

Article

On the Structural Precision Required for Ligand-Specific Elastic Network Dynamics of AI Co-Folded GPCR Structures

Benjamin R. Cummins ^{1,*} , Christina G. Fragel ²  and Joyita Faruk ³

¹ Department of Pharmacology, Biochemistry and Experimental Therapeutics, LSU Health Sciences Center, New Orleans, LA 70112, USA

² School of Biological Sciences, University of Nebraska-Lincoln, Lincoln, NE 68588, USA; cfragel2@huskers.unl.edu

³ Bioinformatics and Computational Biology, Chulalongkorn University, Bangkok 10330, Thailand

* Correspondence: bcumm3@lsuhsc.edu or bcummins96@gmail.com

Abstract

Background/Objectives: AI-based protein–ligand co-folding methods such as Boltz-2 are able to reproduce experimentally determined ligand binding poses at G protein-coupled receptors (GPCRs) with increasing accuracy and precision. As these models advance, they raise several natural questions: assuming the correct placement of a ligand, do current methods generate structural ensembles that are sufficient as a starting point for predictive analysis of receptor dynamics? Does the noise introduced between predictions render these unreliable, or can they be trusted to a degree comparable with experimental structures (which themselves carry inherent noise)? We investigate this at the serotonin 5-HT_{2A} receptor (5-HT_{2A}R), leveraging a rich panel of experimentally determined structures and prior benchmarking of Boltz-2 at this receptor. **Methods:** First, a reference dynamic landscape is established by performing anisotropic network model (ANM) analysis across 11 published cryo-EM and crystal structures trimmed to a uniform 236-residue analysis mask spanning the structured receptor (TM1-H8). We then ask whether Boltz-2 co-folding ensembles recapitulate these experimentally derived dynamics. **Results:** All nine Gq-coupled agonist structures converge to near-identical dynamic subspaces (mean RMSIP = 0.970 ± 0.019), discriminating activation state but not agonist identity, chemotype, or hallucinogenic status. Principal component analysis of the Boltz-2 ensemble yields moderate, highly uniform overlap with the experimental ANM modes (mean RMSIP = 0.599 ± 0.007), with the first two principal components capturing 56.9% of structural variance. **Conclusions:** Replicate predictions of the same ligand are highly self-consistent yet show no usable ligand-specific discrimination (within- versus between-ligand effect size is negligible). These findings suggest a quantitative benchmark for the precision required to derive elastic network dynamics from predicted structures and support the distinction between learned pose recapitulation and genuine biophysical prediction.



Academic Editor: Alain Couvineau

Received: 30 June 2026

Revised: 27 July 2026

Accepted: 22 September 2026

Published: 9 October 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: 5-HT_{2A} receptor; normal mode analysis; elastic network model; Boltz-2; AI co-folding; cryo-EM; GPCR dynamics; conformational ensemble; structural prediction

1. Introduction

AI-based methods for predicting protein–ligand complex structures have advanced rapidly, with tools such as Boltz-2 [1], AlphaFold 3 [2], and Chai-1 [3], among others, achieving remarkable accuracy in reproducing experimentally determined binding poses. These co-folding approaches differ fundamentally from classical docking methods: rather

than placing a ligand into a pre-determined rigid receptor structure, they predict the receptor–ligand complex *de novo* from sequence and ligand information, allowing the receptor backbone to adopt a conformation influenced by the bound ligand. This flexibility raises an intriguing possibility: that co-folded structural ensembles might capture not only the static binding pose but also aspects of ligand-dependent receptor dynamics.

Normal mode analysis (NMA) using elastic network models (ENMs) provides a computationally efficient framework for characterizing protein dynamics from structure. Anisotropic network models (ANMs) and Gaussian network models (GNMs) derive the intrinsic flexibility of a protein from its C α contact topology represented as the matrix of pairwise inter-residue distances within a defined cutoff [4,5]. NMA approximates the essential dynamics obtained from molecular dynamics (MD) simulations, typically achieving root mean square inner product (RMSIP) values of 0.7–0.8 against MD principal components for well-characterized systems [6]. These methods have been extensively applied to GPCRs for comparing activation states, identifying allosteric pathways, and characterizing conformational transitions [7,8]. The ProDy 2.0 framework enables systematic ensemble NMA across panels of related structures [9,10]. Throughout this work, NMA is used for the general approach, ANM when describing the specific calculations and quantities reported, and GNM only where it is used as an independent secondary check.

ENM-based dynamics are sensitive to precise C α positioning. Sub-angstrom shifts can alter which contacts fall within the distance cutoff, reshuffling the Hessian matrix and producing different eigenvectors. This sensitivity raises a practical question: do AI co-folding methods provide sufficient structural precision for the resulting ENM dynamics to be meaningful, and meaningful relative to what reference? Experimentally determined structures can be thought of as the closest available approximation to a static posture, but they are not noise-free; cryo-EM and crystallographic coordinates carry their own resolution-limited uncertainty, and any ENM analysis inherits it. The relevant comparison is therefore not whether predicted structures are perfect, but whether they support dynamics inferences as reliably as the experimental structures already utilized for this purpose. Classical docking cannot enter this comparison at all because it embeds ligands into a rigid receptor backbone, meaning any dynamics analysis would simply recapitulate the single input structure.

Instead, co-folding methods, which allow the receptor to organize around the ligand, offer the potential to capture ligand-dependent dynamics from predictions, given their ability to incorporate receptor flexibility. The serotonin 5-HT_{2A} receptor (5-HT_{2AR}) provides an ideal system for this investigation. As the primary target of classical psychedelics and a leading neuropsychiatric drug target [11,12], it has been the subject of extensive structural characterization. Inactive-state crystal structures [13] have been complemented by active-state experimental structures spanning tryptamines (the endogenous agonist serotonin, psilocin, DMT), ergolines (LSD, BOL-148), phenethylamines (mescaline, 25CN-NBOH, RS130-180), and the non-hallucinogenic ergoline lisuride [14–16]. This pharmacological diversity within a single receptor offers an exceptionally well-controlled testbed.

A prior study benchmarked Boltz-2, Uni-Mol Docking v2, and AutoDock Vina 1.2.1 with a range of tryptamine and phenethylamine test ligands against the recently published cryo-EM structures of psilocin, DMT, and mescaline bound to 5-HT_{2AR}. Boltz-2 demonstrated not only superior pose accuracy (quantified as the RMSD of core scaffold features between the predicted poses and the published cryo-EM ligand orientations) but also notable precision across independent predictions, with replicate structures appearing nearly superimposable over the transmembrane domain [17]. This advantage was attributed to the flexibility of co-folding, but must be interpreted cautiously, as Boltz-2 was trained on several 5-HT_{2AR} structures deposited before its training cutoff, and recent evidence suggests that co-folding models may recapitulate training-set conformations

rather than learning transferable biophysical principles. The present study extends this previous static structural evaluation into the dynamic domain by asking: does Boltz-2's pose accuracy translate into structural ensembles that recapitulate the dynamics encoded in experimental structures?

Because elastic network modes are derived deterministically from a structure's contact topology, its dynamics are strictly encoded in its coordinates. Comparing dynamics is therefore not independent of comparing structures, but neither is it equivalent. A raw coordinate comparison such as backbone RMSD is dominated by displacements in flexible loops and termini and weights all atomic differences equally. Normal mode analysis instead extracts the collective, low-frequency directions in which the contact network is most deformable, isolating the global motions most relevant to activation and allostery while down-weighting local coordinate scatter. Two structures can be similar in coordinate RMSD yet differ in their soft-mode directions, or the reverse; the RMSIP between their mode subspaces reports whether they would deform along the same functional axes. This is information that static distance alone does not provide. It is this collective-motion view, not a restatement of atomic difference, that motivates layering dynamics analysis on top of the static pose accuracy established previously.

An important caveat frames this investigation. Boltz-2's training data include essentially all qualifying PDB structures deposited before its training cutoff [1]. Several 5-HT_{2A}R structures were available before this cutoff (including 6A93, 6WHA, 6WGT, 7VOD, 7WC4, and 7RAN, among others), providing the model with a variety of receptor conformations and ligand poses at this target. While many of the specific psychedelic agonist complexes used in this study were deposited after the cutoff and are therefore excluded from the training set, the receptor itself is well represented. Growing evidence suggests that diffusion-based co-folding models may, in many cases, recapitulate poses from their training data rather than learning transferable biophysical principles [18]. This distinction between memorization and generalization is critical for interpreting any dynamics derived from these predictions and is addressed directly in the Section 4.

