Temporal Perturbation Scanning: AI-Driven Deconstruction of Universal Biomolecular Recognition Mechanisms

Fodil Azzaz[1] & Jacques Fantini[1]

Aix-Marseille Univ, INSERM UA16, Marseille, France

Correspondence to: azzaz.fodil@gmail.com

Abstract

Understanding the physicochemical principles governing intermolecular recognition remains a fundamental challenge across biochemistry, drug discovery, and synthetic biology. While molecular dynamics simulations capture interaction dynamics and machine learning models predict binding affinities, neither provides a systematic approach to deconstruct recognition mechanisms into their fundamental physical components across time. Here we present Temporal Perturbation Scanning (TPS), an AI framework that universally decomposes intermolecular interactions into electrostatic, hydrophobic, and steric contributions across binding stages. Our method: i) combines graph neural networks with interpretable, multi-faceted *in silico* mutagenesis to automatically process biomolecular complexes, ii) applies targeted physicochemical perturbations, and iii) quantifies residue-specific importance through statistical effect sizes. Crucially, TPS achieves this mechanistic insight with orders-of-magnitude greater computational efficiency than traditional mutation simulations, extracting equivalent information from single trajectories that would otherwise require dozens of independent simulations.

Applied to neurotoxin-receptor and other biomolecular systems, TPS reveals fundamentally different engagement strategies and stage-dependent interaction patterns, demonstrating its capacity to capture mechanistic variation across complexes. Beyond resolving these biological insights, our framework provides a generalizable computational platform for mechanistic analysis across diverse molecular interactions, enabling researchers to move beyond predicting whether molecules interact to understanding how and why they recognize one another through interpretable, physicochemical decomposition across time.

Keywords: Temporal Perturbation Scanning (TPS); Molecular Recognition; Graph Neural Networks; Residue-Level Mechanistic Analysis; interface analysis; physicochemical decomposition

Introduction

Understanding the physicochemical principles that determine molecular recognition requires moving beyond structural observation to systematic interrogation. While molecular dynamics simulations (MDS) capture interaction dynamics [1–3], and machine learning models predict binding affinities, these approaches remain fundamentally observational: they scribe what occurs during binding but cannot actively test which specific residues drive recognition through causative physical forces.

The critical gap lies in transitioning from correlation to causation. Structural snapshots[4] and binding affinity predictions can identify functional importance. Methods like free energy calculations[5] provide ensemble averages but lack the resolution to decompose interfaces into their fundamental electrostatic, hydrophobic, and steric components or track their evolution across binding stages.

This limitation has practical consequences across multiple domains. In drug discovery[6], the inability to distinguish critical from peripheral interactions hinders rational design. In synthetic biology[7] engineering novel protein interfaces remain challenging without understanding the physical determinants of recognition. Even in fundamental studies of systems like neurotoxin-receptor interactions[8,9] we can observe differential binding but cannot systematically explain its physical basis.

Here we introduce Temporal Perturbation Scanning (TPS), a framework that transforms molecular interactions analysis from passive observation to active interrogation. By combining graph neural networks[10] with multi-faceted *in silico* mutagenesis, we systematically perturb electrostatic, hydrophobic and steric properties across binding stages, measuring the causal effects of each modification. This approach attains its full significance when viewed through the lens of biology's fundamental parameters and biophysical constraints[11], moving the fundamental question from "which residue are nearby?" to "which residues actually matter, and through which physical forces?".

We illustrate and validate framework through analysis of neurotoxin-receptor systems, revealing not just interaction patterns but the causative physical mechanisms underlying differential binding specificity. More broadly, our method provides a generalizable platform for moving beyond describing molecular interactions to understanding and engineering them through interpretable physicochemical decomposition.

Material and Methods

Graph Representation of Biomolecular Structures.

Feature engineering. We converted molecular dynamics simulation frames into graph neural network-compatible representations using a 13èdimensional feature vector per atom. (1) Electrostatic features (partial charges from force field topology (Charmm36m[12]), (2) mass properties (normalized atomic masses), (3) element encoding (atomic numbers: H=1, C=6, N=7, O=8, S=16 and P=15), (4) residue hydrophobicity (Kyte-Doolittle scale[13], normalized to [-1, 1]), (5) charge classification (positive=1, negative=-1, neutral=0), (6) structural indicators (backbone vs sidechain), (7) segment encoding (one-hot for biomolecule component referred as segid).

Edge construction. Two edge types were established:

Covalent edges: Directly from PSF topology files.

Non-covalent edges: Spatial proximity edges (< 6 Å) identified using k-d tree algorithms[14]
.

The edge features included:

. Radial basis functions (16 dimensions for distance encoding)

. Bond type indicators (covalent vs non-covalent: note that only inter-molecular non-covalent bonds are accepted, the intra-molecular are rejected (not necessary for our approach))

. Directional vectors (normalized atomic displacements)

Neural Network Architecture

We designed a graph convolutional network with frame interleaving. As shown in figure 1, the architecture includes 3 GCN layers (128→64→32 neurons)[10] with batch normalization[15], 30% dropout for regularization, multi-head attention for interface-specific feature pooling and interface predictor linear layer for stability scoring. Interface detection was implemented using k-d tree spatial algorithms[14] with a 6.0 Å threshold to identify contacting atoms between molecular segments.

Training protocol

We employed frame interleaving to ensure each data subset spanned the entire simulation trajectory: training frames (0, 3, 6...), validation frames (1, 4, 7...) and test frames (2, 5, 8...), each representing 33% of the data. Model optimization used AdamW[16] (learning rate=0.001, weight decay=1e-4) with mean squared error loss and early stopping (patience=15 epochs). Statistical robustness was assessed through 10 independent training trials with different random seeds to evaluate model variance.

Interface Perturbation Analysis

We implemented six distinct biological perturbation strategies to assess interface criticality: (1) electrostatic charge reversal of Arg/Lys/Asp/Glu residues, (2) hydrophobicity inversion of hydrophobic core, (3) introduction of bulky sidechain substitutions, (4) disruption of π-stacking networks, (5) breaking of hydrogen bonding networks, and (6) structural displacement of interface atoms. Interface importance was quantified as $\Delta = |P(\text{perturbed}) - P(\text{original})|$, where P represents the model-predicted interface stability score. Residues were ranked by Sum Δ, which weights contributions by both interaction strength and frequency.

Configurable Temporal Staging

Temporal analysis in TPS is performed across user-defined stages. For clarity and consistency in this study, we devided trajectories into three equal-length regimes: early (initial binding events), mid (equilibrium interactions), and late (stabilized complex formation). However, TPS is designed with flexible stage definition: users may specify any number of stages, uneven time intervals and employ unsupervised clustering to identify natural kinetic phases directly from trajectory data. This flexibility allows TPS to model a wide range of dynamic processes including stepwise complex assembly, induced-fit binding, and dissociation pathways, by aligning the analysis with the underlying biochemical timeline.

Statistical Validation

We performed multi-trial statistical validation with 10 independent training runs to assess model variance. Statistical significance was evaluated using bootstrap confidence intervals (CI) (95% CI) across all trials. Performance metrics included R^2 coefficient (predictive accuracy), Spearman correlation (rank correlation with true stability), mean absolute error (absolute prediction error), and within-tolerance metrics (percentage of predictions within 10/15/20% of true values). The final model was selected based on balanced criteria of validation loss and test R^2 performance to ensure both training stability and predictive power.

