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.

Dynamic Network Context Predicts Mutation-Induced Reshaping of Protein Conformational Free-Energy Landscapes

Laura Orellana

laura.orellana@ki.se

Karolinska Institute https://orcid.org/0000-0003-1927-555X

Byung Ho Lee

Karolinska Institute https://orcid.org/0000-0001-6515-7656

Domenico Scaramozzino

Karolinska Institute https://orcid.org/0000-0002-6235-8070

Serena Piticchio

Karolinska Institute

Article

Keywords:

Posted Date: November 5th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11467619/v1

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

Additional Declarations: There is **NO** Competing Interest.

Dynamic Network Context Predicts Mutation-Induced Reshaping of Protein Conformational Free-Energy Landscapes

Byung Ho Lee[1], Domenico Scaramozzino[1], Serena Piticchio[1], Laura Orellana[1,*]

[1] Department of Oncology-Pathology, Karolinska Institutet, Stockholm 171 77, Sweden.

[*] Corresponding author. Email: laura.orellana@ki.se

Abstract

Missense substitutions can reshape protein conformational free-energy landscapes, but predicting such effects at scale remains challenging. We develop a mutation-impact framework based on the internal-coordinate, essential-dynamics-refined elastic network model (ICed-ENM). Changes in vibrational entropy quantify substitution-level dynamic perturbation; aggregation across substitutions and neighboring positions yields a position-level dynamic sensitivity score. Across three conformationally diverse proteins, alanine substitutions at high-sensitivity sites change molecular-dynamics-derived conformational landscapes more than low-sensitivity controls selected in broadly comparable structural contexts. Analysis of more than 560,000 modeled substitutions reveals systematic relationships with side-chain volume, amino-acid identity, solvent accessibility, and secondary structure. In a PTEN functional benchmark, ICed-ENM position-level scores outperform those of two comparator methods in distinguishing activity-selective loss (reduced phosphatase activity with preserved intracellular abundance) from

abundance-related loss, achieving an area under the receiver operating characteristic curve of 0.745. Here, we show that structure-derived dynamic sensitivity provides a functionally informative basis for prioritizing mutations.

Introduction

Missense mutations that alter individual amino acids can profoundly affect protein function and contribute to diseases ranging from cancer[1,2] to neurodegenerative disorders such as Alzheimer's and Parkinson's disease[3,4]. Such mutations can perturb active sites, alter molecular interactions, compromise structural stability, or shift the balance among functional conformational states[5]. However, determining which of these mechanisms underlies the effect of a given mutation often remains challenging. Dynamic mechanisms can be particularly difficult to resolve when substitutions occur far from functional sites, where their effects may propagate through long-range structural and dynamic coupling[6,7]. Characterizing this conformational-dynamic component of mutation impact may therefore provide mechanistic insight that is not readily apparent from static structural changes alone.

Molecular dynamics (MD) simulations provide a powerful approach for characterizing protein conformational dynamics and mutation-induced changes in conformational landscapes[8]. Advances in computational power[9,10] and conformational sampling strategies[11–14] have greatly improved the ability of MD to capture large-scale transitions and rare conformational states. Nevertheless, intrinsic limitations remain, including strong dependence on force-field accuracy and substantial computational cost[15]. These limitations hinder routine and large-scale characterization of mutation-induced changes in conformational dynamics, particularly when such effects involve rare states, high-energy barriers, or long-range allosteric coupling.

Coarse-grained models offer a computationally efficient alternative for studying protein dynamics over larger structural and mutational scales. Among these approaches, elastic network models (ENMs), which represent proteins as networks of nodes connected by harmonic springs,

have been widely used to describe collective low-frequency motions[16–18] and conformational transitions[19–21]. Mutation-oriented extensions such as ENCoM introduced atom-type contact interactions to estimate amino-acid-specific effects on protein dynamics and stability[22]. DynaMut2 combines normal-mode-derived dynamics descriptors with graph-based representations of the structural environment to predict mutation-induced changes in protein stability[23], while Rhapsody-2 demonstrated that intrinsic dynamic properties, together with evolutionary, structural, energetic, and physicochemical features, can contribute to variant pathogenicity prediction[24]. These studies establish that protein dynamics provides valuable information about mutation effects. However, it remains less clear whether mutation-specific dynamic perturbations predicted by coarse-grained models correspond to actual reshaping of conformational landscapes, how this dynamic component relates to other molecular features associated with mutation effects, and whether it is associated with experimentally observed functional consequences.

Here, we developed a mutation-impact analysis framework based on an internal-coordinate, essential-dynamics-refined elastic network model (ICed-ENM) to quantify the structure-derived conformational-dynamic component of amino acid substitution effects (Fig. 1). The framework combines ICed-ENM with systematic in silico substitution and vibrational-entropy-based analysis. By formulating protein motions in internal coordinates, ICed-ENM preserves covalent geometry while capturing collective dynamic responses encoded by protein structure. For each substitution, the absolute change in vibrational entropy ($|\Delta S|$) provides a substitution-specific measure of perturbation in intrinsic conformational dynamics, while aggregation across substitutions and spatial integration of neighboring effects yield a position-level mutation-impact score (Ω) that reflects dynamic sensitivity within the surrounding structural network. We first

benchmark ICed-ENM against iMOD[25], edENM[17], and ANM[18] to evaluate agreement with experimental conformational transitions and preservation of stereochemical geometry. We then test whether substitutions at predicted high-sensitivity sites produce greater changes in MD-derived conformational landscapes than low-sensitivity controls selected in broadly comparable structural and physicochemical contexts. We also examine how the resulting mutation-impact measures relate to structural and physicochemical features across a curated protein dataset. Finally, we assess the functional relevance of ICed-ENM-based mutation-impact analysis using an integrated PTEN abundance–activity dataset[26]. This benchmark distinguishes activity-selective loss, defined by reduced phosphatase activity with WT-like intracellular abundance, from abundance-related loss, defined by reduced abundance regardless of activity. We compare ICed-ENM-based scores with ENCoM |ΔS| and DynaMut2 stability predictions at the substitution level, and with ENCoM-derived Ω and dynamics-only Rhapsody-2(dyn) scores at the position level. ICed-ENM Ω preferentially identifies activity-selective loss positions and provides greater phenotype separation across the full ranking than either position-level comparator, connecting predicted dynamic sensitivity to experimentally measured functional impairment despite preserved intracellular abundance.

Results

Physical realism of ICed-ENM collective motions

We first evaluated whether ICed-ENM captures experimentally relevant collective protein motions while preserving stereochemical realism. To this end, we compared ICed-ENM with three representative ENM approaches: iMOD[25], edENM[17], and ANM[18], using a benchmark

dataset of 41 protein pairs comprising 82 open- and closed-state structures. Predicted low-frequency modes were compared with experimental transition vectors defined between paired endpoint structures, and the resulting overlap metrics are summarized in Table 1.

Across all methods, the first normal mode showed substantially greater average overlap with the experimental transition vector (O_1) than the other modes, indicating that transition-relevant motion was preferentially captured by the dominant low-frequency mode. Among the tested methods, ICed-ENM showed the highest O_1 values across all structural groups: 0.54 for all structures, 0.64 for open-state structures, and 0.44 for closed-state structures. In comparison, ANM yielded the lowest scores of 0.44, 0.55, and 0.33, respectively. Together with its higher O_1, the lower O_2 of ICed-ENM suggests that a greater fraction of transition-relevant motion is concentrated in the leading mode.

We next examined whether transition-relevant motions were captured at low mode indices. ICed-ENM achieved the highest average maximum overlap (O_{max}) and the lowest corresponding mode indices ($N(O_{max})$), indicating that experimentally relevant motions appeared earlier in its mode spectrum than in the other models. ICed-ENM also showed higher cumulative squared overlap (CSO) across most mode subsets and structural groups, indicating that a small number of ICed-ENM low-frequency modes more effectively spanned the experimental transition vectors.

We then asked whether this improved agreement with experimental conformational transitions was achieved without compromising stereochemical integrity. Analysis of mode-deformed structures showed that the internal-coordinate-based models, ICed-ENM and iMOD, produced consecutive Cα–Cα distance distributions tightly concentrated around the expected backbone spacing of approximately 3.8 Å (Fig. 2A). In contrast, the Cartesian-coordinate-based models,

edENM and ANM, showed broader distributions, indicating larger local backbone distortions during mode deformation.

Consistent with this backbone-distance analysis, the internal-coordinate-based models also exhibited smaller changes in pairwise Cα–Cα contact networks than the Cartesian-coordinate-based models (Fig. 2B). Among the tested models, ICed-ENM showed the strongest overall preservation of both backbone geometry and contact-network integrity.

Conformational landscapes at predicted hot and cold spots

Having established the physical realism of ICed-ENM collective motions, we next asked whether ICed-ENM-based mutation-impact analysis could identify residues whose substitutions reshape conformational dynamics. In each protein, position-level mutation-impact scores (Ω) were used to identify candidate mutation-sensitive hot spots. Selected high-Ω residues were paired with low-Ω cold-spot controls chosen to provide broadly comparable local structural and physicochemical contexts, considering solvent-accessible surface area (SASA), secondary structure, and amino-acid properties. Alanine substitutions were then introduced at the selected positions, and wild-type (WT) and mutant systems were analyzed using all-atom MD simulations. The resulting trajectories were projected onto subspaces defined by the first two principal components (PC1 and PC2) derived from experimental structural ensembles to construct projected conformational free-energy landscapes and evaluate mutation-induced redistribution of conformational sampling.

For this validation, we selected three well-characterized conformational-switching proteins: ribose-binding protein (RBP), 5′-nucleotidase (5′NTase), and sarco/endoplasmic reticulum Ca^{2+}-ATPase 1 (SERCA). Their experimentally determined structural ensembles span distinct

functional conformations and therefore provide reference coordinates for evaluating mutation-induced changes in conformational sampling[27]. In each system, PC1 and PC2 derived from the experimental ensembles captured more than 70% of the structural variance and separated the major functional states, enabling MD trajectories to be evaluated within experimentally defined rather than simulation-derived conformational coordinates.

RBP, a member of the periplasmic binding protein family, undergoes a hinge-like opening-and-closing motion between its N- and C-terminal domains associated with ribose binding and transport[28]. Using chain A of the apo open structure 1URP (1URP:A) as the WT reference, the position-level Ω profile highlighted the canonical hinge region and interdomain interface as mutation-sensitive regions (Fig. 3A). Q235 and E255 were selected as high-Ω hot spots, with Q184 and E122, respectively, selected as low-Ω controls with broadly comparable solvent exposure and local structural contexts (Supplementary Table 1). Notably, Q235 lies within the canonical hinge region, whereas E255 is located in a solvent-exposed hinge-proximal loop. In the experimentally derived PC1–PC2 subspace, the open and closed states were primarily separated along PC1, which captured the dominant interdomain opening-and-closing motion. The WT trajectory exhibited a dominant basin near the open reference conformation, and the low-Ω mutants Q184A and E122A retained broadly similar landscapes (Fig. 3A). In contrast, the high-Ω mutants Q235A and E255A showed more pronounced redistribution of conformational sampling: Q235A populated an additional basin shifted along both PC1 and PC2, whereas E255A exhibited broader sampling across the PC subspace. These differences were also apparent in the corresponding PC1 and PC2 frequency distributions (Fig. 4A).

5′NTase, a periplasmic enzyme involved in nucleotide metabolism, consists of N- and C-terminal domains connected by an α-helical linker. Its catalytic cycle involves large-scale interdomain

rearrangements, in which the C-terminal domain rotates relative to the N-terminal domain, giving rise to open, intermediate, and closed conformational states[29]. Using 1OID:A as the apo open WT reference, the Ω profile showed elevated dynamic sensitivity at the interdomain interface and in several solvent-exposed regions (Fig. 3B). E182 and R396 were selected as high-Ω hot spots, with K134 and R102, respectively, selected as low-Ω controls with comparable solvent exposure and local structural contexts (Supplementary Table 1). E182 is positioned at an interdomain-facing surface of the N-terminal domain, whereas R396 lies in an interdomain-facing α-helical segment of the C-terminal domain. In the experimental PC1–PC2 subspace, PC1 primarily separated the open and closed states, whereas PC2 distinguished the intermediate state. The WT trajectory sampled a primary basin displaced toward the closed-state region together with a nearby secondary basin (Fig. 3B). The low-Ω mutants K134A and R102A largely retained this landscape, although R102A showed additional sampling at positive PC2. In contrast, the high-Ω mutants E182A and R396A redistributed conformational populations over broader regions of the PC subspace, a distinction also evident in the corresponding PC1 and PC2 frequency distributions (Fig. 4B). Notably, these redistributions were not aligned with the canonical open-to-closed direction.