2. Materials and Methods

2.1. Experimental Structure Panel

Eleven experimentally determined structures of human 5-HT_{2A}R were selected to maximize pharmacological diversity across activation state and ligand chemotype (Table 1). All agonist-bound cryo-EM structures were determined in complex with mini-Gαq heterotrimers.

Table 1. Experimental structure panel for comparative NMA.

PDB	Chain	Method	Ligand	Chemotype	Reference
6A93	A	X-ray	Risperidone	Antipsychotic (inactive)	Kimura et al. 2019 [13]
6WH4	A	X-ray	Methiothepin	Inverse agonist	Kim et al. 2020 [14]
9ARX	A	Cryo-EM	Serotonin	Tryptamine (endogenous)	Gumpper et al. 2025 [15]
9AS7	A	Cryo-EM	Psilocin	Tryptamine	Gumpper et al. 2025 [15]
9AS1	A	Cryo-EM	DMT	Tryptamine	Gumpper et al. 2025 [15]
9AS3	A	Cryo-EM	LSD	Ergoline	Gumpper et al. 2025 [15]
9ARZ	A	Cryo-EM	BOL-148	Ergoline	Gumpper et al. 2025 [15]
8UWL	A	Cryo-EM	Lisuride	Ergoline (non-hallucinogenic)	Lyu et al. 2024 [16]
6WHA	A	Cryo-EM	25CN-NBOH	Phenethylamine	Kim et al. 2020 [14]
9AS5	A	Cryo-EM	Mescaline	Phenethylamine	Gumpper et al. 2025 [15]
9AS9	A	Cryo-EM	RS130-180	Phenethylamine	Gumpper et al. 2025 [15]

2.2. Analysis Mask and Structure Preparation

A critical methodological consideration in comparative ENM studies is ensuring that all structures are analyzed over the identical set of residues. Structures from different laboratories resolve different subsets of receptor residues, particularly in loop regions, and naive inclusion of heterogeneous coordinate sets can introduce artifacts. We therefore defined a uniform analysis mask as the intersection of resolved C α positions across all 11 structures, yielding 236 common residues spanning positions 84–390 of the receptor sequence. The mask comprises the structured transmembrane core through helix 8; it is not contiguous, excluding intracellular loop 3 (ICL3; residues 263–314) and a small number of short, variably resolved surface loops that are not resolved in every structure. The conserved elements critical for ligand recognition and activation (the orthosteric pocket anchored by Asp155^{3,32}, the Trp336^{6,48} toggle switch, and the P-I-F activation motif) fall within the mask.

All structure preparation was performed programmatically using ProDy 2.0 [9,10]. For each structure, the 5-HT_{2A}R receptor chain was selected explicitly (Table 1). Once the receptor chain was identified, G-protein subunits, nanobodies, single-chain variable fragments, ligands, and non-protein components were excluded, as were crystallographic fusion proteins (e.g., T4 lysozyme/BRIL in ICL3) present in the X-ray structures. Only protein C α atoms of the receptor chain were retained. All structures were trimmed to the 236-residue analysis mask and superimposed onto the risperidone-bound inactive structure (PDB: 6A93) using least-squares C α alignment. Boltz-2 structures were loaded from local CIF files and processed identically, trimmed to the same mask so that every comparison operates on an identical set of C α positions.

2.3. Boltz-2 Co-Folding Ensembles

Boltz-2 [1] was selected on the basis of a prior benchmarking study [17], in which it demonstrated superior pose accuracy and notable inter-run precision compared with Uni-Mol Docking v2 and AutoDock Vina 1.2.1 for psychedelic ligands at 5-HT_{2A}R. Five independent co-folding predictions were generated per ligand (55 structures total) using Boltz-2 in UCSF ChimeraX version 1.11.dev202511072120 [19]. These were generated from the human 5-HT_{2A}R sequence (NCBI Gene ID 3356) and each ligand's SMILES code. Each set of five predictions samples independent trajectories through Boltz-2's diffusion-based denoising process, functionally equivalent to running five independent random seeds.

Notably, 5-HT_{2A}R is represented in Boltz-2's training set through several structures deposited before the training cutoff, including complexes with risperidone (6A93), zotepine (6WGT), 25CN-NBOH (6WHA), LSD (7WC4), and others. Many of the specific psychedelic agonist complexes used here were deposited after the cutoff and are therefore excluded from the training data.

2.4. Normal Mode Analysis

ANM calculations were performed on C α -only representations using a distance cutoff of 15.0 Å with 20 non-trivial modes. GNM calculations used a cutoff of 7.3 Å; the Kirchhoff matrix was constructed using SciPy 1.17.1 spatial-distance functions for cross-platform reproducibility [20]. Dynamic similarity between structures was quantified using the RMSIP of the 10 lowest-frequency ANM modes [21], which ranges from 0 (orthogonal subspaces) to 1 (identical subspaces) and provides a global, rotation-invariant measure of low-frequency subspace overlap. Per-residue mean-square fluctuations, cross-correlation matrices, eigenvalue spectra, and mode collectivity indices were computed for all structures. All reported numerical values derive from ANM; GNM was used only as an independent corroboration of the scalar flexibility trends.

Normal modes are mathematical constructs rather than physical trajectories. They form an orthogonal basis that decomposes a structure's accessible deformations into independent components ordered by frequency; a protein does not move along any single mode in isolation. Real conformational change is a superposition of many modes, weighted by thermal energy and external constraints. The low-frequency modes analyzed here are therefore best understood as the directions along which the elastic network is softest (the motions most readily excited), not as discrete movements the receptor executes one at a time. RMSIP comparisons accordingly assess whether two structures share the same set of soft directions, not whether they undergo identical motions.

2.5. Ensemble Principal Component Analysis

To characterize the conformational landscape sampled by Boltz-2 ensembles, principal component analysis (PCA) was performed on the flattened C α coordinate arrays (N structures $\times$ N atoms $\times$ 3 dimensions) after centering. The resulting principal components (PCs) represent the dominant axes of structural variation across the ensemble. To assess whether the Boltz-2 ensemble samples the same motions that ANM predicts from experimental structures, the overlap between Boltz-2 PCs and each experimentally determined structure's ANM eigenvectors was quantified by RMSIP. A high RMSIP indicates that the structural variance in the co-folding ensemble aligns with the intrinsic flexibility encoded in the experimental structure's contact topology. Joint PCA and t-distributed stochastic neighbor embedding (t-SNE) were performed across all structures (experimental and Boltz-2 combined) to visualize clustering by ligand identity, chemotype, and structure source.

2.6. Replicate Variance Analysis

To assess whether the small structural variability among visually near-identical co-folding predictions is large enough to affect dynamics inferences, RMSIP values from the replicate dataset were partitioned into within-ligand pairs (same ligand, different Boltz-2 seeds; $n = 110$) and between-ligand pairs (different ligands; $n = 1375$). The rationale is as follows: if between-ligand dynamic differences represent genuine ligand-specific signal, the between-ligand RMSIP distribution should be shifted measurably below the within-ligand distribution, because replicates of the same ligand should share a dynamic signature distinct from other ligands. If inter-prediction variance (between the multiple predictions for the same ligand-receptor complex) is comparable to inter-ligand differences, the two distributions overlap. Statistical comparison used the one-sided Mann–Whitney U test with Cohen's d as an effect-size measure; the replicate variance analysis was performed on the same 236-residue mask used throughout, with effect size (rather than the p -value alone) treated as the operative quantity given the large number of between-ligand pairs.

3. Results

3.1. Experimental Structure ANM: Agonist Dynamics Converge in the Ternary Complex

ANM analysis of the 11 experimental structures, trimmed to the uniform 236-residue analysis mask, revealed a striking convergence of dynamic subspaces among all Gq-coupled agonist-bound states. The mean RMSIP among the nine agonists (serotonin, psilocin, DMT, LSD, BOL-148, lisuride, 25CN-NBOH, mescaline, RS130-180) was 0.970 ± 0.019 (Figure 1A). This convergence was independent of chemotype: cross-chemotype comparisons such as LSD (ergoline) versus mescaline (phenethylamine) yielded RMSIP = 0.983, while serotonin (tryptamine) versus LSD reached 0.985 and lisuride (ergoline) versus mescaline reached 0.963. The non-hallucinogenic partial agonists BOL-148 and lisuride, the endogenous agonist serotonin, and the β -arrestin-biased agonist RS130-180 were all mutually indistinguishable by ANM (pairwise RMSIP range across the agonist set: 0.928–0.994). The most

distinct agonist, 25CN-NBOH, still overlapped its peers at a mean RMSIP of 0.937, well within the converged cluster. Hallucinogenic status, chemotype, and signaling bias are not reflected in the coarse-grained dynamics of the Gq-coupled ternary complex.

