Supplementary Information

Azadirachtin-Driven Allosteric Basin Squeezing at the Ecdysone Receptor–Ultraspiracle Heterodimer

Nalgirkar, Lobhi, Jangam, Choure

Supplementary Methods S1: Extended Computational Methods

S1.1 Force Field Parameterisation Details

For azadirachtin (MW 720 Da; 6+ hydroxyl groups; rigid lactone-epoxide core), the CGenFF web server parameterisation yielded charge penalty scores below 50 and bonded penalty scores below 25 for all atom types, within the acceptable range for production-quality MD without quantum mechanical re-optimisation. For eugenol (MW 164 Da; one hydroxyl, one methoxy group), all charge and bonded penalty scores were below 10, confirming high-confidence parameterisation. Ligands with charge penalty scores above 50 or bonded penalty scores above 25 would ordinarily require geometry optimisation at the QM level (B3LYP/6-31G*) prior to production simulation; neither compound met this threshold.

S1.2 System Construction Details

Four molecular systems were constructed using CHARMM-GUI Solution Builder v3.7: the apo EcR–USP ligand-binding domain heterodimer (APO; chain A = EcR LBD, chain B = USP LBD), azadirachtin-bound (AZA), eugenol-bound (EUG), and dual-ligand complex (DOUBLE). Each system was solvated in a rectangular TIP3P water box with minimum solvent padding of 10 Å from the protein surface. Sodium and chloride ions were added to neutralise each system and achieve a final ionic concentration of 0.15 M NaCl.

Steepest-descent energy minimisation was performed until the maximum force fell below 1000 $\mathrm{kJ\,mol^{-1}\,nm^{-1}}$, followed by sequential NVT and NPT equilibration of 100 ps each at 300 K. Position restraints generated by the CHARMM-GUI workflow were maintained during equilibration and released prior to production simulations. For each system, three independent replicates were initiated from the same equilibrated CHARMM-GUI starting structure using different random initial velocity seeds.

Note on replication: While APO, AZA, and DOUBLE complexes were simulated in triplicate ($N = 3$), the EUG-alone system was extended to five independent replicates ($N = 5$) to comprehensively sample the stochastic nature of eugenol surface contact-loss and unbinding trajectories. Replicate-level outcomes for the extended EUG-alone ensemble are reported in the main text (Section 3.6).

S1.3 GNINA CNN Rescoring Protocol

To provide an independent, force-field-agnostic validation of the eugenol surface-site binding pose, the production EUG docking pose was submitted to GNINA v1.1 for CNN rescoring. The receptor was prepared as a static PDBQT file; the ligand was provided in PDBQT format identical to the AutoDock Vina input. A docking box centred at $(45.2, 52.8, 48.6)$ Å (size $22 \times 22 \times 22$ Å, exhaustiveness 24) was used to generate 20 alternative modes ranked by CNN pose score and affinity.

S1.4 Boltzmann Inversion Analysis

The order parameter ξ was defined as the instantaneous distance between the centre of mass of eugenol (all atoms) and the centroid of the Cα atoms of the eugenol-site residues. The site comprised every protein residue with at least one heavy atom within 5 Å of eugenol in the initial docked structure of the EUG-alone system (ASP400, ARG404, GLU410, PHE474, ARG470, and LEU474; PDB 1R20 numbering). The identical residue set was applied to both EUG-alone and DOUBLE trajectories for consistency. Because a centre-to-centre separation is invariant under rigid-body superposition, no trajectory alignment was performed.

A replicate was classified as reversibly sampled, and therefore eligible to contribute to the PMF, if its ligand-RMSD trajectory did not undergo a sustained, irreversible departure from the bound-state basin over the full 400 ns window (operationally, the same sustained 5 Å-crossing criterion described in Section S1.5 below). Replicate-level outcomes for each system are reported in the main text (Section 3.4). Uncertainty bands denote ±1 SD across 500 moving-block bootstrap resamples (10 ns block length). These profiles characterise bound-state basin shape and breadth and should not be interpreted as absolute binding free energies; no escape barrier or ΔG_{bind} is implied.

S1.5 EUG-Site Residue Contact-Loss Analysis (REP_4 and REP_5)

To test whether eugenol lability in the single-ligand system was reproducible rather than an artefact of limited replication, two additional EUG-alone replicates (REP_4 and REP_5) were initiated under the identical protocol described in Section 2.2 of the main text. Replicate-level dissociation outcomes and onset times for the full five-replicate EUG-alone ensemble are reported in the main text (Section 3.6).

