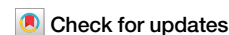


<https://doi.org/10.1038/s42004-025-01763-0>

Challenging AlphaFold in predicting proteins with large-scale allosteric transitions



Brooks H. Perkins-Jechow¹, Juan Pablo Iglesias Ahualli¹, Huyen Thuc Nhu¹, Alireza Omid¹, Chunchao Li¹, Jorge A. Holguin-Cruz¹, Daeahn Cho², Dokyun Na², Nawar Malhis¹, Jennifer M. Bui¹ & Jörg Gsponer^{1,3} ✉

Many proteins function by toggling between distinct conformations, yet most structure predictors have been trained on data that do not capture this conformational diversity. Here, we benchmarked AlphaFold2, AlphaFold3, and recent variants on autoinhibited proteins, a class of allosterically regulated, often multi-domain proteins that exist in equilibrium between active and autoinhibited states. Our analyses show that AlphaFold2 fails to reproduce the experimental structures of many autoinhibited proteins, which is reflected in reduced confidence scores. This contrasts sharply with its high-accuracy, high-confidence predictions of non-autoinhibited multi-domain proteins. When tested for its ability to capture conformational diversity, we found that AlphaFold2 performs better when combined with uniform subsampling of sequence alignments rather than local subsampling. BioEmu and AlphaFold3 improve upon these results, yet still struggle to accurately reproduce details of experimental structures. Together, our study underscores the persistent challenges of predicting protein structures shaped by complex energy landscapes.

Predicting a protein's tertiary structure from its amino acid sequence is a major challenge in biology. The development of AlphaFold2¹, a deep learning model capable of near-experimental precision, marked a breakthrough. Its release spurred the creation of similar tools and enhancements that improved prediction accuracy and scope^{2–6}. However, a significant challenge for all of these tools is the fact that proteins are inherently dynamic and that a single model never captures the structural diversity of any protein state^{7,8}. Furthermore, many protein sequences encode multiple functional states that have distinct structural characteristics, which makes structure predictions for these proteins particularly challenging.

Several AlphaFold-based approaches have been proposed as solutions to this problem. Initial studies have manipulated the evolutionary information provided to AlphaFold either through subsampling of the multiple sequence alignment (MSA) or rational in-silico mutagenesis^{9–11}. Although these approaches generated structures resembling the different experimentally determined conformations, their generalizability remained uncertain due to small validation sets. To address this caveat, a critical study tested these approaches on a set of 92 fold-switching proteins¹², previously identified based on having multiple PDB entries exhibiting distinct secondary structures¹³. Using AF-Cluster, SPEACH-AF, and iterative

AlphaFold runs, the authors found that accurate prediction of alternative conformations was achieved for only a subset of fold-switching proteins. In parallel, more elaborate methods have been developed with the explicit goal of predicting conformational diversity. In one such approach, CFold, AlphaFold was retrained on a conformational split of the PDB, using only one conformation per protein for training and reserving the others for testing, which demonstrated success in accurately predicting alternative structures for some proteins¹⁴. In another approach, BioEmu, a deep-learning biomolecular emulator trained on large-scale molecular dynamics simulations, AlphaFold structures, and stability data, was designed to generate diverse conformations during inference¹⁵. BioEmu shows promising results, including for systems that undergo large-scale conformational rearrangements. These recent findings underscore the potential and limitations of recently developed methods in predicting alternative conformations.

Here, we present a detailed evaluation of AlphaFold's and related structure prediction tools' performance on a dataset of 128 autoinhibited proteins, i.e., proteins that adopt at least two functionally distinct conformations. Autoinhibition is a widespread allosteric regulatory mechanism in signaling proteins that prevents spurious activity^{16–20}. Dysregulation of

¹Michael Smith Laboratories, University of British Columbia, Vancouver, BC, Canada. ²School of Integrative Engineering, Chung-Ang University, Dongjak-gu, Seoul, Republic of Korea. ³Department of Biochemistry and Molecular Biology, University of British Columbia, Vancouver, BC, Canada.