The two non-agonist structures were clearly separated from the agonist cluster: each of risperidone (inactive) and methiothepin (inverse agonist) overlapped the agonists at a mean RMSIP of 0.740 ± 0.015 , well below the 0.970 internal convergence of the agonists, demonstrating that ANM discriminates activation state even when it cannot discriminate agonist identity. The two non-agonist structures were not themselves closely similar (risperidone–methiothepin RMSIP = 0.676), consistent with their being two distinct, non-agonist ligand classes rather than members of a single tight cluster. Per-residue fluctuation profiles confirmed these findings: agonist-bound structures were virtually superimposable, with the primary distinction being elevated flexibility in TM6 and the intracellular coupling interface relative to the inactive state (Figure 1B). GNM analysis independently corroborated the ANM trends (Supplementary Figure S1).

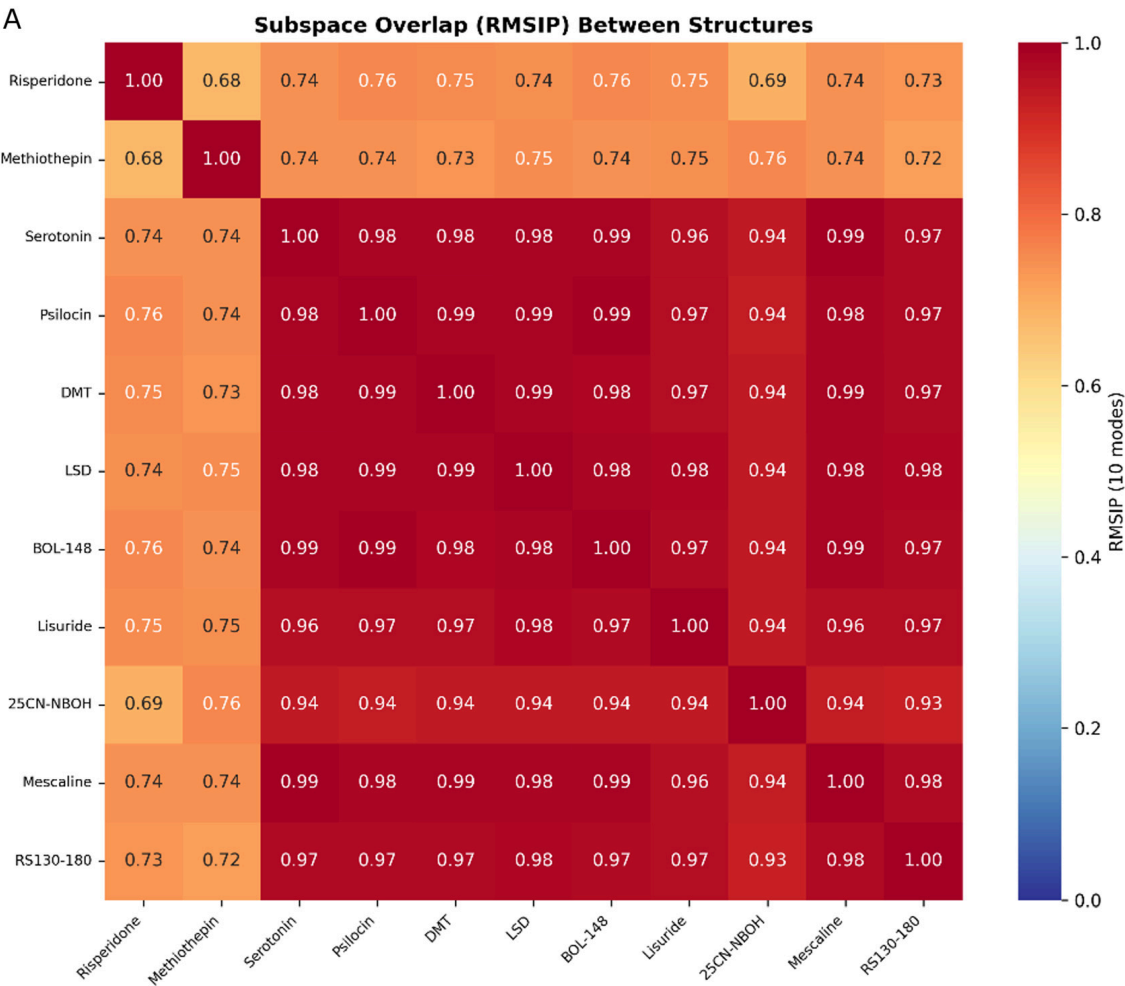


Figure 1. Cont.

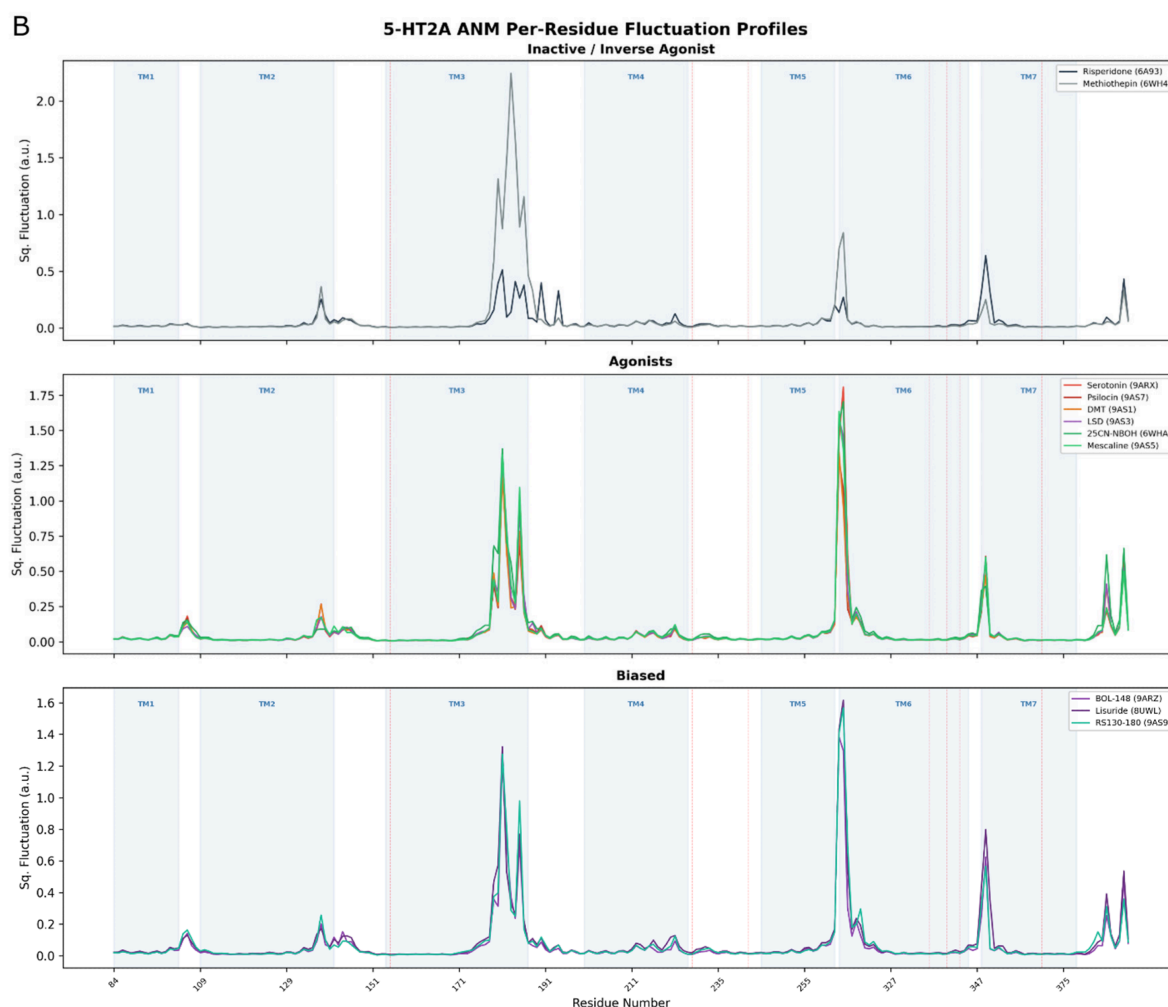


Figure 1. Experimental structure reference dynamics. (A) ANM RMSIP heatmap across all 11 experimentally determined structures; (B) per-residue ANM fluctuation profiles grouped by pharmacological category. Y-axis represents mean-square fluctuation at each C α position; residues with higher values are predicted to be more flexible. Red lines orient pharmacological landmarks D155, L229, S242, W336, F340, N343, and Y370.