SERCA, a P-type ATPase involved in Ca^{2+} transport across the sarcoplasmic/endoplasmic reticulum membrane, consists of a transmembrane domain (TM) and three cytoplasmic domains: the actuator (A), phosphorylation (P), and nucleotide-binding (N) domains[30]. Its transport cycle involves large-scale rearrangements of the cytoplasmic and TM domains, giving rise to representative conformational states, including the E1 open, E1 closed, and E2 closed states. The E1 open-to-E1 closed transition involves closure of the cytoplasmic headpiece, whereas the E2 closed state is associated with larger A-domain rotation and TM-domain rearrangement. Using

the Ca^{2+}-bound E1 open structure 2C9M:A as the reference, the Ω profile highlighted several interdomain regions, particularly the TM–A and N–A interfaces (Fig. 3C). K47 and N428 were selected as high-Ω hot spots, with K169 and N453, respectively, selected as low-Ω controls with broadly comparable solvent exposure and local structural contexts (Supplementary Table 1). K47 is located in the TM–A linker, whereas N428 lies on an A-domain-facing surface of the N domain. In the experimental PC1–PC2 subspace, PC2 primarily described E1 open-to-E1 closed headpiece closure, whereas PC1 separated E2 closed conformations from E1-like states through rotational rearrangements of the cytoplasmic domains. The WT trajectory sampled two dominant basins near the E1 open reference, with diffuse sampling extending toward the E1 closed region (Fig. 3C). Although all SERCA mutants exhibited relatively broad conformational sampling, the high-Ω mutants K47A and N428A showed dominant basins shifted toward the E1 closed region relative to WT and their low-Ω controls. This redistribution was most evident along PC2 and was also reflected in the corresponding frequency distributions (Fig. 4C). Across the three validation systems, alanine substitutions at selected high-Ω residues produced stronger redistribution of conformational sampling than substitutions at low-Ω controls selected in broadly comparable structural environments.

Structural and physicochemical context of dynamic sensitivity

To examine how the mutation-impact scores from the ICed-ENM-based analysis relate to structural and physicochemical features, we applied the framework across our curated PDB dataset, excluding ACLY structures containing missing regions. This generated a large-scale mutation-impact dataset containing substitution-level |ΔS| scores and position-level Ω scores, enabling global statistical analyses across diverse structural and dynamic contexts. For analyses

based directly on substitution-level |ΔS| scores, substitutions at low-connectivity residues were excluded to reduce potential artifacts from poorly connected regions. In contrast, position-level analyses were performed across all residues, because Ω already incorporates the local-connectivity criterion during smoothing. First, we evaluated whether mutation-impact scores were associated with structural deviations of mutant models relative to WT, as quantified by RMSD from the corresponding WT structures (Fig. 5A). To this end, mutations at valid residues were grouped into the top 10% by |ΔS| (56,045 samples) and the remaining 90% (504,094 samples). Despite pronounced differences in |ΔS|, the RMSD distributions of the two groups were highly overlapping. Moreover, the inset of the raw RMSD distribution shows that absolute structural deviations remain exceedingly small, with a median RMSD of approximately 0.006 Å. These results indicate that high-|ΔS| scores are not trivially explained by the magnitude of modeled structural deviations.

Next, we investigated how side-chain volume changes induced by single-point mutations affected |ΔS| scores (Fig. 5B). To facilitate interpretation, amino acids were grouped into five classes according to side-chain volume[31], and WT-to-mutant class transitions were systematically evaluated. The heatmap reports the probability of substitutions belonging to the high-|ΔS| group, calculated as the fraction (%) of substitutions within each WT-to-mutant class that fell within the top 10% of |ΔS| values; therefore, 10.0% corresponds to the background expectation. A prominent pattern emerged: large-magnitude volume changes were generally associated with elevated probabilities of high-|ΔS| responses. Transitions from smaller to larger volume classes generally yielded higher high-|ΔS| fractions than the corresponding reverse transitions, particularly for substitutions into the very large class. In particular, mutations from the very large class to smaller volume classes, except those to the very small class, showed high-

|ΔS| fractions close to or below the background level. This analysis revealed systematic biases in high-|ΔS| responses associated with both the magnitude and direction of side-chain volume changes. Beyond side-chain volume, additional analyses based on broad physicochemical class and individual amino-acid identities revealed further substitution-specific patterns (Supplementary Fig. 1). High-|ΔS| enrichment varied with substitution direction across amino-acid classes (Supplementary Fig. 1A). Same-class transitions ranged from below the 10% baseline for polar-to-polar and negative-to-negative substitutions to above it for positive-to-positive substitutions; hydrophobic-to-hydrophobic substitutions were near baseline. Positive-to-hydrophobic substitutions showed the strongest enrichment. At the individual amino-acid level, the WT-to-mutant matrix showed broadly elevated high-|ΔS| fractions for transitions to or from glycine and arginine; proline-involving transitions were more heterogeneous but included several above-baseline cases (Supplementary Fig. 1B). For all three residues, some of these cases occurred despite modest side-chain-volume changes. When aggregated by WT amino-acid identity, these three residues showed overall enrichment in the high-|ΔS| group (Supplementary Fig. 1C).

We next analyzed position-level Ω scores. Fig. 5C shows the distribution of position-level SASA values computed with FreeSASA[32] for the high-Ω group, defined as the top 10% of Ω-ranked residues within each structure (3,118 samples), and the remaining group (27,714 samples). The remaining group was concentrated at low normalized SASA values, whereas the high-Ω group had a broader distribution shifted toward higher normalized SASA. High-Ω residues therefore showed a relative tendency toward greater solvent exposure while spanning a broad range of solvent accessibilities. A similar trend was observed for per-residue contact counts, with the

high-Ω group showing a shift toward lower values than the remaining group (Supplementary Fig. 2).

To further characterize the structural context of high-Ω residues, we examined their secondary-structure distribution using a three-state secondary-structure classification (SS3), in which DSSP-assigned secondary structures were grouped into helix (α-helical regions), strand (β-strand segments), and coil states (loops or irregular regions) (Fig. 5D)[33]. Compared with the remaining group, the top 10% Ω group contained a higher proportion of coil residues (58.1% vs 37.6%) and lower proportions of helix (29.8% vs 39.5%) and strand residues (12.1% vs 22.9%). Although coils predominated, helices and strands together accounted for 41.9% of high-Ω residues, showing that high dynamic sensitivity was represented in all three secondary-structure classes. Together, these analyses show that substitution-level dynamic perturbation is associated with the physicochemical nature of amino-acid changes and that position-level dynamic sensitivity varies with the local structural environment.

Functional benchmarking of dynamic sensitivity in PTEN

To assess the functional relevance of ICed-ENM-based mutation-impact analysis in the tumor suppressor PTEN[34], we compared its substitution-level |ΔS| and position-level Ω scores with outputs from methods capturing related but distinct aspects of mutation impact. ENCoM[22] provides a comparison based on changes in vibrational entropy, using a Cα Cartesian-coordinate model with atom-type-dependent contacts, whereas ICed-ENM uses internal coordinates without explicit atom-type-dependent contact weights. DynaMut2[23] provides a stability-oriented comparison, combining structural-environment and normal-mode-derived features in a machine-learning model that predicts changes in folding stability (ΔΔG). Rhapsody-2(dyn)[24] is a machine-

learning pathogenicity predictor using only descriptors of intrinsic dynamics derived from the WT structure.

Using the integrated PTEN abundance–activity dataset[26], we compared activity-selective loss (reduced phosphatase activity with WT-like abundance) with abundance-related loss (reduced abundance regardless of activity). Mapping the two position-level phenotype classes onto the apo open PTEN structure (1D5R:A) showed that residues associated with activity-selective loss were distributed across the phosphatase domain and several exposed loop regions of the C2 domain, whereas residues associated with abundance-related loss were distributed more broadly throughout the structure (Fig. 6A). This benchmark therefore evaluates association with loss phenotypes rather than pathogenicity. Using 1D5R:A as the common structural background, top 10% score groups were defined across 5,643 substitutions or 297 positions; phenotype comparisons used 956 substitutions or 193 positions assigned exclusively to one phenotype (Methods; Table 2).

At the substitution level, the top 10% group ranked by $|\Delta S|$ scores from the ICed-ENM-based analysis included 13.0% of activity-selective and 9.2% of abundance-related substitutions (odds ratio [OR] = 1.47; Table 2), compared with 26.0% and 16.5% for ENCoM (OR = 1.78). Both methods therefore showed modest preferential inclusion of activity-selective substitutions in their respective top 10% score groups. DynaMut2, ranked by predicted destabilization, showed the opposite association: its top group included 6.0% of activity-selective and 22.3% of abundance-related substitutions (OR = 0.22). To assess phenotype separation across the full score rankings, we examined the receiver operating characteristic (ROC) curves for ICed-ENM, ENCoM, and DynaMut2 (Supplementary Fig. 3). The ENCoM curve generally lay above the ICed-ENM curve, although both remained close overall to the diagonal reference line, indicating

limited phenotype separation. The DynaMut2 curve lay predominantly below the diagonal, indicating preferential ranking of abundance-related substitutions over activity-selective substitutions, consistent with the top-group analysis. The corresponding areas under the ROC curve (AUROCs) were 0.510 for ICed-ENM, 0.567 for ENCoM, and 0.375 for DynaMut2.

At the position level, several high-Ω positions identified by ICed-ENM-based mutation-impact analysis coincided with activity-selective positions, whereas abundance-related positions were more broadly distributed and frequently had lower Ω values (Fig. 6B). The top 10% of positions ranked by Ω included 30.0% of activity-selective but only 3.5% of abundance-related positions (OR = 11.93; Table 2). The corresponding fractions were 15.0% and 12.1% for ENCoM-derived Ω (OR = 1.28) and 25.0% and 8.1% for Rhapsody-2(dyn) scores (OR = 3.79). The ROC curve for ICed-ENM Ω lay above those for ENCoM-derived Ω and Rhapsody-2(dyn) scores over most of the range, indicating greater recovery of activity-selective positions at comparable false-positive rates (Fig. 6C). The corresponding AUROCs were 0.745 for ICed-ENM, 0.530 for ENCoM, and 0.490 for Rhapsody-2(dyn). Thus, ICed-ENM Ω showed the strongest observed preference for activity-selective positions both within the top 10% score group and across the full ranking, whereas the top-group association of Rhapsody-2(dyn) was not apparent across its full ranking.

Direct comparisons showed no significant substitution-level difference between ICed-ENM and ENCoM in top-group association (ratio of ORs = 0.825; nominal 95% confidence interval [CI] = 0.213 to 3.201; $P = 0.781$) or AUROC (ΔAUROC = −0.056; CI = −0.176 to 0.075; $P = 0.371$; Supplementary Table 2). At the position level, the ratios of ORs comparing ICed-ENM with ENCoM and Rhapsody-2(dyn) were 9.34 (CI = 1.34 to 65.17; Holm-adjusted $P = 0.0484$) and 3.15 (CI = 0.46 to 21.47; Holm-adjusted $P = 0.241$), respectively. The corresponding ΔAUROCs

were 0.214 (CI = 0.012 to 0.412; Holm-adjusted $P = 0.0478$) and 0.254 (CI = 0.035 to 0.462; Holm-adjusted $P = 0.0478$). Thus, in this PTEN benchmark, ICed-ENM Ω showed significantly stronger top-group association than ENCoM, although the comparison with Rhapsody-2(dyn) remained inconclusive. Its full-ranking AUROC was significantly higher than those of both comparators after Holm correction.

Discussion

Interpreting mutation-induced perturbations in protein dynamics requires a physically realistic representation of the collective motions encoded by protein structure. In this context, a key feature of ICed-ENM is that its interaction parameters were obtained through systematic parameterization against MD-derived effective stiffness patterns from simulations performed with multiple atomistic force fields. The resulting model was then benchmarked against experimentally observed conformational transitions and evaluated for stereochemical preservation during mode deformation. These analyses showed that ICed-ENM normal modes achieved improved agreement with experimental transition vectors while generating mode-deformed structures with strong preservation of both backbone geometry and contact-network integrity (Table 1 and Fig. 2). This behavior likely arises, at least in part, from the internal-coordinate formulation, which removes redundant bond-length and bond-angle degrees of freedom while retaining physically meaningful torsional motions. Although the improvement relative to iMOD was moderate, consistent with the fact that both methods share the fundamental advantages of internal-coordinate representations, this benchmark was important for establishing ICed-ENM as a physically grounded dynamic reference. This is particularly relevant for

mutation-impact analysis, where subtle mutation-induced perturbations must be distinguished from artifacts caused by unrealistic mode deformations.

Pairing high-Ω sites identified by ICed-ENM-based mutation-impact analysis with low-Ω controls selected in broadly comparable local contexts provided a test of whether the position score identifies sites whose substitutions alter MD-derived conformational sampling. These effects were evaluated within conformational coordinates derived from experimental structural ensembles, allowing the mutation-induced changes to be interpreted relative to biologically meaningful functional coordinates rather than simulation-defined reaction coordinates.