✉ e-mail: gsponer@msl.ubc.ca

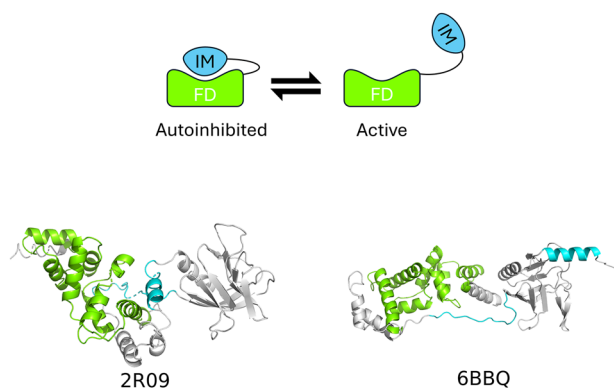


Fig. 1 | Simplest model of autoinhibition. This model involves intramolecular interactions between an inhibitory module (IM) and a functional domain (FD). This concept is illustrated with structures of the mouse Cytohesin-3. In the autoinhibited form (left; PDB ID 2R09⁷²), the Sec7-PH linker and C-terminal helix (cyan) occlude the catalytic site of the Sec7 domain (light green). In the active form (right; PDB ID 6BBQ⁷³), autoinhibition is relieved by binding of the activating partner Arf-GTP (not pictured) to the PH domain.

autoinhibition has been implicated in several diseases, including cancer^{21–24}. In its simplest form, autoinhibition arises from transient interactions between a functional domain (FD) and an inhibitory module (IM) (Fig. 1), placing the protein in equilibrium between an open, active state and a closed, inactive state. However, many autoinhibited proteins follow more complex regulatory schemes with multiple partially active states²⁴. Autoinhibited proteins are often multi-domain with activity cycles that typically involve large rearrangements in domain positioning, frequently without major secondary structure changes. Nonetheless, secondary structure changes can occur, particularly when the IM is intrinsically disordered²⁰. Intramolecular interactions suppress FD activity either sterically or allosterically, most often by blocking partner binding or catalysis. Many autoinhibited proteins default to the inactive conformation^{25–27}. External signals, such as post-translational modifications (PTMs) or binding to partner molecules, then trigger conformational transitions that shift the protein to an active form, either a single open state in the simplest case or multiple active states in more complex scenarios.

The modern view of protein allostery suggests that all possible states are embedded within a protein's energy landscape as multiple significant minima, each with distinct statistical weights^{17,28,29}. This means that a protein's sequence inherently encodes all relevant conformations, including active and inhibited states in autoinhibited proteins³⁰. Because many autoinhibited proteins are multi-domain and undergo large inter-domain rearrangements when transitioning between active and inactive states, they provide an additional benchmark for evaluating the structure prediction capabilities of AlphaFold and related tools.

Results

AlphaFold predicts structures of autoinhibited proteins less accurately than those of proteins with permanent inter-domain contacts

We first evaluated how well AlphaFold2 (AF2) and AlphaFold3 (AF3) reproduce experimentally determined structures of autoinhibited proteins. For this purpose, we assembled a dataset of 128 autoinhibited proteins from a recently curated database³¹, which includes proteins in which autoinhibition has been experimentally demonstrated through deletion-construct assays. Our dataset is restricted to entries with high-quality experimental structures available in the Protein Data Bank (PDB) (see Methods). Comparison with two earlier benchmarks of conformationally dynamic proteins^{12,14} revealed only minimal overlap, three proteins in common with the Chakravarty set and 12 with the Bryant set, highlighting the complementary nature of our dataset. As a control for our assessment,

we collected a set of 40 proteins with available PDB structures that have two UniProt-annotated domains and no available information indicating they are regulated via autoinhibition (“two-domain” data set). Seven of these two-domain proteins have permanent intramolecular domain-domain interactions (obligate subset). This control was chosen because autoinhibited proteins, by definition, contain at least a FD and an IM, which are often independently folded domains that change their relative positioning when transitioning between active and autoinhibited states. In contrast, two-domain proteins with permanent (obligate) domain-domain interactions typically require these stable interactions for proper function and therefore do not undergo regulatory rearrangements as autoinhibited proteins do. For 71 and 29 of the autoinhibited and two-domain proteins, respectively, at least two PDB files are available. Using proteins with multiple PDBs, we confirmed that proteins in the autoinhibited data set are structurally more heterogeneous than those in the two-domain data set (Fig. S1).