TPS-Lite: Accessible Static Structure Analysis

For researchers without access to molecular dynamics simulations, we provide TPS-Lite, a simplified implementation requiring only a PDB structure and a PSF topology file. TPS-Lite performs static interface perturbation analysis using the same physicochemical feature engineering as full TPS (charge, hydrophobicity, structural indicators), by replaces the trained neural network with heuristic importance scoring based on residue properties and contact counts.

Given a protein complex structure, TPS-Lite: (1) identifies interface residues through spatial proximity (default cutoff: 6.0 Å), (2) calculates importance scores combining

electrostatic magnitude (weight: 0.5), hydrophobicity (0.3), and classifies interaction types (electrostatic, hydrophobic, aromatic, hydrogen-bond, or general) based on residue chemistry. While lacking the temporal resolution and causal precision of full TPS, TPS-Lite enables rapid interface mapping suitable for mutagenesis guidance, educational use, and initial screening. The complete TPS-Lite implementation is included in my github.

Code Implementation

All analyses were implemented in Python 3.9 using PyTorch Geometric 2.3.0 for graph neural networks[10], MDAnalysis 2.4.0 for trajectory processing[17], and scipy 1.10.0 for spatial algorithms. Visualization utilized matplotlib 3.7.0 and seaborn 0.12.0.

Code Availability

The full implementation of Temporal Perturbation Scanning (TPS) and TPS-Lite, including scripts for graph construction, neural network training, perturbation analysis, and visualization, is publicly available on GitHub: https://github.com/fodil13. All codes are released under a custom license permitting free academic use; commercial use requires a license.

Results

BoNT/B1-SYT1 and BoNT/B1-SYT2 Systems

Model Performance Across SYT1 and SYT2 Systems Reveals Distinct Dynamic Landscapes. To evaluate how the neural network interprets the interaction landscape across simulation time of each neurotoxin-receptor complex, we trained the TPS on BoNT/B1-SYT1 and BoNT/B1-SYT2 MD trajectories using identical computational protocol. For both systems, frames were processed into molecular graphs and split using the frame interleaving strategy (figure 1), ensuring that training, validation and test sets sampled the entire conformational ensemble without temporal leakage. Each system was subjected to 10 independent training trials, and performance was quantified using R^2 and Spearman rank correlation. Across the ten trials, SYT1 system exhibited higher variability in both R^2 and Spearman correlation. The mean test R^2 was 0.81 ± 0.035 with a 95% confidence interval (CI) of [0.78, 0.83] and the mean Spearman correlation reached 0.9 ± 0.013 (CI: [0.902, 0.919]). Training convergence curves (figure 2 E) showed a progressive and stable increase in validation performance, although the dispersion across trials indicated sensitivity to initial weights. These results suggest that SYT1 undergoes broader conformational fluctuations during the MD trajectory, creating a more diverse landscape for the model to learn. In the SYT2 system, test R^2 values collapsed to approximately zero across all trials (0.027 ± 0.041). This pattern reflects the SYT2 complex's intrinsic temporal stability rather than any limitations of the model. The

BoNT/B1-SYT2 complex maintains remarkable stability throughout the trajectory, exploring a significantly narrower conformational landscape than SYT1. Due to the minimal variance in target stability scores, even minor prediction errors result in high residual-to-total variance ratios, naturally driving R^2 toward zero. Critically, the model maintains substantial rank correlation (Spearman $\rho = 0.704 \pm 0.031$), demonstrating preserved capacity to capture relative temporal ordering of subtle fluctuations despite the absence of meaningful absolute variance. Thus, the R^2 collapse is a direct consequence of the complex's intrinsic temporal stability and does not indicate model underperformance.

Temporal Perturbation Scanning Reveals Distinct SYT1 and SYT2 Assembly Architectures. To dissect BoNT/B1 ternary complex assembly, we applied Temporal Perturbation Scanning (TPS) to quantify residue-level contributions across all interfacial components. Each residue's perturbation effect (Δ) reflects the change in predicted complex stability across six physicochemical perturbations. Interface strengths were calculated as the sum of Δ values (Sum Δ), with percentile ranks indicating relative importance across early, mid, and late stages.

SYT1 Employs Compensatory Assembly via GT1b Bridging. SYT1 shows compensatory binding (figure 3 A). The BoNT/B1-SYT1 interface is strong (Sum Δ = 3.447) and stable across assembly stages, while SYT1-GT1B engagement declines over time (0.472 → 0.144). This deficit is offset by BoNT/B1-GT1b strengthening (0.647 → 1.123), indicating a handoff mechanism where GT1b transitions from SYT1 recruitment to BoNT/B1 stabilization (figure 3 B). Key residues driving SYT1 compensation include LYS-1187/1188 and GLU1189 at the BoNT/B1-SYT1 interface, and R1242 engaging GT1b's ANE5AC-6 (sialic acid) and BGALNA-4 (β-D-N-acetylgalactosamine), forming a tripartite recognition system (figure 3 D).

SYT2 Features Modular, Autonomous Protein-Protein Binding. SYT2 exhibits a modular architecture (figure 3 A). The BoNT/B1-SYT2 interface is stronger (Sum Δ = 4.020) and largely autonomous, while SYT2-GT1b engagement shows a recoverable U-shaped profile (0.745 → 0.333 → 0.742), complementing but not compensating protein-protein interactions (figure 3 B). BoNT/B1-GT1b contact aids early recruitment but declines as the protein-protein interface matures. Residue-level analysis reveals that SYT2's autonomous binding is stabilized by K1187/1188, K46 and N56, with ganglioside engagement maintained through a hydrophobic patch (A64, L65 and I68) (figure 3 E).

BoNT/A1-SV2A and BoNT/A1-SV2C System

To further benchmark our TPS framework, we applied it to the BoNT/A1-SV2A and BoNT/A1-SV2C complexes. In contrast to BoNT/B1, which recognizes and binds with high affinity to an α-helix structure in SYT1 and SYT2, BoNT/A1 engages its membrane receptors SV2A and SV2C through a robust hydrogen bonds network between two β-strand structure [8,18.] : one from the toxin (residues 1141-1145) and the other from either SV2A (570-578) or SV2C (557-563).

TPS Reveals Fundamental Differences in Temporal Engagement Strategies. TPS uncovered strikingly different binding dynamics between the two receptor complexes (figure 4 A). The SV2C interface demonstrated progressive stabilization, with perturbation sensitivity increasing from 2.962 to 3.105 in mid-stages, reaching 3.368 in late stages. This pattern indicates a mature, continuously strengthening interaction that becomes more optimized over time. In contrast, the SV2A complex exhibited a non-monotonic stabilization, starting strong at 3.260 in early stages, dropping sharply to 2.508 during mid-stage engagement, before partially recovering to 2.714 in late stages. This temporal instability suggests that SV2A undergoes transient rearrangement or competitive displacement before establishing weaker final contacts. By deconstructing the temporal evolution of interface properties, our TPS framework reveals the physical mechanism underlying SV2C's established role as the high affinity receptor[19], showing how progressive stabilization and optimized hydrogen bond networks enable superior engagement compared to SV2A's transient interface instability.

Residue-Level Analysis Reveals Key Interface Contributors. Examination of the most influential BoNT/A1 residues further distinguished the two interfaces (figure 4 B). The top five SV2C-contributing residues showed 2-4× higher perturbation sensitivity than their SV2A counterparts. Thr-1145 emerged as the dominant SV2C contributor (Sum Δ = 0.4938), forming critical hydrogen bonds that were substantially weaker in SV2A binding (0.1416). Other key differential contributors included Met-1143 (0.2365 vs 0.1145) and Arg-1293 (0.1998 vs 0.1106), with all showing enhanced roles in the high-affinity complex.