The three validation systems illustrate that high-Ω substitutions can redistribute conformational sampling in different ways (Fig. 3). However, these results should not be interpreted to mean that high-Ω substitutions generally drive conformational sampling along native functional transitions. In 5′NTase, for example, hot-spot substitutions induced pronounced redistribution of conformational populations without aligning with the canonical open-to-closed axis. Thus, Ω profiles should primarily be interpreted as identifying mutation-sensitive sites whose perturbations can redistribute conformational populations relative to WT, rather than as predicting a specific direction of functional motion. The contrasting behavior of hot- and cold-spot mutants supports the ability of Ω-guided analysis to prioritize mutation-sensitive sites, including non-obvious positions outside canonical hinge or interface regions, such as RBP E255 and SERCA K47, whose effects may propagate through distributed dynamic coupling networks rather than solely through localized structural disruption. Such behavior is consistent with the recognized role of long-range allosteric communication in mediating mutation-dependent functional regulation[6,7].

Beyond the individual validation cases, we examined global mutation-impact patterns across the curated PDB dataset and compared the observed trends with previously reported patterns from large-scale public mutation datasets (Fig. 5). These comparisons were not intended as direct pathogenicity prediction, but as a way to assess whether residues and substitutions identified as dynamically sensitive by our framework overlap with physicochemical and structural features known to distinguish disease-associated mutations from more neutral variation.

Substitution-level $|\Delta S|$ was associated with both the magnitude and direction of side-chain volume changes (Fig. 5B). The directional asymmetry indicates that volume-change magnitude alone does not account for the observed scores. Amino-acid-class patterns likewise show that chemically similar substitutions need not produce similar responses and that transitions involving charged residues can yield elevated scores (Supplementary Fig. 1A). The enrichment of charged-residue transitions may partly reflect the loss of charged interactions, including salt bridges, which can affect protein conformational dynamics and have been implicated in disease-related mutations[35,36]. Although ENM-based models do not explicitly encode specific chemical interactions or residue charges, our workflow uses MODELLER[37] to generate mutant structures from the input conformation before rebuilding the elastic network. This structure-reconstruction step may allow substitution-induced local structural adjustments to be propagated into the reconstructed network. Thus, chemically disruptive substitutions may be indirectly represented as mechanical perturbations at the network level, suggesting a potential strength of the framework in translating local physicochemical changes into detectable changes in conformational dynamics.

These physicochemical trends qualitatively resemble patterns reported in large-scale analyses of disease-associated and polymorphic missense variants, in which disease-associated substitutions

tend to involve larger changes in amino-acid physicochemical properties, including side-chain volume and charge/hydrophobicity[38]. However, the enrichment pattern observed in our analysis should not be interpreted as a direct pathogenicity signal. Rather, it suggests that substitution types frequently associated with pathogenic variation can overlap with those that produce detectable perturbations in structure-derived dynamic networks.

At the individual amino-acid level, arginine-associated enrichment was observed even for some substitutions with modest side-chain-volume changes (Supplementary Fig. 1B). This pattern is consistent with prior observations that arginine-related mutations are disproportionately associated with pathogenic variants[38,39]. Given that arginine residues frequently participate in salt bridges and extensive hydrogen-bond networks, these findings suggest that arginine substitutions may induce network-level mechanical perturbations detectable by the framework, extending beyond simple steric effects. The elevated responses observed for selected glycine- and proline-involving substitutions may reflect their distinct effects on backbone conformational preferences, beyond differences in side-chain volume or broad physicochemical class.

At the position level, we next examined how Ω-based mutation sensitivity relates to structural context and to trends reported in curated variant datasets. In terms of solvent accessibility, previous studies have shown that disease-causing mutations are often enriched in buried or partially exposed regions compared with polymorphic mutations[38,40–42]. The high-Ω group included both accessible and more buried sites, but its SASA distribution was shifted toward greater exposure than that of the remaining group (Fig. 5C), consistent with its lower contact counts (Supplementary Fig. 2). This pattern suggests that Ω identifies mutation-sensitive positions across diverse packing environments rather than being limited to densely packed structural cores.

Secondary-structure context provided a complementary pattern. Previous work on curated variant datasets reported that sites of both disease-causing and polymorphic variants occurred most frequently in coil regions, followed by helices and strands[35,38]. The high-Ω group showed the same ordering and a higher coil fraction than the remaining group (Fig. 5D). This identifies loops and irregular regions as prominent contexts of predicted dynamic sensitivity, while the presence of high-Ω residues in helices and strands shows that the signal is not confined to coils. Because the ordering was reported for both variant classes, coil enrichment alone does not establish a pathogenicity association. The SASA, contact-count, and secondary-structure analyses describe complementary group-level patterns but do not establish how these features combine at individual sites. Together with the substitution-level physicochemical trends, these position-level structural associations motivate evaluating whether predicted dynamic sensitivity relates to measured PTEN function.

The PTEN benchmark provided a direct assessment of how scores from ICed-ENM-based mutation-impact analysis relate to experimentally measured loss phenotypes. The clearest association emerged at the position level: high-Ω positions preferentially coincided with activity-selective loss, linking predicted dynamic sensitivity to impaired phosphatase activity despite WT-like intracellular abundance. Substitution-level |ΔS|, however, showed limited separation between the two phenotype groups. These findings support using Ω to prioritize positions associated with activity-selective loss, while indicating that position-level performance should not be taken as evidence of accurate phenotype prediction for individual substitutions. ICed-ENM Ω showed greater phenotype separation across the full ranking than ENCoM-derived Ω and Rhapsody-2(dyn) scores, whereas its advantage in top-group association was supported against ENCoM but remained inconclusive against Rhapsody-2(dyn). The preferential ranking of

abundance-related substitutions by DynaMut2 was consistent with its stability-oriented output, illustrating the distinct phenotype associations of scores that capture different components of mutation impact. Together, these results show that ICed-ENM Ω provides as a functionally informative ranking of PTEN positions, preferentially identifying those associated with activity-selective loss. The mechanistic basis of this association and its generalizability across proteins remain to be established.

Several important limitations remain. First, ENMs are known to exhibit non-physical fluctuations in poorly connected regions or terminal residues[43], which may lead to an overestimation of predicted mutation impacts in these regions. Although we mitigated this issue by applying local-connectivity criteria during mutation-impact scoring, this intrinsic limitation of coarse-grained ENMs cannot be entirely eliminated. Consequently, some residual noise may still be present in our mutation-impact estimates. Second, our framework may underprioritize substitutions whose functional consequences arise primarily from specific chemical interactions beyond the level of detail represented by the coarse-grained model. Cysteine substitutions provide a potential example because disulfide bonding and metal coordination can be critical to their effects[38,44,45]. Most WT-cysteine-to-mutant transition types had high-|ΔS| fractions below the 10% baseline (Supplementary Fig. 1B), and WT cysteine had one of the lowest aggregate fractions among source amino acids (Supplementary Fig. 1C). Together, these observations raise the possibility that the substitution-level |ΔS| score underprioritizes cysteine substitutions with chemistry-dependent functional effects. Third, our framework quantifies mutation-induced perturbations of conformational dynamics but does not directly estimate changes in folding free energy, a widely used measure of mutation effects on protein stability[5,46,47]. The contrasting phenotype associations observed in the PTEN benchmark suggest that dynamics- and stability-

based scores emphasize different, potentially complementary aspects of mutation impact, both of which may be informative when interpreting functional consequences.

Together, these findings support structure-derived dynamic sensitivity as a functionally informative component of mutation impact. ICed-ENM-based mutation-impact analysis provides a computationally tractable framework for characterizing substitution-induced perturbations of conformational dynamics and prioritizing positions for functional investigation. The publicly available implementation enables researchers to generate mutation-impact maps from protein structures and formulate experimentally testable hypotheses about how substitutions influence function through changes in conformational dynamics. Its coarse-grained formulation supports systematic analysis across large structural datasets, providing a scalable basis for investigating the contribution of conformational dynamics to mutation effects alongside other structural and physicochemical determinants.

Methods

MD datasets for ICed-ENM parameterization

We built two complementary datasets of all-atom MD simulation trajectories covering diverse protein structures and force-field models, including proteins with nearly complete conformational landscapes captured experimentally[20,27]. The first dataset was generated by performing unbiased MD simulations for five proteins—RBP (271 residues), 5′NTase (525 residues), RNaseIII (438 residues), SERCA (994 residues), and ACLY (4384 residues)—with simulations initiated from two experimentally determined structures representing distinct

conformational states of each protein. Simulations were conducted using three force fields: AMBER99SB-ILDN[48], CHARMM36m[49], and OPLS-AA/L[50].

System preparation, energy minimization, and equilibration were performed using force-field-specific input files generated by the CHARMM-GUI web interface[51], followed by production MD. Production simulations were performed under NPT conditions at 303.15 K and 1 bar using force-field-specific standard GROMACS protocols. All simulations used a 2-fs time step, periodic boundary conditions, PME electrostatics, and constraints on bonds involving hydrogen atoms. Short-range cutoff schemes, thermostat settings, and van der Waals treatments followed the corresponding force-field-specific protocols. System information, force fields, replica numbers, and simulation lengths are provided in Supplementary Table 3. The second dataset was compiled from MD simulation data of various protein systems indexed in MDverse, a searchable platform designed to aggregate and organize publicly shared MD simulation data[52] (Supplementary Table 4). These datasets served as the foundation for model development, and their specific roles are described in the following sections.

PDB dataset for conformational-transition benchmarking

We assembled a PDB structure dataset comprising 82 endpoint conformations from 41 proteins, which was used to evaluate how well the model captures experimental conformational transitions. The proteins span a broad range of sizes and structural classes. To define the relative conformational states of each endpoint pair, we assessed structural compactness using the radius of gyration and classified the two structures as either ‘open’ or ‘closed’. Details of the PDB structure dataset, including protein names, corresponding PDB identifiers, and structural annotations are provided in Supplementary Table 5.

PCA of conformational ensembles

Principal component analysis (PCA) is a dimensionality-reduction method that identifies the dominant axes of structural variation within an ensemble of protein conformations. After structural alignment, the covariance matrix of atomic positional fluctuations is constructed and diagonalized. The resulting eigenvectors define principal directions of collective structural variation, whereas the corresponding eigenvalues quantify the variance captured along each direction. The leading principal components (PCs) therefore provide a reduced-dimensional representation of the dominant conformational changes sampled by the ensemble.

In this study, PCA was used in two distinct contexts. First, PCA was applied to MD trajectories to extract essential dynamics (ED) modes for ICed-ENM parameterization[53]. In this MD-derived ED analysis, the leading PCs represent the dominant collective fluctuations sampled during MD simulations and were used to guide the refinement of ICed-ENM normal modes. Second, PCA was applied to experimental PDB structural ensembles to define reference conformational axes for validation[27]. Unlike MD-derived ED, PCA of experimental ensembles captures structural variation among experimentally resolved conformational states. The resulting experimental PC1–PC2 subspaces were used as reference collective variables for projecting WT and mutant MD trajectories and comparing mutation-induced redistribution within projected conformational free-energy landscapes.

ICed-ENM model and parameterization

ICed-ENM is a coarse-grained elastic network model formulated in internal-coordinate space and subsequently refined using dynamic information extracted from all-atom MD trajectories. It

represents a protein structure as a mass-spring mechanical system, in which selected atoms or pseudo-atoms are treated as nodes, and inter-node interactions are modeled as harmonic springs. To describe molecular geometry in internal-coordinate space and to represent mutation-induced geometric changes, we defined nodes using key heavy atoms: N, Cα, and C from the backbone, together with the side-chain center of mass (SCOM), calculated from side-chain heavy atoms (Fig. 1A). Atomic masses were assigned to the backbone heavy-atom nodes, whereas the total mass of the corresponding side-chain heavy atoms was assigned to each SCOM. The spring constants (k) are defined using a nonlinear model combining two types of interactions defined by sequential and spatial distances. The nonlinear model is given by:

$$k_{i,j} = \begin{cases} \frac{C_{seq}}{(d_{seq})_{i,j}^{n_{seq}}} + \frac{C_{cart}}{(d_{cart})_{i,j}^{n_{cart}}}; & if\ (d_{cart})_{i,j} < d_{cutoff} \\ 0; & otherwise \end{cases}, \quad (1)$$

where $k_{i,j}$ represents the spring constant value between i^{th} and j^{th} nodes. $\left(d_{seq}\right)_{i,j}$ represents the sequential distance between them, indicating how far apart they are in the primary sequence of the molecule. $(d_{cart})_{i,j}$ represents the spatial distance between two nodes in three-dimensional space. d_{cutoff} is a spatial distance cutoff, defining the range within which neighboring nodes are connected by springs. The parameters C_{seq} and C_{cart} represent the prefactors for the sequential and spatial interaction terms, respectively, while n_{seq} and n_{cart} denote the corresponding power-law exponents.