For AF2, structure predictions were retrieved from the public AlphaFold Database, while AF3 predictions were generated with the AlphaFold Server using full-length protein sequences. As an initial measure of prediction accuracy we used root mean square deviations (RMSDs) calculated after aligning different sequence segments: the “global”, full available coordinate region (gRMSD); the functional domain ($_{fd}$ RMSD); and the inhibitory module ($_{im}$ RMSD). In addition, we computed the RMSD of the IMs when the structures were aligned on the FDs ($_{fd}^{im}$ RMSDs), providing insights into the correct positioning of inhibitory modules relative to the functional domain (see Methods). For the control set of two-domain proteins, we performed similar RMSD calculations, aligning the first (D1) and second (D2) domain in the sequences. For proteins with multiple PDB entries, we selected the structure pair with the lowest gRMSD, capturing the best overall agreement between the prediction and available experimental structures for each protein. The distributions of gRMSDs indicate that AlphaFold predicts structures of two-domain proteins significantly closer to experimentally determined structures than autoinhibited proteins (Fig. 2a). Using a 3 Å cutoff, slightly more than half of the autoinhibited protein predictions by AF2 match an experimental structure, compared to nearly 80% for two-domain proteins. This discrepancy does not stem from poor individual domain predictions. Although RMSD distributions for individual domains differ between autoinhibited and two-domain proteins, more than 75% of proteins in both data sets have domain RMSDs smaller than 3 Å. Instead, the key difference lies in domain positioning, specifically, the placement of inhibitory modules relative to functional domains. $_{fd}^{im}$ RMSDs are significantly higher than $_{d2}^{d1}$ RMSDs, with about half of the predicted IMs misaligned relative to experimental structures (using a 3 Å cutoff). We tested whether AF3 predictions for full-length sequences of the autoinhibited proteins are more accurate (Fig. 2a), specifically with respect to the placement of inhibitory modules relative to functional domains. Although prediction accuracies of autoinhibited proteins, as assessed by RMSDs, are marginally better for AF3 than AF2, they are not statistically better.

These differences between predictions of autoinhibited and two-domain proteins become even more pronounced when focusing on two-domain proteins with experimentally confirmed permanent domain interactions (obligate subset) (Fig. S2a). AF2 accurately predicts all proteins in this subset, including their relative domain placement. Given that inhibitory modules (IMs) are sometimes disordered²⁰ and AF2/3 are primarily trained on folded domains, we examined whether the prediction discrepancies between the data sets are due to IM disorder. However, excluding IMs with more than 50% predicted disorder does not alter observed trends (Fig. S2c). Since most structures in the autoinhibited protein set predate AF2 and AF3's release, it is likely that at least some of these proteins were included in training. It is possible that AlphaFold has memorized proteins known to adopt multiple conformations, while treating autoinhibited proteins with a single conformation in the PDB as simpler prediction cases. To test this, we examined whether autoinhibited proteins with only one experimentally determined conformation (see Methods) are more accurately predicted than those with two or more known conformations (Fig. S3a). However, RMSD distributions show no significant differences