Interaction Type Distribution Reveals Distinct Binding Chemistries. Analysis of physicochemical interaction types revealed striking differences in binding mechanisms. SV2C engagement was dominated by hydrogen bonds (45% of total interface strength), with electrostatic (25%) and hydrophobic (20%) contributions playing secondary roles (figure 4 C). Conversely, SV2A exhibited a more balanced interaction profile, with hydrogen bonds (35%), hydrophobic (30%), and electrostatic (20%) forces all contributing significantly. This suggests that SV2C achieves robust binding through focused, optimized hydrogen bond networks, while SV2A relies on a more distributed, less specialized interface.

By deconstructing the BoNT/A1-receptor interfaces temporally and mechanistically, TPS revealed not merely quantitative binding differences but fundamentally distinct engagement strategies. SV2C is progressively stabilized through focused hydrogen bond networks, while SV2A exhibits transient instability with more balanced force contributions. This analysis demonstrates TPS's capacity to move beyond describing binding interfaces to understanding their dynamic formation and physical determinants, providing a framework for mechanistic analysis of diverse biomolecular interactions.

Plasmolipin systems

Finally, we applied the TPS framework to plasmolipin embeded in two distinct membrane environments: a pure POPC bilayer and a lipid-raft membrane composed of POPC, cholesterol, and GM1[20]. In agreement with our previous observations, TPS Δ values were markedly higher in the lipid-raft system than in the POPC bilayer across all stages (Table

1). Specifically, the summed Δ values for the plasmolipin- lipid raft system reached 33.99, 32.03 and 31.75 for early, mid, and late simulations windows, respectively, while the corresponding values in the POPC system remained much lower, at 8.37, 8.66, and 8.47.

The enhanced interactions in the lipid raft system can be attributed to the specific lipid composition within plasmolipin's 6 Å contact shell (Table 2). Despite constituting only 13.1% of contact frequency, ceramide (CER160) contributed 35.1% of the total interaction score (16.48 Δ), demonstrating exceptionally strong binding to plasmolipin. In contrast, POPC-although representing 44.3% of membrane's contact—accounted for only 21.4% of the total interactions (10.06 Δ). Ganglioside polar components (β-glucose (BGLC), β-galactose (BGAL), β-D-N-acetylgalactosamine (BGALNA): collectively 19.6%) and cholesterol (CHL1: 15.4%) contributed 39.3% of total Δ score. This lipid-specific interaction profile—particularly ceramide's 5.6-fold higher Δ per contact compared to POPC (0.262 vs 0.047)—highlights its stabilizing role in anchoring plasmolipin within lipid raft microdomains.

Collectively, these results demonstrate that the TPS framework effectively captures and quantifies protein-lipid interactions, revealing significant differences in plasmolipin stability across distinct membrane environments. The method successfully identifies: i) Enhanced stabilization in lipid raft membranes (4x higher TPS Δ values compared to POPC), ii) Specific lipid contribution—particularly the disproportionate role of ceramide despite its low abundance, and iii) Temporal persistence of these interactions across simulation windows. This sensitivity to both lipid composition and interaction dynamics establishes TPS as a powerful tool for characterizing membrane protein stability in complex lipid environments.

Accessibility Extension: TPS-Lite for Static Structures

To extend accessibility to researchers without molecular dynamics capabilities, we implemented TPS-Lite—a heuristic, non-AI version requiring only static PDB coordinates and PSF topology files. TPS-Lite applies the same physicochemical feature engineering as full TPS but replaces the neural network with rule-based scoring based on contact density and residue properties.

Barnase-Barstar Validation. Applied to the classic Barnase-Barstar complex (PDB: 1BRS) with 6 Å cutoff, TPS-Lite successfully identified experimentally validated critical interface residues including Barnase's Arg-59 (score=0.804) and Arg-87 (score=0.800), and Barstar's Asp-35 (score=0.790) and Asp-39 (score=0.790).

BoNT/A1-SV2C Interface Mapping. For the neurotoxin-receptor complex (PDB: 6ES1), TPS-Lite identified key interface residues including SV2C's Lys-558 (score=0.790), Phe-563 (score=0.785), and Asn-559 (score=0.760)—forming the β-strand engagement motif. On the toxin side, key contributors included Tyr-1122 (score=0.787), Tyr-1149 (score=0.787), and Thr-1145/1146 (score=0.720).

Discussion

Stage-Resolved Perturbation Scanning as a New Computational Paradigm. The development of Temporal Perturbation Scanning (TPS) establishes a new computational paradigm for analyzing biomolecular interactions that moves beyond descriptive observation to causal interrogation. Traditional approaches in computational structural biology—whether molecular dynamics simulations or machine learning predictions—operate as passive observers, recording what occurs during binding but lacking the capacity to actively test causative mechanisms. TPS transforms this paradigm through its integration of three general computational principles: (1) graph-based representations that preserve both covalent and non-covalent relationships, (2) neural networks trained with strategic frame interleaving to sample conformational ensembles, and (3) multi-faceted in silico mutagenesis that systematically probes different physical interaction types. This integrated framework enables not only the identification of important residues, but the deconstruction of their specific stage-wise physicochemical contributions to binding.

Comprehensive Benchmarking Across Diverse Neurotoxin Systems. TPS demonstrates robust performance across four distinct neurotoxin-receptor systems spanning two different toxin families with distinct recognition mechanisms. For BoNT/B1 systems, TPS successfully captured the inherent dynamic differences between SYT1 and SYT2 complexes—revealing SYT1's compensatory assembly via ganglioside bridging versus SYT2's modular autonomy—despite dramatically different conformational landscapes ($R^2 = 0.79$ for flexible SYT1 vs $R^2 \approx 0$ for rigid SYT2). This demonstrates TPS's capacity to adapt to systems with varying degrees of conformational freedom while maintaining high rank correlation (Spearman $\rho = 0.93$ and 0.71, respectively). For BoNT/A1 systems, TPS revealed fundamentally different stage-dependent engagement strategies: SV2C's progressive stabilization versus SV2A's transient instability. This four-system validation across toxin families with different structural recognition elements (α-helix for BoNT/B1 vs β-strand for BoNT/A1) establishes TPS as a generalizable framework rather than a system-specific tool.

General Computational Principles with Broad Applicability. The core computational innovations of TPS—stage segmentation, perturbation scanning, and effect size attribution—represent generalizable principles applicable beyond neurotoxin systems. Notably, the stage division (early, mid, late) is not a framework limitation but a user-configurable parameter; researchers can define any number of stages or use clustering approaches to identify natural binding regimes specific to their systems. This flexibility ensures that TPS adapts to diverse binding kinetics, from rapid enzyme-substrate encounters to slow protein folding events. The framework's architecture is agnostic to specific molecular systems, requiring only molecular dynamics trajectories as input. This generality enables TPS to address fundamental questions across diverse domains:

in drug discovery, it can identify not just binding sites but the critical interaction windows where small molecules exert their effects; in enzyme engineering, it can reveal how mutations alter catalytic mechanisms across reaction coordinates; in membrane biology, it can decode the stage-wise orchestration of lipid-protein interactions. The discovery of four distinct engagement strategies across our benchmark systems (compensatory, modular, progressive, transient) suggests that binding kinetics represent a rich design space with evolutionary optimization along multiple axes, expanding traditional structure-function relationships to include stage-dependent design principles that computational frameworks must now capture.

