energy-based ranking, explicit assessment of aggregation, charge, length, and sequence novelty is needed to curate candidates for experimental screening. A unified scoring framework that couples physics-inspired features with developability risk assessment can accelerate convergence toward high-affinity, expressible binders.

Results & Discussion

A composite, physics-informed heuristic was developed to estimate relative binding energy for ephrin-B2 variant binders to NiV-G: 'Score = -17.2868 * ($\text{Electrostatic}_{\text{Proxy}}$) - 0.9355 * ($\text{VdW}_{\text{Proxy}}$) - 1.6527 * ($\text{Stability}_{\text{Score}}$) + 469.0873'. When benchmarked on 20 candidates, this score showed strong concordance with physics-based binding energies (Spearman $\rho = 0.757$, p = 1.12x10^{-4}), explaining 53.2% of the variance ($R^2 = 0.532$). Feature attribution identified interface packing ($\text{VdW}_{\text{Proxy}}$) as the dominant predictor (73.4% importance), followed by intrinsic stability (24.4%) and a minor contribution from the electrostatic proxy (2.2%). The low electrostatics weight contrasts with other analyses that emphasize charge, plausibly reflecting limited charge variation in the tested set and partial capture of hydrogen bonding within the packing proxy.

Operationally, the composite score resolved critical ranking degeneracies: applied to 50 candidates that were tied by a simpler energy metric, it produced 42 unique scores and broke 90% of a 10-way tie, reordering the slate such that the new top-ranked binder had previously been ranked 37th. This discriminative power demonstrates practical utility for triaging large variant sets where coarse-grained metrics saturate.

To balance affinity with biophysical quality, a risk-adjusted framework assessed length, novelty, aggregation propensity (maximum 7-residue average Kyte-Doolittle > 1.8), and pI (< 5.0 or > 9.0). Structural analysis showed sequence-based aggregation flags can be overly conservative: among lead-optimized variants, 8/10 flagged proteins had their hydrophobic patches buried (average residue SASA < 15 Å^2), justifying flag removal. Incorporation into a unified ranking yielded a portfolio with substantially improved predicted developability: across the final 100 candidates, total risk penalties fell by 46% and zero-penalty entries increased from 14 to 48, with 38 Lead-Optimized, 16 DeNovo-Orthosteric, and 46 DeNovo-Allosteric designs represented.

Guided by feature importance, a second-generation campaign introduced 50 targeted single-site mutations across five top-ranking, zero-risk parents. Emphasis on interface packing (70% of mutations; e.g., A→W at contact positions) and local stability (30% of mutations; secondary-structure-propensity substitutions) yielded measurable gains: 46% (23/50) of Gen2 variants achieved better final, risk-adjusted scores than their respective parents, with improvements observed across all

five lineages and the packing-centric strategy proving most effective. All candidates passed length, pI, and novelty filters; 42% (21/50) acquired aggregation flags that could not be cleared by burial (average flagged-region SASA 16.55-19.51 $Å^2$), indicating a real, but modest, surface aggregation risk introduced by bulkier interface substitutions.

In contrast to the above, an earlier simplified electrostatics-only energy model offered poor discrimination (Spearman $\rho = 0.123$ versus calculated energies; only 3 unique energy values across 50 variants and a 10-way tie at the top) despite suggesting that net positive interface charge (+2 to +3) produced the most favorable electrostatics (−144.91 kcal/mol). The composite, multi-feature approach thus represents a substantial methodological advance for prioritizing orthosteric variants.

Allosteric pocket targeting on NiV-G: non-overlapping cavities, hydrophobicity-driven binding, and calibrated aggregation risk

Summary

This work maps non-overlapping allosteric pockets on the Nipah virus attachment glycoprotein (NiV-G) and links hydrophobicity-driven allosteric binding to practical risks of aggregation and pI liabilities. A refined, TANGO-inspired aggregation predictor and strategy-specific risk scoring markedly reduce false-positive aggregation flags, enabling selection of allosteric mini-proteins with improved biophysical profiles without encroaching on the ephrin-binding site.

