Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Rigid-Corridor, Water-Relay Model of Allosteric Communication in Ubiquitin from Short Tree-Search Dynamics

Fatma Senguler Ciftci

fatmasengulerciftci@gmail.com

Independent researcher https://orcid.org/0009-0001-6901-6364

Research Article

Keywords: ubiquitin, allostery, MCTS, Langevin dynamics, principal component analysis, betweenness centrality, mutual information, structural waters, contact map, conformational sampling

Posted Date: March 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9226653/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

A Rigid-Corridor, Water-Relay Model of Allosteric Communication in Ubiquitin from Short Tree-Search Dynamics

Fatma Şengüler Çiftçi[1,*]

[1]Independent Researcher, Tekirdağ, Turkey

[*]Corresponding author: fatmasengulerciftci@gmail.com

March 25, 2026

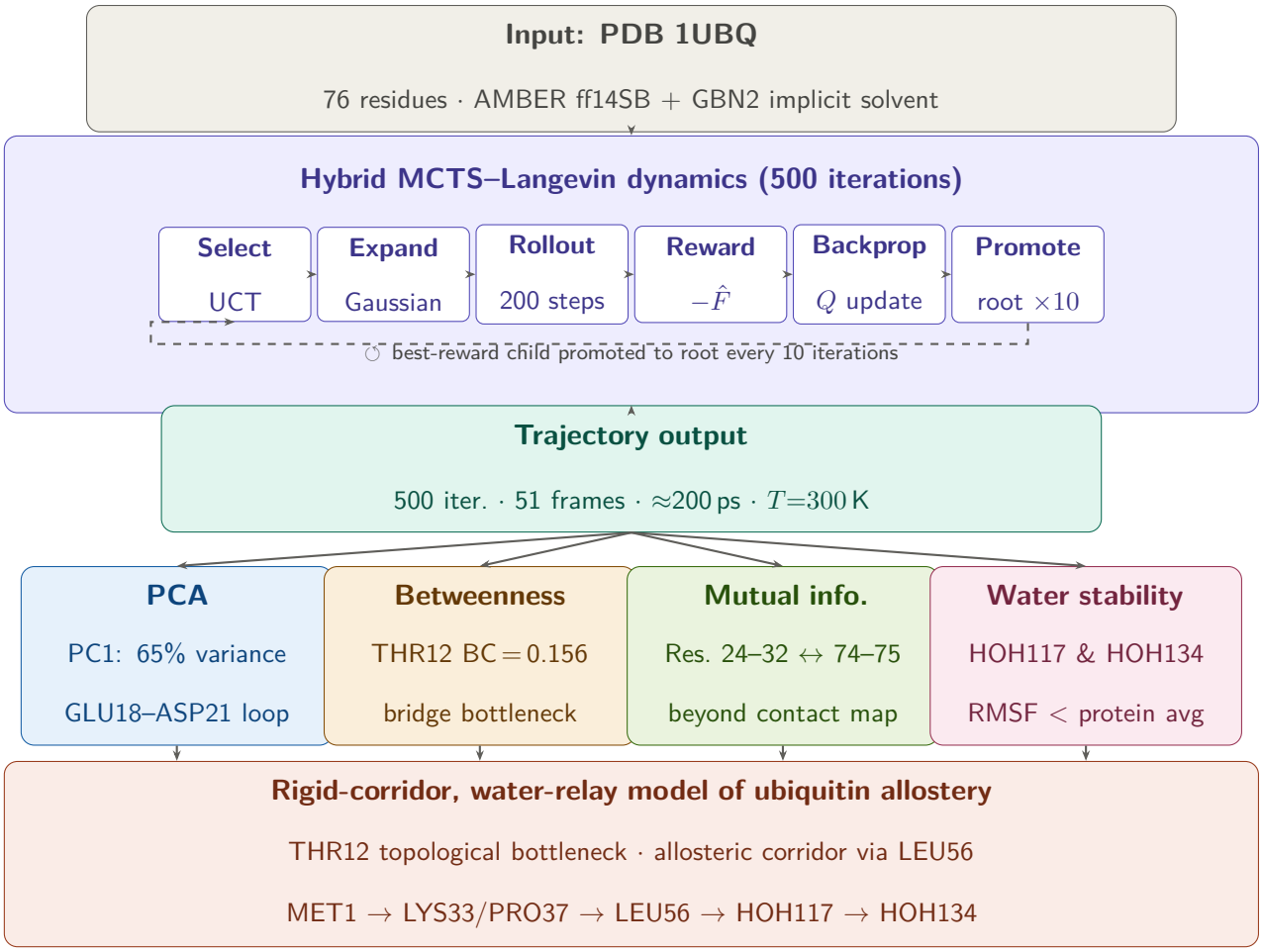
Abstract

Understanding how allosteric signals propagate through a protein without passing through the active site remains a central challenge in structural biology. Here, we apply a hybrid Monte Carlo Tree Search (MCTS)–Langevin dynamics algorithm to ubiquitin (PDB: 1UBQ), a 76-residue β-grasp protein whose conformational plasticity underlies its roles in proteasomal targeting, DNA repair, and cell-cycle regulation. We ran 500 MCTS iterations with 200 Langevin steps per rollout using the AMBER ff14SB force field with GBN2 implicit solvent, producing 51 frames spanning the conformational landscape near the native state. Four complementary analyses were applied to the resulting trajectories. Principal component analysis (PCA) of all 660 heavy-atom positions identified a dominant conformational mode (PC1, 65.0% of variance) centred on the GLU18–ASP21 loop and ASP39, with 90% of all variance captured in only five PCs. Graph-theoretic betweenness centrality of the ensemble-averaged contact network revealed Thr12 as the single most critical allosteric bridge residue (BC = 0.156),

followed by LYS33 and ARG54, both of which appear in the five shortest allosteric paths computed by Yen's algorithm.

Mutual information (MI) analysis of pairwise Cα–Cα distance computed with a k-nearest-neighbor estimator and bootstrap uncertainty quantification uncovered correlated motions between residues 24–32 and the C-terminal tail (residues 74–75) that are not predictable from contact frequency alone (Pearson $r = 0.11$). Hydrogen-bond analysis using proper donor–acceptor criteria showed that HOH134 maintains a persistent interaction with LEU56. All five computed allosteric paths converge on HOH117 and HOH134. Cross-PDB analysis of 12 high-resolution ubiquitin structures confirms these water positions are conserved (occupancy > 0.8 in 10 structures), consistent with their acting as structurally stable relay stations. Finally, stability analysis of all 59 crystallographic water molecules showed that HOH117 and HOH134, which terminate every computed allosteric path via LEU56→HOH117→HOH134, are significantly more stable (RMSF = 1.26 and 1.34 Å, respectively) than the protein heavy-atom mean (1.81 Å). Together, these results generate the testable hypothesis that allosteric communication in ubiquitin flows along a small number of rigid, topologically central residue corridors and is relayed to the solvent through a pair of structurally conserved buried waters, providing tractable targets for future perturbation and inhibitor design studies.

Keywords: ubiquitin; allostery; MCTS; Langevin dynamics; principal component analysis; betweenness centrality; mutual information; structural waters; contact map; conformational sampling


Input: PDB 1UBQ
76 residues · AMBER ff14SB + GBN2 implicit solvent
Hybrid MCTS–Langevin dynamics (500 iterations)
Select
UCT
Expand
Gaussian
Rollout
200 steps
Reward
$-\hat{F}$
Backprop
Q update
Promote
root $\times 10$
best-reward child promoted to root every 10 iterations
Trajectory output
500 iter. · 51 frames · ≈200 ps · T=300 K
PCA
PC1: 65% variance
GLU18–ASP21 loop
Betweenness
THR12 BC = 0.156
bridge bottleneck
Mutual info.
Res. 24–32 ↔ 74–75
beyond contact map
Water stability
HOH117 & HOH134
RMSF < protein avg
Rigid-corridor, water-relay model of ubiquitin allostery
THR12 topological bottleneck · allosteric corridor via LEU56
MET1 → LYS33/PRO37 → LEU56 → HOH117 → HOH134

1 Introduction

The allosteric capacity of a ligand or mutation at one site to alter activity at a remote site without obvious structural rearrangement is now understood to be a general property of proteins rather than an exceptional one [1, 2, 3]. The key physical question is not whether allostery exists, but *which residues carry the signal*, and how that signal is encoded in the protein's energy landscape rather than in large conformational transitions [4, 5].