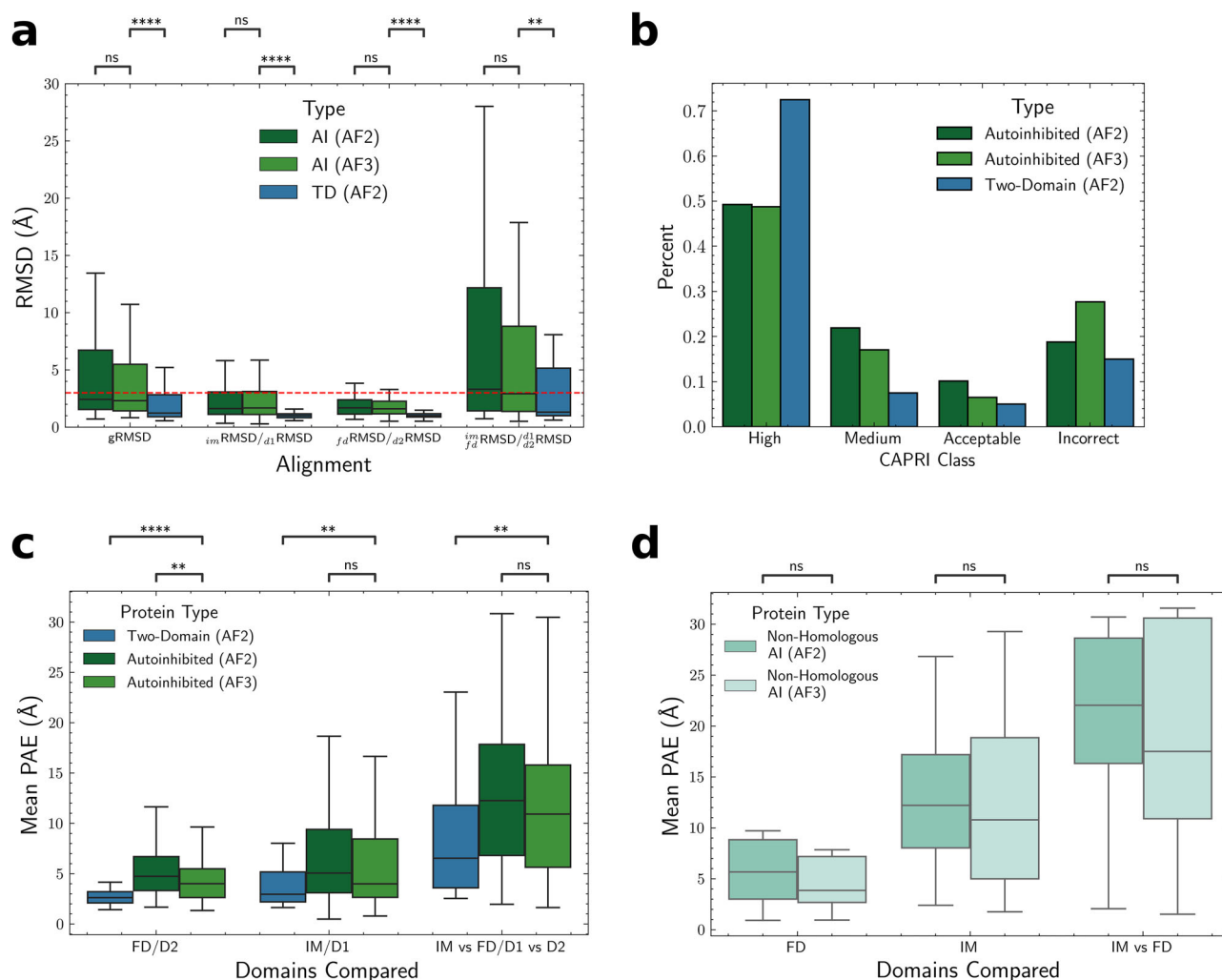


Fig. 2 | Predictions of autoinhibited proteins deviate more from experimental data than predictions of non-autoinhibited proteins. **a** Distributions of the best (lowest) RMSD between predictions and experimental structures, broken down by region and model. AI, autoinhibited; TD, two-domain. $N = 128$ for AF2 and AF3 autoinhibited protein predictions, $N = 40$ for AF2 two-domain structure predictions. The red line indicates a 3 Å cutoff. **b** Percentage of proteins in each CAPRI

class based on the best prediction/experimental structure fit. **c** Distributions of mean PAEs per and across domains. **d** Mean PAEs per and across domains compared for AF2 and AF3 predictions of autoinhibited proteins with no homologous templates in the PDB. $N = 32$ proteins. Statistical differences were assessed using the Mann-Whitney test (****: p -value ≤ 0.0001 , ***: p -value ≤ 0.001 , **: p -value ≤ 0.01 , and *: p -value ≤ 0.05).

between these subgroups. Some of the autoinhibited proteins in our dataset contain PTMs or are bound to interaction partners in the available experimental structures. These factors may stabilize conformations that differ from those of the unmodified protein in isolation, potentially complicating AlphaFold's prediction task. When splitting autoinhibited proteins into two groups, with and without partners or PTMs, we find no statistically significant difference in prediction performance for AF2 as measured by RMSD (Fig. S3b). AF3 does significantly better for autoinhibited proteins without partners or PTMs, suggesting that additional information is needed and may help in matching some experimental structures (see below).