Deciphering Complex Lipid-Protein Interactions. Beyond protein-protein systems, TPS provides unprecedented resolution in analyzing membrane environments through per-lipid-type interaction profiling. Where conventional methods report aggregate membrane-protein interactions or bulk binding energies, TPS dissects the individual contributions of each lipid species within heterogeneous bilayers. In our analysis of plasmolipin, this capability revealed how ceramide—despite constituting only 13.1% of contact frequency—dominated interactions (35.1% of total Δ score), while abundant POPC (44.3% contact frequency) contributed proportionally less (21.4%). This lipid-resolution insight moves beyond phenomenological descriptions of "lipid raft stabilization" to quantify how specific lipid chemistries—glycolipids, cholesterol, phospholipids—orchestrate membrane protein stability through stage-dependent interactions. The framework thus establishes a quantitative basis for understanding lipid selectivity, raft composition effects, and therapeutic lipid targeting in membrane biology.

Computational Efficiency and Practical Advantage. A key practical advantage of TPS lies in its dramatic computational efficiency compared to traditional approaches. Conventional methods for identifying critical residues typically require multiple independent MD simulations of mutant variants—each consuming hundreds of GPU hours and significant researcher time for setup and analysis. For comprehensive analysis of interfaces like those studied here (20-30 residues across multiple binding stages), this translates to months of computational time and substantial financial cost. In contrast, TPS extracts equivalent mechanistic information through in silico perturbations applied to a single wild-type trajectory, avoiding the need for separate mutant simulations and reducing computational requirements by approximately N-fold (where N is the number of residues studied), with statistical uncertainty quantified through bootstrap confidence intervals. This efficiency gain enables rapid comparative analysis across multiple systems—as demonstrated by our four-system benchmark—that would be computationally prohibitive through brute-force mutation simulations, democratizing detailed mechanistic analysis for laboratories without access to supercomputing resources.

Validation and Explanatory Power Beyong Recaputulation. TPS successfully recapitulates experimentally validated critical residues across all four neurotoxin systems—including Lys-1187/1188 and Glu-1189 for BoNT/B1-SYT1, Lys-1187/1188 for BoNT/B1-SYT2, and Thr-1145, Met-1143, Arg-1293 for BoNT/A1-SV2C—demonstrating its predictive accuracy across diverse recognition mechanisms. More significantly, TPS provides explanatory power beyond mere identification: it reveals why these residues matter through their specific roles in compensatory bridging (SYT1)[3], modular autonomy (SYT2), progressive stabilization (SV2C), or transient engagement (SV2A)[19]. This transition from correlative to causative understanding across multiple biological contexts represents a fundamental advance over existing computational approaches, which typically identify important residues without elucidating their system-specific mechanistic roles or stage-dependent contributions.

Biological Implications and Medical Relevance. The differential binding strategies identified by TPS across four systems have direct implications for understanding botulism pathogenesis and developing therapeutic interventions. The discovery that SYT1 employs compensatory assembly while SYT2 features modular autonomy, combined with SV2C's progressive optimization versus SV2A's transient instability, reveals that BoNT serotypes have evolved distinct stage-dependent strategies for host receptor engagement. These differences suggest serotype-specific therapeutic approaches: SYT1's ganglioside dependence may render it vulnerable to glycolipid synthesis inhibitors, while SYT2's protein-centric interface might be better targeted by direct protein-protein interaction disruptors. Similarly, SV2C's progressive hydrogen bond network presents different vulnerabilities than SV2A's transient interface. More broadly, our findings across four systems contribute to the growing recognition that molecular recognition represents not merely complementary shapes or charge distributions but orchestrated sequential processes with multiple evolutionary solutions. This perspective aligns with emerging views in systems biology emphasizing stage coordination in biological function and suggests that binding sequence—the specific order of interaction formation—may prove as important as binding affinity for understanding biological specificity and developing targeted interventions.

Limitations and Framework Extensions. While TPS represents a significant computational advance validated across four diverse systems, several limitations merit consideration. First, our perturbations represent artificial computational modifications; future work should validate key predictions through experimental mutagenesis and binding assays across multiple systems. Second, the framework currently analyzes individual trajectories; integration across multiple simulation replicas could provide more robust statistical estimates. Third, the stage flexibility—while a strength—requires users to make informed decisions about stage definition, though automated clustering approaches could address this. Fourth, while TPS successfully captured diverse binding mechanisms (protein-protein, protein-glycolipid), future applications to more

challenging systems—particularly those involving large conformational changes, disordered regions, or multi-domain rearrangements—would further stress-test the framework's robustness. These limitations suggest clear directions for framework extension: (1) incorporating experimental data through multi-modal learning, (2) developing adaptive stage segmentation algorithms that automatically identify natural binding phases, (3) extending perturbation strategies to include post-translational modifications and allosteric effects, and (4) creating standardized benchmarks for stage-resolved interaction analysis across broader molecular classes. Importantly, the modular architecture of TPS enables these extensions without fundamental redesign, demonstrating the framework's capacity for evolution as computational and experimental methods advance.

Conclusion

Temporal Perturbation Scanning provides a generalizable computational framework for moving beyond describing molecular interactions to understanding their dynamic formation and physical determinants. By decomposing interactions into fundamental physical components across user-definable binding stages, TPS reveals not simply which residues participate but how and when they contribute through specific forces—transforming our capacity to decode the complex language of molecular recognition. The framework's computational efficiency (extracting mechanistic insight from single trajectories) and comprehensive validation (across four distinct neurotoxin systems) make it both practical and broadly applicable across natural sciences.

As cryo-EM captures ever more transient states and AI generates ensembles of plausible conformations, frameworks like TPS will become increasingly essential for extracting mechanistic understanding from complex dynamic datasets. The principles established here—configurable stage analysis, systematic perturbation, and interpretable attribution—represent general computational strategies applicable across natural sciences. By making stage-dependent mechanisms interpretable and computationally accessible across diverse systems, TPS transforms molecular recognition from observational science to engineering discipline, where researchers can not only understand but potentially orchestrate biomolecular interactions through sequential stages across diverse biological contexts.

Author Contributions

F.A.: Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Data Curation, Writing – Original Draft, Writing – Review & Editing, Visualization.

J.F.: Supervision, Writing – Original Draft, Writing – Review & Editing, Resources.

Competing Interests

F.A is the copyright holder of the TPS & TPS-Lite software. J.F declares no competing interests.

References

1. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J Comput Chem* **26**, 1781–1802 (2005).
2. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).
3. Azzaz, F., El Far, O. & Fantini, J. Membrane constraints reshape synaptotagmin recognition by botulinum neurotoxin B1. *bioRxiv* 2025.08.27.672333 (2025) doi:10.1101/2025.08.27.672333.
4. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res* **28**, 235–242 (2000).
5. Park, S., Khalili-Araghi, F., Tajkhorshid, E. & Schulten, K. *Free Energy Calculation from Steered Molecular Dynamics Simulations Using Jarzynski's Equality*.
6. Congreve, M., Carr, R., Murray, C. & Jhoti, H. A 'Rule of Three' for fragment-based lead discovery? *Drug Discov Today* **8**, 876–877 (2003).
7. Benner, S. A. & Sismour, A. M. Synthetic biology. *Nat Rev Genet* **6**, 533–543 (2005).
8. Azzaz, F., Hilaire, D. & Fantini, J. Structural basis of botulinum neurotoxin serotype A1 binding to human SV2A or SV2C receptors. *Chem Biol Interact* **373**, 110384 (2023).
9. Azzaz, F., Hilaire, D. & Fantini, J. Structural Basis of Botulinum Toxin Type F Binding to Glycosylated Human SV2A: In Silico Studies at the Periphery of a Lipid Raft. *Biomolecules* **12**, (2022).
10. Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O. & Dahl, G. E. Neural Message Passing for Quantum Chemistry. http://arxiv.org/abs/1704.01212 (2017).
11. Fantini, J., Matveeva, M., Lefebvre, M. & Chahinian, H. What Is life? Rethinking Biology in Light of Fundamental Parameters. *Life* **14**, (2024).
12. Huang, J. *et al.* CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods* **14**, 71–73 (2017).
13. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J Mol Biol* **157**, 105–132 (1982).
14. Maneewongvatana, S. & Mount, D. M. *Analysis of Approximate Nearest Neighbor Searching with Clustered Point Sets ⋆*. (1999).
15. Ioffe, S. & Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. http://arxiv.org/abs/1502.03167 (2015).
16. Loshchilov, I. & Hutter, F. Decoupled Weight Decay Regularization. http://arxiv.org/abs/1711.05101 (2019).
17. Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J Comput Chem* **32**, 2319–2327 (2011).