Examination of the eigenvalue spectra (Figure S2) reveals that the lowest-frequency modes of all agonist-bound structures share similar eigenvalue magnitudes, indicating that the energetic cost of the dominant collective motions is conserved across agonist chemotypes. In contrast, the inactive risperidone-bound structure exhibits a shift in its lowest eigenvalues, with modes 1–3 occurring at higher frequencies than their agonist counterparts. This indicates that the global breathing and hinge motions most accessible in the active state become stiffer, and therefore less thermally accessible, in the inactive conformation. The redistribution is most pronounced in modes corresponding to TM5–TM6 outward displacement and intracellular cavity opening, consistent with these motions being characteristic of the activation pathway. While individual ANM modes should be interpreted cautiously as they do not capture the full complexity of conformational transitions in isolation, this mode-level analysis reinforces the RMSIP finding: the active-state structures share not only the same mode subspace but also a similar mode hierarchy, with the inactive state diverging in both the directions and the energetic accessibility of its dominant motions.

3.2. Boltz-2 Ensemble PCA: Conformational Landscape of Co-Folded Structures

PCA of the 55 Boltz-2-predicted structures (5 replicates $\times$ 11 ligands), trimmed to the same 236-residue analysis mask, showed that the first principal component captured 38.1% of structural variance, with the first two PCs together explaining 56.9% and the first 10 PCs capturing 86.1% (Figure 2A). The two leading components thus account for the majority of structural variation across the ensemble; given that the panel includes both agonist- and antagonist-bound predictions, the dominant axes plausibly correspond to activation-state-related reorganization rather than to fine ligand-specific differences (Figure 2B). Joint t-SNE embedding recapitulated the RMSIP-based clustering: the active-state experimental structures formed a tight grouping, while the non-agonist structures separated, and the Boltz-2 predictions were broadly distributed across the embedding space rather than co-localizing with the experimental active-state group (Figure 2C).

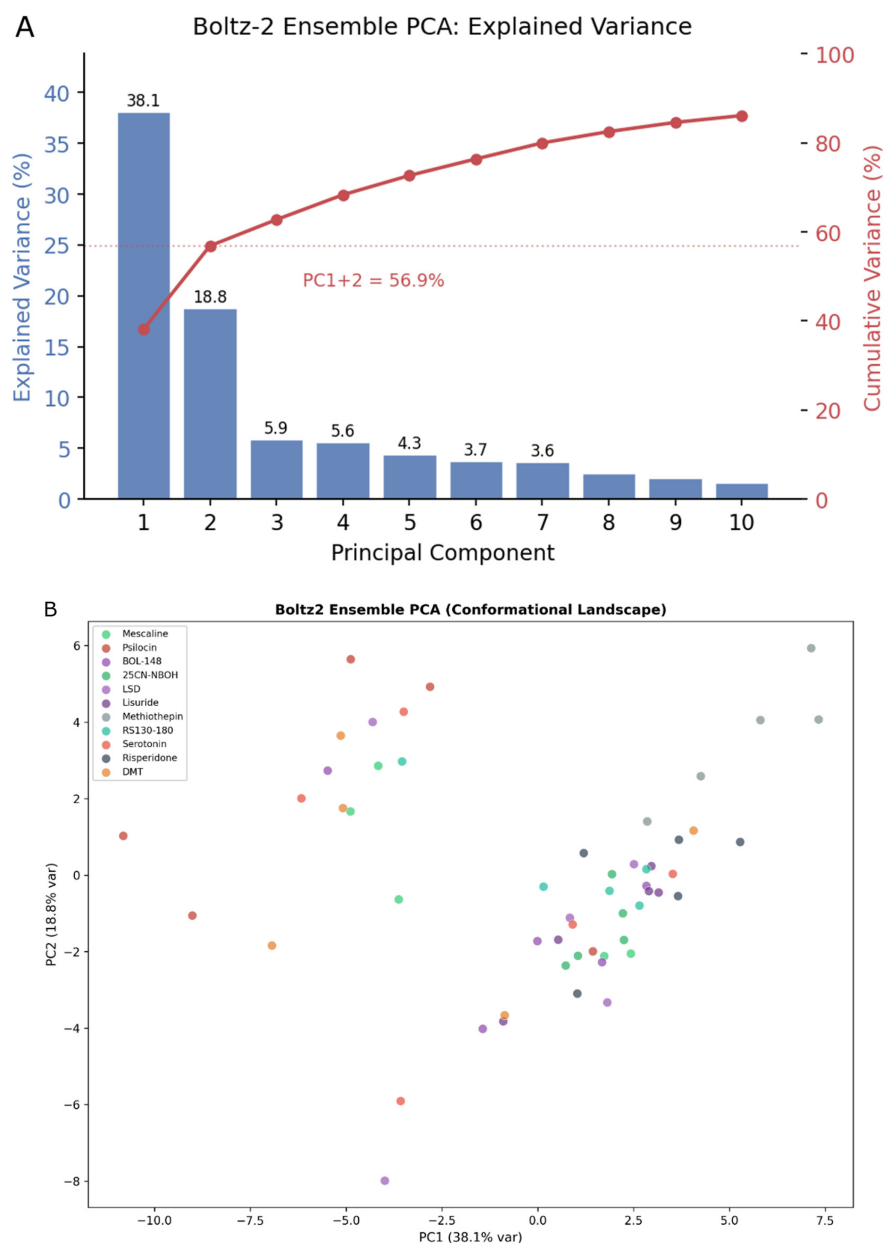


Figure 2. Cont.

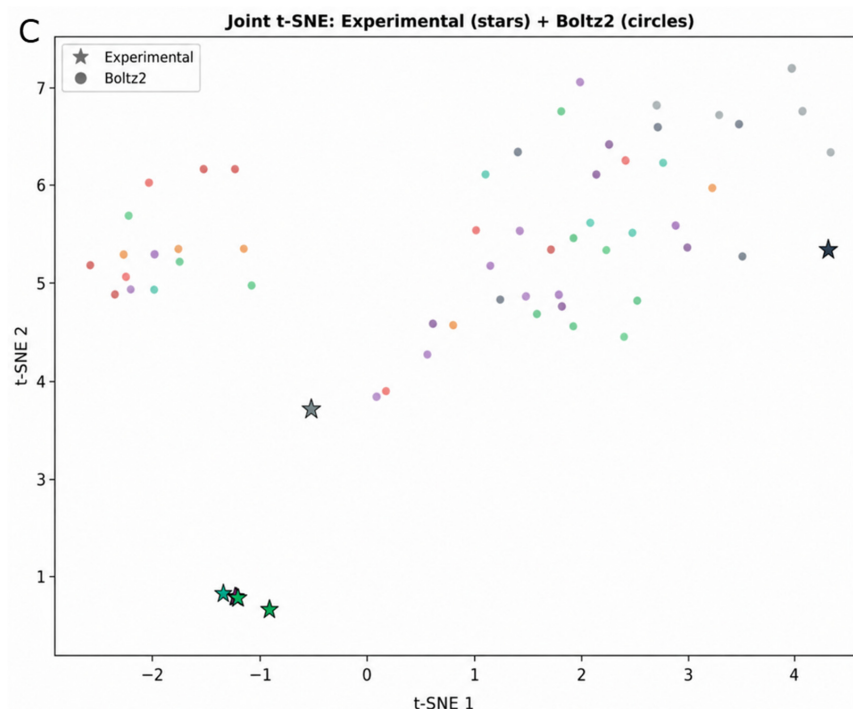


Figure 2. Boltz-2 ensemble conformational landscape. (A) PCA explained-variance (scree) plot. (B) PC1 vs. PC2 scatter colored by ligand. (C) Joint t-SNE of experimental + Boltz-2 structures.