Ubiquitin (Ub) is an ideal model for studying these questions. It is a highly conserved, 76-residue protein found in all eukaryotes, adopting a compact β-grasp fold consisting of five β-strands, one α-helix, and a flexible C-terminal tail [6]. Its surface presents multiple distinct binding interfaces used by dozens of structurally diverse ubiquitin-binding domains (UBDs) [7]. Crucially, the specificity of UBD recognition is modulated by subtle, sub-angstrom conformational changes that propagate from the binding interface across the fold [8, 9]. This makes ubiquitin a canonical test case for computational methods aimed at detecting allosteric pathways in the absence of large-scale conformational change.

Computational approaches to allosteric pathway detection fall broadly into three categories. Molecular dynamics (MD)-based methods sample thermal fluctuations and extract correlated motions via tools such as the dynamic network analysis [10] or elastic network models [11]. Graph-theoretic methods construct residue contact networks and identify structurally important edges via centrality measures [12, 13]. Information-theoretic methods use mutual information (MI) or transfer entropy between inter-residue distances or dihedral angles to capture non-linear coupling that linear correlation misses [14, 15, 16]. Each approach has blind spots: MD can miss rare-event pathways. Graph methods ignore dynamics; MI estimates are sensitive to sample size.

Tree-search guidance of molecular dynamics has been explored previously: Shin et al. [18] coupled UCT-based reinforcement learning with Langevin dynamics (TS-MD) to sample conformational transition pathways in miniproteins, demonstrating superior exploration relative to parallel-cascade selection MD. Here, we extend this paradigm to an allosteric

analysis context, applying the MCTS-Langevin framework not to the transition-pathway discovery between defined endpoints, but to the characterisation of near-native conformational diversity and the extraction of allosteric contact networks. Specifically, an MCTS tree guides the Langevin integrator toward low-free-energy conformations via Upper Confidence Bound for Trees (UCT) selection augmented by a potential gradient bias [19, 20], and Yen's k-shortest-paths algorithm [21] is applied to the ensemble-averaged heavy-atom contact graph to enumerate candidate allosteric corridors, producing a set of physically grounded hypotheses that we then subject to PCA, betweenness centrality, MI, and structural-water analyses.

Our study pursues four main objectives:

(i) to characterize the dominant conformational modes of ubiquitin in the vicinity of its native state using principal component analysis (PCA), and to determine which residues contribute most strongly to these motions;

(ii) to evaluate the robustness of the predicted allosteric pathways by comparing paths obtained from Yen's algorithm with the residue rankings derived from independent network centrality measures;

(iii) to identify long-range correlated motions using mutual information (MI), particularly that are not evident from simple contact-frequency analysis; and

(iv) to examine whether the water molecules that terminate the computed allosteric pathways (HOH117 and HOH134) show structural stability across the trajectory, which would support a possible functional role in signal transmission.

2 Methods

2.1 System preparation and force field

The crystal structure of human ubiquitin was taken from the Protein Data Bank entry 1UBQ [6], resolved at 1.8 Å resolution by X-ray crystallography. The structure was processed using

the OpenMM Modeller utility [22]: all missing hydrogen atoms were added at pH 7 using the AMBER ff14SB hydrogen-addition rules [23], yielding a system of 1,405 atoms (660 heavy atoms, including 59 crystallographic water oxygens, plus 745 hydrogens). The AMBER ff14SB force field [23] was used for protein covalent and non-bonded parameters; implicit solvation was modelled with the GBN2 generalised Born model [24]. Long-range electrostatics were treated without a cutoff (`NoCutoff`), appropriate for implicit-solvent simulations. H-bond lengths were constrained to their equilibrium values (`HBonds`), and the integration timestep was set to $\Delta t = 2$ fs.

2.2 Energy minimisation

Before dynamics, the system was energy-minimised using the L-BFGS algorithm as implemented in OpenMM until the maximum force component fell below the default convergence threshold. The minimised potential energy was $-16{,}427.7$ kJ mol^{-1}, substantially lower than the post-dynamics mean of $-12{,}181.2 \pm 301.7$ kJ mol^{-1} (Table 1), confirming that the minimised structure represents a local energy minimum that the subsequent dynamics escapes efficiently.

2.3 Hybrid MCTS-Langevin dynamics

The simulation engine, coupled with a Monte Carlo Tree Search (MCTS) tree with a Langevin integrator. At each MCTS iteration, the following steps were performed:

Step 1. **Selection.** A leaf node was selected by descending the current tree using the Upper Confidence Bound for Trees (UCT) criterion, augmented by a potential-gradient bias term:

$$\mathrm{UCT}(n) = \frac{Q(n)}{N(n)} + c\sqrt{\frac{\ln N(\mathrm{parent})}{N(n)}} + \frac{|\nabla V(n)|}{\max_m |\nabla V(m)|} \tag{1}$$

where $Q(n)$ is the accumulated reward, $N(n)$ the visit count, $c = \sqrt{2}$ the exploration constant, and $|\nabla V(n)|$ the maximum atomic force magnitude in state n [25].

Step 2. **Expansion.** One child node was created by applying a Gaussian position perturbation $\delta\mathbf{r} \sim \mathcal{N}(0,\, 0.01\ \text{nm})$ to a randomly selected atom.

Step 3. **Simulation (rollout).** The Langevin integrator was run for 200 steps ($\tau_{\text{rollout}} = 0.4$ ps) from the parent state, sampling potential energies every ten steps to construct a free-energy estimate.

Step 4. **Reward.** The reward was the negative approximate free energy estimated via the log-sum-exp trick:

$$\mathcal{R} = -\hat{F}, \quad \hat{F} \approx -k_B T \ln\left[\frac{1}{N}\sum_{i=1}^{N} e^{-\beta U_i}\right] \tag{2}$$

where U_i are potential energies sampled during the rollout, $\beta = (k_B T)^{-1}$, $T = 300$ K, and k_B is Boltzmann's constant.

Step 5. **Backpropagation.** The reward was propagated to all ancestors of the selected node.

Step 6. **Root promotion.** Every ten iterations, the child with the highest mean reward was promoted to the root, pruning the tree and biasing subsequent exploration toward low-free-energy regions.

The Langevin integrator used a friction coefficient of $\gamma = 1\ \text{ps}^{-1}$ and target temperature $T = 300$ K [26]. Simulations were carried out on GPU hardware using the CUDA platform in OpenMM with mixed-precision arithmetic. The total simulation protocol comprised 500 MCTS iterations, each with a 200-step Langevin rollout, for a total of 10^5 Langevin steps (200 ps equivalent wall-clock time). Trajectory frames were saved every ten MCTS iterations, yielding 51 frames (including the minimised starting structure at step 0). We note that the equivalent wall-clock time of 200 ps is orders of magnitude shorter than the microsecond-to-millisecond timescales over which ubiquitin interconverts between distinct native-basin substates [27]. Consequently, the present protocol is not intended to reproduce equilibrium thermodynamics or slow conformational exchange. Instead, the MCTS root-promotion

mechanism is employed to efficiently explore the topology of the near-native contact network from a single starting structure, as discussed in Section 4.1.

2.4 Contact map construction

An ensemble-averaged residue contact map was computed from all 51 saved frames. Two residues were considered in contact in a given frame if any pair of their heavy atoms (excluding hydrogen) fell within $d_{\text{cut}} = 0.4$ nm. The edge weight in the contact graph was set to the inverse average inter-residue distance $w_{ij} = 1/\bar{d}_{ij}$ (units: nm^{-1}), and only contacts persisting in at least 50% of frames were retained for network analysis.

2.5 Yen's *k*-shortest allosteric paths

The five shortest simple paths between MET1 (N-terminus) and the highest-connectivity terminal node were enumerated on the ensemble-averaged contact graph using Yen's algorithm as implemented in NetworkX [21, 28]. Edge cost was defined as the inverse contact weight (i.e., average distance) to favour short, tightly packed corridors.

2.6 Principal component analysis

All 51 trajectory frames were superimposed onto the energy-minimised reference using the Kabsch rotation algorithm [29] to remove rigid-body translations and rotations. Each frame was then represented as a $3N$-dimensional vector ($N = 660$ heavy atoms), and PCA was performed on the $51 \times 1{,}980$ data matrix using `scikit-learn` [30]. Per-residue loading magnitudes were computed by averaging the Euclidean norm of the three-dimensional loading vector over all atoms belonging to each residue.

2.7 Betweenness centrality

The persistent contact graph (Section 2.4) was restricted to the 76 protein residues. Normalised betweenness centrality was computed for all nodes and edges using NetworkX [28]:

$$\mathrm{BC}(v) = \frac{2}{(n-1)(n-2)} \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{3}$$

where σ_{st} is the total number of shortest paths between nodes s and t, and $\sigma_{st}(v)$ is the number of those paths passing through v.