18. Benoit, R. M. *et al*. Crystal structure of the BoNT/A2 receptor-binding domain in complex with the luminal domain of its neuronal receptor SV2C. *Sci Rep* **7**, 43588 (2017).
19. Dong, M. *et al*. SV2 Is the Protein Receptor for Botulinum Neurotoxin A. *Science (1979)* **312**, 592–596 (2006).
20. Azzaz, F. *et al*. Structure of the Myelin Sheath Proteolipid Plasmolipin (PLLP) in a Ganglioside-Containing Lipid Raft. *Front. Biosci. (Landmark Ed)* **28**, 157 (2023).

Figure legends

Figure 1. The Temporal Perturbation Scanning (TPS) Framework. (Upper pannel) Worflow for AI-driven deconstruction of biomolecular recognition. Molecular dynamics trajectories of a biomolecular complex (e.g, a neurotoxin bound (BoNT/B1 1150-1280, colored in white) to its receptor (synaptotagmin 1 (SYT1 30-85) or 2 (SYT2 30-85), colored in blue) are converted into a series of molecular graphs. Our graph creation code engineers 13-dimensional node features (including electrostatic, hydrophobic and steric properties) and constructs both covalent and non-covalent edges. (Lower panel) **Architecture:** The model processes the molecular graphs using a multi-layer Graph Convolutional Network (GCN). The architecture performs message passing along both covalent and non-covalent edges to learn a complex, hierarchical representation of the molecular system, culminating in a predicted stability score. **Frame Interleaving Training Strategy:** To respect the temporal correlation of MDS data, frames are split into training, validation and test sets using a consistent interleaving pattern (every 3rd frame). This strategy prevents data leakage by avoiding contiguous blocks of time in a single split and crucially, forces the model to learn from and generalize across the entire ensemble of conformational states sampled by the simulation, rather than memorizing a specific temporal segment. **Multi-run validation:** The training process is repeated over multiple independent trials (n=10). Performance metrics (R^2, Spearman correlation) are aggregated, with results reported as mean ± standard deviation and confidence intervals, quantifying the pure variance of the model and guaranteeing the reliability of the findings.

Figure 2. Statistical performance comparison of the TPS stability predictor for the BoNT/B1-SYT1 and BoNT/B1-SYT2 systems. For both systems, MD trajectories were converted into molecular graphs and processed using the same Temporal Interleaving training protocol. Each model was trained over 10 independent trials, and performance metrics were aggregated to quantify model variance. (A, G) Boxplots showing the distribution of test-set R^2 scores across 10 independent runs for SYT1 (A) and SYT2 (G). Individual data points represent each trial. (B, H) Histograms of the R^2 distributions for SYT1 (B) and SYT2 (H), with dashed lines marking the mean performance. (C, I) Boxplots summarizing Spearman rank correlation coefficients (ρ) across trials for SYT1 (C) and SYT2 (I). (D, J) Histograms of Spearman correlations for SYT1 (D) and SYT2 (J). (E, K) Training convergence curves showing the mean validation R^2 ± standard deviation across trials for SYT1 (E) and SYT2 (K). These curves reflect the stability and reproducibility of the optimization process. (F, L) Statistical summary panels reporting the means ± standard deviation and the 95% confidence intervals for R^2 and Sperman ρ for SYT1 (F) and SYT2 (L) systems. Together, these panels compare how the TPS learns stability patterns for each receptor. The same protocol was used for SYT1 and SYT2; differences in model

performance arise solely from differences in the underlying MD conformational ensembles.

Figure 3. Comparative Interface for BoNT/B1-SYT1-GT1b and BoNT/B1-SYT2-GT1b systems. Total interface strength comparison showing SYT2's stronger autonomous protein-protein binding (blue) versus SYT1's compensatory GT1b bridging (orange) (A). Temporal dynamics reveal SYT1's compensatory handoff (dashed lines) versus SYT2's modular stability (solid lines) (B). Network architecture models illustrate compensatory (SYT1) versus modular (SYT2) assembly strategies; edge thickness ∝ interface strength (C). Top residue contributors for SYT1 and SYT2 systems, ranked by Sum Δ and colored by perturbation type sensitivity (D-E).

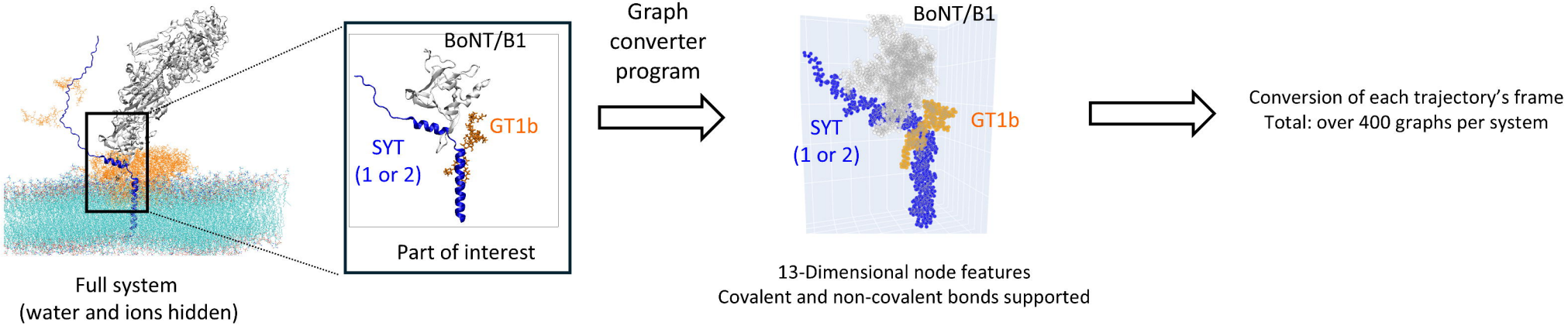
Figure 4. Temporal Perturbation Scanning Analysis of BoNT/A1 Binding to SV2A vs SV2C. Bar plots show the Sum Δ (interface perturbation sensitivity) for SV2C (blue) and SV2A (purple) complexes across early, mid and late binding stages (A). Horizontal bar chart siplaying the mean Sum Δ values for the top five contributing residues from BoNT/A1 in each complex (B). Pie charts comparing the percentage contribution of different physicochemical interaction types to total interface strength (C). Hsd is for histidine, BGALNA is for β-D-N-acetylgalactosamine, and BGAL is for β-galactose. The numbers reflect their position in the sequence (from 1 to 7).