Per-residue protein–ligand contact occupancy (any protein heavy atom within 4 Å of any eugenol heavy atom; MDAnalysis) was computed for the bound replicate over the full 400 ns trajectory and for each dissociating replicate over its pre-dissociation window (first sustained crossing of a 5 Å ligand-RMSD threshold, requiring ≥70% of the following 10 ns to remain above threshold). Three kinetically distinct contact classes were defined a priori from this occupancy comparison: deep-pocket contacts expected to be lost on dissociation, electrostatic anchors expected to be retained during departure, and surface contacts expected to be gained along the exit pathway. The resulting per-residue occupancies and class assignments are tabulated in Supplementary Table S7.

S1.6 Water Radial Distribution Function Analysis

The radial distribution function of water oxygen atoms around eugenol was computed using `gmx rdf` with the centre of mass of eugenol as the reference group and all TIP3P water oxygen atoms within a 12 Å shell as the selection group. Analysis was averaged over post-equilibration frames (100–400 ns, stride = 5) for the EUG-alone bound replicate (REP_1) and all three DOUBLE replicates independently. A primary peak at $r \approx 2.8$ Å corresponds to the direct H-bond distance between eugenol's hydroxyl oxygen and the nearest water oxygen; a second-shell peak at $r \approx 3.6$ Å reflects the first hydration shell. Comparative results across the EUG-alone and DOUBLE replicates are reported in the main text (Section 3.5) and shown in Supplementary Figure S5.

S1.7 H-Bond Autocorrelation Analysis

Hydrogen-bond persistence was quantified by computing the autocorrelation function (ACF) of the binary H-bond existence time-series. For each protein–ligand polar contact pair, a bond was defined as present when donor–acceptor distance ≤ 3.5 Å and donor–H–acceptor angle $\geq 120°$. The ACF $C(\tau) = \langle h(t)h(t+\tau)\rangle / \langle h(t)\rangle^2$ was computed using MDAnalysis v2.x with

post-equilibration frames (stride = 5) for the dominant contacts in REP_1 of the DOUBLE and EUG-alone systems. The 1/e decay time τ was estimated by exponential fit. Comparative decay behaviour for the two systems is discussed in the main text (Section 3.5) and shown in Supplementary Figure S6.

S1.8 Configurational Entropy Estimation

The configurational entropy of the eugenol-site residues was estimated using Schlitter's quasi-harmonic approximation: $S = \frac{1}{2} k_{\mathrm{B}} \ln \det\left[\mathbf{1} + (k_{\mathrm{B}} T e^2/\hbar^2)\, \mathbf{M}\boldsymbol{\sigma}\right]$, where $\boldsymbol{\sigma}$ is the mass-weighted covariance matrix of the Cartesian fluctuations of the pocket atoms and $\mathbf{M}$ the corresponding mass matrix. Covariance matrices were accumulated over the post-equilibration trajectory (100–400 ns) after least-squares superposition on the pocket atoms to remove overall translation and rotation. Entropy was evaluated at $T = 300$ K for the bound EUG-alone replicate and for each DOUBLE replicate independently; values are reported as TS in kcal mol^{-1}. Interpretation of the resulting entropy difference, together with the relevant caveats regarding sampling convergence and atom-set-size dependence, is given in the main text (Section 3.5).

S1.9 Residue Interaction Network Construction and Centrality Analysis

Dynamical residue networks were constructed from post-equilibration trajectories (100–400 ns, replicate-pooled). Each residue was represented as a node. An undirected, unweighted edge was placed between two residues when at least one inter-residue heavy-atom pair remained within 4.5 Å in $\geq$75% of analysed frames; sequence-adjacent residues ($|i - j| \leq 1$) were excluded to ensure edges report tertiary rather than trivial backbone connectivity. Shortest-path betweenness centrality was computed with NetworkX and normalised by $(N-1)(N-2)/2$ for cross-system comparability. Closeness centrality, eigenvector centrality, and node degree were computed analogously and used to corroborate betweenness-based hub assignments. Network size statistics (residue and edge counts per system) and the resulting cross-state betweenness comparison are reported in the main text (Section 3.7).

Supplementary Tables

Table S1: Full Docking Energetics and Redocking Validation.

Parameter	**Production**	**Validation (Deep Search)**
AutoDock Vina version	1.2.0	1.2.7
Exhaustiveness	8	64
AZA best-pose score ($\mathrm{kcal\,mol^{-1}}$)	-8.7 ± 0.3	-8.40
AZA top-6 cluster range ($\mathrm{kcal\,mol^{-1}}$)	—	-8.40 to -8.05
AZA poses within $\sim$3 Å of best	—	18/19
AZA redocking RMSD (Å)	—	0.412
EUG redocking RMSD (Å)	—	0.744
Ponasterone A (1R1K) RMSD (Å)	—	1.546

Table S2: GNINA CNN Rescoring of the Eugenol Surface-Site Binding Pose. Docking box centred at $(45.2, 52.8, 48.6)$ Å, size $22 \times 22 \times 22$ Å, exhaustiveness 24.