3.3. Boltz-2 Ensemble PCs Show Moderate, Uniform Overlap with Experimental ANM Modes

The central question is whether Boltz-2 structural ensembles recapitulate the dynamics encoded in experimental structures. This was addressed by computing the RMSIP between the Boltz-2 ensemble PCs and each experimental structure's ANM eigenvectors. The mean RMSIP across all 11 reference structures was 0.599 ± 0.007 (Figure 3A). This moderate overlap indicates that the Boltz-2 ensemble captures a portion of the same conformational motions that ANM derives from experimental topology, but falls substantially short of the 0.7–0.8 RMSIP typically observed between NMA and MD essential dynamics [6].

The overlap was remarkably uniform across experimental reference structures (standard deviation 0.007): agonist references ranged narrowly from 0.590 (psilocin) to 0.609 (lisuride), and the non-agonist references were comparable at 0.601 (risperidone) and 0.617 (methiothepin). This near-constant overlap, irrespective of the activation state or ligand of the reference, suggests that the Boltz-2 ensemble samples a generic conformational neighborhood that aligns comparably with all experimental ANM mode sets, rather than preferentially reproducing any one structure's dynamics (Figure 3B).

Reciprocal co-projection analyses were performed for better understanding of the spatial relationship between experimental and predicted conformational landscapes (Figure 4). PCA was fitted independently on the Boltz-2 ensemble and on the experimentally determined structures alone, and each dataset was projected into the other's principal component space. When experimental structures were projected into Boltz-2-defined PCA space (Figure 4A), the active-state experimental structures occupied the upper portion of the Boltz-2 distribution, while the two inactive-state structures (risperidone and methiothepin) separated. The experimental structures thus span a broader region of the Boltz-2 PC landscape than the predictions themselves occupy, suggesting that the conformational differences between activation states captured by cryo-EM and crystallography exceed the structural variance that Boltz-2 co-folding samples across its replicate predictions. When Boltz-2 predictions were projected into experimental structure PCA space (Figure 4B), the predicted structures grouped predominantly toward the inactive risperidone region along

PC1 rather than overlapping with the tightly clustered active-state experimental structures on the left. This asymmetry would suggest that the Boltz-2 ensemble's conformational variance is not centered on the active-state conformation but is instead distributed in a direction that more closely resembles the inactive state in cryo-EM PCA space. The cryo-EM PCA alone (Figure 4C) confirms that PC1 separates active from inactive states, with methiothepin and risperidone at opposite extremes along PC2, consistent with the RMSIP-based clustering. Together, these projections indicate that the Boltz-2 ensemble samples a conformational region that only partially overlaps with the experimentally observed active-state landscape, with the predicted variance axis oriented differently than the activation-state axis that dominates the experimental structures.

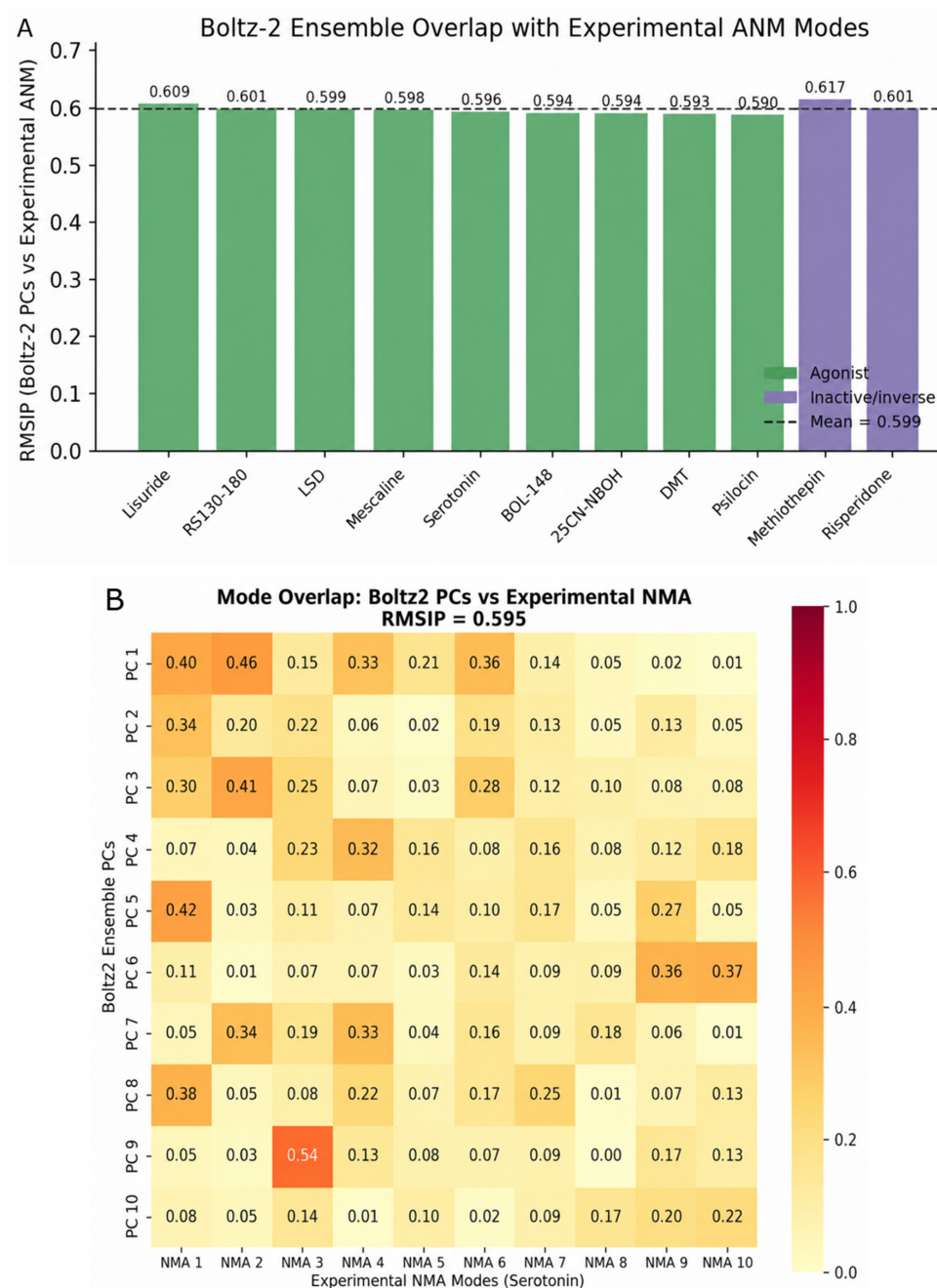


Figure 3. Boltz-2 ensemble versus experimental structure dynamics. **(A)** RMSIP bar chart, Boltz-2 ensemble PCs versus each experimental ANM mode set. **(B)** Mode-overlap heatmap, Boltz-2 PCs versus a representative experimental ANM mode set (e.g., serotonin).

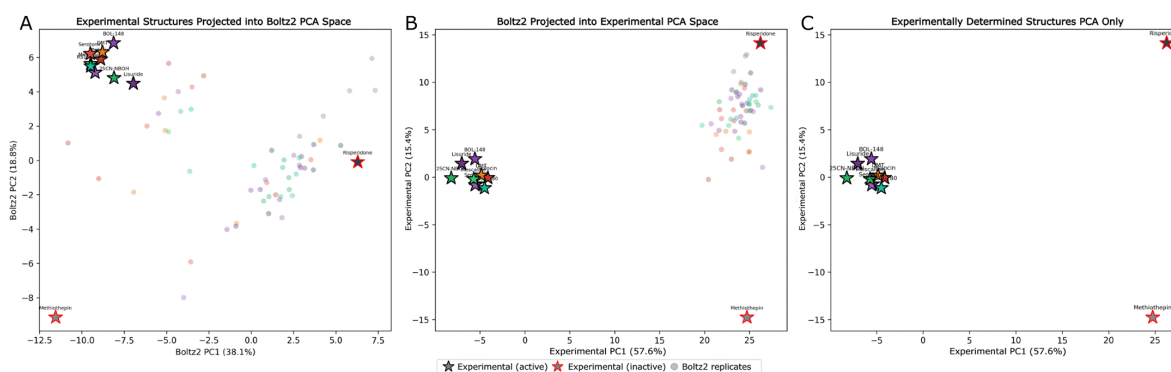


Figure 4. Co-projection of cryo-EM and Boltz-2 structures in shared PCA spaces. (A) Cryo-EM structures (stars) projected into PCA space defined by the Boltz-2 ensemble (circles). Active-state structures are outlined in black; inactive-state structures in red. (B) Boltz-2 ensemble projected into PCA space defined by the cryo-EM structures alone. (C) Cryo-EM PCA in isolation, showing separation of active and inactive states along PC1.