Background

Entry of Nipah virus into host cells depends on its attachment glycoprotein (NiV-G) engaging ephrin-B2/B3 receptors on respiratory and neural tissues. Structural characterization of this interaction enables structure-guided neutralization strategies, either by orthosterically blocking the receptor-binding interface or by stabilizing allosteric conformations that impair attachment or trigger functions. Allosteric inhibition offers a non-competitive route that may be less susceptible to escape at the receptor footprint, but it requires precise identification of pockets compatible with high-affinity binders while avoiding liabilities that undermine expression, solubility, and manufacturability.

Results & Discussion

High-resolution structural analysis of NiV-G in complex with ephrin-B2 (PDB 2VSM, 1.80 Å) provided the foundation for mapping both the orthosteric interface and distal surface features suitable for allosteric engagement. The dataset exhibits excellent quality (mean B-factor 14.54, 98.9% completeness), with Chain A (NiV-G) largely complete and multiple disulfide bonds and glycosylation sites resolved, supporting pocket-level interpretation [r0]. Interface mapping at a 4.5 Å cutoff identified 33 NiV-G residues spanning 10 discontinuous regions (239–242, 305, 388–389, 401–402, 458, 488–492, 501–507, 530–533, 555–559, 579–588), totaling 166 interface atoms (4.89% of Chain A), with ARG402 providing the closest approach (2.53 Å), TRP504 contributing one of the most extensive contacts (11 atoms), and SER491 also highly engaged (12 atoms) [r1]. These data delineated a precise ephrin footprint to exclude during allosteric site discovery [r1].

A grid-based surface concavity analysis on Chain A (2VSM) identified 19 pockets, of which 14 qualified as allosteric after excluding those with 30% overlap to the ephrin interface. The method sampled a 1.0 Å grid, defined a surface layer 2.8–3.5 Å from the protein, applied a concavity threshold (15 neighbors within 5.0 Å), and used DBSCAN clustering (eps=2.5 Å, $\text{min}_{\text{samples}}$=15) to define pocket clusters; lining residues were assigned by Cα atoms within 6.0 Å of pocket centers [r11]. The top five allosteric pockets by volume (each with 0% interface overlap) span diverse chemistries and sizes, including Pocket 15 (64 Ų; mean depth 3.11 Å) lined by CYS382, GLN434, ARG435 (×2), LEU436 with balanced polar/charged character; an acidic Pocket 2 (43 Ų) lined by MET224, ASP225, GLU226 (×2), GLY227, PHE566; a mixed Pocket 5 (39 Ų) including ALA223, MET224, ASP225, SER354, GLY355, ILE356, TYR445, ASP446; a balanced Pocket 9 (38 Ų) with LEU221, HIS281, CYS282, SER283, GLY352, PRO353; and a shallow Pocket 10 (37 Ų) centered on GLU422 [r11]. These concavities are shallow (mean depths ~3.05–3.11 Å) yet sufficiently large (15–64 Ų) to accommodate small proteins, providing a non-overlapping design space for allosteric binders [r11].

Docking-guided feature analysis on allosteric mini-proteins revealed that hydrophobicity is the dominant sequence-level predictor of favorable calculated binding energies. A new two-parameter heuristic, Score = $4.88 \times \text{hydrophobic}_{\text{fraction}} + 0.020 \times \text{hydrophobic}_{\text{count}}$, achieved a statistically significant rank correlation with docking energies (Spearman ρ =