Mode	**Affinity ($\mathrm{kcal\,mol^{-1}}$)**	**CNN Pose Score**	**CNN Affinity (pK units)**
1	-4.57	0.690	3.934
2	-4.44	0.544	3.994
3	-4.41	0.504	3.944
4	-4.41	0.446	3.508
5	-4.41	0.357	3.420
6	-4.30	0.487	3.390
Top-6 range	-4.57 to -4.30	—	—

Table S3: Orthosteric (AZA) and Distal (EUG) Contact Networks from Docking.

Residue	**Helix**	**Interaction Type**	**Ligand**
LYS283	H3	Hydrogen bond / electrostatic	AZA
TRP286	H3	Hydrophobic / π-stacking	AZA
GLU288	H3	Electrostatic	AZA
PHE432	H11	Hydrophobic packing	AZA
GLU433	H11	Electrostatic	AZA
LEU500	H12	Hydrophobic anchoring	AZA
ASP400	H9	Hydrogen bond / electrostatic	EUG
ARG404	H9	Electrostatic	EUG
GLU410	H9	Electrostatic	EUG
PHE474	H10	Hydrophobic	EUG
ARG470	H11	Electrostatic	EUG
LEU474	H11	Hydrophobic packing	EUG

Table S4: MM-GBSA Binding Free Energies: Full Per-Replicate Data. Values in kcal mol^{-1}. Dissociated replicates excluded from bound-state means. Entropy not included; values are comparative.

System	**REP_1**	**REP_2**	**REP_3**	**Mean ± SEM**
AZA (alone)	-23.60	-4.13[a]	-21.99	-22.80 ± 0.81[b]
EUG (alone)	-17.84	-7.12[c]	-11.69[c]	-12.22 ± 3.11
AZA in DOUBLE	-23.06	-24.09	-19.03	-22.06 ± 1.54
EUG in DOUBLE	-15.81	-17.43	-11.05	-14.76 ± 1.91

[a] AZA dissociated (RMSD $\sim$42 Å; vdW term -8.2 kcal mol^{-1}); excluded from mean. [b] Mean of REP_1 and REP_3 only. [c] EUG drifts or weakens at surface site when bound alone.

Table S5: Top Per-Residue MM-GBSA Contributions: Azadirachtin.

Residue	**Subunit**	**Alone** (kcal mol^{-1})	**In DOUBLE** (kcal mol^{-1})	**Interaction Mode**
PHE436	EcR (A)	-2.64	-1.85	vdW / CH–π
PHE437	EcR (A)	-1.50	-0.72	vdW / CH–π
TYR476	USP (B)	-1.30	-0.62	Ring packing + weak H-bond
ASN567	USP (B)	-1.18	-0.79	H-bond (carboxamide)
PHE432	EcR (A)	-1.01	-1.74	vdW / CH–π
LEU429	EcR (A)	-0.45	-0.54	vdW / hydrophobic
GLN564	USP (B)	-0.35	-0.48	H-bond (carboxamide)
MET502	USP (B)	-0.29	—	vdW / hydrophobic
ILE509	USP (B)	—	-0.49	vdW / hydrophobic

Table S6: Top Per-Residue MM-GBSA Contributions: Eugenol.

Residue	**Subunit**	**Alone (kcal mol^{-1})**	**In DOUBLE (kcal mol^{-1})**
ARG383	USP (B)	−1.85	—
ARG387	USP (B)	−1.08	—
PHE397	USP (B)	−0.91	—
VAL384	USP (B)	−0.43	—
TYR308	USP (B)	−0.41	—
ARG470	USP (B)	—	−1.39
PHE487	USP (B)	—	−0.88
PHE474	EcR (A)	—	−0.67
ALA475	EcR (A)	—	−0.56
SER483	USP (B)	—	−0.51
LEU474	USP (B)	—	−0.35

Table S7: Per-Residue Contact Occupancy at the EUG Site: Bound vs. Dissociating Replicates. Any protein heavy atom within 4 Å of any eugenol heavy atom (MDAnalysis). REP_1 = full 400 ns; REP_2–REP_5 scored over pre-dissociation windows (∼3.6, 10.7, 2.0, 4.1 ns). **Δ** = bound occupancy minus mean of four dissociating replicates. Crystallographic PDB 1R20 numbering; all residues on USP LBD (chain B).