RMSD is a common metric for comparing protein structures but is alignment-dependent, which can be problematic for multi-domain proteins. Small alignment differences in one domain can cause large RMSD variations in another. To mitigate this, we also used DockQ³², a metric traditionally applied to protein complex predictions. DockQ integrates local quality measures, including native interface contacts and local RMSDs³³. While it typically evaluates protein docking, it can assess multi-domain protein predictions by treating one domain (e.g., the IM) as “docked” to another (e.g., the FD) intramolecularly. Based on the calculated DockQ scores, predicted structures can then be placed in critical assessment of

prediction of interactions (CAPRI) classes that identify predictions as inaccurate or accurate, with the latter further subdivided into high, medium, or acceptable quality. Applying this prediction classification to autoinhibited and two-domain proteins with inter-domain contacts (Fig. 2b) revealed that two-domain protein predictions are more often of high quality. The difference in CAPRI class distributions is even more pronounced between proteins in the autoinhibited and obligate data sets (Fig. S2b). Recent assessments of AlphaFold and AlphaFold-Multimer's ability to predict binary protein complexes demonstrated high accuracy, provided the exact interaction sequences are used³³. However, prediction accuracy declines when full-length protein sequences are paired. Therefore, we tested AF2 and AF3 using trimmed protein sequences, specifically the shortest continuous segments containing both the FD and IM or D1 and D2. This approach is justified, as autoinhibited proteins are often large and multi-domain, making prediction more complex. Focusing on relevant sequence regions could potentially improve accuracy. However, trimming sequences leads to only marginal improvements (Fig. S3c, d). Overall, our analyses show that AF2 and AF3 predictions for autoinhibited proteins are less consistent with experimentally determined structures compared to non-autoinhibited proteins, especially those with permanent inter-domain contacts.

AlphaFold has limited confidence in predicting structures of autoinhibited proteins

We evaluated whether AlphaFold's confidence scores reflected the discrepancies between autoinhibited and two-domain predictions. AlphaFold provides two key self-confidence measures: the Predicted Local Distance Difference Test (pLDDT) and the Predicted Aligned Error (PAE). pLDDT evaluates the confidence in local structure predictions at the residue level, while PAE assesses the reliability of the predicted relative positions between pairs of residues. PAE serves as a strong measure of the relative placement of domains³⁴. pLDDT scores are significantly lower for autoinhibited proteins, both at the whole-protein and individual domain levels (Fig. S4a, b), even when considering the subset with only structured inhibitory modules (Fig. S4c, d). PAE distributions are generally shifted toward lower values for two-domain proteins, regardless of residue pair location (Fig. 2c). However, median PAE values remain around 5 for intra-domain residue pairs in both datasets, suggesting reliable domain predictions. Critically, PAE values of AF2 and AF3 for inter-domain residue pairs, specifically between the inhibitory module and functional domain in autoinhibited proteins, indicate significantly lower confidence in domain positioning, which aligns with the observed higher $^{im}_{fd}$ RMSDs. This effect persists when trimming the sequences (Fig. S5a) or excluding disordered IMs (Fig. S5b), reinforcing that AlphaFold struggles to orient the domains similarly to the experimental structures.

Our results suggest that AlphaFold may “recognize” the complexity of the energy landscape of autoinhibited proteins and thus the prediction task, particularly as it pertains to the positioning of inhibitory modules relative to functional domains. If true, this should also be reflected in the confidence scores of predictions for autoinhibited proteins without available structures. The curated database of autoinhibited proteins³¹ contains many cases where IM and FD boundaries are mapped but no experimental structures exist. We selected the subset of these proteins that had no experimental structures, even for homologs, and collected their AF2 and AF3 predictions as described above (see Methods). PAE scores of AF2 predictions for these autoinhibited proteins without homologous structures are slightly higher than for those with known structures (Fig. 2c, d), with IM-FD residue pairs showing particularly high PAEs, mirroring trends seen in autoinhibited proteins with available structures. Predictions with AF3 show marginally improved but not significantly different PAE distributions (Fig. 2d). These findings hold even when excluding proteins with highly disordered IMs (Fig. S5c) and when trimming the sequences (Fig. S5d), reinforcing that AlphaFold has consistently limited confidence in predicting autoinhibited protein structures.