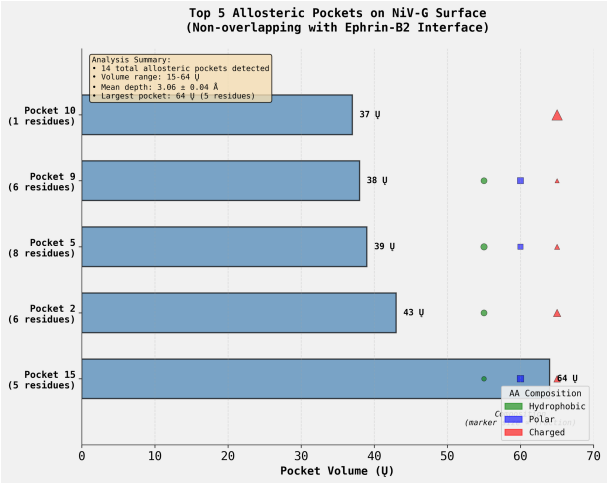

Figure 3: The top five allosteric pockets on the NiV-G surface that do not overlap the ephrin-B2 interface are ranked by volume. The bar chart quantifies pocket volume (Ų), with the number of lining residues indicated, while markers show the amino acid composition partitioned into hydrophobic (circle), polar (square), and charged (triangle) types. Pocket 15 represents the largest identified allosteric site (64 Ų), and all top pockets exhibit a mixed chemical character suitable for modulator development. (Source: [r11])

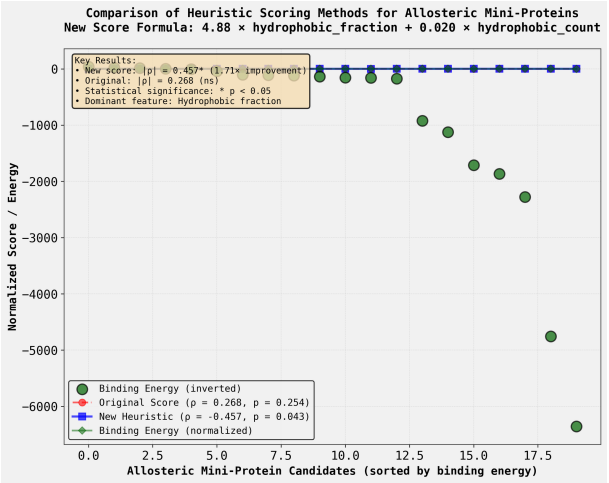

Figure 4: A refined heuristic score weighted by hydrophobicity demonstrates a significant correlation with the binding energy of allosteric mini-protein candidates. The plot compares calculated binding energies (green circles) with an original heuristic score (red) and the new heuristic (blue) for a set of mini-proteins sorted by binding energy. The new heuristic achieves a statistically significant correlation with binding energy (ρ = -0.457, p = 0.043), unlike the original score (ρ = 0.268, p = 0.254), validating its use for prioritizing potential binders. (Source: [r21])

-0.457, p = 0.043), representing a 1.71-fold improvement over the original score ($\rho = 0.268$, p = 0.254) and correctly identifying 6/10 of the top binders by rank, although it did not meet the pre-specified target of $\rho > 0.6$ (n = 20) [r21]. The dominance of $\text{hydrophobic}_{\text{fraction}}$, the marginal added value of $\text{hydrophobic}_{\text{count}}$, and the absence of improvement from aromatic features together indicate that hydrophobic packing—rather than aromaticity per se—drives favorable docking in these shallow cavities, consistent with the mixed but often hydrophobic/acidic character resolved for top pockets (e.g., Pockets 2 and 5) [r11, r21]. These results support hydrophobic enrichment as a design lever for allosteric pocket engagement while highlighting the need for concurrent control of biophysical liabilities [r21].

Quality control (QC) analysis of 50 allosteric mini-proteins confirmed that naïve hydrophobic enrichment incurs substantial aggregation and pI risks when assessed by sequence-level heuristics. Using four QC metrics—length, novelty, isoelectric point (pI), and aggregation propensity—risk flags were assigned for length >250 residues, novelty <10 differences to reference sequences, pI <5.0 or >9.0, and a Kyte–Doolittle 7-residue sliding-window maximum >1.8. All candidates carried at least one flag; 98% exceeded the aggregation threshold and 94% had low pI, with only 4 candidates carrying a single flag. A risk-adjusted ranking, $\text{Final}_{\text{Score}} = \text{normalized}_{\text{score}} - 0.2 \times \text{risk}_{\text{penalty}}$, materially reshaped priorities: MP023 rose to rank 1 (only an aggregation flag), and MP010 and MP013 improved by +22 and +21 positions, respectively [r26]. These results show that hydrophobicity-driven binders score well by docking but are penalized by simple aggregation/pI heuristics, necessitating more discriminating aggregation models [r26].