Residue	**REP_1 (bound) %**	**REP_2 %**	**REP_3 %**	**REP_4 %**	**REP_5 %**	**Δ (%)**
THR346†	84	0	0	0	0	+84
LEU349†	70	0	0	0	0	+70
PHE397†	85	32	10	43	39	+53
VAL384†	52	0	0	1	0	+51
TYR403†	46	0	0	0	0	+46
GLY321†	41	0	0	0	0	+41
TYR322†	40	0	3	9	0	+36
MET342†	32	0	0	0	0	+32
LEU345†	28	0	0	0	0	+28
VAL395†	23	0	0	0	0	+23
ARG383‡	92	62	41	97	96	+18
ARG387‡	84	68	23	86	61	+25
GLN324§	20	24	14	93	77	−32
GLU320§	34	68	85	83	49	−38
GLN319§	38	84	99	98	95	−56

† Deep-pocket contacts: lost on dissociation (signature of unseated EUG). ‡ Electrostatic anchors: retained during departure. § Surface contacts: gained on exit pathway, tracing the dissociation route.

Table S8: Backbone RMSD, R_g, and SASA Summary Across Systems.

System	**Mean RMSD (nm)**	**SD RMSD (nm)**	R_g **(nm)**	**SASA (nm^2)**
APO	0.190	Low	2.22	220–225
AZA	0.2246	±0.0362	2.3065 ± 0.0136	223.6
EUG	0.2121	±0.0221	2.2962 ± 0.0103	223.5
DOUBLE	0.2172	±0.0217	2.29–2.30	220–225

Table S9: PC1 Maximum Cα Displacement and Top Displacement Residues (REP_1). EcR LBD = trajectory residues 1–239; USP LBD = trajectory residues 240–467. USP crystallographic number = trajectory number + 61.

System	**PC1 (%)**	**PC2 (%)**	**Max Cα disp. (Å)**	**Top PC1 residues (traj. numb.)**
APO	61.9	38.1	66.96	35, 36, 37
AZA	77.2	22.8	75.90	17, 18, 19 (EcR N-term loop)
EUG	85.1	14.9	94.46	387, 388, 389 (USP domain)
DOUBLE	74.9	25.1	93.90	390, 330, 265 (USP domain)

Table S10: Allosteric Communication Hub Betweenness Centrality Across Systems. Betweenness normalised by $(N-1)(N-2)/2$. Values averaged over three replicates per system. DOUBLE-emergent residues exceed their single-ligand counterparts by 1.7–7-fold.

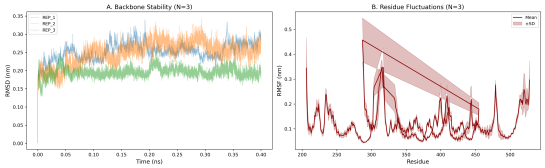
Residue	**APO**	**AZA**	**EUG**	**DOUBLE**	**Class**
ARG209 (EcR)	0.092	0.087	0.089	0.098	Conserved
LEU219 (EcR)	0.071	0.068	0.074	0.075	Conserved
GLU496 (USP)	0.084	0.079	0.081	0.085	Conserved
THR495 (USP)	0.078	0.091	0.082	0.089	Conserved
ARG131 (EcR)	0.032	0.041	0.038	0.086	DOUBLE-emergent
LEU494 (USP)	0.045	0.057	0.042	0.073	DOUBLE-emergent
PRO450 (USP)	0.009	0.031	0.018	0.067	DOUBLE-emergent

Supplementary Figures

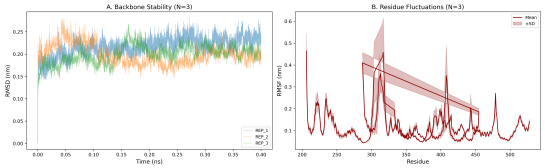
(a) Azadirachtin — orthosteric site

(b) Eugenol — distal site

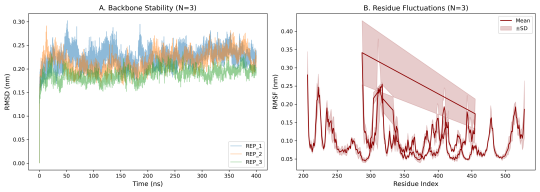
Figure S1: Two-dimensional protein–ligand interaction maps for azadirachtin (A, orthosteric site) and eugenol (B, distal site). No residue appears in both panels, confirming that the two binding regions are spatially non-overlapping. Generated using BIOVIA Discovery Studio.