To optimize these parameters for reproducing MD-derived effective stiffness between nodes, we conducted a two-step parameterization procedure based on the MD simulation datasets described above. In the first step, four scalar parameters of the nonlinear model in equation (1), C_{seq}, C_{cart}, n_{seq}, and n_{cart}, were calibrated using apparent force constants (K_{app}) derived from distance

fluctuations in MD trajectories[17]. K_{app} represents the effective stiffness between node pairs, including backbone N, Cα, and C atoms and SCOM beads:

$$(K_{app})_{i,j} = \frac{k_B T}{\langle \Delta r_{i,j}^2 \rangle}, \quad (2)$$

where k_B is the Boltzmann constant and T is the absolute temperature (303.15 K in this study). The term $\langle \Delta r_{i,j}^2 \rangle$ represents the trajectory average of the squared deviation of the inter-node distance $r_{i,j}$ from its mean value. All spatial distances were expressed in Å, and K_{app} was expressed in kcal mol^{-1} Å^{-2}. The nonlinear model was trained using 10-fold cross-validation on the locally generated MD dataset, comprising 18,480,049 samples from 26 MD trajectories. Supplementary Fig. 4A shows the distribution of K_{app} as a function of d_{seq} and d_{cart} for this dataset. The model performance was then evaluated on the MDverse dataset, comprising 13,807,275 samples from 21 trajectories. This procedure yielded the following optimal parameter values: C_{seq} = 286.4, C_{cart} = 147.6, n_{seq} = 1.92, and n_{cart} = 1.66.

In the second step, we optimized the distance cutoff, d_{cutoff}. To identify the cutoff value that best reproduces MD-derived dominant dynamics subspaces, we compared the subspace spanned by ICed-ENM normal modes with the ED subspace from the local and MDverse datasets. For each candidate d_{cutoff}, the normal modes were computed using the optimized nonlinear-model parameters described above. The similarity between the normal-mode and ED subspaces was quantified using the root-mean-square inner product ($RMSIP$), which measures the overlap between two dynamic subspaces[54]. $RMSIP$ is computed as:

$$RMSIP_k = \sqrt{\frac{1}{k}\sum_{i=1}^{k}\sum_{j=1}^{k}(\mathbf{v}_i^A \cdot \mathbf{v}_j^B)^2}, \quad (3)$$

where k is the number of modes included in each subspace, and $\mathbf{v}_i^A$ and $\mathbf{v}_j^B$ denote the i^{th} and j^{th} normalized eigenvectors from the mode sets A and B, respectively. Supplementary Fig. 4B shows the mean and median $RMSIP$ values for subspaces composed of the lowest 3, 5, and 10 modes as a function of d_{cutoff}. Based on these results, we selected d_{cutoff} = 8 Å as the optimal cutoff value, as it maximized the $RMSIP$ values and therefore best captured the dominant flexibility patterns sampled in all-atom MD simulations.

NMA based on ICed-ENM

NMA based on ENMs is a widely used computational technique for characterizing the vibrational properties of proteins at equilibrium[18]. It employs a harmonic approximation of the potential energy surface and solves an eigenvalue problem derived from the Hessian matrix, which contains the second derivatives of the potential energy with respect to node displacements with a mass matrix incorporated when mass-weighted modes are considered. Each normal mode consists of an eigenvector and its corresponding eigenvalue, representing the displacement pattern and the associated squared frequency, respectively. In particular, the low-frequency normal modes are often associated with large-scale conformational changes that are functionally relevant.

In this study, we performed NMA using ICed-ENM, which characterizes structural dynamics based solely on backbone dihedral angles (i.e., ϕ and ψ angles). To carry out the NMA, we first expressed both the potential and the kinetic energy functions in internal-coordinate space, enabling the construction of the Hessian and mass matrices, corresponding to the second derivatives of the potential and kinetic energy functions, respectively[21,55]. Let $\boldsymbol{\theta} =$

$(\theta_1, \theta_2, \cdots, \theta_n)^T$ denote the column vector of internal coordinates. The Hessian matrix (**H**) in internal-coordinate space is given by:

$$V = \frac{1}{2}\sum_{(i,j)\in\varepsilon} k_{i,j}\left(\left\|\boldsymbol{r}_i(t) - \boldsymbol{r}_j(t)\right\| - \left\|\boldsymbol{r}_i(0) - \boldsymbol{r}_j(0)\right\|\right)^2 = \frac{1}{2}\boldsymbol{\theta}^T\mathbf{H}\boldsymbol{\theta}\,, \tag{4}$$

where,

$$\mathbf{H}_{\alpha,\beta} = \frac{\partial^2 V}{\partial\theta_\alpha \partial\theta_\beta} = \left(\sum_{(i,j)} \left(\frac{\partial \boldsymbol{r}_i}{\partial\theta_\alpha}\right)^T \mathbf{H}^C_{i,j} \frac{\partial \boldsymbol{r}_j}{\partial\theta_\beta}\right), \tag{5}$$

i and j index nodes, whereas α and β index internal-coordinate variables, ε denotes the set of spring-connected node pairs, $\mathbf{r}_i(t)$ is the position vector of the i^{th} node at time t, and $\mathbf{H}^C_{i,j}\left(= \frac{\partial^2 V}{\partial \mathbf{r}_i\, \partial \mathbf{r}_j}\right)$ is the Hessian matrix in Cartesian coordinates. Similarly, the global mass matrix (**M**) is obtained from the kinetic energy:

$$T = \frac{1}{2}\sum_i^N m_i \parallel \dot{\mathbf{r}}_i(t) \parallel^2 = \frac{1}{2}\dot{\boldsymbol{\theta}}^T\mathbf{M}\dot{\boldsymbol{\theta}}, \tag{6}$$

where N is the number of nodes and m_i is the mass of the i^{th} atom. The mass matrix elements are given by

$$\mathbf{M}_{\alpha,\beta} = \frac{\partial^2 T}{\partial\theta_\alpha\, \partial\theta_\beta} = \sum_i m_i \left(\frac{\partial \mathbf{r}_i}{\partial\theta_\alpha}\right)^T \frac{\partial \mathbf{r}_i}{\partial\theta_\beta}. \tag{7}$$

Finally, the generalized eigenvalue problem $\mathbf{Hv} = \lambda\mathbf{Mv}$ is solved to obtain normal modes in internal-coordinate space.

Benchmarking of ENM-predicted collective motions

To evaluate how well ICed-ENM captures experimentally observed collective motions, we benchmarked its normal modes against those obtained from three existing ENM approaches: iMOD[25], edENM[17], and ANM[18]. These models were chosen to compare the effects of coordinate representation and force-field parameterization. For iMOD, we used the Cα-only internal-coordinate representation with the edENM force-field option. edENM was used as a Cartesian-coordinate model with its original essential-dynamics-based parameterization. ANM was implemented as a conventional Cα-only Cartesian-coordinate model, in which Cα pairs within a 12 Å cutoff were connected by uniform springs with force constant set to unity. These settings were used consistently for all structures in the benchmark dataset to assess how coordinate representation and parameterization influence the experimental relevance and stereochemical realism of predicted motions.

Using the PDB structure dataset described above, we evaluated how well each ENM model captured experimentally observed conformational transitions. For each paired endpoint structure, the experimental transition vector was defined as the Cartesian displacement vector between the aligned open- and closed-state conformations. Normal mode analysis was then performed for each endpoint structure using the four ENM models, and the resulting low-frequency modes were compared with the corresponding experimental transition vector.

For each normal mode, the overlap with the experimental transition vector was computed as the normalized dot product between the mode vector and the transition vector. The individual overlap score for the i^{th} mode was calculated as

$$O_i = \frac{|\Delta \mathbf{r} \cdot \mathbf{v}_i|}{\|\Delta \mathbf{r}\| \|\mathbf{v}_i\|}, \tag{8}$$

where $\Delta\mathbf{r}$ denotes the conformational transition vector, and $\mathbf{v}_i$ denotes the i^{th} normal mode vector. We reported the overlap scores for the first and second modes, O_1 and O_2, as measures of how strongly the leading low-frequency modes align with the experimentally observed transition. We also calculated the maximum overlap, O_{max}, among the analyzed low-frequency modes and recorded the corresponding mode index, $N(O_{max})$, to determine how early the most transition-relevant motion appears in each model's mode spectrum.

To evaluate how well a set of low-frequency modes collectively spans the experimental transition, we calculated the cumulative squared overlap (CSO). For a subspace composed of the k lowest-frequency modes, CSO was defined as

$$CSO_k = \sum_{i=1}^{k} \left(\frac{\Delta\mathbf{r} \cdot \mathbf{v}_i}{\|\Delta\mathbf{r}\|\|\mathbf{v}_i\|}\right)^2, \quad (9)$$

A CSO value of 1 indicates that the experimental transition vector lies entirely within the selected normal-mode subspace. We reported CSO values for mode subsets composed of the lowest 3, 5, and 10 modes and used them to compare the extent to which each ENM model captures transition-relevant collective motions within its low-frequency mode subspace.

To assess whether predicted collective motions preserve stereochemical realism, mode-deformed structures were generated for each ENM model. Each original structure was displaced by 5 Å along the positive and negative directions of each of the lowest three normal modes. For the resulting deformed structures, local backbone geometry was evaluated using the distribution of consecutive Cα–Cα distances along the protein backbone, which are expected to remain close to 3.8 Å in chemically reasonable conformations.

As a complementary network-level measure, we quantified preservation of the Cα–Cα contact network after mode deformation. Pairwise Cα–Cα contacts were defined using an 8 Å distance

cutoff. For each deformed structure, the relative change in the total number of pairwise Cα–Cα contacts was calculated with respect to the corresponding original structure. These backbone-distance and contact-network analyses were used to compare the stereochemical integrity of mode-deformed structures generated by ENM models.

Difference in vibrational entropy

Vibrational entropy differences derived from normal modes provide a measure of changes in structural flexibility and stability between two conformations or systems[22,56]. In this study, we computed this difference between the WT and mutant systems to quantify the impact of single-residue substitutions on conformational dynamics. The entropy change from conformation A to conformation B ($\Delta S_{A\to B}$) is calculated as:

$$\Delta S_{A\to B} = \ln\left(\frac{\prod_{n=1}^{N_s}\lambda_{n,A}}{\prod_{n=1}^{N_s}\lambda_{n,B}}\right), \tag{10}$$

where N_s is the number of modes included in the calculation, and $\lambda_{n,i}$ denotes the n^{th} eigenvalue from the NMA calculation for conformation i. When $\Delta S_{A\to B} < 0$, it indicates that conformation B is less flexible than conformation A (and vice versa). Because the mutation-impact analysis focuses on the magnitude rather than the direction of the entropy change, |ΔS| was used as the substitution-level impact score.

Mutation-impact analysis by amino-acid substitution scanning

Because ICed-ENM represents both backbone atoms and approximate side-chain geometry, it can incorporate side-chain-mediated structural perturbations. Building on this capability, we applied the model to evaluate mutation-induced perturbations in conformational dynamics. Fig.

1B illustrates the overall workflow of the mutation-impact analysis. Starting from an input three-dimensional conformation (typically a PDB structure), all possible single-residue mutants were generated by substituting each residue with the 19 other standard amino acids using MODELLER 10.7[37]. For the WT structure and all generated mutants, ICed-ENM-based NMA was performed to estimate mutation-induced perturbations in conformational dynamics. For each single-residue substitution, the substitution-level impact score was defined as the absolute vibrational entropy difference, $|\Delta S|$, between the WT and mutant normal-mode spectra.

To derive the position-level mutation-impact score, Ω, additional post-processing steps were applied to the resulting $|\Delta S|$ profile. First, for each residue, the maximum $|\Delta S|$ value among the 19 substitutions was selected to capture position-specific mutation sensitivity. Next, to incorporate local structural context, these position-specific scores were smoothed using a Gaussian kernel (σ = 5.0 Å) over neighboring residues within a 15 Å inter-residue Cα distance cutoff.

In constructing the smoothing kernel, we introduced a local-connectivity criterion to determine which residues could contribute to each smoothed estimate. Local connectivity was quantified as the number of neighboring Cα atoms within an 8 Å cutoff. Residues with at least six neighboring Cα atoms were considered valid residues, whereas residues with fewer than six neighboring Cα atoms were defined as low-connectivity residues. Only valid residues were used as contributors to the smoothing kernel, including the self-contribution when the target residue itself was valid. This criterion was motivated by the fact that ENM-based estimates can be less reliable in poorly connected local environments, where local dynamic responses may be dominated by network-connectivity artifacts. Accordingly, low-connectivity target residues could still receive Ω scores

when valid neighboring residues were available within the smoothing neighborhood. Finally, the resulting Ω profile was mapped onto the protein structure.

Validation systems and robustness of PC axes

To validate the ICed-ENM-based mutation-impact analysis, we used experimentally derived PC1–PC2 subspaces as reference conformational coordinates for projecting WT and mutant MD trajectories and comparing mutation-induced redistribution of conformational sampling.