3.4. Replicate Predictions Are Self-Consistent but Carry No Usable Ligand-Specific Signal

Because the practical value of co-folded ensembles for dynamics rests on the assumption that replicate-to-replicate variation is small relative to ligand-dependent variation, this was tested directly. The within-ligand comparison (same ligand, different Boltz-2 seeds; $n = 110$) measures pure prediction noise: whether the structural variability among visually near-identical predictions is large enough to compromise dynamics inferences. The between-ligand comparison (different ligands; $n = 1375$) measures whether co-folding registers ligand-specific differences: if the model captures distinct receptor organization for distinct ligands, the between-ligand distribution should sit measurably below the within-ligand distribution.

Over the 236-residue mask, replicate predictions of the same ligand were highly self-consistent (within-ligand RMSIP = 0.917 ± 0.038): high in absolute terms, but notably below the 0.970 internal convergence of the experimental agonist structures. Crucially, this self-consistency did not translate into ligand discrimination: the between-ligand RMSIP (0.912 ± 0.037) was essentially identical, with an absolute separation of 0.005 RMSIP. Although a one-sided Mann–Whitney test returned a nominally significant difference ($p = 0.037$), the effect size was negligible (Cohen’s $d = 0.142$): a difference detectable only because of the large number of between-ligand pairs, and far below any threshold that would permit one ligand’s co-folded dynamics to be distinguished from another’s (Figure 5). Replicates of the same ligand diverge in their ANM dynamics as much as structures of different ligands. This pattern is interpreted in Section 4.2.

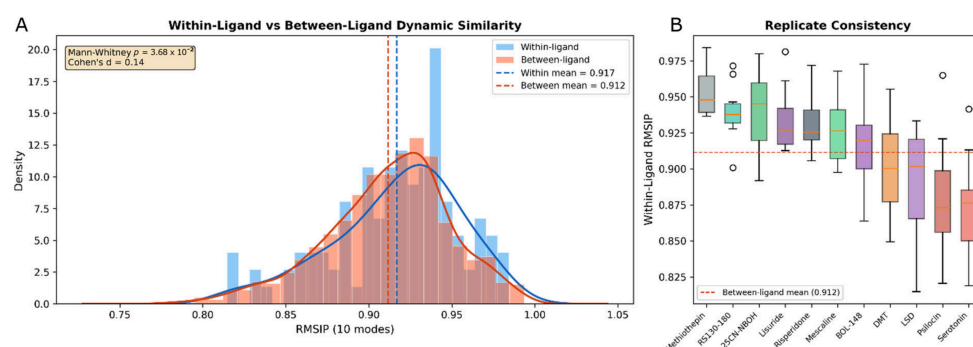


Figure 5. Replicate variance. (A) Within-ligand versus between-ligand RMSIP distributions (kernel density or histogram) with the Mann–Whitney/effect-size annotation. (B) Per-ligand replicate-consistency box plot.

4. Discussion

4.1. The Central Question: Can Co-Folding Be Trusted for Dynamics?

The question motivating this study is whether AI co-folded structural ensembles can serve as a trustworthy basis for receptor dynamics analysis, and specifically, whether they can be trusted to a degree comparable with the experimental structures already used for this purpose. This question is distinct from, and more demanding than, whether co-folding produces accurate static poses. Classical docking methods, despite achieving reasonable pose accuracy in many systems, cannot address receptor dynamics at all, because they embed ligands into a rigid, pre-determined backbone; any downstream dynamics analysis would simply return the input structure. Co-folding represents a qualitative advance because it allows the receptor backbone to reorganize around the ligand, creating the possibility, in principle, that different ligands produce structurally distinct conformations with correspondingly distinct dynamics.

The results indicate that this possibility is realized only partially. The Boltz-2 ensemble PCs overlap moderately and uniformly with experimental structure ANM modes (RMSIP = 0.599), capturing some of the conformational motions encoded in experimental structures while missing a substantial fraction. For comparison, NMA modes typically achieve RMSIP of 0.7–0.8 against MD-derived essential dynamics [6], and the experimental structures themselves show near-perfect internal convergence among agonists (RMSIP = 0.970). Whether Boltz-2 generated ensembles can be trusted as much as the experimental structures is answered most directly by comparing their respective structural variances. Importantly, all of these are statements about precision relative to an experimentally derived reference (which itself contains imperfection and inconsistency), not an absolute truth.

4.2. Prediction Noise, Replicate Consistency, and the Memorization Question

A direct comparison of structural variances quantifies the reliability gap between co-folding and experiment. Among the experimental structures, different agonist complexes are highly similar but not identical (mean agonist-agonist RMSIP = 0.970). Boltz-2's replicate-to-replicate variation for a single ligand is larger: repeated predictions of the same ligand are less similar to one another (within-ligand RMSIP = 0.917) than experimentally derived structures of different agonists are to each other (0.970). Expressed as dispersion ($1 - \text{RMSIP}$), the model's pure prediction noise (≈ 0.083) is roughly threefold the entire spread observed across distinct agonist complexes in cryo-EM (≈ 0.030). This is the crux of the reliability question: because the run-to-run noise of the co-folding model exceeds the structural scale on which agonist complexes actually differ, co-folding cannot resolve ligand-specific receptor organization at that scale; any genuine ligand-dependent differences are necessarily smaller than the prediction noise that obscures them. The experimental structures, whose inter-agonist variation falls below this noise floor, are correspondingly more reliable as a substrate for dynamics analysis. This is not because cryo-EM reveals large ligand-specific differences (it does not; the agonists converge) but because its effective structural resolution is finer than the co-folding noise. The experimental structures are evidently capturing small, consistent differences in receptor organization that the co-folding model cannot recapitulate, placing them on meaningfully firmer ground for dynamics inference. Consistent with this, the between-ligand RMSIP (0.912) is statistically indistinguishable from the within-ligand value (0.917; Cohen's $d = 0.142$), confirming that the apparent inter-ligand spread in the Boltz-2 ensemble reflects prediction variance rather than ligand-specific signal; the nominal significance of the within-versus-between difference is an artifact of the 1375 between-ligand comparisons and carries no practical weight.

Why does co-folding behave this way, internally self-consistent for a given ligand, yet blind to ligand identity and coarser than experiment? The most reasonable explanation lies in how these models relate to their training data. Boltz-2's training set includes structures deposited before its cutoff, and several 5-HT_{2A}R complexes fall within this window (6A93, 6WHA, 6WGT, 7VOD, 7WC4, 7RAN, among others). While many of the specific psychedelic complexes used here were deposited after the cutoff, the receptor itself is well represented in the training data across multiple conformational states and binding partners. Recent analyses have demonstrated that diffusion-based co-folding models frequently recapitulate poses present in their training data rather than learning transferable biophysical principles that would generalize to novel systems [18]. These models allow the receptor to organize around the ligand, which is a genuine advance over rigid docking, but appear, in many cases, to do so from memory rather than from an explicit representation of physics. For a receptor the model has seen extensively, repeated predictions therefore converge tightly because they reconstruct the same memorized region of conformational space, but that convergence reflects the consistency of recall rather than sensitivity to the particular ligand. This is highlighted by the results in Section 3.4.

If Boltz-2's predictions for 5-HT_{2A}R are primarily drawing on memorized receptor conformations, then the 0.599 RMSIP may reflect similarity between the training set's conformational distribution and the experimental structure ANM modes, rather than a genuine ability to predict dynamics from first principles. Two features of the data reinforce this reading. First, the overlap is nearly constant across active and inactive experimental references (0.590–0.617), as expected if the ensemble samples a generic neighborhood of the training distribution rather than a physically specific dynamic landscape (Figure 6). Second, the within-versus-between difference, though statistically detectable, is negligible in magnitude (Cohen's $d = 0.142$; ≈ 0.005 RMSIP), far below the threshold for practical ligand discrimination, and best read as residual prediction noise rather than a ligand-specific signal.