(a) AZA

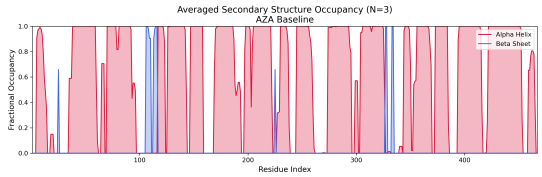


(b) EUG

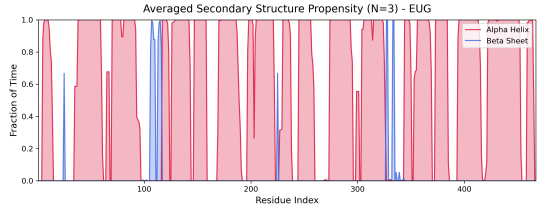


(c) DOUBLE

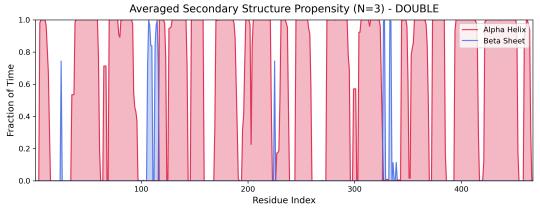
Figure S2: Backbone RMSD (top), ligand RMSD (middle), and radius of gyration (bottom) trajectories for all three ligand-bound systems. All systems achieve equilibration within 100 ns and maintain stable conformations throughout 400 ns production simulations.



(a) AZA-bound EcR

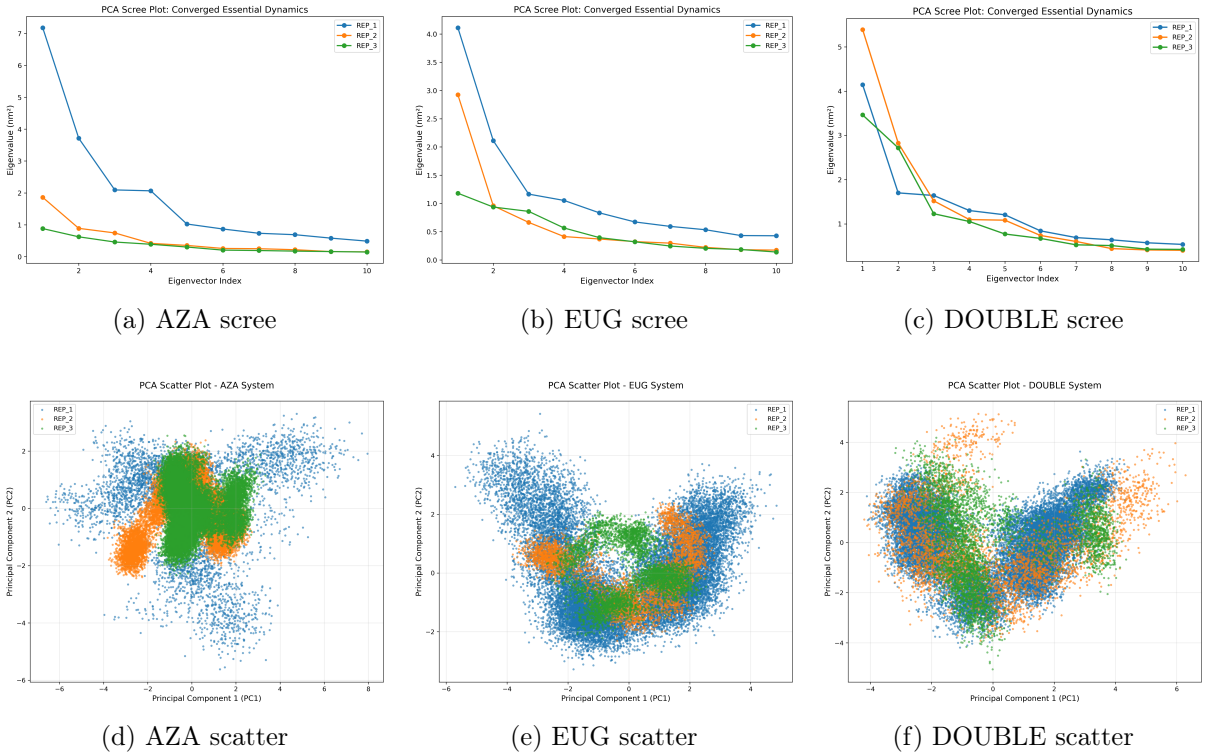


(b) EUG-bound EcR



(c) DOUBLE-bound EcR

Figure S3: Secondary structure propensity averaged across $N = 3$ replicates for (A) AZA-bound, (B) EUG-bound, and (C) DOUBLE-bound systems. All systems preserve the characteristic nuclear receptor α-helical fold throughout the simulations.