For this analysis, we selected three representative proteins: RBP, 5′NTase, and SERCA, whose PDB ensembles satisfy a structural richness criterion originally proposed by Weiss and Levitt and subsequently formalized in our previous work[20,27,57]. Each ensemble contains multiple interconnected conformational states spanning the functional transition and yields principal components that capture more than 70% of the structural variance while separating major functional states along physically interpretable conformational axes. Under these conditions, PC1 and PC2 provide robust reference coordinates for evaluating mutation-induced changes in MD conformational sampling. The robustness of these conformational axes against reduced structural sampling has been demonstrated previously through systematic ensemble-reduction analysis[27]. In particular, PCA analyses of SERCA based on substantially different structural ensembles have yielded similar conformational spaces across multiple studies[20,27,58]. RBP and 5′NTase satisfy the same structural-richness criterion and have likewise served as benchmark systems in experimental-ensemble PCA analysis[27]. The structural ensembles used for PCA are listed in Supplementary Table 6.

Selection of hot- and cold-spot mutations for MD validation

For each validation system, position-level mutation-impact scores (Ω) were used to identify candidate dynamic hot spots. High-Ω residues were first ranked within each protein, and validation targets were selected to sample diverse structural and physicochemical environments. For each selected high-Ω residue, a lower-Ω control site was chosen to provide a broadly comparable local environment, considering solvent exposure, secondary structure, and amino-acid properties where possible. SASA was calculated using FreeSASA[32], and secondary-structure assignments were obtained using DSSP[33]. The selected hot- and cold-spot pairs, together with their Ω values, SASA values, and structural contexts, are summarized in Supplementary Table 1. All selected positions were subjected to single-point alanine substitution for all-atom MD validation.

All-atom MD simulations of selected mutants

All-atom MD simulations were performed to evaluate whether alanine substitutions at high-Ω sites produced stronger changes in conformational sampling than substitutions at low-Ω sites selected in broadly comparable local structural environments. For each validation system, the reference structure used for mutation-impact analysis was also used to construct the WT simulation system: 1URP:A for RBP, 1OID:A for 5′NTase, and 2C9M:A for SERCA. For 5′NTase, the engineered disulfide-forming mutations S228C and P513C in 1OID were reverted to their native residue identities before system preparation. For SERCA, the Ca^{2+}-bound E1 open structure was simulated with two bound Ca^{2+} ions and embedded in a POPC membrane.
WT and mutant systems were prepared using CHARMM-GUI[51]. Single-point alanine substitutions were introduced at the selected hot- and cold-spot positions for each protein. Protein interactions were described using the CHARMM36m force field[49]. For the SERCA

membrane system, POPC lipids were described using the CHARMM36 lipid force field[59]. Each system was solvated and neutralized according to the CHARMM-GUI setup protocol, followed by energy minimization and equilibration before production simulations.

For each WT and mutant system, three independent 100-ns production MD simulations were performed. Replicates were initialized with different starting velocities to assess reproducibility of mutation-dependent conformational sampling. Production simulations were performed under NPT conditions at 303.15 K and 1 bar, using a 2-fs time step, periodic boundary conditions, PME electrostatics, and constraints on bonds involving hydrogen atoms, consistent with the general MD protocol described above. The resulting trajectories were subsequently projected onto experimentally derived PC1–PC2 subspaces for conformational free-energy landscape analysis.

MD projection onto experimental PC subspaces and free-energy landscapes

To quantify mutation-dependent conformational redistribution, potentials of mean force (PMFs) were constructed in the experimentally derived PC1–PC2 subspace. A PMF provides a free-energy profile along selected collective variables by integrating over the remaining degrees of freedom[60]. In this study, conformations sampled from WT and mutant MD trajectories were projected onto the first two PCs derived from the experimental PDB ensembles to characterize mutation-induced reshaping of projected conformational free-energy landscapes. Before projection, MD frames were Cα-aligned to the corresponding reference structure used to construct the experimental PC basis.

The PMF, expressed as a function of the collective variables s_1 and s_2 corresponding to PC1 and PC2, respectively, was computed using the Boltzmann relation:

$$F(s_1, s_2) = -k_B T \ln P\,(s_1, s_2), \tag{11}$$

where $P(s_1, s_2)$ is the probability density of conformations at position (s_1, s_2) in the PC1–PC2 subspace. $P(s_1, s_2)$ was estimated from the projected MD conformations by merging the three independent 100-ns replicas for each system and constructing a two-dimensional histogram on a common 80 × 80 grid for the WT and mutant trajectories of each protein. Histogram counts were smoothed using a Gaussian filter with σ = 1.2 bins, and a pseudocount of 10^{-12} was added before applying the Boltzmann relation at 303.15 K. The resulting PMF was shifted so that its minimum value was zero, enabling comparison of relative landscape shapes and mutation-dependent redistribution of conformational sampling among WT and mutant systems.

PTEN variant phenotypes and comparative benchmarking

PTEN variant phenotypes were obtained from the integrated abundance–activity dataset reported by Matreyek and colleagues[26]. The tabular file used in this analysis contained 4,177 unique missense substitutions with intracellular abundance and lipid phosphatase activity measurements. Of the 4,177 substitutions, 3,218 mapped to resolved positions in the apo open PTEN structure 1D5R:A. Among these mapped substitutions, 2,931 had high-confidence activity scores. Of these, 992 belonged to one of three source-defined loss categories: “loss_of_activity_only” (reduced activity with WT-like abundance), “loss_of_abundance_only” (reduced abundance with WT-like activity), or “loss_of_both” (reduced activity and abundance). To limit potential boundary effects near the unresolved internal segment (residues 282–312), we excluded five additional positions on each side (277–281 and 313–317). After this exclusion, 956 substitutions remained: 100 “loss_of_activity_only” substitutions were assigned to the activity-selective loss group, whereas 419 “loss_of_abundance_only” and 437 “loss_of_both” substitutions were

combined into the abundance-related loss group (n = 856). Substitutions annotated as WT-like or intermediate/unclassified, as well as those without a high-confidence activity score, were excluded from the two-group comparison.

The 956 substitutions mapped to 215 PTEN positions. A position was assigned to the activity-selective loss group if it had at least one activity-selective substitution and no abundance-related substitution, and to the abundance-related loss group if it had at least one abundance-related substitution and no activity-selective substitution. The 22 positions with substitutions from both groups were excluded from the position-level phenotype comparison, leaving 20 activity-selective and 173 abundance-related positions.

The computational ranking background comprised the 297 retained resolved positions of 1D5R:A and all 19 non-WT amino-acid substitutions at each position (5,643 substitutions). ICed-ENM and ENCoM[22] were evaluated using the same WT structure and modeled mutant structures. Their substitution-level scores were the absolute changes in vibrational entropy between the WT and mutant models, $|\Delta S|$. For the position-level comparison, ENCoM $|\Delta S|$ scores were aggregated and spatially smoothed using the procedure defined above for ICed-ENM Ω. The ten positions flanking the unresolved segment were excluded both as scored targets and as smoothing contributors. DynaMut2[23] $\Delta\Delta G$ predictions were used as a substitution-level stability comparator, with more negative values ranked higher to prioritize predicted destabilization; these predictions were not aggregated into position-level scores. Rhapsody-2(dyn)[24] was selected to provide a comparator based solely on intrinsic dynamics, consistent with our focus on the dynamic component of mutation impact. Scores were generated from the experimental PTEN structure (1D5R:A) using a model trained after excluding PTEN variants. Because its dynamics descriptors depend on the WT structure rather than the substituted amino

acid, the model assigns the same score to all substitutions at a given position. These scores were therefore used directly at the position level.

Each method independently defined its top 10% score group over the complete computational background: the 565 highest-ranked substitutions among 5,643 or the 30 highest-ranked positions among 297. Score ties at the cutoff were resolved by ascending variant identifier at the substitution level or ascending residue number at the position level. We then calculated top-group inclusion rates separately for the two phenotype groups. The OR was the odds of inclusion for activity-selective cases divided by those for abundance-related cases. ROC curves were constructed using all phenotype-annotated substitutions or positions, with activity-selective loss as the positive class and abundance-related loss as the negative class. Substitutions or positions were progressively included in descending order of the ranking score. AUROC was calculated as the area under each ROC curve.

Direct method comparisons contrasted ICed-ENM with ENCoM at the substitution level and with ENCoM and Rhapsody-2(dyn) at the position level. For top-group inclusion, we fitted logistic generalized estimating equation models with inclusion in the top 10% score group as the binary outcome and phenotype, method, and their interaction as predictors. Observations were clustered by PTEN position at both the substitution and position levels, using an independence working correlation and position-clustered robust sandwich standard errors; no small-sample correction was applied. The interaction term yielded the ICed-ENM-to-comparator ratio of ORs and its nominal 95% CI. Two-sided Wald *P* values tested the null hypothesis that the ratio of ORs was 1.

For full-ranking discrimination, ΔAUROC was defined as the ICed-ENM AUROC minus the comparator AUROC. We generated 100,000 paired bootstrap replicates (random seed 20261006)

by resampling PTEN positions with replacement. In each replicate, AUROCs for both methods were recalculated using the same sampled positions and their precomputed scores. At the substitution level, positions were sampled with replacement without phenotype stratification, and each selected position contributed all its phenotype-annotated substitutions. At the position level, the 20 activity-selective and 173 abundance-related positions were sampled with replacement separately within their phenotype groups, preserving group sizes. Nominal 95% CIs were the 2.5th and 97.5th percentiles of the bootstrap ΔAUROC distribution. Two-sided bootstrap *P* values were calculated as twice the smaller estimated tail probability of the bootstrap ΔAUROC distribution, capped at 1. Each tail probability was estimated as $(k + 1)/(B + 1)$, where k was the number of estimates in the corresponding tail (≤ 0 or ≥ 0) and $B = 100{,}000$. The two position-level *P* values for ratio-of-OR comparisons and the two for ΔAUROC comparisons were Holm-adjusted as separate families; the sole substitution-level comparison for each metric was unadjusted. CIs were not adjusted for multiple comparisons.

Data availability

The numerical source data for the figures and selected tables in this study are available on Zenodo at https://doi.org/10.5281/zenodo.23260125. The dataset contains figure- and table-level input files for structural and conformational benchmarks, mutation-impact analyses, the PTEN phenotype benchmark including ROC curves and method comparisons, and apparent-force-constant and RMSIP analyses. File descriptions, column definitions, and selection criteria are provided in the accompanying README files.

Code availability

The custom code used for ICed-ENM normal-mode analysis and mutation-impact analysis is archived on Zenodo at https://doi.org/10.5281/zenodo.23259671. The archived release includes the ICed-ENM NMA Fortran engine, mutation-impact analysis pipeline, example inputs, walkthrough scripts, documentation, and Conda environment files. The code is distributed under the BSD 3-Clause License. The development repository is available at https://github.com/BBHH-byb/ICed-ENM-mutation-impact.

References

1. Ostroverkhova, D., Przytycka, T. M. & Panchenko, A. R. Cancer driver mutations: predictions and reality. *Trends Mol. Med.* **29**, 554–566 (2023).
2. Bailey, M. H. *et al.* Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell* **173**, 371-385.e18 (2018).
3. Guerreiro, R. & Hardy, J. Genetics of Alzheimer's Disease. *Neurotherapeutics* **11**, 732–737 (2014).
4. Klein, C. & Westenberger, A. Genetics of Parkinson's Disease. *Cold Spring Harbor Perspectives in Medicine* **2**, a008888–a008888 (2012).
5. Stefl, S., Nishi, H., Petukh, M., Panchenko, A. R. & Alexov, E. Molecular Mechanisms of Disease-Causing Missense Mutations. *Journal of Molecular Biology* **425**, 3919–3936 (2013).
6. Naganathan, A. N. Modulation of allosteric coupling by mutations: from protein dynamics and packing to altered native ensembles and function. *Current Opinion in Structural Biology* **54**, 1–9 (2019).
7. Ose, N. J. *et al.* Dynamic coupling of residues within proteins as a mechanistic foundation of many enigmatic pathogenic missense variants. *PLoS Comput Biol* **18**, e1010006 (2022).
8. Dror, R. O., Dirks, R. M., Grossman, J. P., Xu, H. & Shaw, D. E. Biomolecular Simulation: A Computational Microscope for Molecular Biology. *Annu. Rev. Biophys.* **41**, 429–452 (2012).
9. Shaw, D. E. *et al.* Anton 3: twenty microseconds of molecular dynamics simulation before lunch. in *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis* 1–11 (ACM, St. Louis Missouri, 2021). doi:10.1145/3458817.3487397.
10. Buch, I., Harvey, M. J., Giorgino, T., Anderson, D. P. & De Fabritiis, G. High-Throughput All-Atom Molecular Dynamics Simulations Using Distributed Computing. *J. Chem. Inf. Model.* **50**, 397–403 (2010).