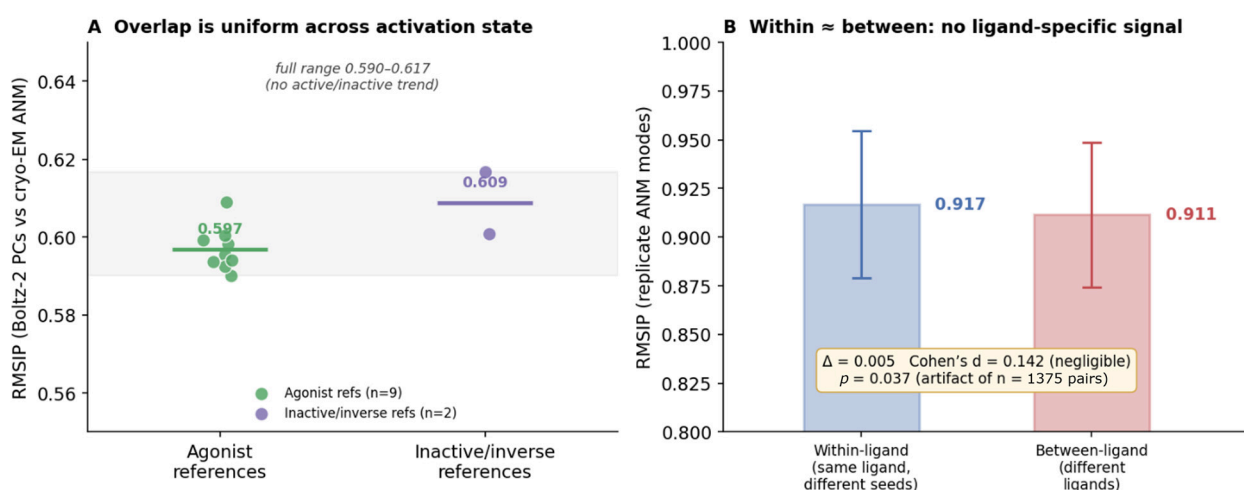


Figure 6. (A) Boltz-2 overlap with experiment is the same whether the reference is active (0.597) or inactive (0.609), with the whole set inside a 0.027-wide band. (B) Within/between are nearly equal with overlapping means. $p = 0.037$ is an artifact of 1375 pairwise comparisons, and the effect itself ($d = 0.142$) is negligible.

This concern does not diminish the practical utility of co-folded ensembles for pose prediction, virtual screening, and structure-based design. It does, however, urge caution when extending these predictions to biophysical inferences such as dynamics, allostery, or conformational selection, where the distinction between memorized and learned conformational diversity is consequential. Targets absent from training data, and decoy ligands with

no experimental reference, will be needed to disentangle these contributions by isolating the model's capacity for de novo biophysical prediction from its capacity for recall. These controls would help to definitively distinguish whether the moderate dynamic overlap and absence of ligand-specific signal reflect fundamental limitations of co-folding precision, or are specific to a target whose conformational landscape the model has already memorized.

One interpretive caveat deserves explicit statement. We refer to the spread among replicate predictions as noise, but it can also be read as signal: the model's sampling of conformationally accessible states. In this case, the spread does not depend on the ligand: within- and between-ligand variances are statistically indistinguishable (Section 3.4), so whatever the replicate diversity represents, it is not ligand-specific conformational availability. We therefore describe it as noise relative to the question of ligand-specific dynamics, while acknowledging that determining whether it encodes any physical sampling would require direct comparison with MD-derived ensembles.

4.3. Experimental Structures as an Imperfect Reference

The experimental structure ANM convergence (RMSIP = 0.970 among all agonists) is itself a notable observation: hallucinogenic tryptamines, non-hallucinogenic ergolines, and a β -arrestin-biased phenethylamine all produce indistinguishable coarse-grained dynamics in the Gq-coupled ternary complex. This is consistent with the G-protein dominating the receptor's conformational landscape once coupling occurs, constraining the receptor into a shared active conformation regardless of which agonist initiated activation. The clean separation is between active and non-agonist states, demonstrating activation-state discrimination in the absence of agonist discrimination.

It is important, however, not to treat these experimental structures as an absolute standard against which predictions simply pass or fail. Cryo-EM coordinates carry resolution-limited uncertainty, and the conformational differences relevant to ligand-specific dynamics (on the order of tenths of an ångström in backbone C α positioning) may lie below the resolution of current structural biology altogether. In that regime, neither predicted nor experimental structures can be said to resolve a definitive ligand-specific dynamic signature, and the appropriate framing is comparative and exploratory rather than absolute: all of these methods, computational and experimental alike, are tools for generating hypotheses about dynamics, each with its own noise floor. The relevant conclusion is therefore not that Boltz-2 "fails" against a perfect reference, but that current co-folding does not yet supply dynamics inferences with precision comparable to what the experimental structures are already considered acceptable for this purpose (a bar that the structures themselves meet only modestly at this scale).

4.4. Future Directions

Several extensions of this work are envisioned. First, MD trajectories would provide essential dynamics via PCA that can be compared directly with both the experimental structure ANM modes and the Boltz-2 ensemble PCs, establishing a three-way comparison between experimental dynamics, NMA predictions, and co-folding ensembles. Second, AlphaFold2-based conformational sampling via MSA-depth modulation [22,23] offers an alternative route to generating structural ensembles for dynamics analysis; preliminary evidence suggests that shallow MSA depths can sample both active and inactive GPCR conformations, and the NMA of such ensembles could be benchmarked against the reference established here. Third, the replicate variance framework used here (comparing within-condition to between-condition structural variability) provides a general-purpose quality-control step for any study deriving dynamics from predicted structures. Fourth, testing co-folding on GPCR targets absent from the training data would help disentangle

learned biophysics from pose memorization. Fifth, methods designed explicitly for generating protein conformational ensembles (such as AlphaFlow, DynaProt, and BioEmu) represent particularly relevant comparators, as these approaches aim to capture conformational heterogeneity or protein dynamics rather than static co-folded structures.

4.5. Limitations

NMA captures only the lowest-frequency collective motions and cannot resolve fine-grained conformational differences. The cryo-EM structures all employed mini-Gαq as the coupling partner; structures coupled to other transducers may reveal transducer-dependent dynamic signatures not examined here. The 236-residue analysis mask, while necessary for uniform comparison and now spanning the full structured receptor through helix 8, still excludes ICL3, the receptor termini, and a few variably resolved loops. The Boltz-2 results are specific to this co-folding method, this receptor, and this ENM formulation. Boltz-2 was developed primarily for the prediction of static structures of protein–ligand complexes, and was not intended for theoretical investigation of protein dynamics. The absence of ligand-specific dynamics in the replicate predictions does not represent a shortcoming of this method within its intended domain of application. The training-data overlap between Boltz-2 and the evaluation structures limits the generalizability of the RMSIP values reported here, as emphasized throughout.

RMSIP itself quantifies the overlap between two mode subspaces (whether two structures share the same accessible low-frequency directions), but is insensitive to how those directions are populated. Two ensembles can span the same subspace while sampling it with very different amplitudes or weights, producing a high RMSIP despite representing different conformational distributions. RMSIP is thus a necessary but not sufficient condition for dynamic equivalence: a high value indicates shared accessible motions, not identical sampling of them. This distinction is especially pertinent when comparing a generative ensemble such as the Boltz-2 principal components, which reflect how the prediction set is distributed, against single-structure elastic network modes, which encode accessible directions but no population information.

This study relies primarily on ANM-derived dynamics quantified by RMSIP, which captures overlap in low-frequency mode subspaces but does not encompass all dimensions of structural comparison. Complementary metrics such as root mean square deviation of inter-residue contact maps, spectral comparison of Kirchhoff matrix eigenvalue distributions, or distance-based similarity measures that do not depend on mode decomposition, could reveal aspects of ligand-dependent structural variation that RMSIP misses. The independent corroboration of ANM trends by GNM analysis (which uses a distinct mathematical formulation and captures scalar flexibility without directional information) provides some assurance that the findings are not specific to one ENM variant. Nevertheless, the most definitive comparison would involve essential dynamics from molecular dynamics trajectories, where principal components derived from all-atom simulations can be directly compared to both ANM modes and co-folding ensemble PCs. Such MD trajectories are currently in progress for select ligands at this receptor and will be the subject of a subsequent investigation.

A final point of concern is the applicability of elastic network models to a receptor of this size. ANM assumes the protein fluctuates within a single harmonic energy well centered on the input structure. This approximation is well supported for small structures, but is a bigger assumption for large membrane proteins, which populate multiple conformational basins rather than a single minimum. For such systems the single-well approximation may capture only part of the functional dynamics, and confidence in NMA-derived motions is appropriately tempered until corroborated by methods that sample

multiple states. The present NMA results should be read as a structurally grounded but harmonic-approximation view of the dynamics.