(a) AZA scree (b) EUG scree (c) DOUBLE scree

(d) AZA scatter (e) EUG scatter (f) DOUBLE scatter

Figure S4: (A–C) Scree plots showing variance explained by the first 10 principal components for AZA, EUG, and DOUBLE systems. (D–F) PC1–PC2 scatter plots for all three replicates per system. The DOUBLE system shows extensive replicate overlap, indicative of a reproducible, well-defined conformational ensemble.

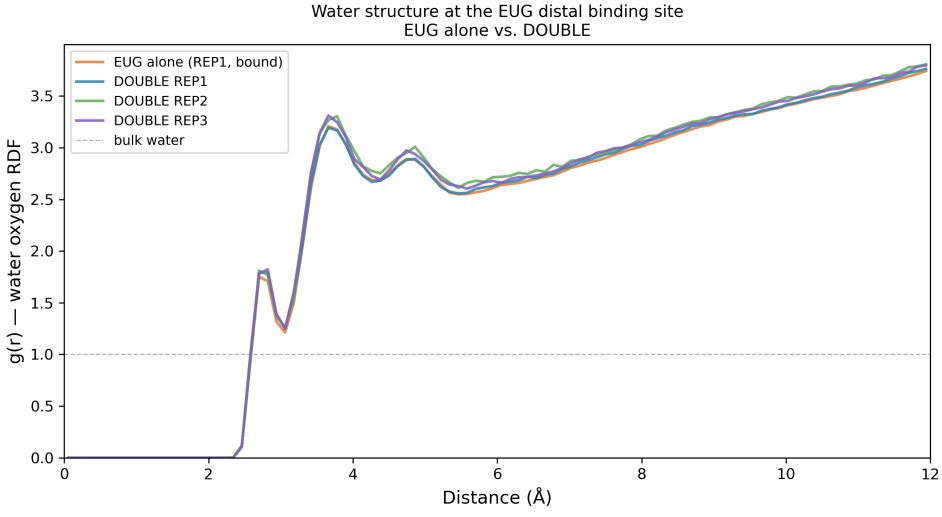


Figure S5: Radial distribution function (RDF) of water oxygen atoms around eugenol at the distal binding site. All four curves (EUG-alone bound replicate and three DOUBLE replicates) are superimposable within trajectory noise, consistent with the main-text negative control for the basin-squeezing model (Section 3.5).

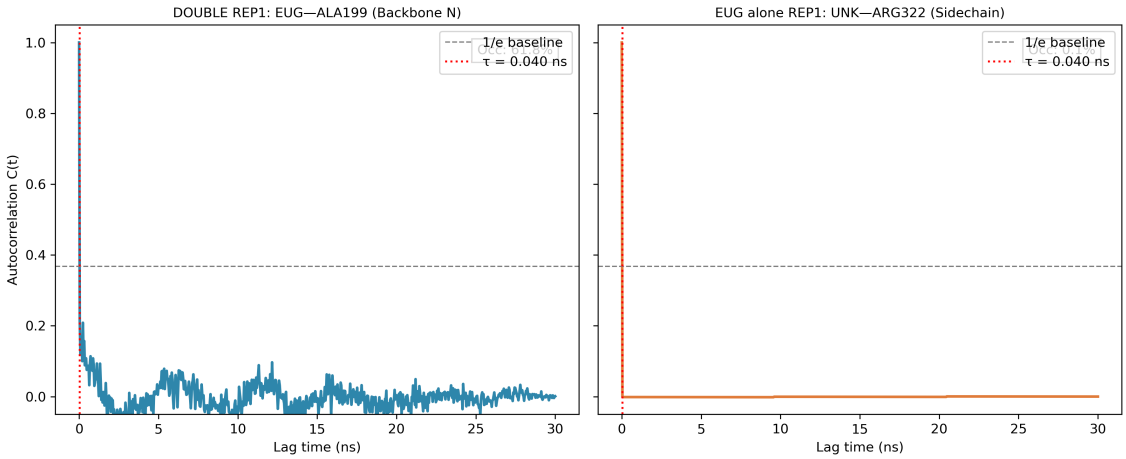
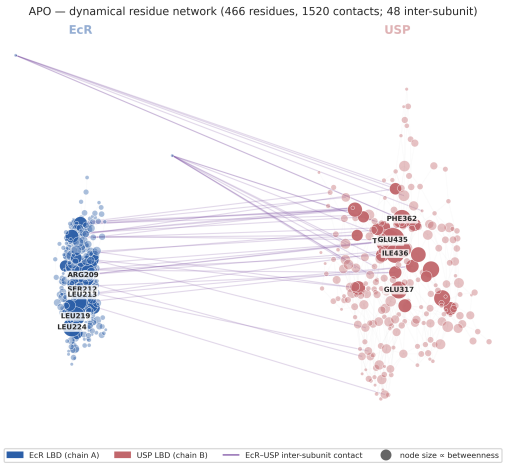
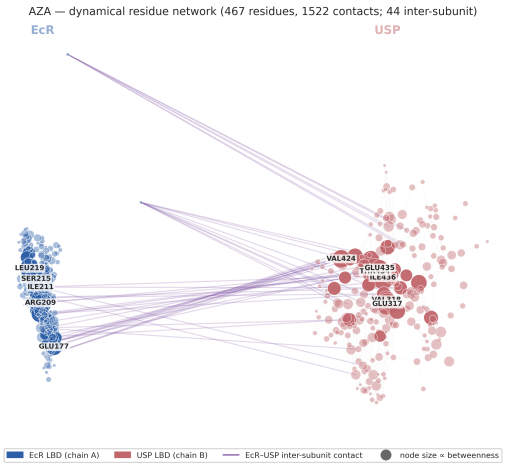