11. Park, S. W. *et al.* An accelerated molecular dynamics study for investigating protein pathways using the bond-boost hyperdynamics method. *Protein Science* **34**, e70073 (2025).
12. Hamelberg, D., Mongan, J. & McCammon, J. A. Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *The Journal of Chemical Physics* **120**, 11919–11929 (2004).
13. Laio, A. & Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12562–12566 (2002).
14. Sugita, Y. & Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters* **314**, 141–151 (1999).
15. Childers, M. C. & Daggett, V. Validating Molecular Dynamics Simulations against Experimental Observables in Light of Underlying Conformational Ensembles. *J. Phys. Chem. B* **122**, 6673–6689 (2018).
16. 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.* **5**, 104–118 (2023).
17. Orellana, L. *et al.* Approaching Elastic Network Models to Molecular Dynamics Flexibility. *J. Chem. Theory Comput.* **6**, 2910–2923 (2010).
18. Atilgan, A. R. *et al.* Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model. *Biophysical Journal* **80**, 505–515 (2001).
19. Gur, M., Madura, J. D. & Bahar, I. Global Transitions of Proteins Explored by a Multiscale Hybrid Methodology: Application to Adenylate Kinase. *Biophysical Journal* **105**, 1643–1652 (2013).
20. Scaramozzino, D., Lee, B. H. & Orellana, L. Efficient sampling of large-scale transition pathways and intermediate conformations in sub-mesoscopic protein complexes. *Nat Commun* **17**, 2202 (2026).

21. Lee, B. H., Park, S. W., Jo, S. & Kim, M. K. Protein conformational transitions explored by a morphing approach based on normal mode analysis in internal coordinates. *PLoS ONE* **16**, e0258818 (2021).
22. Frappier, V. & Najmanovich, R. J. A Coarse-Grained Elastic Network Atom Contact Model and Its Use in the Simulation of Protein Dynamics and the Prediction of the Effect of Mutations. *PLoS Comput Biol* **10**, e1003569 (2014).
23. Rodrigues, C. H. M., Pires, D. E. V. & Ascher, D. B. DynaMut2 : Assessing changes in stability and flexibility upon single and multiple point missense mutations. *Protein Science* **30**, 60–69 (2021).
24. Banerjee, A., Bogetti, A. T. & Bahar, I. Accurate identification and mechanistic evaluation of pathogenic missense variants with *Rhapsody-2*. *Proc. Natl. Acad. Sci. U.S.A.* **122**, e2418100122 (2025).
25. Lopéz-Blanco, J. R., Garzón, J. I. & Chacón, P. iMod: multipurpose normal mode analysis in internal coordinates. *Bioinformatics* **27**, 2843–2850 (2011).
26. Matreyek, K. A., Stephany, J. J., Ahler, E. & Fowler, D. M. Integrating thousands of PTEN variant activity and abundance measurements reveals variant subgroups and new dominant negatives in cancers. *Genome Med* **13**, 165 (2021).
27. Orellana, L., Yoluk, O., Carrillo, O., Orozco, M. & Lindahl, E. Prediction and validation of protein intermediate states from structurally rich ensembles and coarse-grained simulations. *Nat Commun* **7**, 12575 (2016).
28. Björkman, A. J. & Mowbray, S. L. Multiple open forms of ribose-binding protein trace the path of its conformational change. *Journal of Molecular Biology* **279**, 651–664 (1998).
29. Schultz-Heienbrok, R., Maier, T. & Sträter, N. Trapping a 96° domain rotation in two distinct conformations by engineered disulfide bridges. *Protein Science* **13**, 1811–1822 (2004).

30. Toyoshima, C., Nakasako, M., Nomura, H. & Ogawa, H. Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 Å resolution. *Nature* **405**, 647–655 (2000).
31. Pommié, C., Levadoux, S., Sabatier, R., Lefranc, G. & Lefranc, M. IMGT standardized criteria for statistical analysis of immunoglobulin V-REGION amino acid properties. *J of Molecular Recognition* **17**, 17–32 (2004).
32. Mitternacht, S. FreeSASA: An open source C library for solvent accessible surface area calculations. *F1000Res* **5**, 189 (2016).
33. Hekkelman, M. L., Salmoral, D. Á., Perrakis, A. & Joosten, R. P. DSSP 4: FAIR annotation of protein secondary structure. *Protein Science* **34**, e70208 (2025).
34. Lee, Y.-R., Chen, M. & Pandolfi, P. P. The functions and regulation of the PTEN tumour suppressor: new modes and prospects. *Nat Rev Mol Cell Biol* **19**, 547–562 (2018).
35. Kucukkal, T. G., Petukh, M., Li, L. & Alexov, E. Structural and physico-chemical effects of disease and non-disease nsSNPs on proteins. *Current Opinion in Structural Biology* **32**, 18–24 (2015).
36. Thomas, L. *et al.* Assessment of the potential pathogenicity of missense mutations identified in the GTPase-activating protein (GAP)-related domain of the neurofibromatosis type-1 (*NF1*) gene. *Hum. Mutat.* **33**, 1687–1696 (2012).
37. Šali, A. & Blundell, T. L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *Journal of Molecular Biology* **234**, 779–815 (1993).
38. Petukh, M., Kucukkal, T. G. & Alexov, E. On Human Disease-Causing Amino Acid Variants: Statistical Study of Sequence and Structural Patterns. *Human Mutation* **36**, 524–534 (2015).
39. Casadio, R., Vassura, M., Tiwari, S., Fariselli, P. & Luigi Martelli, P. Correlating disease-related mutations to their effect on protein stability: A large-scale analysis of the human proteome. *Hum. Mutat.* **32**, 1161–1170 (2011).

40. Yue, P., Li, Z. & Moult, J. Loss of Protein Structure Stability as a Major Causative Factor in Monogenic Disease. *Journal of Molecular Biology* **353**, 459–473 (2005).
41. Savojardo, C., Manfredi, M., Martelli, P. L. & Casadio, R. Solvent Accessibility of Residues Undergoing Pathogenic Variations in Humans: From Protein Structures to Protein Sequences. *Front. Mol. Biosci.* **7**, 626363 (2021).
42. Gao, M., Zhou, H. & Skolnick, J. Insights into Disease-Associated Mutations in the Human Proteome through Protein Structural Analysis. *Structure* **23**, 1362–1369 (2015).
43. Dubanevics, I. & McLeish, T. C. B. Optimising Elastic Network Models for Protein Dynamics and Allostery: Spatial and Modal Cut-offs and Backbone Stiffness. *Journal of Molecular Biology* **434**, 167696 (2022).
44. Wiedemann, C., Kumar, A., Lang, A. & Ohlenschläger, O. Cysteines and Disulfide Bonds as Structure-Forming Units: Insights From Different Domains of Life and the Potential for Characterization by NMR. *Front. Chem.* **8**, 280 (2020).
45. Poole, L. B. The basics of thiols and cysteines in redox biology and chemistry. *Free Radical Biology and Medicine* **80**, 148–157 (2015).
46. Aledo, P. & Aledo, J. C. Proteome-Wide Structural Computations Provide Insights into Empirical Amino Acid Substitution Matrices. *IJMS* **24**, 796 (2023).
47. Pandey, P. & Alexov, E. Most Monogenic Disorders Are Caused by Mutations Altering Protein Folding Free Energy. *IJMS* **25**, 1963 (2024).
48. Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**, 1950–1958 (2010).
49. Huang, J. *et al.* CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods* **14**, 71–73 (2017).

50. Kaminski, G. A., Friesner, R. A., Tirado-Rives, J. & Jorgensen, W. L. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B* **105**, 6474–6487 (2001).
51. Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *J Comput Chem* **29**, 1859–1865 (2008).
52. Tiemann, J. K. *et al.* MDverse, shedding light on the dark matter of molecular dynamics simulations. *eLife* **12**, RP90061 (2024).
53. Daidone, I. & Amadei, A. Essential dynamics: foundation and applications. *WIREs Comput Mol Sci* **2**, 762–770 (2012).
54. 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* **36**, 419–424 (1999).
55. Braun, W., Yoshioki, S. & Gō, N. Formulation of Static and Dynamic Conformational Energy Analysis of Biopolymer Systems Consisting of Two or More Molecules. *J. Phys. Soc. Jpn.* **53**, 3269–3275 (1984).
56. Karplus, M. & Kushick, J. N. Method for estimating the configurational entropy of macromolecules. *Macromolecules* **14**, 325–332 (1981).
57. Weiss, D. R. & Levitt, M. Can Morphing Methods Predict Intermediate Structures? *Journal of Molecular Biology* **385**, 665–674 (2009).
58. Sankar, K., Liu, J., Wang, Y. & Jernigan, R. L. Distributions of experimental protein structures on coarse-grained free energy landscapes. *The Journal of Chemical Physics* **143**, 243153 (2015).
59. Klauda, J. B. *et al.* Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *J. Phys. Chem. B* **114**, 7830–7843 (2010).

60. Frenkel, D. & Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*. (ELSEVIER ACADEMIC PRESS, Amsterdam, 2023).

Funding

L.O. discloses support for the research of this work from Cancerfonden through a Junior Investigator Award [CF 21 0305 JIA] and Project Grants [CF 21 1471 Pj and CF 24 3801 Pj], Vetenskapsrådet through a Starting Grant [VR 2021-02248], and the Karolinska Institutet Interdisciplinary Research Incubator, jointly with S.P. D.S. discloses support for the research of this work from Cancerfonden through a Postdoctoral Fellowship [CF 24 0908 PT].

Author contributions

Conceptualization: B.H.L., L.O., Methodology: B.H.L., D.S., L.O., Formal analysis: B.H.L., Investigation: B.H.L., D.S., S.P., Data curation: B.H.L., Visualization: B.H.L., Supervision: L.O., Writing—original draft: B.H.L., Writing—review & editing: B.H.L., D.S., S.P., L.O.

Competing interests statement

The authors declare no competing interests.

Figures and Tables

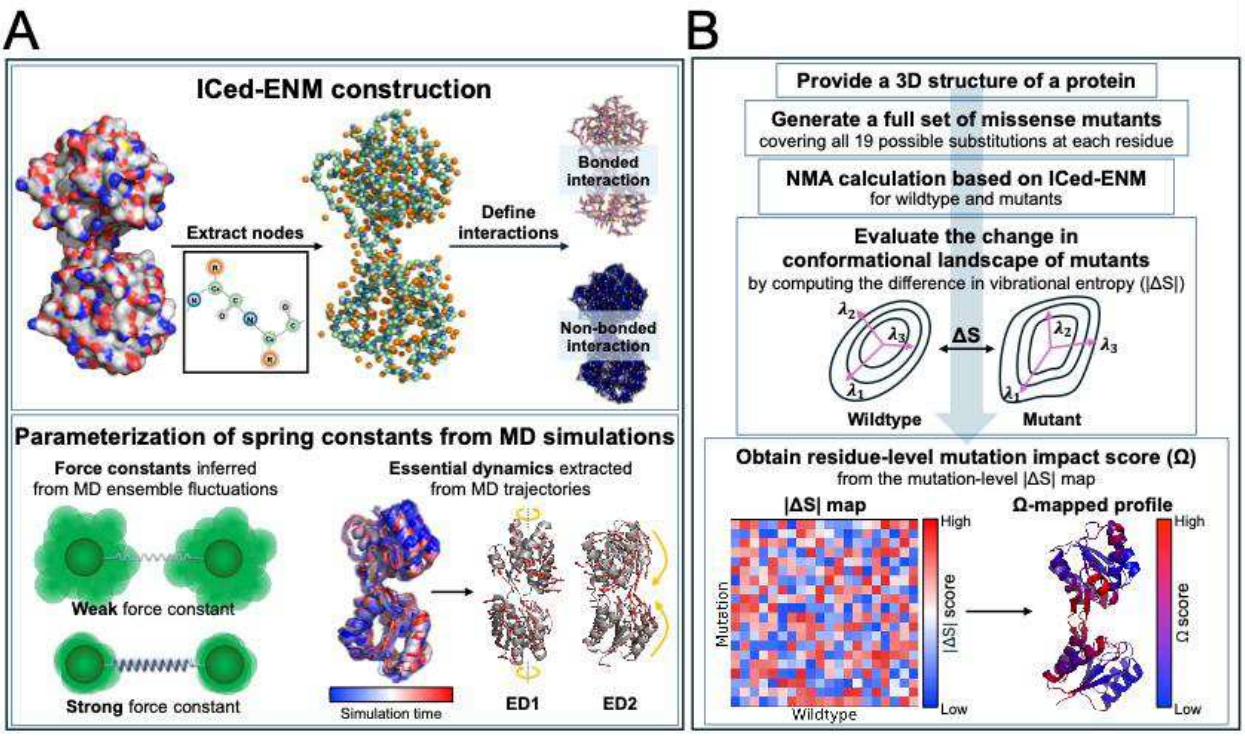


Fig. 1. Overview of the internal-coordinate, essential-dynamics-refined elastic network model (ICed-ENM) framework for mutation-impact analysis. (A) Model construction and parameterization. An all-atom structure is coarse-grained into a node-spring network, in which backbone atoms (N, Cα, and C) and the centers of mass of side chains serve as nodes. These nodes are connected by effective springs whose stiffness is defined as the sum of two classes: bonded and non-bonded interactions, defined by sequential and spatial distances, respectively. Model parameterization was performed in two stages using molecular dynamics (MD) simulation data. First, inter-node apparent force constants derived from time-ensemble fluctuations in MD simulations were used to train the model, yielding effective spring constants that reproduce the structural flexibility encoded in the MD ensemble (the lower-left panel). Second, essential dynamics extracted from MD trajectories were used as a reference to optimize

the non-bonded interaction cutoff distance, by maximizing the agreement between ICed-ENM normal-mode subspaces and the MD-derived essential subspace (the lower-right panel). **(B)** Mutation-impact analysis through the amino-acid scanning approach. In the first step, a full set of single-point missense mutants is generated from the input three-dimensional conformation. Next, normal mode analysis based on ICed-ENM is performed for both the wild type and the mutants. From these results, the differences in vibrational entropy ($|\Delta S|$) are computed and mapped across amino-acid substitution types and residue positions. By integrating this map with the protein structural information, position-level mutation-impact scores (Ω) are finally obtained.