Table 1. Total TPS Δ values for plasmolipin in different membrane environments

System	Early Sum Δ	Mid Sum Δ	Late Sum Δ
Plasmolipin-Lipid Raft	33,987	32,03	31,753
Plasmolipin-POPC	8,37	8,657	8,465

Table 2. Lipid composition within 6 Å of plasmolipin and corresponding TPS Δ interaction scores in the lipid raft system. Percentages indicate relative abundance of each lipid type within the contact shell.

Lipid	Contact Frequency (%)	TPS Δ (sum)	Δ per Contact
POPC	44.3	10.06	0.047
CER160	13.1	16.48	0.262
CHL1	15.4	6.07	0.082
BGLC	9.8	7.74	0.165
BGAL	5.8	3.25	0.116
ANE5AC	5.0	1.43	0.060
BGALNA	4.0	1.34	0.071
BGAL	2.7	0.65	0.050



Architecture

INPUT: [13D Features]

↓

GCN LAYERS: 128 → 64 → 32

(BatchNorm + ReLU + Dropout)

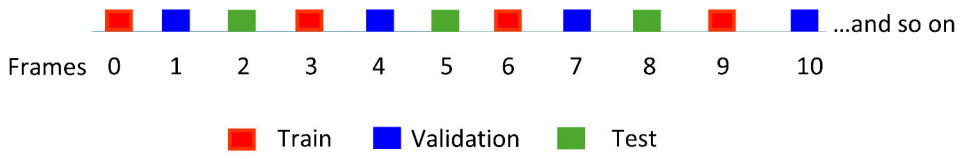
↓

INTERFACE FOCUS: BoNT/B1-SYT and BoNT/B1-GT1b and SYT-GT1b

↓

OUTPUT: Trained Model for Comparative Binding Mechanism Analysis

Frame Interleaving Strategy



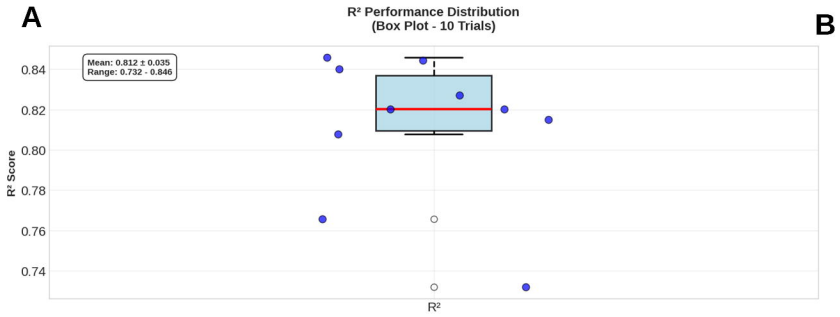
Multi-run validation

Trial 1: Different Random Seed $R^2 = 0.XX \pm 0.YY$ Spearman = 0.XX
Trial 2: Different Random Seed $R^2 = 0.XX \pm 0.YY$ Spearman = 0.XX
Trial 3: Different Random Seed $R^2 = 0.XX \pm 0.YY$ Spearman = 0.XX
...
Trial 10: Different Random Seed $R^2 = 0.XX \pm 0.YY$ Spearman = 0.XX

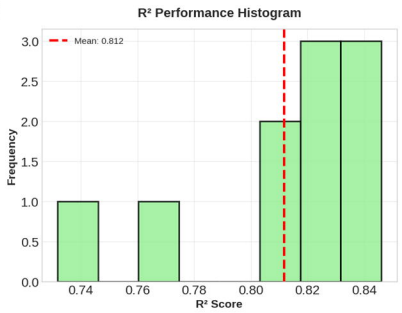
FINAL: $R^2 = 0.XX \pm 0.YY$ (95% CI) - Pure Model Variance

BoNT/B1-SYT1 System

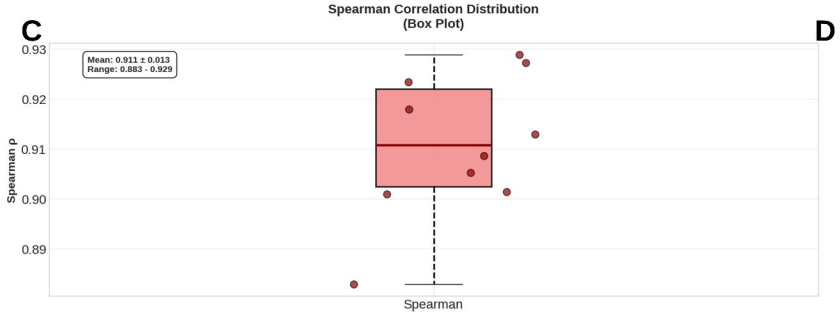
A



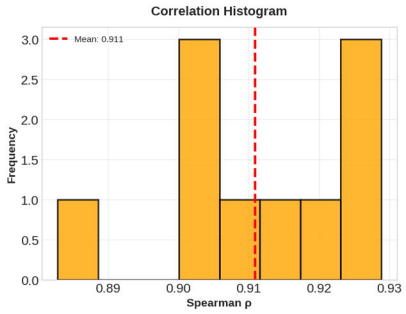
B



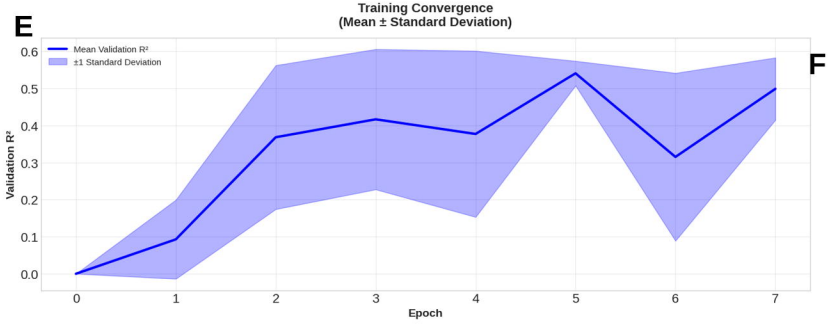
C



D



E



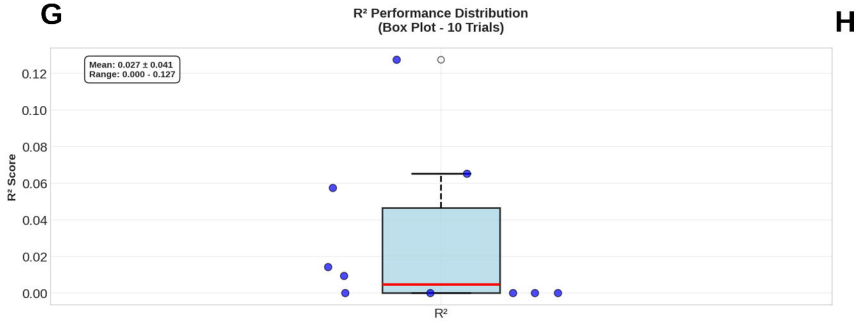
F

Statistical results summary
Performance (n=10)

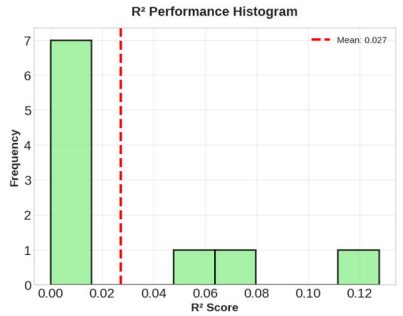
- R^2 : 0.812 ± 0.035
- 95 % CI : [0.789, 0.830]
- Spearman p : 0.911 ± 0.013
- 95 % CI : [0.902, 0.919]