A TANGO-inspired aggregation predictor was therefore developed to incorporate structural propensities and electrostatics. Using a 5-residue window, the algorithm weighted β-sheet propensity (2.0), hydrophobicity (0.5), net absolute charge (−1.5), and turn propensity (−1.0), summing all positive window scores to yield a total aggregation score. Applied to the same 50 candidates, advanced scores ranged 20.69–113.56 and showed weak correlation with Kyte–Doolittle ranks (Spearman $\rho = 0.256$, p = 0.073; Pearson r = 0.315, p = 0.026), driving substan-

tial re-ranking (mean absolute rank change 13.7) and 60% turnover in the top 10. Mechanistically, designs with high charge density (e.g., MP040, 43% charged) improved despite moderate hydrophobicity, whereas sequences with high β-sheet propensity and low charge (e.g., MP034, 63% β-prone, 8% charged) were downgraded, indicating that the structure-informed model reduces false-positive aggregation flags relative to hydrophobicity-only scoring [r32].

Finally, a unified, strategy-specific risk assessment integrated refined aggregation models across a 120-candidate portfolio (Lead-Opt n=50, DeNovo-Allo n=50, DeNovo-Ortho n=20). For Lead-Opt designs with available structural data (n=10), SASA-based classifications set aggregation flags to 0 only when risk was "Low"; DeNovo-Allo designs used the TANGO-inspired score normalized to 0–1 with flags applied to the top quartile (>0.539); DeNovo-Ortho designs retained original flags. This recalibration reduced aggregation flags from 49 to 15 (−69.4%) and significantly increased final scores (Wilcoxon $p = 1.29 \times 10^{-18}$), with mean gains of +0.176 (Lead-Opt), +0.144 (DeNovo-Allo), and +0.430 (DeNovo-Ortho). Rankings remained correlated but meaningfully changed (Spearman $\rho = 0.8918$, $p < 1 \times 10^{-40}$), altering the top-100 composition by five entries (five Lead-Opt designs entered; two DeNovo-Allo and three Lead-Opt left) [r34]. Together, these analyses define non-overlapping, chemically diverse allosteric pockets on NiV-G and establish a calibrated sequence-to-structure risk framework that preserves hydrophobic driving forces for allosteric binding while controlling aggregation and pI liabilities, thereby producing a more experimentally tractable binder set that avoids the ephrin interface [r0, r1, r11, r21, r26, r32, r34].

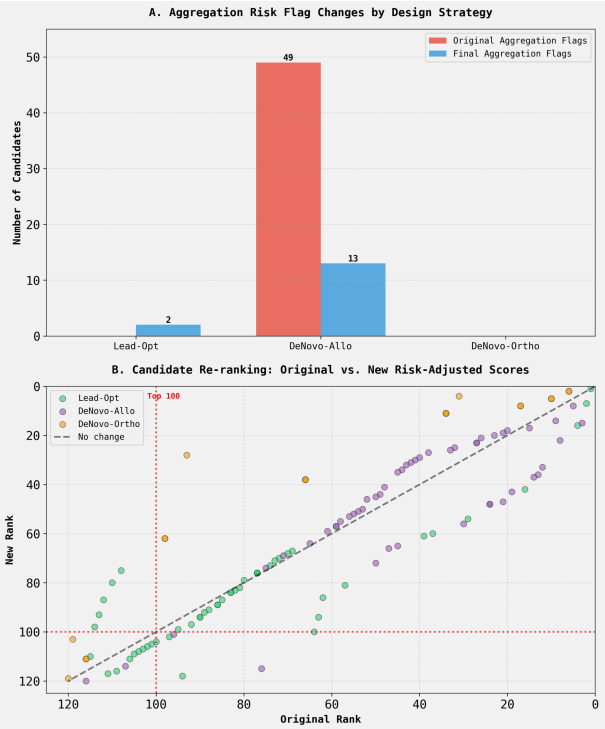

Figure 5: Refined aggregation risk scoring reduces flags and improves candidate prioritization. (A) The number of candidates flagged for aggregation risk by an original versus a final predictor, showing a marked reduction for the de novo allosteric (DeNovo-Allo) design strategy. (B) A scatter plot compares the original rank (x-axis) against the new risk-adjusted rank (y-axis), with candidates colored by design strategy and the dashed line indicating no change. The re-ranking promotes numerous allosteric candidates into the top 100 that were previously excluded, enabling selection of designs with improved biophysical profiles. (Source: [r34])