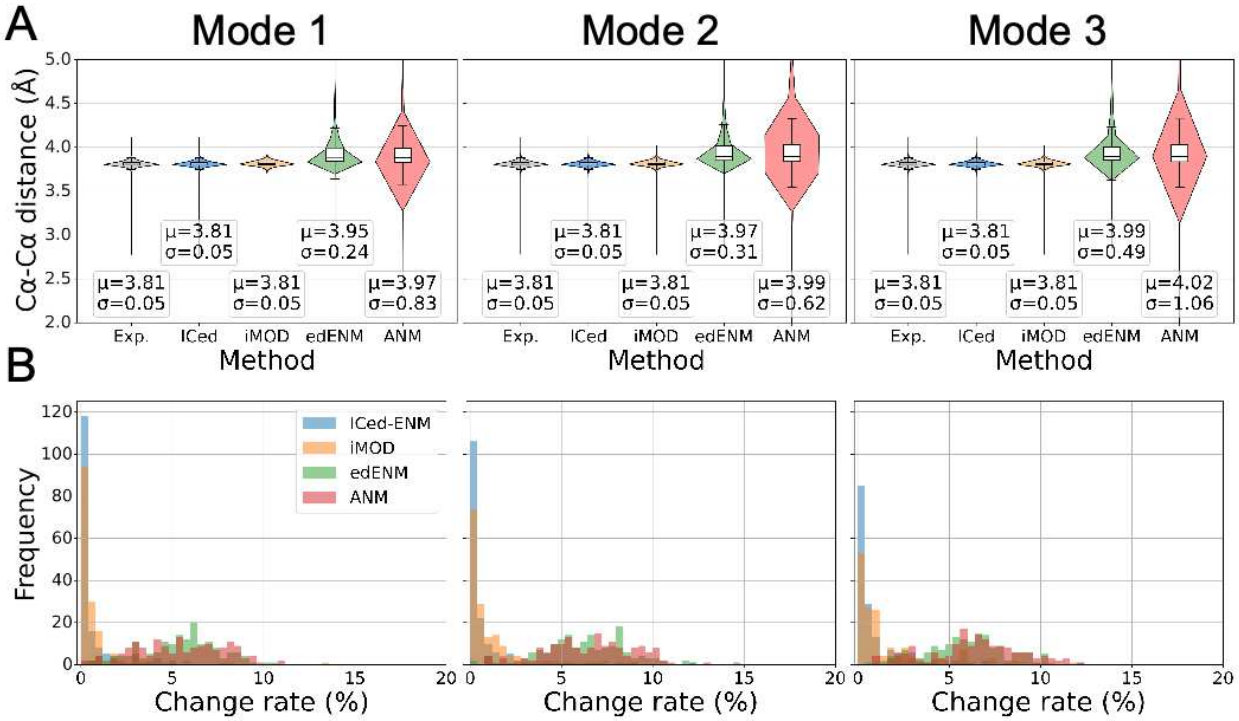


Fig. 2. Evaluation of stereochemical stability of low-frequency normal modes. (A) Violin plots of consecutive Cα–Cα distance distributions for the experimental structures (Exp., gray) and structures deformed along the first three low-frequency normal modes using different ENM models: ICed-ENM, iMOD, edENM, and ANM (blue, orange, green, and red, respectively). The three columns correspond to Mode 1, Mode 2, and Mode 3, respectively. **(B)** Histograms of contact-count change rates for Cα–Cα contacts in the corresponding mode-deformed structures. The change rate was calculated by comparing the number of pairwise Cα–Cα contacts within an 8 Å cutoff in the deformed structures with those in the corresponding experimental structures.

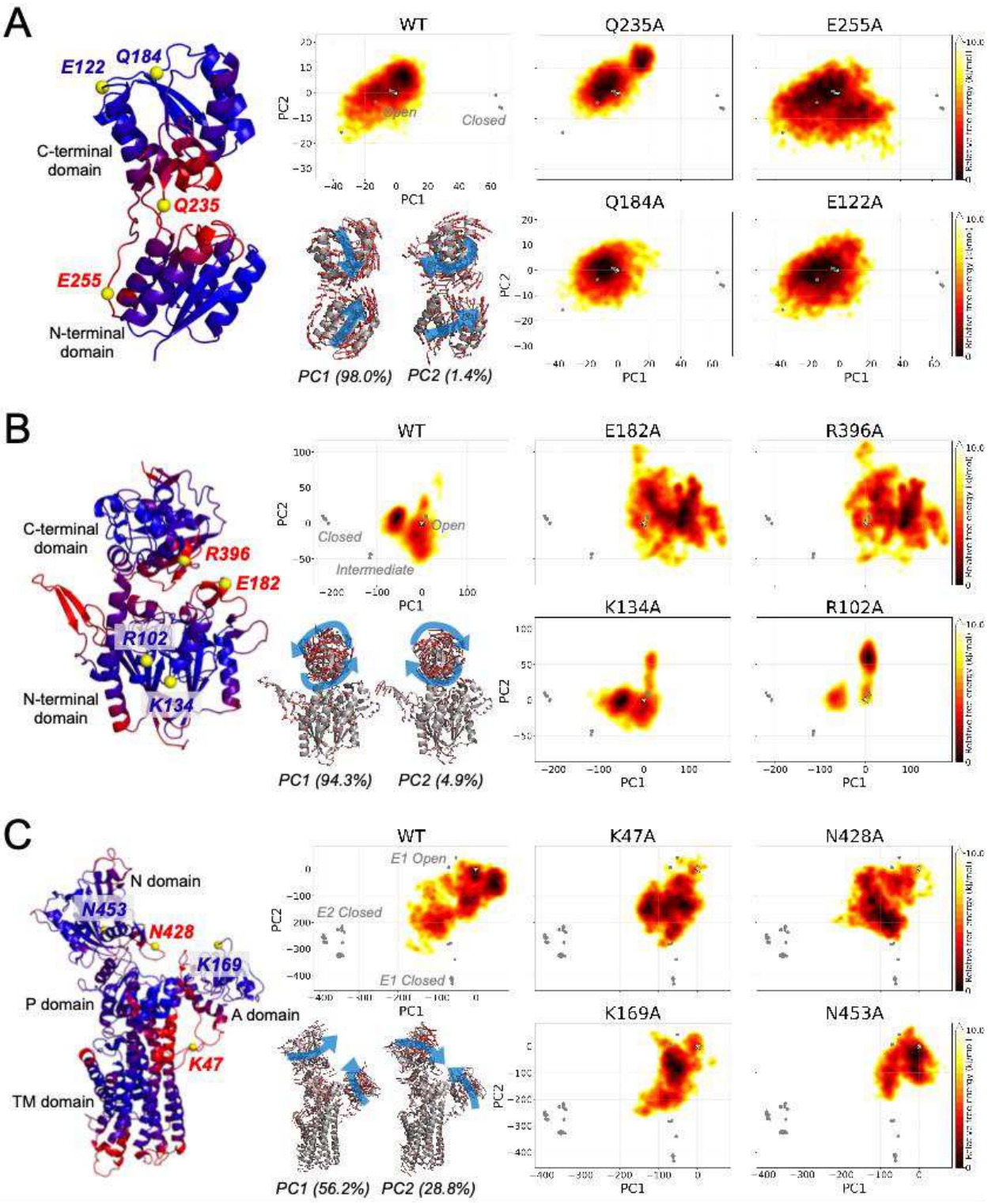

A
E122
Q184
C-terminal domain
Q235
E255
N-terminal domain
WT
Q235A
E255A
Q184A
E122A
Open
Closed
PC1 (98.0%)
PC2 (1.4%)
B
C-terminal domain
R396
E182
R102
N-terminal domain
K134
WT
E182A
R396A
K134A
R102A
Closed
Open
Intermediate
PC1 (94.3%)
PC2 (4.9%)
C
N domain
N453
N428
K169
P domain
A domain
K47
TM domain
WT
K47A
N428A
K169A
N453A
E1 Open
E2 Closed
E1 Closed
PC1 (56.2%)
PC2 (28.8%)
PC1
PC2

Fig. 3. Mutation-induced reshaping of conformational free-energy landscapes projected onto experimental principal component (PC1–PC2) subspaces. The left panels display Ω profiles mapped onto the three-dimensional structures, with domain labels shown in black. High-Ω hot spots and structurally comparable low-Ω cold spots were selected for alanine substitution and all-atom MD validation. The selected hot and cold spots are marked in red and blue, respectively. The small structural panels illustrate the dynamic shapes of PC1 and PC2 derived from the experimental ensembles. Projected conformational free-energy landscapes for wild-type (WT) and mutant systems were estimated from MD trajectories and represented as potentials of mean force, with the minimum free energy shifted to zero. Gray points denote projected experimental PDB structures, with representative states labeled in gray, and the light-green star marks the reference structure. **(A)** In RBP, using 1URP:A as the reference structure, Q235, located in the canonical hinge loop, and E255, located in a solvent-exposed loop, were selected as hot spots, with Q184 and E122 selected as their respective cold-spot counterparts. PC1 describes the dominant interdomain opening-and-closing motion between the N- and C-terminal domains, while PC2 represents a minor twisting motion around the hinge axis. **(B)** In 5′NTase, using 1OID:A as the reference structure, E182, located in a solvent-exposed loop region of the N-terminal domain at the interdomain interface, and R396, located in an α-helical segment of the C-terminal domain at the interdomain interface, were selected as hot spots, with K134 and R102 as their respective cold-spot counterparts. PC1 and PC2 represent distinct rotational motions of the C-terminal domain relative to the N-terminal domain. **(C)** In SERCA, using 2C9M:A as the reference structure, K47, located in the linker between the actuator (A) and transmembrane (TM) domains, and N428, positioned on an A-domain-facing surface of the nucleotide-binding (N) domain, were selected as hot spots, with K169 and N453 as their respective cold-spot

counterparts. PC1 describes a rotational rearrangement involving the A, N, and phosphorylation (P) domains, whereas PC2 represents an opening-and-closing motion mainly between the A and N domains.