5. Conclusions

We have investigated whether AI co-folded structural ensembles can recapitulate the receptor dynamics encoded in experimental structures, using 5-HT_{2A}R as a testbed. Cryo-EM ANM establishes that all nine Gq-coupled agonist structures converge to a single dynamic attractor (RMSIP = 0.970), discriminating activation state but not agonist identity. Boltz-2 co-folding ensembles show moderate, highly uniform overlap with these experimental dynamics (RMSIP = 0.599), capturing some conformational motions but falling short of the precision that would make them a substitute for experimentally derived dynamics; replicate predictions, though internally consistent, carry no usable ligand-specific signal. This overlap must be interpreted cautiously given Boltz-2's training-data overlap with 5-HT_{2A}R structures, and the experimental reference is itself an imperfect, resolution-limited standard. Co-folding represents a necessary advance over classical docking for dynamics applications by allowing the receptor backbone to reorganize around the ligand, but current methods have not crossed the precision threshold required for the elastic network contact topology to yield ligand-specific dynamic information. These findings establish a quantitative benchmark and underscore the importance of rigorous validation, including replicate variance testing and careful chain handling, before deriving biophysical insight from AI-predicted structures.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/receptors5040032/s1>, Figure S1. GNM flexibility profiles (validation); Figure S2. ANM eigenvalue spectra.

Author Contributions: Conceptualization, B.R.C.; methodology, B.R.C. and J.F.; validation, B.R.C., C.G.F. and J.F.; formal analysis, B.R.C.; data curation, B.R.C.; writing—original draft preparation, B.R.C.; writing—review and editing, C.G.F. and J.F.; visualization, B.R.C. and C.G.F.; All authors have read and agreed to the published version of the manuscript.

Funding: All software used was run locally and is freely available through the developers. External review and project guidance was provided by the organizers of CompBio Asia 2026, a computational workshop supported by NSF awards 1953405 and 2420222.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All cryo-EM and crystal structures are publicly available from the Protein Data Bank under the accession codes listed in Table 1. The NMA pipeline, ensemble dynamics pipeline, and replicate variance pipeline, with all configuration parameters, are available at <https://github.com/bcummins96/> (accessed on 21 September 2026). Boltz-2 co-folded structures and complete numerical results are available from the corresponding author upon request.

Acknowledgments: During the preparation of this manuscript, the authors used Claude Opus 4.6 for the purposes of visual figure generation. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Passaro, S.; Corso, G.; Wohlwend, J.; Reveiz, M.; Thaler, S.; Somnath, V.R.; Getz, N.; Portnoi, T.; Roy, J.; Stark, H.; et al. Boltz-2: Towards accurate and efficient binding affinity prediction. *bioRxiv* **2025**, bioRxiv:2025.06.14.659707. [CrossRef] [PubMed]
2. Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A.J.; Bambrick, J.; et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **2024**, *630*, 493–500. [CrossRef] [PubMed]

3. Chai Discovery Team; Boitreaud, J.; Dent, J.; McPartlon, M.; Meier, J.; Reis, V.; Rogozhonikov, A.; Wu, K. Chai-I: Decoding the molecular interactions of life. *bioRxiv* **2024**, bioRxiv:2024.10.10.615955. [\[CrossRef\]](#)
4. Park, S.W.; Lee, B.H.; Kim, M.K. Elastic network model: A coarse-grained approach to the study of biomolecular dynamics. *Multiscale Sci. Eng.* **2023**, *5*, 104–118. [\[CrossRef\]](#)
5. Atilgan, A.R.; Durell, S.R.; Jernigan, R.L.; Demirel, M.; Keskin, O.; Bahar, I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.* **2001**, *80*, 505–515. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Amadei, A.; Ceruso, M.A.; Di Nola, A. On the convergence of the conformational coordinates basis set obtained by the essential dynamics analysis of proteins' molecular dynamics simulations. *Proteins* **1999**, *36*, 419–424. [\[CrossRef\]](#)
7. Bahar, I.; Chennubhotla, C.; Tobi, D. Intrinsic dynamics of enzymes in the unbound state and relation to allosteric regulation. *Curr. Opin. Struct. Biol.* **2007**, *17*, 633–640. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Kolan, D.; Fonar, G.; Samson, A.O. Elastic network normal mode dynamics reveal the GPCR activation mechanism. *Proteins* **2014**, *82*, 579–586. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Zhang, S.; Krieger, J.M.; Zhang, Y.; Kaya, C.; Kaynak, B.; Mikulska-Ruminska, K.; Doruker, P.; Li, H.; Bahar, I. ProDy 2.0: Increased scale and scope after 10 years of protein dynamics modelling with Python. *Bioinformatics* **2021**, *37*, 3657–3659. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Krieger, J.M.; Sorzano, C.O.S.; Carazo, J.M.; Bahar, I. Protein dynamics developments for the large scale and cryoEM: Case study of ProDy 2.0. *Acta Crystallogr. Sect. D Struct. Biol.* **2022**, *78*, 399–409. [\[CrossRef\]](#) [\[PubMed\]](#)
11. Vollenweider, F.X.; Preller, K.H. Psychedelic drugs: Neurobiology and potential for treatment of psychiatric disorders. *Nat. Rev. Neurosci.* **2020**, *21*, 611–624. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Cummins, B.R.; Billac, G.B.; Nichols, D.E.; Nichols, C.D. 5-HT_{2A} receptors: Pharmacology and functional selectivity. *Pharmacol. Rev.* **2025**, *77*, 100059. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Kimura, K.T.; Asada, H.; Inoue, A.; Kadji, F.M.N.; Im, D.; Mori, C.; Arakawa, T.; Hirata, K.; Nomura, Y.; Nomura, N.; et al. Structures of the 5-HT_{2A} receptor in complex with the antipsychotics risperidone and zotepine. *Nat. Struct. Mol. Biol.* **2019**, *26*, 121–128. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Kim, K.; Che, T.; Panova, O.; DiBerto, J.F.; Lyu, J.; Krumm, B.E.; Wacker, D.; Robertson, M.J.; Seven, A.B.; Nichols, D.E.; et al. Structure of a hallucinogen-activated Gq-coupled 5-HT_{2A} serotonin receptor. *Cell* **2020**, *182*, 1574–1588.e19. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Gumpfer, R.H.; Jain, M.K.; Kim, K.; Sun, R.; Sun, N.; Xu, Z.; DiBerto, J.F.; Krumm, B.E.; Kapolka, N.J.; Kaniskan, H.Ü.; et al. The structural diversity of psychedelic drug actions revealed. *Nat. Commun.* **2025**, *16*, 2734. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Lyu, J.; Kapolka, N.; Gumpfer, R.H.; Alon, A.; Wang, L.; Jain, M.K.; Barros-Álvarez, X.; Sakamoto, K.; Kim, Y.; DiBerto, J.; et al. AlphaFold2 structures guide prospective ligand discovery. *Science* **2024**, *384*, eadn6354. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Cummins, B.R.; Nichols, C.D. Assessing Modern AI-Driven Protein-Ligand Modeling with Phenethylamine and Tryptamine Psychedelics. *AI Chem.* **2026**, *1*, 4. [\[CrossRef\]](#)
18. Masters, M.R.; Mahmoud, A.H.; Lill, M.A. Investigating whether deep learning models for co-folding learn the physics of protein-ligand interactions. *Nat. Commun.* **2025**, *16*, 8854. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Meng, E.C.; Couch, G.S.; Croll, T.I.; Morris, J.H.; Ferrin, T.E. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **2021**, *30*, 70–82. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods* **2020**, *17*, 261–272. Correction in *Nat. Methods* **2020**, *17*, 352. [\[CrossRef\]](#) [\[PubMed\]](#)
21. Amadei, A.; Linssen, A.B.M.; Berendsen, H.J.C. Essential dynamics of proteins. *Proteins* **1993**, *17*, 412–425. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Del Alamo, D.; Sala, D.; McHaourab, H.S.; Meiler, J. Sampling alternative conformational states of transporters and receptors with AlphaFold2. *eLife* **2022**, *11*, e75751. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Wayment-Steele, H.K.; Ojoawo, A.; Otten, R.; Apitz, J.M.; Pitsawong, W.; Hömberger, M.; Ovchinnikov, S.; Colwell, L.; Kern, D. Predicting multiple conformations via sequence clustering and AlphaFold2. *Nature* **2024**, *625*, 832–839. [\[CrossRef\]](#) [\[PubMed\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.