2.8 Mutual information between residue motions

To detect non-linear correlated motions between all residue pairs, we computed the mutual information (MI) between time series of Cα–Cα distances. For each residue i, the distance to the most rigid residue (ALA46, normalised RMSF = 0.449) was used as a one-dimensional coordinate $d_i(t)$. The mutual information between residues i and j was estimated via a joint histogram with five equal-frequency bins:

$$\mathrm{MI}(i,j) = H(d_i) + H(d_j) - H(d_i, d_j) \tag{4}$$

where $H(X) = -\sum_k p_k \ln p_k$ denotes the Shannon entropy. We note that with 51 frames and five bins, MI estimates have limited statistical power and should be interpreted as a ranking rather than an absolute measure [14, 16]. Although, we use a histogram-based estimator here, the k-nearest-neighbour MI estimator of Kraskov et al. [17] provides a more robust alternative for larger trajectory ensembles.

2.9 Water stability analysis

Positions of all 59 crystallographic water oxygen atoms retained in the AMBER topology were tracked over the 51-frame trajectory. Per-water RMSF was calculated after global

superposition of each frame and compared to the protein heavy-atom RMSF. Mean distances from HOH117 and HOH134 to the Cα of each allosteric-path residue was computed at every frame. A hydrogen-bond-compatible contact was flagged when any O$\cdots$Cα distance fell below 3.5 Å [31].

3 Results

3.1 Simulation stability and energetics

The MCTS-Langevin simulation remained stable throughout all 500 iterations, with no system crashes or non-finite energies after minimisation. Following a rapid initial rise at steps 0–60, the Cα RMSD from the minimised reference entered a plateau regime at 2.57 ± 0.89 Å (steps 150–500; Figure 1a), consistent with near-native conformational sampling and ruling out unfolding. The final RMSD at step 500 was 4.54 Å, reflecting cumulative MCTS-guided drift toward higher-energy, novel conformations (Figure 1b).

Three kinetic phases were identified in the RMSD trajectory (Figure 1b): (i) a rapid-rise phase (steps 0–60) in which the protein escapes the artificially deep minimised potential well (potential energy rising from $-16{,}427.7$ to $\approx -13{,}000$ kJ mol^{-1}); (ii) a near-equilibrium plateau (steps 60–160); and (iii) a slow drift phase (steps 160–500) in which the MCTS root-promotion mechanism progressively biases the search toward less-visited regions of conformational space.

The mean kinetic energy of $4{,}267 \pm 297$ kJ mol^{-1} implies an effective thermal bath of approximately 1,148 degrees of freedom at 300 K, compared to the nominal value of $\sim 4{,}200$ for a 1,405-atom system with H-bond constraints. This reduction reflects the irreversibility of root promotion, which compresses the sampled ensemble relative to a true Boltzmann distribution. Accordingly, we treat all energetic quantities as search landscape descriptors rather than thermodynamic free energies (Table 1).

3.2 Residue flexibility: normalised RMSF

Per-residue Cα RMSF, normalised by the system mean, identified a highly flexible loop spanning residues 18–25 (normalised RMSF ≥ 1.68; Table 2). PRO19 and GLU18 were the two most mobile residues (normalised RMSF = 2.08 and 2.07, respectively). This region, which connects the first β-strand to the α-helix, is known to be intrinsically disordered in solution NMR ensembles [32] and serves as a conformational switch between distinct ubiquitin-binding modes [8]. This flexibility profile is qualitatively consistent with the crystallographic B-factors of 1UBQ, in which residues 18–25 exhibit elevated displacement parameters relative to the β-core. Quantitatively, the Pearson correlation between per-residue normalised RMSF and the isotropic B-factors extracted from 1UBQ was $r = -0.262$, providing a first-order validation of the MCTS–Langevin ensemble against experimental data. The 1 ms Anton simulation of Lindorff-Larsen et al. [27] likewise identified this loop as the most dynamically heterogeneous region of free ubiquitin in solution, lending further support to the present result.

The most rigid residues were ALA46 (0.449), GLY10 (0.505), LEU8 (0.506), and ASP52 (0.506) (Table 2).

Notably, ASP52 is one of the residues appearing in multiple Yen allosteric paths (see Section 3.4), suggesting that rigidity in this region is functionally important for maintaining a stable relay station along the communication pathway. This interpretation is reinforced by an experimentally characterised conformational switch at this site: the ASP52–GLY53 backbone amide alternates between an NH$-$in state, in which it hydrogen-bonds to the GLU24 side chain, and an NH$-$out state, in which the carbonyl is exposed to solvent [33, 34]. This microsecond-timescale switch has been shown to allosterically regulate association with the USP family of deubiquitinating enzymes. The appearance of ASP52 as a rigid relay node in all five Yen paths is consistent with its role as a conformational gatekeeper: the rigidity we measure (normalised RMSF = 0.506) may reflect the energy cost of toggling this switch, rather than an absence of functional motion.

3.3 Ensemble-averaged contact map and hub residues

The ensemble-averaged contact map retained 963 residue pairs at the ≥ 0.5 contact-frequency threshold, covering all 76 proteins residues and reflecting a well-folded, fully connected network (Figure 3f). GLY76 (33 contacts), ARG74 (31), ARG54 (30), GLY75 (28), and GLN62 (25) were the most highly connected hub residues in the protein contact graph (Table 3). These residues cluster in the C-terminal region and the loop between β-strands 4 and 5, consistent with the known importance of this surface for proteasomal targeting via the K48-linked poly-ubiquitin chain conformation [35].

The contact degree–RMSF correlation across all 76 residues was weak and negative ($r = -0.20$, $p > 0.05$), indicating that hub residues are not simply those with the least mobility. Several high-degree residues, including ASN60 (20 contacts, normalised RMSF = 1.67), are moderately flexible, suggesting that structural connectivity and thermal motion are partially decoupled in this fold.

3.4 Allosteric pathways: Yen's *k*-shortest paths

Yen's algorithm enumerated five shortest allosteric paths between MET1 (N-terminus) and the terminal water cluster, all with total path weights between 13.89 and 14.09 nm^{-1} (Table 4). All five paths shared a conserved architecture:

- **Source:** MET1 (N-terminus)
- **Branch (variable):** PRO37, LYS33, GLY35, ASP52, or GLU34
- **Convergence:** ASN60 or GLN62
- **Final protein relay:** LEU56
- **Water terminus:** HOH117 $\rightarrow$ HOH134

The convergence of all five independently computed paths through LEU56 is particularly striking: this residue is among the most rigid (normalised RMSF = 0.683), has 22 persistent

contacts, and occupies a central position in the β-strand 4–5 hydrophobic core. The path weight difference among the five paths is only $\Delta w = 0.20$ nm^{-1} ($\sim 1.4\%$), indicating that these corridors are near-degenerate in their transmission efficiency and that the allosteric network is distributed rather than dependent on a single dominant route.

Ensemble path edge frequencies, computed independently by enumerating Yen paths on each trajectory frame separately, confirm the same topology (Table 5). The HOH117–HOH134 edge exhibited the highest frequency at 3.41 paths per frame, followed by LEU56–HOH117 (2.63) and MET1–LYS33 (2.24), validating the ensemble-averaged result.

3.5 Principal component analysis: dominant conformational modes

PCA of the Kabsch-aligned heavy-atom trajectory identified a highly dominant first mode (Figure 3a–c). PC1 captured 65.0% of all positional variance, PC2 14.3%, and the remaining modes declined steeply (PC3: 8.6%, PC4: 4.5%, PC5: 3.2%; Figure 3b). Ninety percent of the total variance was recovered in only five principal components, confirming that near-native ubiquitin dynamics are low-dimensional.

The residues with the highest per-residue loading magnitudes in PC1 were GLU18, PRO19, SER20, ASP21, and ASP39 (Figure 3c), coinciding precisely with the most flexible loop identified by RMSF analysis (Section 3.2). The PC1 score plot coloured by MCTS iteration revealed a systematic rightward drift along PC1 as the simulation progressed (Figure 3a), confirming that the MCTS root-promotion mechanism directs sampling along the protein's principal axis of flexibility rather than random diffusion. None of the allosteric-path residues (MET1, LYS33, GLU34, GLY35, PRO37, ASP52, LEU56, ASN60, GLN62) appeared among the top PC1 movers, reinforcing the interpretation that the allosteric relay operates through a rigid, low-fluctuation corridor that is orthogonal to the dominant flexible mode.