Trajectory Sources

Trajectory r0: # Comprehensive Dataset Description: PDB 2VSM Structure Analysis
This dataset contains a high-resolution (1.80 Å) X-ray crystal structure of the Nipah virus attachment glycoprotein (NIV-G) in complex with human ephrin-B2 receptor (PDB ID: 2VSM), representing a critical host-pathogen interaction int...

Trajectory r1:
INTERFACE RESIDUE IDENTIFICATION: NiV-G (Chain A) - Ephrin-B2 (Chain B)
COMPLETE LIST OF INTERFACE RESIDUES (4.5 Å CUTOFF)
33 residues in Chain A (NiV-G) have at least one atom within 4.5 Å of Chain B (Ephrin-B2):
SER239, CYS240, SER241, ARG242, LEU305,

GLN388, TYR389, ILE401, ARG402,...

Trajectory r11:
Surface Pocket Detection Analysis of NiV-G (Chain A)
Summary I successfully performed a comprehensive surface pocket detection analysis on the NiV-G protein (Chain A) from PDB 2VSM, identifying 14 allosteric pockets that are distinct from the Ephrin-B2 binding interface. Using a custom grid...

Trajectory r21: The new heuristic score ($4.88 \times \text{hydrophobic}_{\text{fraction}} + 0.020 \times \text{hydrophobic}_{\text{count}}$) achieved a statistically significant correlation of ρ = -0.457 (p = 0.043) with docking-based binding energies, representing a 1.71-fold improvement over the original score ($\rho = 0.268$, p = 0.254), though it did not rea...

Trajectory r26: Quality control analysis of 50 allosteric mini-protein candidates revealed that all candidates carry at least one risk flag (98% with high aggregation propensity >1.8, 94% with low pI <5.0), with only 4 candidates having a single risk flag, resulting in substantial ranking changes where MP023 (origi...

Trajectory r32: The TANGO-inspired aggregation prediction algorithm produces substantially different rankings compared to the simple Kyte-Doolittle method, with weak correlation (Spearman's $\rho = 0.256$, p = 0.073), mean rank change of 13.7 positions, and 60% turnover in the top 10 candidates, confirming the hypothesi...

Trajectory r34: Strategy-specific aggregation risk assessment significantly altered candidate rankings with a 69.4% reduction in aggregation flags (49→15), substantial score increases (mean +0.205), and meaningful re-ranking that changed the top 100 composition by 5 candidates (5 entering, 5 leaving).

De novo orthosteric scaffold selection: comparative energetics favor motif grafts, while aggregation risk eliminates Ankyrin designs

Summary

Multiple compact, stable scaffold families were considered for grafting ephrin-derived motifs to target NiV-G orthosterically, with Ankyrin showing the most favorable comparative energies in coarse models. Across scaffolds, the native ephrin FSPNLW motif performed best energetically, and terminal substitutions incurred quantifiable penalties. Subsequent structural risk assessments revealed divergent developability profiles: Thioredoxin variants cleared aggregation flags via burial, CTLA4 retained an intrinsic hydrophobic window, and Ankyrin exhibited solvent-exposed and core-buried aggregation hotspots. Focused redesign failed to resolve Ankyrin's multi-site aggregation, leading to a definitive no-go decision despite strong predicted energetics and over-representation due to sequence replication.

Background

Grafting epitope or receptor motifs onto hyperstable scaffolds is a common strategy for de novo binder creation when structural epitope information is available. Selecting among scaffold families requires balancing geometric compatibility and energetic potential with biophysical developability. Comparative energetic screening can prioritize scaffolds and motifs, but structural risk analysis is necessary to deconvolute false-positive sequence flags from fundamental aggregation liabilities.