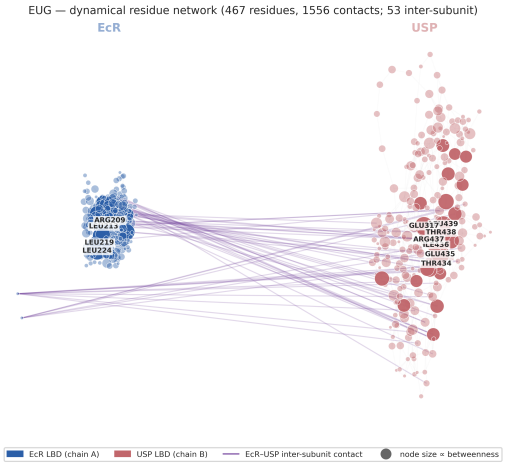
Figure S6: Hydrogen-bond autocorrelation functions for the dominant protein–eugenol polar contacts. Left (blue): EUG–ALA475:N backbone amide contact in the DOUBLE system (τ = 0.040 ns). The contact re-forms on a sub-nanosecond timescale despite rapid H-bond fluctuation (mean occupancy 75.6%). Right (orange): EUG–ARG383:O sidechain contact in the EUG-alone system. The ACF collapses to zero within a single lag step, indicating a transiently formed interaction lacking persistent temporal coherence.



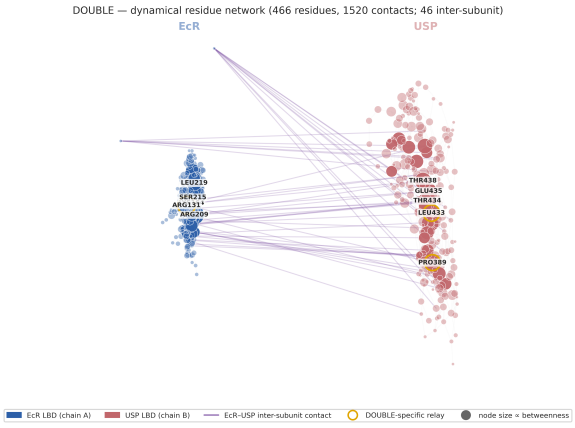
(a) APO



(b) AZA



(c) EUG



(d) DOUBLE

Figure S7: Dynamical residue-interaction networks for the four systems. Each subunit is laid out separately (EcR left, USP right). Nodes are residues sized by betweenness centrality and coloured by subunit; edges represent persistent heavy-atom contacts (occupancy ≥ 0.75, replicate-pooled); EcR–USP inter-subunit contacts are highlighted in purple. The conserved relay (ARG209/LEU219 in EcR; GLU496/THR495 in USP) is present in all systems. In the DOUBLE panel, gold-ringed nodes ARG131, LEU494, and PRO450 emerge as additional high-betweenness relays exclusive to the dual-occupancy state. Cross-state betweenness statistics for these residues are presented as Figure 3 in the main text.