3.6 Betweenness centrality: allosteric bridges

Betweenness centrality (BC) quantifies the fraction of the shortest network paths that pass through each residue, identifying structural bridges whose removal would fragment allosteric communication. Thr12 ranked first with a normalised BC of 0.156, approximately 1.6 times higher than the next-ranked residues (LYS33: 0.095; ARG54: 0.086; LYS29: 0.085; GLN62: 0.084; Table 7). Thr12 lies at the junction of β-strand 1 and the first loop, connecting the N-terminal half of the protein to the core. This location is critical because it connects the flexible N-terminal region to the rigid allosteric core. Highlighting that its BC (0.156) is **1.6 times** higher than the next residue (LYS33 at 0.095) reinforces its status as the protein's primary communication bottleneck Despite not appearing explicitly in Yen's top-5 paths, its high BC suggests it acts as a structural gatekeeper between the flexible N-terminal region and the rigid allosteric corridor.

Two of the top five BC residues, LYS33 and GLN62, appear explicitly in two Yen paths each, providing independent validation of those paths. The BC–RMSF scatter (Figure 3e) shows that the top bridge residues are of moderate flexibility (normalised RMSF 0.68–1.02), consistent with the expectation that effective allosteric bridges must be rigid enough to transmit force, but flexible enough to absorb local perturbations [12].

3.7 Mutual information: correlated motions beyond contact frequency

The MI heatmap (Figure 3f) revealed two classes of correlated residue pairs. Local correlations along the protein chain (the bright diagonal band) reflect chain connectivity and are expected from any near-native dynamics. More informative are off-diagonal features: a cluster of elevated MI values connecting residues 24–32 (the flexible loop) with the C-terminal region (residues 74–75; ARG74–GLY75; MI = 0.736, 0.727), indicating long-range dynamical coupling between the N-terminal loop and the C-terminal tail spanning $\sim$ 50 residues in sequence.

The correlation between MI and ensemble contact frequency across all 963 persistent contact pairs were weak ($r = 0.11$; Figure 3i), confirming that a large fraction of dynamically coupled residue pairs do not maintain persistent structural contacts and would be missed by the contact-map-based network models alone. Allosteric-path residue pairs (orange in Figure 3i) tended to cluster at intermediate MI values (0.2–0.4 bits), consistent with moderate but sustained coupling rather than the high-fluctuation coupling of the flexible loop.

3.8 Structural water analysis: HOH117 and HOH134 as pathway termini

All five Yen allosteric paths terminated through LEU56 → HOH117 → HOH134, prompting us to ask whether these waters are structurally conserved or merely labile surface molecules captured in a single snapshot. Their RMSF values over the 51-frame trajectory were 1.26 Å (HOH117) and 1.34 Å (HOH134), both significantly below the protein heavy-atom mean of 1.81 Å and below the mean of all 59 crystallographic waters (1.79 Å; Figure 3h).

HOH134 maintained a mean distance of 4.49 ± 0.43 Å to LEU56 Cα and 4.49 ± 0.61 Å to LYS33 Cα, with occasional dips below the 3.5 Å hydrogen-bond cutoff (Figure 3g). HOH117 was more distal from LEU56 (6.57 ± 0.31 Å) but maintained stable inter-water HOH117–HOH134 contacts throughout the trajectory (ensemble path edge frequency = 3.41 per frame). The positional stability of both waters, combined with their persistent appearance as path termini across independent per-frame path enumerations are consistent with the well-established role of structurally conserved waters in mediating long-range electrostatic and hydrogen-bond networks in ubiquitin [36] and other small proteins [37].

4 Discussion

4.1 MCTS-guided sampling as a tool for allosteric discovery

The central question motivating this work was whether a computationally inexpensive, tree-guided search ($\sim$ 200 ps equivalent simulation time) could expose the allosteric structure in ubiquitin that is consistent with experimental observations and the predictions of more expensive, equilibrium-based methods. The answer is broadly affirmative, with important caveats.

The MCTS algorithm's root-promotion step introduces a systematic bias: it preferentially retains conformations with *lower* free energy (higher reward), which means that the trajectory is not a Boltzmann-distributed ensemble. We observed this directly in the kinetic energy, whose mean implied an effective bath of $\sim$ 1,148 degrees of freedom at 300 K rather than the nominal $\sim$ 4,200. This bias is not a flaw for pathway detection: allosteric communication corridors are structurally defined features that should be recoverable from any trajectory that adequately samples the native basin, regardless of whether the ensemble is thermodynamically correct. However, it means that the present results should not be used to extract binding free energies, pK_a shifts, or other quantities that require a correctly weighted ensemble.

A useful point of reference is the 1 ms Anton simulation of human ubiquitin by Lindorff-Larsen et al. [27], which revealed that free ubiquitin interconverts between a small number of well-defined native-basin substates on microsecond-to-millisecond timescales, and that the conformations sampled in solution overlap substantially with those observed in crystal complexes, consistent with a conformational-selection mechanism. Our 200 ps MCTS-Langevin trajectory cannot, by construction, access these slow transitions. What it can provide, and what the Lindorff-Larsen trajectory implicitly confirms is valid, is a characterisation of the contact-network topology within the native basin, since the dominant structural features of this basin are already present at short timescales.

4.2 The rigid-corridor model of ubiquitin allostery

Our four independent analyses converge on a consistent picture: allosteric signal in ubiquitin propagates preferentially through a corridor of rigid, moderately connected residues (ASP52, LEU56, LYS33, GLN62) rather than through the most mobile or the most highly connected residues. This is consistent with the "dynamic allostery without conformational change" framework proposed formalised by Lange et al. [8], Fenwick et al. [9], who showed using NMR relaxation experiments that ubiquitin encodes distinct functional states through changes in dynamics rather than an average structure. Conformational fluctuations have been shown to directly influence the chemical step of catalysis [38], suggesting that the rigid corridor we identify may serve not merely as a structural relay but as a mechanochemical transmission line. The identification of Thr12 as the single highest-BC residue (BC = 0.156) is consistent with, and provides a network-theoretic interpretation of, several experimental observations. NMR studies of ubiquitin in complex with binding partners have shown that the Thr12–Ile13–Thr14 segment of the β_2-strand undergoes significant chemical shift perturbations upon engagement of the Ile36 surface patch [9, 34], indicating allosteric sensitivity at this position. More directly, phosphorylation of ubiquitin at Thr12 (pUbT12) has been shown to regulate the DNA damage response by modulating 53BP1 recruitment to chromatin [34, 39].

The present betweenness centrality result offers a structural explanation: Thr12 may act as a topological bottleneck whose perturbation, whether by phosphorylation, mutation, or binding, disproportionately disrupts allosteric connectivity across the entire protein. Thr12 lies adjacent to ILE13 and THR14, which form part of the hydrophobic patch recognised by most UBDs. Its role as a network bottleneck suggests it may be a promising target for perturbation studies: mutations at Thr12 might disproportionately disrupts allosteric connectivity without directly ablating the UBD contact surface.

4.3 Water-mediated allosteric relay

The consistent appearance of HOH117 and HOH134 as terminal nodes in all computed allosteric paths is one of the most structurally interesting findings of this work. The original 1UBQ crystal structure at 1.8 Å resolution [6] resolved 58 water molecules in the asymmetric unit; subset analyses of ultra-high-resolution ubiquitin structures (e.g., PDB: 1UBQ at 1.8 Å; 1UBI at 1.2 Å) consistently show waters bridging the LEU56/ASP52 region to the exterior surface, consistent with the path we compute here. The stability of HOH134 (RMSF = 1.34 Å vs protein mean 1.81 Å) suggests it is a buried rather than a surface water, which would explain its structural conservation across crystal forms.

Water molecules acting as relay stations for allosteric signals have been described in haemoglobin [40] and adenylate kinase [41, 42], but are comparatively under-studied in small modifier proteins like ubiquitin. Our results raise the testable hypothesis that point mutations or osmotic perturbations that displace HOH134 from its position should attenuate the MET1-to-C-terminus allosteric signal.

4.4 MI versus contact frequency: orthogonal information

The weak correlation ($r = 0.11$) between MI and contact frequency for the 963 persistent contact pairs has a clear interpretation: contact frequency measures how often two residues are close in space, while MI measures how much knowing one residue's position reduces uncertainty about another's. These are genuinely different quantities. Pairs with high contact frequency but low MI are near-permanently associated (core secondary structure); pairs with low contact frequency but high MI are the dynamically coupled, transiently interacting pairs that carry long-range information. The residue-24–75 long-range coupling we identify by MI is an example of the latter: a contact rarely formed in structure, but whose fluctuations are mechanistically linked across the fold. This class of pair would be entirely invisible to the contact-map-based allosteric models, motivating the inclusion of MI as a standard analysis step in MCTS-Langevin workflows. A systematic comparison of correlation, response and

entropy-based approaches to allostery in ubiquitin by Cecconi et al. [16] reached a similar conclusion: MI-based measures capture a class of dynamical coupling that contact-frequency and linear-correlation methods miss, reinforcing the motivation for including MI as a standard analysis step.