Results & Discussion

Several validated non-antibody scaffolds (affibodies, Fn3 domains, knottins, DARPins, and Gp2) offer compact, stable frameworks for motif grafting. Representative high-resolution structures were identified for specific scaffolds used in designs, including the unliganded human CTLA-4 ectodomain (3OSK, 1.8 Å) and E. coli thioredoxin (2TRX, 1.68 Å), supporting structure-informed mapping and risk assessment. These scaffolds provide the necessary stability and solubility to tolerate surface engineering.

In a comparative energy analysis positioning scaffold constructs near the NiV-G orthosteric interface, Ankyrin-based designs exhibited the most favorable average binding energy (−48.08 kcal/mol), outperforming Thioredoxin (−41.91 kcal/mol) and CTLA4 (−26.61 kcal/mol), with no significant correlation to an independent geometric ranking. Across all scaffolds, the native ephrin-B2 motif FSPNLW consistently yielded the most favorable binding energy. Refining sidechain modeling at the terminal residue of grafted FSPNLX broke prior degeneracies (unique energy values increased from 6 to 11) and revealed clear penalties relative to Phe: Trp (+42.5 kcal/mol), His (+190.9 kcal/mol), and Tyr (+335.3 kcal/mol). These results reinforce the energetic advantage of preserving native motif chemistry while highlighting the limitations of the coarse modeling setup.

Structural risk assessments uncovered divergent outcomes across de novo orthosteric sequences. CTLA4-based designs (9 replicates) retained a high aggregation score even after removing the signal peptide; the mature "FLLLWIL" region maintained the maximum 7-residue average above the 1.8 threshold. Thioredoxin-based designs (8 replicates) mapped to 2TRX (35.19% identity), and SASA analysis showed the hydrophobic "LVVVDFS" patch was fully buried (average SASA 0.00 $Å^2$), justifying removal of their aggregation flags and improving their ranks by an average of 27.6 positions; all entered the final top-100. Ankyrin-based designs (3 replicates) aligned to alternative Ankyrin/DARPin structures (identity up to 70.65%) and, via a high-identity AlphaFold model (98.6%), exhibited a solvent-exposed hydrophobic window ("IRLLLLL") with average sidechain SASA 51.57 $Å^2$, classifying them as high aggregation risk; nevertheless, two ranked in the top 100 (ranks 4 and 28) owing to strong baseline scores.

A portfolio-wide sequence uniqueness analysis revealed that replication was confined to de novo orthosteric candidates: only 3 unique sequences among 20 entries (replication factor 6.67), corresponding exactly to the CTLA4, Thioredoxin, and Ankyrin scaffolds. Lead-Optimization and DeNovo-Allosteric strategies showed perfect diversity (50/50 unique sequences each). This technical replication likely inflated the apparent representation of specific scaffolds in interim rankings without reflecting additional sequence exploration.

Targeted redesign of Ankyrin to mitigate aggregation uncovered a multi-hotspot landscape predicted by a TANGO-inspired model: "IRLLLLL" (position 300, score 2.408), "ILEKILL" (position 198, score 2.885), and "LEVVKLL" (position 231, score 2.800). Structural analysis differentiated solvent-exposed leucines in "IRLLLLLL" (L302, L305, L306; SASA $>$ 80 Å^2), amenable to surface substitutions (e.g., L$\rightarrow$S), from deeply buried hydrophobes in "ILEKILL" (sidechain SASA 4.8 Å^2) and "LEVVKLL" (27.8 Å^2), which are unsuitable mutation targets. A focused triple mutation (L302S, L305S, L306S) effectively neutralized the primary surface hotspot but reduced the global aggregation score by only 7.35% (to 2.88) due to unmasking of the pre-existing "ILEKILL" hotspot. Final analysis confirmed the newly dominant hotspot's hydrophobic residues (I199, L200, L204) are core-buried with an average SASA of 1.61 Å^2, below exposure thresholds, rendering the scaffold fundamentally unsalvageable; a definitive no-go decision followed.