BoNT/B1-SYT2 System

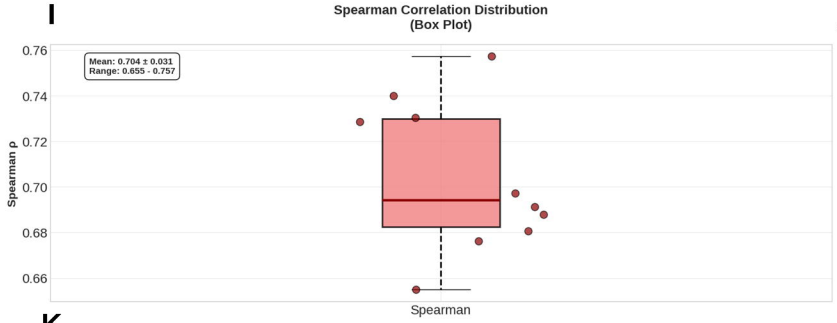
G



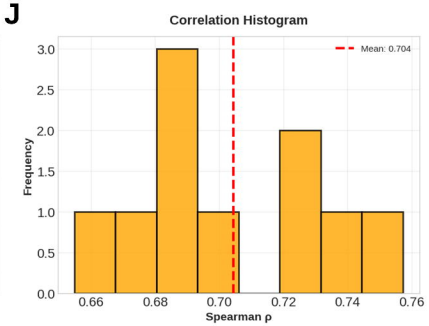
H



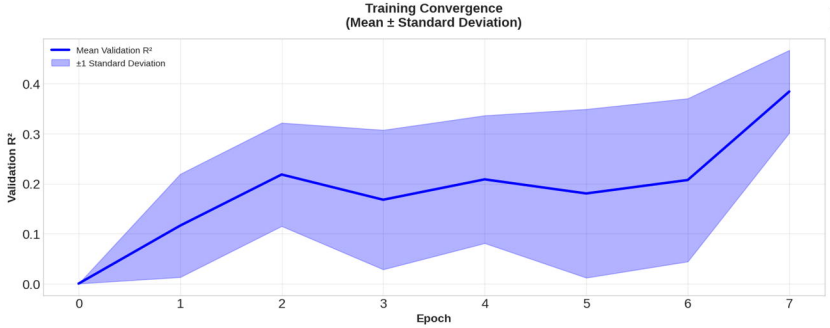
I



J



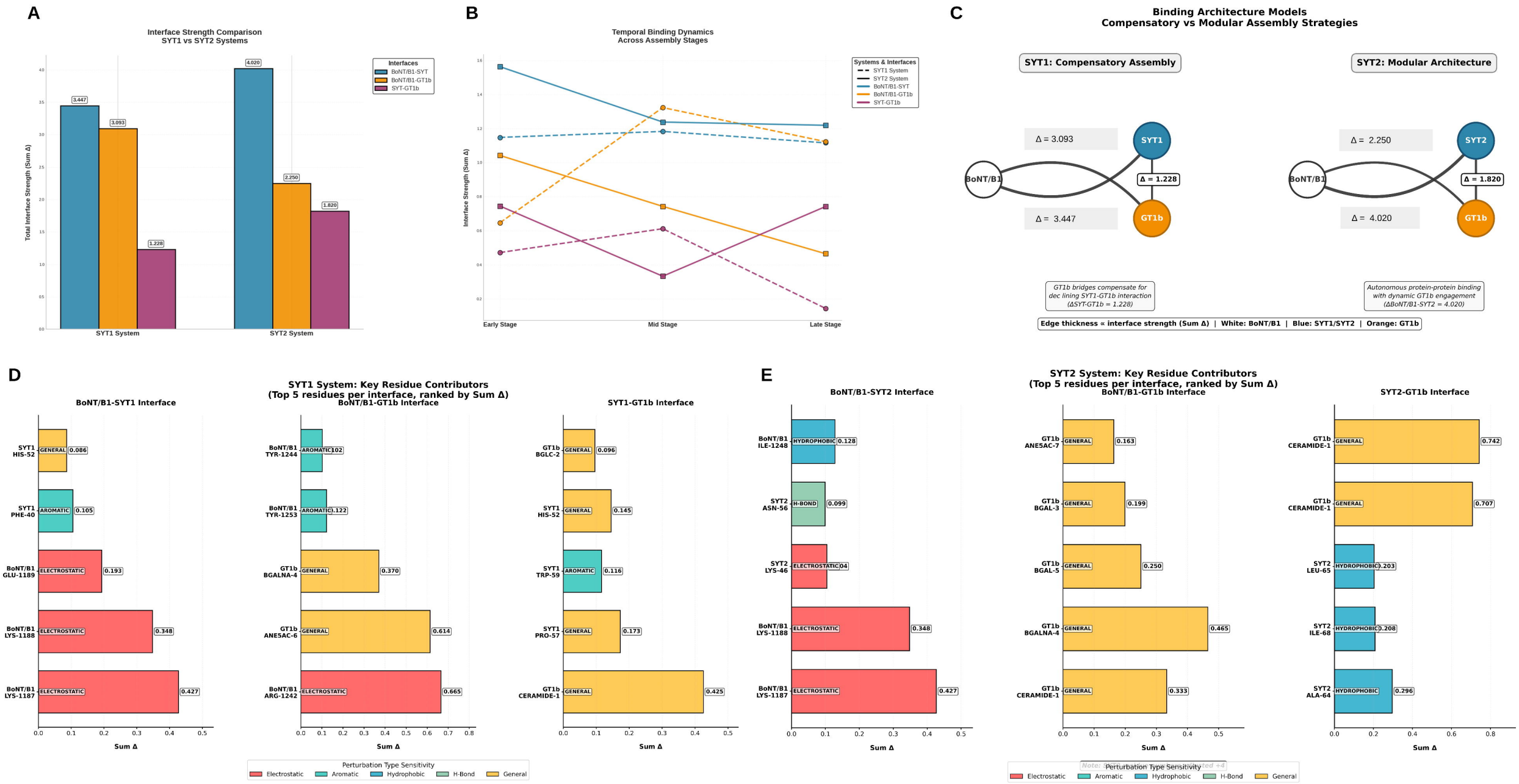
K

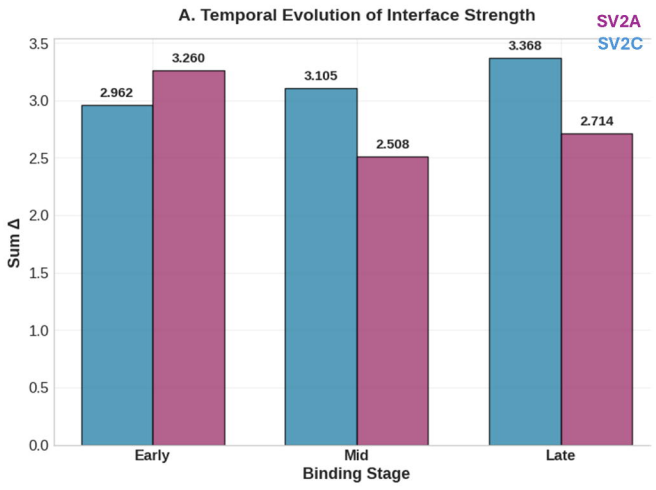


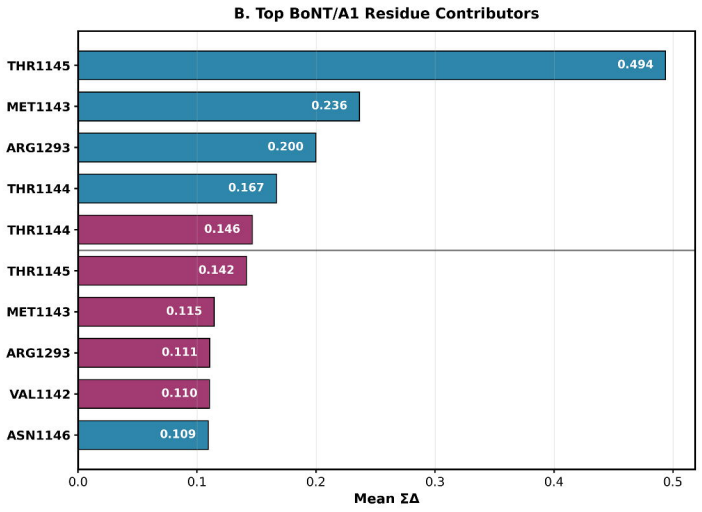
L

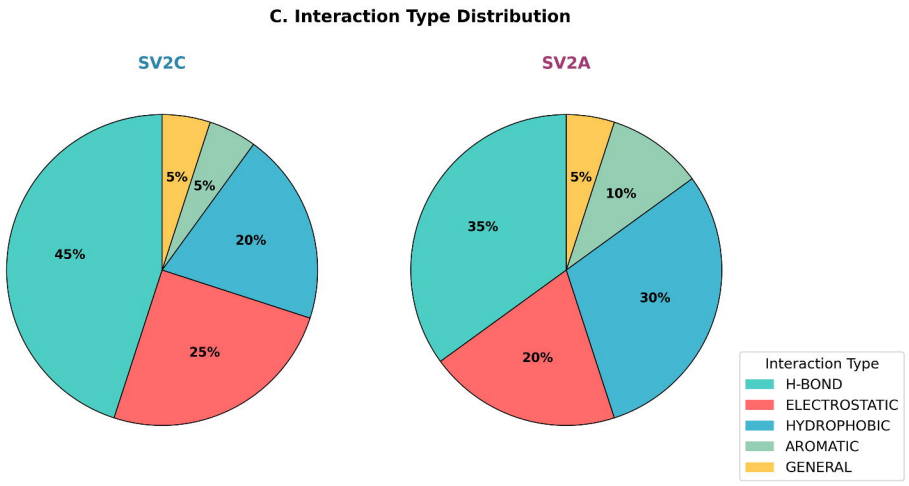
Statistical results summary
Performance (n=10)

- R^2 : 0.027 ± 0.041
- 95 % CI : [0.005, 0.053]
- Spearman p : 0.704 ± 0.031
- 95 % CI : [0.686, 0.723]


A
Interface Strength Comparison
SYT1 vs SYT2 Systems
Total Interface Strength (Sum Δ)
3.447
3.093
1.228
4.020
2.250
1.820
SYT1 System
SYT2 System
Interfaces
BoNT/B1-SYT
BoNT/B1-GT1b
SYT-GT1b
B
Temporal Binding Dynamics
Across Assembly Stages
Interface Strength (Sum Δ)
Early Stage
Mid Stage
Late Stage
Systems & Interfaces
SYT1 System
SYT2 System
BoNT/B1-SYT
BoNT/B1-GT1b
SYT-GT1b
C
Binding Architecture Models
Compensatory vs Modular Assembly Strategies
SYT1: Compensatory Assembly
SYT2: Modular Architecture
Δ = 3.093
SYT1
BoNT/B1
Δ = 1.228
Δ = 3.447
GT1b
Δ = 2.250
SYT2
BoNT/B1
Δ = 1.820
Δ = 4.020
GT1b
GT1b bridges compensate for dec lining SYT1-GT1b interaction (ΔSYT-GT1b = 1.228)
Autonomous protein-protein binding with dynamic GT1b engagement (ΔBoNT/B1-SYT2 = 4.020)
Edge thickness ∝ interface strength (Sum Δ) | White: BoNT/B1 | Blue: SYT1/SYT2 | Orange: GT1b
D
SYT1 System: Key Residue Contributors
(Top 5 residues per interface, ranked by Sum Δ)
BoNT/B1-SYT1 Interface
SYT1 HIS-52 GENERAL 0.086
SYT1 PHE-40 AROMATIC 0.105
BoNT/B1 GLU-1189 ELECTROSTATIC 0.193
BoNT/B1 LYS-1188 ELECTROSTATIC 0.348
BoNT/B1 LYS-1187 ELECTROSTATIC 0.427