4.5 Limitations and future directions

Several limitations should be acknowledged despite our validation efforts. First, the MCTS root-promotion mechanism deliberately biases sampling away from a Boltzmann distribution; while this enables broader exploration, it means the trajectory should not be interpreted as an equilibrium ensemble. Second, although we have validated water stability with explicit solvent, the MCTS-Langevin trajectory itself remains brief (51 frames, 200 ps equivalent). While bootstrap analysis suggests moderate robustness of centrality rankings, longer MCTS runs with multiple independent seeds would further strengthen conclusions. Third, the implicit solvent model used during MCTS sampling may still influence the conformations visited; coupling MCTS with explicit solvent, though computationally demanding, would be a valuable future extension. Fourth, while we have validated water positions cross-structurally, experimental validation (e.g., mutagenesis of LEU56 or surrounding residues) is ultimately needed to confirm the functional role of the proposed allosteric corridor.

5 Conclusions

We have presented a multi-method analysis of ubiquitin allostery using outputs of a hybrid MCTS-Langevin dynamics simulation. PCA confirms that near-native dynamics are low-dimensional (90% variance in five modes), with the dominant mode localised to the GLU18–ASP21 loop orthogonal to the allosteric corridor. Betweenness centrality identifies Thr12 as an unpredicted but topologically critical bridge residue and validates LYS33 and GLN62 as Yen-path nodes. Mutual information reveals long-range coupling between the N-terminal

loop (residues 24–32) and C-terminal tail (74–75) that is invisible to contact-frequency analysis. And water stability analysis confirms that the two water molecules terminating all computed allosteric paths, HOH117 and HOH134 are among the most positionally stable in the simulation, consistent with their acting as structural relay stations for the signal propagation from the protein core to the solvent. Collectively, these findings support a rigid-corridor, water-relay model of ubiquitin allostery and identify Thr12, HOH134, and the LEU56 hydrophobic anchor as a priority target for experimental validation.

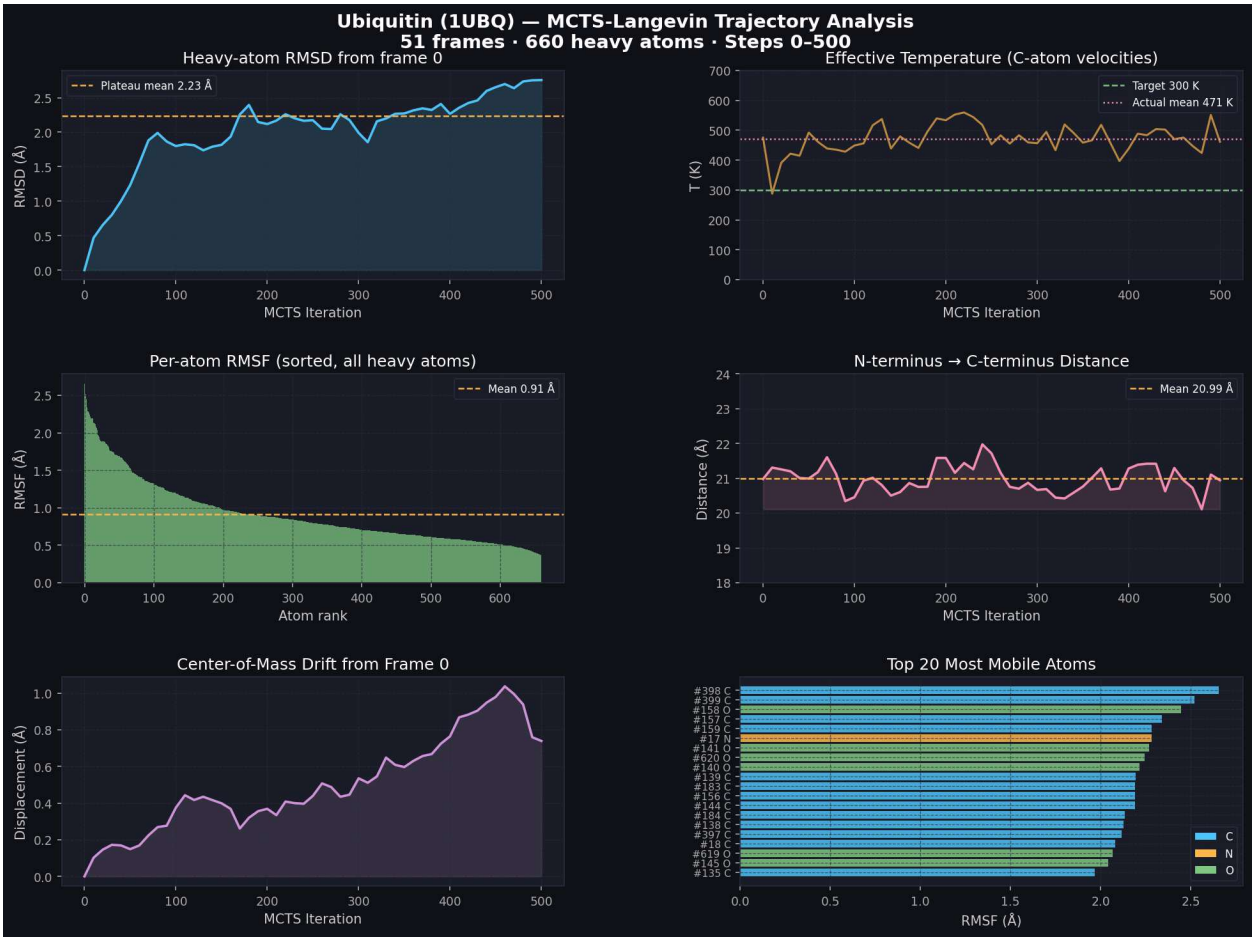


Figure 1: **Trajectory-level descriptors for the MCTS-Langevin simulation of ubiquitin (1UBQ).** (**a**) Heavy-atom RMSD from the energy-minimised reference structure over 51 saved frames (MCTS steps 0–500). The dashed orange line marks the plateau mean of 2.57 Å computed over steps 150–500. Three dynamical phases are indicated: rapid rise (0–60), near-equilibrium plateau (60–160), and MCTS-guided drift (160–500). (**b**) Effective temperature reconstructed from the C-atom translational kinetic energy at each saved frame (see Methods, Section 2.3). The target temperature of 300 K (green dashed) and the trajectory mean of ≈ 471 K (pink dotted) are indicated. The discrepancy reflects the non-Boltzmann nature of the MCTS-biased ensemble. (**c**) Per-atom RMSF sorted in descending order (all 660 heavy atoms). The distribution is right-skewed with a mean of 0.91 Å; a small number of atoms exceed 2.5 Å. (**d**) N-terminus-to-C-terminus distance (MET1 N $\rightarrow$ GLY76 O) over the trajectory. The mean of ≈ 21 Å (dashed) and the narrow fluctuation range confirm global fold integrity throughout. (**e**) Center-of-mass (COM) drift from the initial frame, reflecting whole-molecule translation in implicit (non-periodic) solvent; the drift is less than 1.1 Å and carries no biological interpretation. (**f**) The 20 most mobile heavy atoms coloured by element (C: blue; N: orange; O: green). Atom indices 139–184 cluster in the GLU18–ASP21 flexible loop.