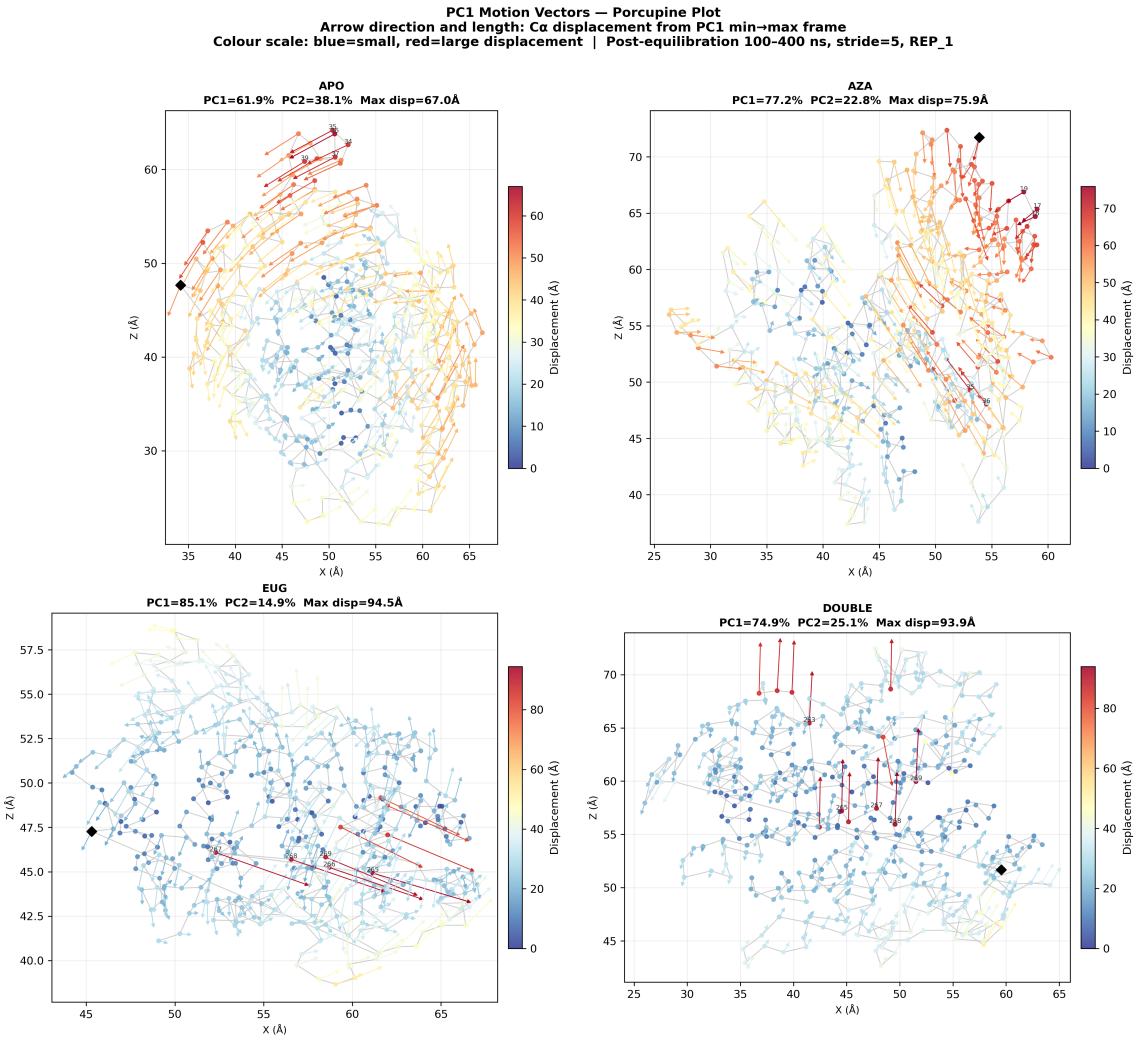


Figure S8: Porcupine plots visualising the dominant conformational motion (PC1) for the APO, AZA, EUG, and DOUBLE systems (REP_1; post-equilibration frames 100–400 ns; stride = 5). Stem length is proportional to displacement magnitude (blue = small, red = large). AZA: PC1 motion concentrated in the EcR N-terminal loop (residues 17–19), consistent with orthosteric anchoring stabilising the helical core. EUG and DOUBLE: PC1 motion peaks in the USP domain (residues 265–269). DOUBLE: despite similar maximum Cα displacement to EUG (93.9 vs. 94.5 Å), inter-replicate PC1 variance is markedly lower (SD = $\pm$3.3% vs. $\pm$7.6%), indicating that AZA co-occupancy constrains the reproducibility of PC1 sampling without reducing its amplitude.