BoNT/B1-GT1b Interface
BoNT/B1 TYR-1244 AROMATIC 0.102
BoNT/B1 TYR-1253 AROMATIC 0.122
GT1b BGALNA-4 GENERAL 0.370
GT1b ANE5AC-6 GENERAL 0.614
BoNT/B1 ARG-1242 ELECTROSTATIC 0.665
SYT1-GT1b Interface
GT1b BGLC-2 GENERAL 0.096
SYT1 HIS-52 GENERAL 0.145
SYT1 TRP-59 AROMATIC 0.116
SYT1 PRO-57 GENERAL 0.173
GT1b CERAMIDE-1 GENERAL 0.425
Sum Δ
Perturbation Type Sensitivity
Electrostatic
Aromatic
Hydrophobic
H-Bond
General
E
SYT2 System: Key Residue Contributors
(Top 5 residues per interface, ranked by Sum Δ)
BoNT/B1-SYT2 Interface
BoNT/B1 ILE-1248 HYDROPHOBIC 0.128
SYT2 ASN-56 H-BOND 0.099
SYT2 LYS-46 ELECTROSTATIC 0.104
BoNT/B1 LYS-1188 ELECTROSTATIC 0.348
BoNT/B1 LYS-1187 ELECTROSTATIC 0.427
BoNT/B1-GT1b Interface
GT1b ANE5AC-7 GENERAL 0.163
GT1b BGAL-3 GENERAL 0.199
GT1b BGAL-5 GENERAL 0.250
GT1b BGALNA-4 GENERAL 0.465
GT1b CERAMIDE-1 GENERAL 0.333
SYT2-GT1b Interface
GT1b CERAMIDE-1 GENERAL 0.742
GT1b CERAMIDE-1 GENERAL 0.707
SYT2 LEU-65 HYDROPHOBIC 0.203
SYT2 ILE-68 HYDROPHOBIC 0.208
SYT2 ALA-64 HYDROPHOBIC 0.296
Sum Δ
Perturbation Type Sensitivity
Electrostatic
Aromatic
Hydrophobic
H-Bond
General


A. Temporal Evolution of Interface Strength
SV2A
SV2C
Sum Δ
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
2.962
3.260
3.105
2.508
3.368
2.714
Early
Mid
Late
Binding Stage


B. Top BoNT/A1 Residue Contributors
THR1145
MET1143
ARG1293
THR1144
THR1144
THR1145
MET1143
ARG1293
VAL1142
ASN1146
0.494
0.236
0.200
0.167
0.146
0.142
0.115
0.111
0.110
0.109
0.0
0.1
0.2
0.3
0.4
0.5
Mean ΣΔ


C. Interaction Type Distribution
SV2C
45%
25%
20%
5%
5%
SV2A
35%
20%
30%
10%
5%
Interaction Type
H-BOND
ELECTROSTATIC
HYDROPHOBIC
AROMATIC
GENERAL