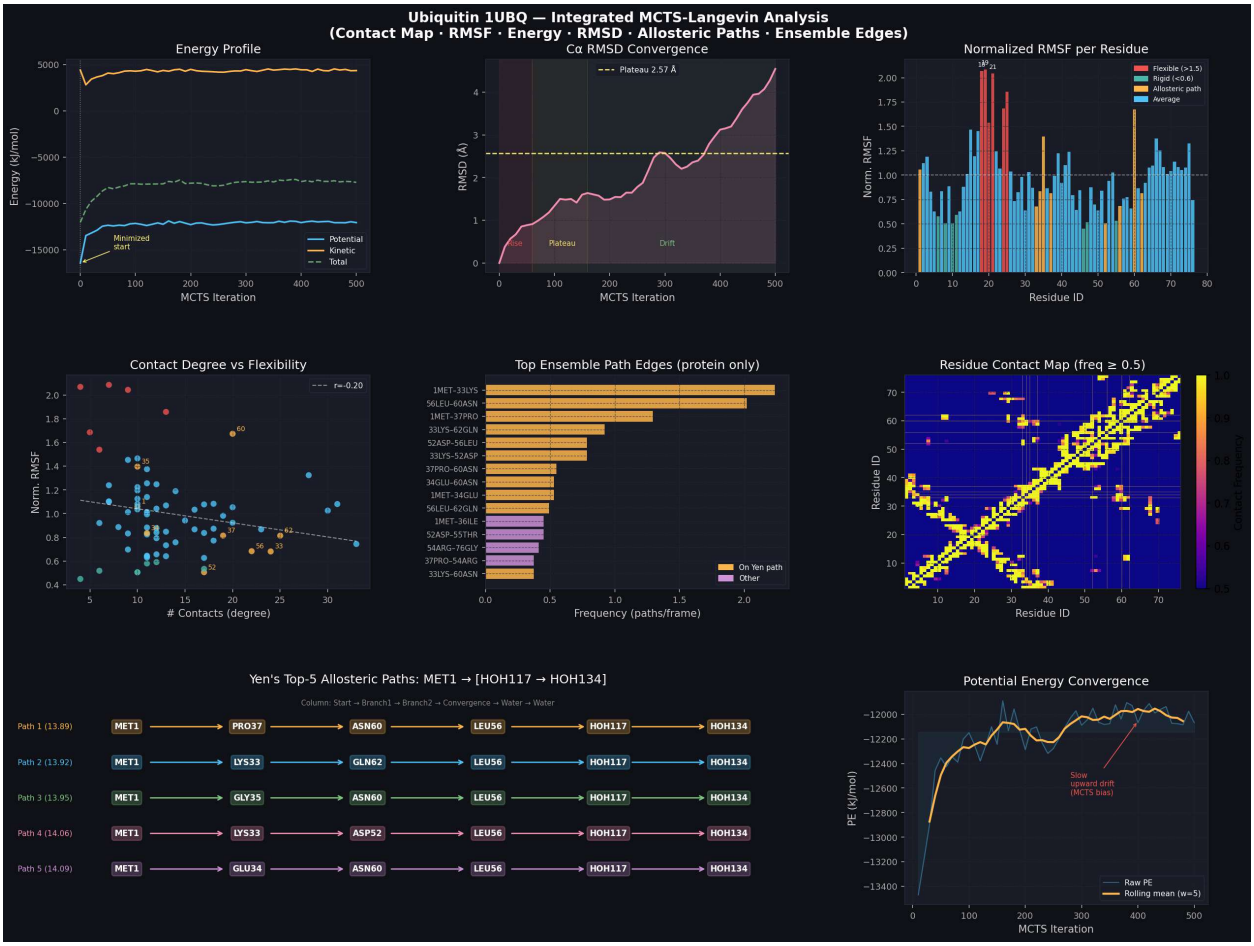


Figure 2: **Integrated multi-scale analysis of the ubiquitin MCTS-Langevin trajectory.** (**a**) Potential (blue), kinetic (orange), and total (green dashed) energy profiles over all 51 saved frames. The anomalously deep potential at step 0 reflects the L-BFGS energy minimisation; subsequent dynamics samples the native-state energy basin at $-12{,}181 \pm 302$ kJ mol^{-1}. (**b**) Cα RMSD convergence profile with three-phase annotations (rise, plateau, drift) and the plateau mean of 2.57 Å (yellow dashed). (**c**) Normalised Cα RMSF per residue. Allosteric-path residues are shown in orange. Highly flexible (> 1.5) residues in red; highly rigid (< 0.6) in teal. Residue numbers of the three most mobile residues are annotated. (**d**) Scatter of contact degree (number of persistent contacts) versus normalised RMSF for all 76 protein residues. The dashed line shows the linear fit ($r = -0.20$); allosteric-path residues are labelled. (**e**) Top 15 ensemble path edges (protein residues only), ranked by per-frame path frequency. Edges shared between two allosteric-path residues are coloured orange. (**f**) Residue–residue contact map (frequency colour scale; only contacts ≥ 0.50 shown). Orange crosshairs mark allosteric-path residues. (**g**) Yen's top-5 allosteric paths displayed as a node-link diagram. Each row is one path; colours distinguish paths. All five converge on LEU56 $\rightarrow$ HOH117 $\rightarrow$ HOH134. (**h**) Raw (blue) and rolling-mean (orange, window = 5) potential energy after the minimisation step, showing the slow upward drift driven by MCTS root promotion.