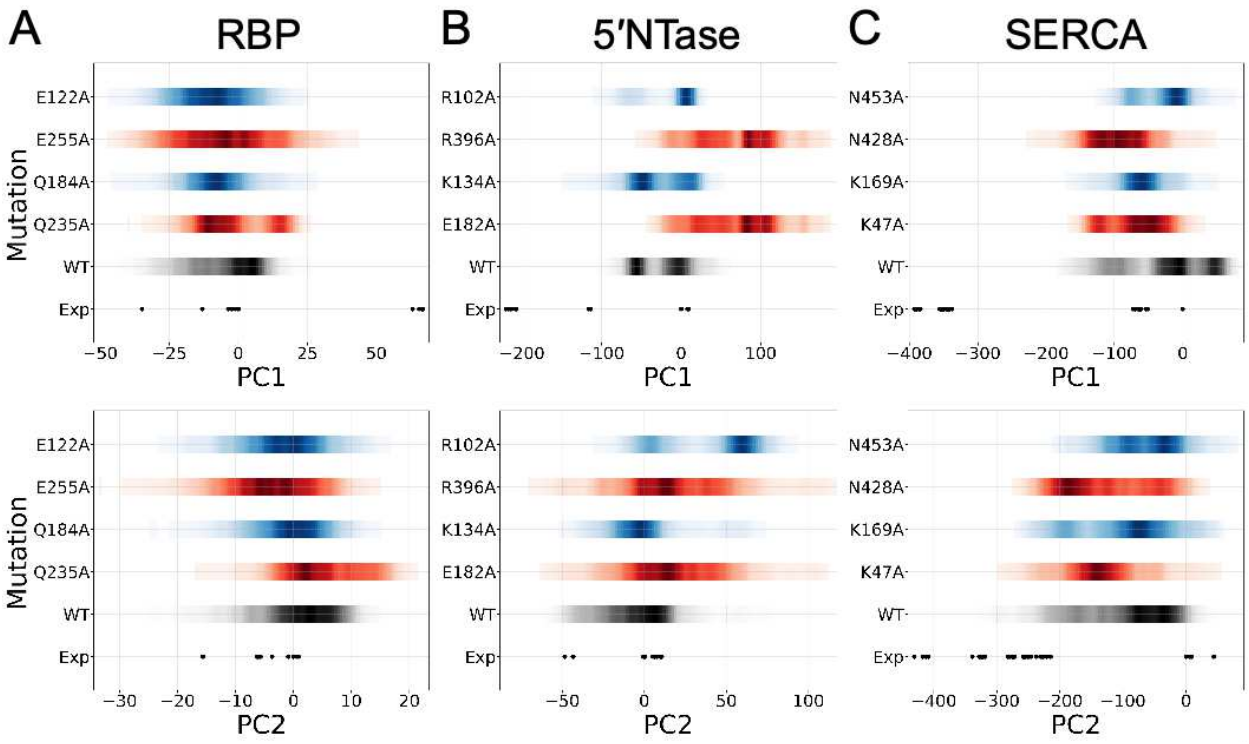


Fig. 4. Frequency distributions of MD trajectories along experimental principal component (PC) subspaces for RBP (A), 5′NTase (B), and SERCA (C). Projections along PC1 and PC2 are shown in the upper and lower panels, respectively. MD simulations were initiated from 1URP:A for RBP, 1OID:A for 5′NTase, and 2C9M:A for SERCA. In each panel, experimental structures (Exp) are shown as black points in the bottom reference strip, with representative states highlighted by circles and labeled accordingly. The remaining strips show color-coded one-dimensional frequency distributions from MD trajectories. Wild-type (WT) distributions are shown in black, whereas selected mutant distributions are colored according to mutation category, with hot-spot mutations in red and counterpart cold-spot mutations in blue. Darker colors correspond to higher frequency.

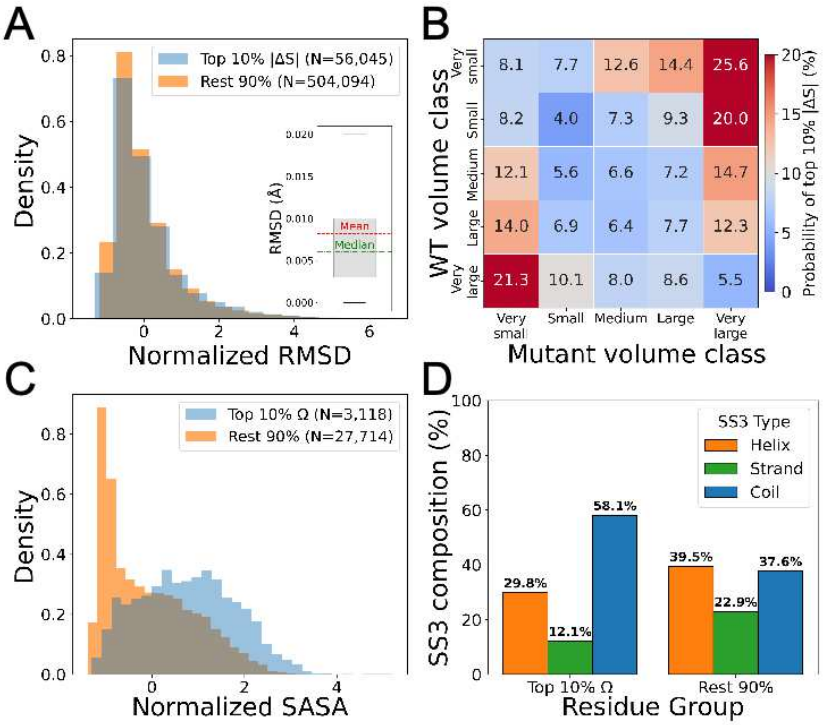


Fig. 5. Statistical analysis for the substitution-level |ΔS| score (A and B) and the position-level Ω score (C and D). (A) Distribution of RMSD values of mutant models relative to the corresponding wild-type (WT) structures, grouped into the top 10% |ΔS| mutations and the remaining 90%. RMSD values were Z-score-normalized per structure. The inset shows the raw RMSD distribution across all mutations. **(B)** Probability of belonging to the top 10% |ΔS| group as a function of WT and mutant side-chain volume classes. The color scale is centered at the global expectation of 10%, such that warmer colors indicate volume-class transitions that generate high-|ΔS| scores more frequently than expected. **(C)** Distribution of position-level solvent-accessible surface area (SASA) for residues ranked within the top 10% of Ω versus the remaining 90%. SASA values were Z-score-normalized per structure. **(D)** Secondary-structure composition of residues in the top 10% Ω group compared with the remaining 90%, based on a

three-state (SS3) classification: helix, strand, and coil. Bars represent the fraction of residues within each group.

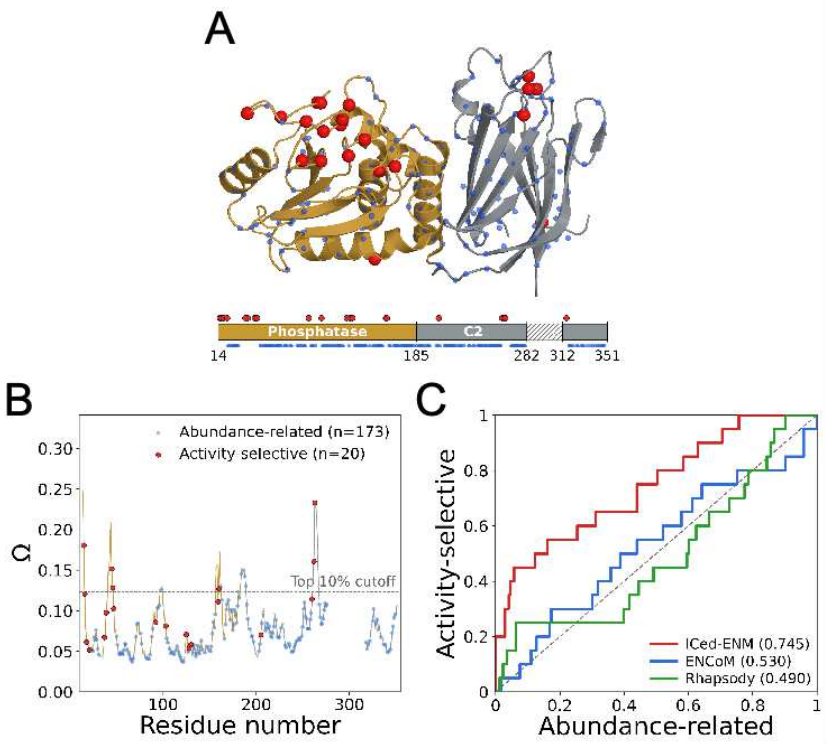


Fig. 6. PTEN loss-phenotype benchmark of position-level scores. (A) Structural domain organization and distribution of phenotype-associated residues in the apo open PTEN structure (1D5R:A). The phosphatase and C2 domains are colored orange and gray, respectively. Residues associated with activity-selective loss (reduced phosphatase activity with wild-type-like abundance) are shown as red spheres; residues associated with abundance-related loss (reduced abundance regardless of activity) are shown as blue spheres. Their positions along the PTEN sequence are shown below the structure. **(B)** Position-level mutation-impact score (Ω) profile obtained from ICed-ENM-based mutation-impact analysis of 1D5R:A. Domain regions and phenotype-associated positions are colored as in panel A. The unresolved segment (residues 282–312) and five flanking positions on each side (277–281 and 313–317) were excluded from scoring to reduce boundary effects. The dashed horizontal line indicates the cutoff defining the top 10% Ω group. **(C)** Receiver operating characteristic (ROC) curves distinguishing activity-

selective loss positions (positive class; n = 20) from abundance-related loss positions (negative class; n = 173). The ICed-ENM curve uses Ω derived from ICed-ENM-based mutation-impact analysis. The ENCoM curve uses the corresponding Ω calculated from ENCoM-derived |ΔS| scores with the same position-level aggregation and spatial smoothing procedure. The Rhapsody curve uses the native Rhapsody-2(dyn) score directly at the position level. The y-axis indicates the proportion of activity-selective positions recovered, whereas the x-axis indicates the proportion of abundance-related positions selected as the score threshold is progressively lowered. The dashed diagonal line represents random classification. The area under the ROC curve (AUROC) is 0.745 for ICed-ENM, 0.530 for ENCoM, and 0.490 for Rhapsody-2(dyn).

Table 1. Overlap analysis between computed normal modes and experimental transition vectors.

Metric	ICed-ENM			iMOD			edENM			ANM		
	A*	O	C	A	O	C	A	O	C	A	O	C
O_1†	**0.54**	**0.64**	**0.44**	0.52	**0.64**	0.40	0.50	0.61	0.39	0.44	0.55	0.33
O_2	0.29	0.28	0.30	0.31	0.31	0.31	0.31	0.32	0.30	**0.34**	**0.37**	**0.32**
O_3	**0.27**	**0.21**	**0.33**	0.26	0.20	0.32	0.25	0.18	0.31	0.26	0.19	**0.33**
O_{max}‡	**0.68**	**0.72**	**0.63**	0.67	0.71	0.62	0.67	**0.72**	**0.63**	0.64	0.70	0.57
$N(O_{max})$§	**1.88**	1.63	**2.12**	1.98	**1.54**	2.41	1.96	1.59	2.34	2.32	1.80	2.83
CSO_3‖	**0.60**	0.66	**0.54**	0.59	**0.67**	0.51	0.59	0.66	0.51	0.55	0.64	0.47
CSO_5	**0.69**	**0.75**	**0.64**	0.68	0.73	0.62	0.68	0.74	0.62	0.64	0.72	0.57
CSO_{10}	**0.77**	**0.81**	**0.74**	0.76	**0.81**	0.71	**0.77**	**0.81**	0.72	0.74	0.79	0.69

Values represent averages across structures within each group. Bold indicates the highest value among methods for overlap and CSO metrics, and the lowest value for $N(O_{max})$.

* 'A' indicates all analyzed PDB structures; 'O' and 'C' represent the open and closed conformations, respectively.

† O_n denotes the overlap value between the n^{th} computed normal mode and the experimental transition vector.

‡ O_{max} represents the highest overlap value among all computed modes.

§ $N(O_{max})$ indicates the mode number corresponding to O_{max}.

‖ CSO_n refers to the CSO score for the lowest n normal modes.

Table 2. Association of benchmark scores with PTEN loss phenotypes at substitution and position levels.

Level	Method	Activity-selective, n/N (%)	Abundance-related, n/N (%)	Odds ratio*	AUROC†
Substitution	ICed-ENM	13/100 (13.0%)	79/856 (9.2%)	1.47	0.510
	ENCoM	26/100 (26.0%)	141/856 (16.5%)	1.78	0.567
	DynaMut2	6/100 (6.0%)	191/856 (22.3%)	0.22	0.375
Position	ICed-ENM	6/20 (30.0%)	6/173 (3.5%)	11.93	0.745
	ENCoM	3/20 (15.0%)	21/173 (12.1%)	1.28	0.530
	Rhapsody-2(dyn)	5/20 (25.0%)	14/173 (8.1%)	3.79	0.490

For each method, the top 10% score group comprised the 565 highest-ranked substitutions among 5,643 non-WT PTEN substitutions or the 30 highest-ranked positions among 297 PTEN positions. Entries show the number and percentage of substitutions or positions from each phenotype group (activity-selective loss or abundance-related loss) included in the top 10% score group. Substitutions were ranked by $|\Delta S|$ for ICed-ENM and ENCoM. For DynaMut2, more negative predicted stability changes ($\Delta\Delta G$), indicating greater predicted destabilization, were ranked higher. Positions were ranked by Ω for ICed-ENM and ENCoM and by the native Rhapsody-2(dyn) score.

* Odds ratios (ORs) compare inclusion in the top 10% score group for activity-selective loss versus abundance-related loss; values above 1 favor activity-selective loss, whereas values below 1 favor abundance-related loss.

† Area under the receiver operating characteristic curve (AUROC) uses the full score rankings of phenotype-annotated substitutions or positions, with activity-selective loss as the positive class and abundance-related loss as the negative class. Values below 0.5 indicate that substitutions or

positions in the abundance-related loss group tend to rank above those in the activity-selective loss group.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryinformationNatCommunapeal.docx