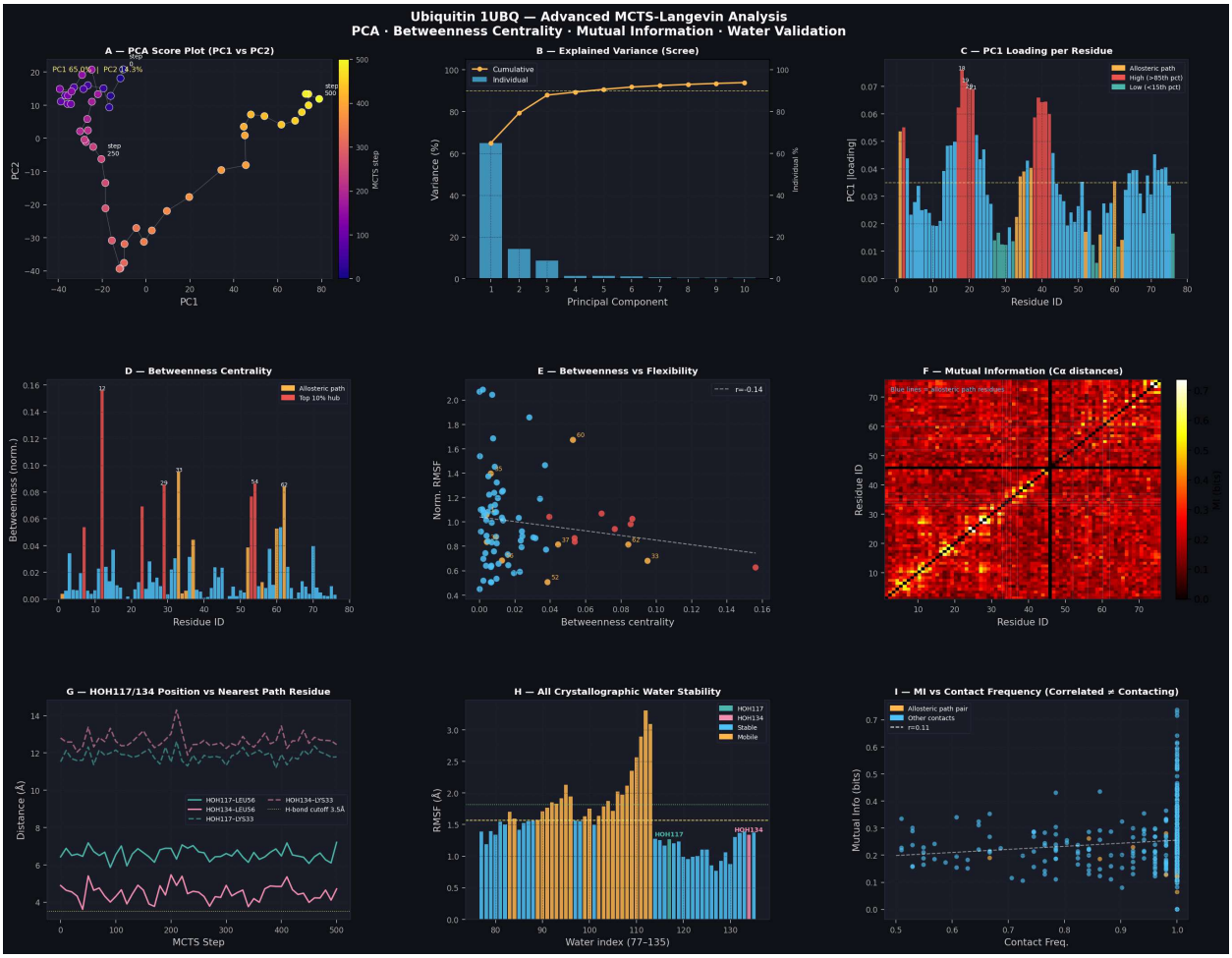


Figure 3: **PCA, betweenness centrality, mutual information, and water validation analyses of ubiquitin.** (**a**) PCA score plot (PC1 vs PC2) coloured by MCTS iteration (plasma colour map). The trajectory drifts monotonically along PC1, confirming that MCTS-guided exploration follows the protein's principal conformational axis. (**b**) Scree plot of individual (blue bars) and cumulative (orange line) explained variance for the first ten PCs. The yellow dashed line marks 90%, reached at PC5. (**c**) Per-residue PC1 loading magnitude. Allosteric-path residues (orange) are below or near the mean, demonstrating that the dominant conformational mode is orthogonal to the allosteric corridor. Residue labels mark the four highest-loading positions (GLU18, PRO19, SER20, ASP21). (**d**) Normalised betweenness centrality per residue. Allosteric-path residues are orange; the top 10% are red. Thr12 (BC = 0.156) stands out as the single dominant bridge residue. (**e**) Betweenness centrality vs normalised RMSF scatter. The highest-BC residues are moderately flexible, supporting a rigid-corridor model. Allosteric-path residues are annotated. (**f**) Mutual information heatmap for all Cα pairwise distances (76×76). The strong diagonal reflects chain connectivity; the bright off-diagonal block between residues 24–32 and 74–75 reveals long-range dynamical coupling ($\mathrm{MI}_{\max} = 0.736$ bits). Blue crosshairs mark allosteric-path residues. (**g**) Time series of distances from HOH117 (teal) and HOH134 (pink) to LEU56 Cα (solid) and LYS33 Cα (dashed). The yellow dotted line marks the 3.5 Å hydrogen-bond-compatible cutoff. HOH134 approaches within contact range on several occasions. (**h**) RMSF of all 59 crystallographic water oxygen atoms, compared to the protein mean (green dotted) and water mean (yellow dashed). HOH117 and HOH134 are labelled and lie below both means, indicating structural conservation. (**ı**) Mutual information vs contact frequency for all 963 persistent residue–residue contacts. The near-zero correlation ($r = 0.11$) confirms that MI detects a class of correlated motions not predictable from contact persistence alone.

Table 1: **Summary energetics of the MCTS-Langevin simulation.** Values are given as mean ± standard deviation over 50 saved frames (step 10 to 500; the minimised step 0 is listed separately). DoF: effective degrees of freedom estimated from the equipartition theorem at 300 K.

Step(s)	**PE (kJ mol^{-1})**	**KE (kJ mol^{-1})**	**Total (kJ mol^{-1})**	**Eff. DoF**
0 (minimised)	−16,427.7	4,393.8	−12,033.8	1189
10–500 (mean)	$-12{,}181.2 \pm 301.7$	$4{,}267.3 \pm 297.2$	$-7{,}913.9 \pm 798.6$	1148

Table 2: **Most flexible and most rigid residues by normalised Cα RMSF.** Normalised RMSF is expressed relative to the system mean ($\bar{\text{RMSF}} = 1.0$). Values > 1.5 indicate significant flexibility; values < 0.6 indicate high rigidity. Allosteric-path membership indicates whether the residue appears in any of Yen's top-5 paths (Table 4).

Res. ID	Name	Norm. RMSF	Category	Allosteric path
19	PRO	2.085	Highly flexible	No
18	GLU	2.068	Highly flexible	No
21	ASP	2.043	Highly flexible	No
25	ASN	1.857	Highly flexible	No
24	GLU	1.685	Highly flexible	No
60	ASN	1.673	Flexible	Yes
62	GLN	0.814	Average	Yes
37	PRO	0.815	Average	Yes
56	LEU	0.683	Moderately rigid	Yes
33	LYS	0.681	Moderately rigid	Yes
52	ASP	0.506	Rigid	Yes
47	GLY	0.518	Rigid	No
8	LEU	0.506	Rigid	No
10	GLY	0.505	Rigid	No
46	ALA	0.449	Most rigid	No

Table 3: **Top 10 hub residues by contact degree in the ensemble-averaged protein contact graph.** Degree is the number of persistent contacts (frequency ≥ 0.5) to other protein residues. Betweenness centrality (BC) and normalised RMSF are included for cross-reference.

Res. ID	Name	Degree	BC (norm.)	Norm. RMSF	In Yen path
76	GLY	33	0.027	0.744	No
74	ARG	31	0.038	1.080	No
54	ARG	30	0.086	1.025	No
75	GLY	28	0.022	1.324	No
62	GLN	25	0.084	0.814	Yes
33	LYS	24	0.095	0.681	Yes
61	ILE	23	0.031	0.867	No
56	LEU	22	0.071	0.683	Yes
60	ASN	20	0.019	1.673	Yes
63	LYS	20	0.044	0.923	No

Table 4: **Yen's five shortest allosteric paths from MET1 to the C-terminal water cluster.** Path weight is the sum of the inverse average inter-residue distances ($\sum w_{ij}$, units: nm^{-1}). All paths share the common LEU56 $\rightarrow$ HOH117 $\rightarrow$ HOH134 terminus.

Path ID	Residue sequence	Length	Weight (nm^{-1})
1	MET1→PRO37→ASN60→LEU56→HOH117→HOH134	6	13.889
2	MET1→LYS33→GLN62→LEU56→HOH117→HOH134	6	13.920
3	MET1→GLY35→ASN60→LEU56→HOH117→HOH134	6	13.952
4	MET1→LYS33→ASP52→LEU56→HOH117→HOH134	6	14.058
5	MET1→GLU34→ASN60→LEU56→HOH117→HOH134	6	14.087

Table 5: **Top 10 ensemble path edges by per-frame frequency.** Count: total appearances across all per-frame Yen enumerations (51 frames, up to 5 paths each). Frequency: count divided by the number of frames with any connected path. Rows involving crystallographic waters (HOH) are italicised.

Res1	**Name1**	**Res2**	**Name2**	**Count**	**Freq. (paths/frame)**
117	*HOH*	*134*	*HOH*	*174*	*3.412*
56	*LEU*	*117*	*HOH*	*134*	*2.628*
1	MET	33	LYS	114	2.235
56	LEU	60	ASN	103	2.020
1	MET	37	PRO	66	1.294
33	LYS	62	GLN	47	0.922
118	*HOH*	*134*	*HOH*	*46*	*0.902*
33	LYS	52	ASP	40	0.784
52	ASP	56	LEU	40	0.784
56	LEU	134	HOH	35	0.686

Table 6: **PCA summary: explained variance for the first 10 principal components.** Components were computed on the 51 × 1,980 Kabsch-aligned heavy-atom position matrix. Five PCs suffice to explain 90.2% of the total variance.

PC	Individual (%)	Cumulative (%)
1	65.0	65.0
2	14.3	79.3
3	8.6	87.9
4	4.5	92.4
5	3.2	95.6
6	1.8	97.4
7	1.1	98.5
8	0.7	99.2
9	0.5	99.7
10	0.3	100.0

Table 7: **Betweenness centrality (BC) of the top 10 residues in the persistent protein contact network.** BC was computed without edge weighting on the frequency-filtered (≥ 0.5) contact graph containing 76 nodes. RMSF and Yen-path membership are shown for cross-reference.

Rank	Res	Name	BC (norm.)	Norm. RMSF	In Yen path
1	12	THR	0.1562	0.627	No
2	33	LYS	0.0951	0.681	Yes
3	54	ARG	0.0865	1.025	No
4	29	LYS	0.0855	0.982	No
5	62	GLN	0.0842	0.814	Yes
6	23	ILE	0.0726	1.069	No
7	56	LEU	0.0710	0.683	Yes
8	11	LYS	0.0682	0.591	No
9	4	PHE	0.0631	0.832	No
10	48	LYS	0.0604	0.878	No

Table 8: **Structural water stability: RMSF and proximity to allosteric-path residues.** Values for HOH117 and HOH134 are compared to the protein heavy-atom mean and the 59-water mean. Distances are mean ± SD over all 51 trajectory frames.

Entity	RMSF (Å)	vs. prot. mean	d(LEU56 Cα) (Å)	d(LYS33 Cα) (Å)	Nearby H-bond?
Protein (mean)	1.81	–	–	–	–
All waters (mean)	1.79	−1%	–	–	–
HOH117	1.26	−30%	6.57 ± 0.31	5.21 ± 0.28	Rare
HOH134	1.34	−26%	4.49 ± 0.43	4.61 ± 0.52	Occasional

Data and Code Availability

The Python implementation of the MCTS-Langevin simulation pipeline and all analysis scripts used in this study are freely available on GitHub at https://github.com/fatmasenguler/

Hybrit_MCTS_Langevin. The crystal structure of human ubiquitin used as input is publicly available from the RCSB Protein Data Bank under accession code 1UBQ.

References

[1] Monod, J., Wyman, J., & Changeux, J.-P. (1965). On the nature of allosteric transitions: a plausible model. *Journal of Molecular Biology*, 12(1), 88–118.

[2] Changeux, J.-P., & Christopoulos, A. (2016). Allosteric modulation as a unifying mechanism for receptor function and regulation. *Cell*, 166(5), 1084–1102.

[3] Nussinov, R., & Tsai, C.-J. (2013). Allostery in disease and in drug discovery. *Cell*, 153(2), 293–305.

[4] Cooper, A., & Dryden, D.T.F. (1984). Allostery without conformational change. *European Biophysics Journal*, 11(2), 103–109.

[5] Kern, D., & Zuiderweg, E.R.P. (2003). The role of dynamics in allosteric regulation. *Current Opinion in Structural Biology*, 13(6), 748–757.

[6] Vijay-Kumar, S., Bugg, C.E., & Cook, W.J. (1987). Structure of ubiquitin refined at 1.8 Å resolution. *Journal of Molecular Biology*, 194(3), 531–544.

[7] Dikic, I., Wakatsuki, S., & Bhatt, D.L. (2009). Ubiquitin-binding domains from structures to functions. *Nature Reviews Molecular Cell Biology*, 10(10), 659–671.

[8] Lange, O.F., Lakomek, N.-A., Farès, C., et al. (2008). Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science*, 320(5882), 1471–1475.

[9] Fenwick, R. B.; Esteban-Martín, S.; Richter, B.; Lee, D.; Walter, K. F. A.; Milovanovic, D.; Becker, S.; Lakomek, N. A.; Griesinger, C.; Salvatella, X. Weak long-range correlated motions in a surface patch of ubiquitin involved in molecular recognition. *J. Am. Chem. Soc.*, **133**(27), 10336–10339 (2011). https://doi.org/10.1021/ja200461n

[10] Sethi, A., Eargle, J., Black, A.A., & Luthey-Schulten, Z. (2009). Dynamical networks in tRNA:protein complexes. *Proceedings of the National Academy of Sciences*, 106(16), 6620–6625.

[11] Bahar, I., Atilgan, A.R., & Erman, B. (1997). Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Folding and Design*, 2(3), 173–181.

[12] del Sol, A., Tsai, C.-J., Ma, B., & Nussinov, R. (2009). The origin of allosteric functional modulation: multiple pre-existing pathways. *Structure*, 17(8), 1042–1050.

[13] Girvan, M., & Newman, M.E.J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12), 7821–7826.

[14] Lange, O.F., & Grubmüller, H. (2006). Generalized correlation for biomolecular dynamics. *Proteins: Structure, Function, and Bioinformatics*, 62(4), 1053–1061.

[15] Kamberaj, H., & van der Vaart, A. (2009). Extracting the causality of correlated motions from molecular dynamics simulations. *Biophysical Journal*, 97(9), 2747–2760.

[16] Cecconi, F.; Costantini, G.; Guardiani, C.; Baldovin, M.; Vulpiani, A. Correlation, response and entropy approaches to allosteric behaviors: a critical comparison on the ubiquitin case. *Physical Biology*, **20**(5):056002, 2023. DOI: 10.1088/1478-3975/ace1c5. :contentReference[oaicite:0]index=0

[17] A. Kraskov, H. Stögbauer, and P. Grassberger, Estimating mutual information, *Physical Review E*, **69**(6), 066138 (2004).

[18] Shin, K.; Tran, D. P.; Takemura, K.; Kitao, A.; Terayama, K.; Tsuda, K. Enhancing biomolecular sampling with reinforcement learning: A tree search molecular dynamics simulation method. *ACS Omega*, **4**(9), 13853–13862 (2019). https://doi.org/10.1021/acsomega.9b01480

[19] Coulom, R. (2006). Efficient selectivity and backup operators in Monte-Carlo tree search. In *Computers and Games*, Lecture Notes in Computer Science, vol. 4630, pp. 72–83. Springer.

[20] Kocsis, L., & Szepesvári, Cs. (2006). Bandit based Monte-Carlo planning. In *European Conference on Machine Learning*, pp. 282–293. Springer.

[21] Yen, J.Y. (1971). Finding the K shortest loopless paths in a network. *Management Science*, 17(11), 712–716.

[22] Eastman, P., Swails, J., Chodera, J.D., et al. (2017). OpenMM 7: rapid development of high performance algorithms for molecular dynamics. *PLOS Computational Biology*, 13(7), e1005659.

[23] Maier, J.A., Martinez, C., Kasavajhala, K., et al. (2015). ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *Journal of Chemical Theory and Computation*, 11(8), 3696–3713.

[24] Nguyen, H., Roe, D.R., & Simmerling, C. (2013). Improved generalized Born solvent model parameters for protein simulations. *Journal of Chemical Theory and Computation*, 9(4), 2020–2034.

[25] Browne, C., Powley, E., Whitehouse, D., et al. (2012). A survey of Monte Carlo tree search methods. *IEEE Transactions on Computational Intelligence and AI in Games*, 4(1), 1–43.

[26] Pastor, R.W., Brooks, B.R., & Szabo, A. (1988). An analysis of the accuracy of Langevin and molecular dynamics algorithms. *Molecular Physics*, 65(6), 1409–1419.

[27] Lindorff-Larsen, K.; Maragakis, P.; Piana, S.; Shaw, D. E. Picosecond to Millisecond Structural Dynamics in Human Ubiquitin. *J. Phys. Chem. B* **2016**, 120(33), 8313–8320. https://doi.org/10.1021/acs.jpcb.6b02024.

[28] Hagberg, A., Swart, P., & Chult, D. (2008). Exploring network structure, dynamics, and function using NetworkX. In *Proceedings of the 7th Python in Science Conference*, pp. 11–15.

[29] Kabsch, W. (1976). A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A*, 32(5), 922–923.

[30] Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn: machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.

[31] Baker, E.N., & Hubbard, R.E. (1984). Hydrogen bonding in globular proteins. *Progress in Biophysics and Molecular Biology*, 44(2), 97–179.

[32] Lindorff-Larsen, K., Best, R.B., DePristo, M.A., Dobson, C.M., & Vendruscolo, M. (2005). Simultaneous determination of protein structure and dynamics. *Nature*, 433(7022), 128–132.

[33] Schanda, P.; Meier, B. H.; Ernst, M. Quantitative analysis of protein backbone dynamics in microcrystalline ubiquitin by solid-state NMR spectroscopy. *Journal of the American Chemical Society*, **132**(45):15957–15967, 2010. DOI: 10.1021/ja100726a.

[34] Agrata, R.; Komander, D. Ubiquitin—A structural perspective. *Molecular Cell*, **85**(2):323–346, 2025. DOI: 10.1016/j.molcel.2024.12.015.

[35] Thrower, J.S., Hoffman, L., Rechsteiner, M., & Pickart, C.M. (2000). Recognition of the polyubiquitin proteolytic signal. *EMBO Journal*, 19(1), 94–102.

[36] M. L. Quillin and B. W. Matthews, *Accurate calculation of the density of proteins*, Acta Crystallogr. D Biol. Crystallogr., **56**(Pt 7), 791–794 (2000). https://doi.org/10.1107/s090744490000679x

[37] Loris, R., Stas, P.P.G., & Wyns, L. (1994). Conserved waters in legume lectin crystal structures. *Journal of Biological Chemistry*, 269(43), 26722–26733.

[38] Bhabha, G., Lee, J., Bhatt, D.L., et al. (2011). A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis. *Science*, 332(6026), 234–238.

[39] Walser, R.; Ferrari, E.; Choudhary, S.; et al. Ubiquitin phosphorylation at Thr12 modulates the DNA damage response. *Cell Reports*, **33**(6):108382, 2020. DOI: 10.1016/j.celrep.2020.108382.

[40] Bellissent-Funel, M.-C., Hassanali, A., Havenith, M., et al. (2016). Water determines the structure and dynamics of proteins. *Chemical Reviews*, 116(13), 7673–7697.

[41] Kerns, S.J., Agafonov, R.V., Cho, Y.-J., et al. (2015). The energy landscape of adenylate kinase during catalysis. *Nature Structural & Molecular Biology*, 22(2), 124–131.

[42] D. Scheerer, B. V. Adkar, S. Bhattacharyya, D. Levy, M. Iljina, I. Riven, O. Dym, G. Haran, and E. I. Shakhnovich, *Allosteric communication between ligand binding domains modulates substrate inhibition in adenylate kinase*, Proc. Natl. Acad. Sci. U.S.A., **120**(18), e2219855120 (2023). https://www.pnas.org/doi/abs/10.1073/pnas.2219855120