Alternative conformations are predicted only in few of the studied cases with high accuracy

Several strategies have been proposed to predict alternative protein conformations using AlphaFold. These efforts have predominantly focused on structural changes within domains and secondary structure changes, although some of the systems previously studied also undergo substantial tertiary rearrangements^{12,14}. Autoinhibited proteins, which frequently experience complex interdomain rearrangements, provide a complementary benchmark for evaluating the effectiveness of such predictive approaches. To assess the performance of recently developed tools and strategies for predicting conformational diversity, we selected a subset of 15 autoinhibited proteins from our dataset for which structural cluster analysis of available experimental structures confirmed the existence of at least two distinct conformations (see Methods). These distinct conformations correspond to the active and inactive states of the proteins in most cases. For the 15 autoinhibited proteins analyzed, 9 of the AF2 predictions using full-depth MSAs are closest, based on gRMSD, to the autoinhibited state, while 6 are closest to the active state. We refer to the conformation that the full-depth MSA prediction most closely resembles (autoinhibited or active) as the base conformation, and we define the conformation representing the alternative functional state, the one less closely matched by the AF2 predictions using full-depth MSAs, as the alternate conformation.

We first evaluated the ability of the AF-Cluster approach to generate structures resembling both the base and alternate conformations of the chosen proteins. AF-Cluster produces two sets of subsampled multiple sequence alignments (MSAs): locally clustered (LC) MSAs, which are clustered by the algorithm DBScan³⁵ based on sequence similarity; and uniformly clustered (UC) MSAs, which consist of 10 or 100 sequences evenly distributed across the full-depth MSA and are intended as control sets. We used the subsampled MSAs along with ColabFold⁶ to generate predictions and evaluated their accuracy with the RMSD measures described previously. From the set of predictions generated using UC and LC MSAs, we selected those with the lowest gRMSD values, representing the best fits between predicted and experimental structures (Fig. 3a, b). Both approaches yield structurally different predictions; however, only a limited number of these match the base or alternate conformations at stringent cutoffs of 2 or 3 Å. Similarly, in only a few cases are both the base and alternate conformations accurately predicted. Given that correct IM placement is critical for accurately modeling the systems under investigation, we repeated our prediction assessment using $^{im}_{fd}$ RMSDs instead of gRMSD as the evaluation metric, and observed slightly worse results (Fig. S6a, b).

Although the predicted structures are not always closely aligned with the experimentally determined ones, visual inspection suggests that the overall topologies characteristic of different functional states are often accurately reproduced for many of the 15 autoinhibited proteins. A notable example is the tyrosine-protein phosphatase non-receptor type 6 (PTPN6). For this protein, predictions generated through MSA subsampling successfully capture the relative positioning of the IM and FD (Fig. 4a–c). However, the gRMSD of the closest predicted conformation to the experimentally determined autoinhibited state remains relatively high at 6.7 Å. One contributing factor is the linker domain between the IM and FD: while it is correctly structured in the full-depth MSA prediction representing the active state (Fig. 4a), it is misfolded in the MSA subsampling prediction that most closely resembles the autoinhibited conformation (Fig. 4c). Given these insights, a more lenient cutoff may provide a more informative measure of prediction performance (Fig. 3a, b). Uniform MSA subsampling yields predictions within 10 Å of both conformations for about half of the autoinhibited proteins. When focusing specifically on the inhibitory module and its positioning relative to the functional domain ($^{im}_{fd}$ RMSDs, Fig. S6a, b), results are, again, worse. Next, we used CFold to predict alternative conformations. CFold is a retrained version of AlphaFold2, developed using a conformationally split version of the PDB, that is, only one conformation per protein was included in the training set. Among the 15 autoinhibited proteins tested, none yielded predicted structures with a gRMSD or $^{im}_{fd}$ RMSDs below 3 Å, and only a few systems have predictions below the 10 Å threshold (Fig. 3d and Fig. S6d). Finally, we evaluated BioEmu, an MD-informed deep learning model¹⁵ trained using more than 200 milliseconds of MD simulations, AF-predicted structures, and experimental protein stability data. Overall, BioEmu achieved better prediction performance than MSA uniform subsampling (Fig. 3e and Fig. S6e). However, at a gRMSD threshold of 5 Å, both the base and alternate conformations were accurately predicted for only two autoinhibited proteins. Notably, this number increased to 11 when a more permissive cutoff of 10 Å was applied. The trends in prediction performance across the different approaches are further supported by scatter plots showing the best gRMSD with respect to base and alternate conformations for each system (Fig. S7). Predictions generated with AF2 using uniform subsampling and with BioEmu perform substantially better than those from local subsampling and CFold. However, the differences between prediction methods are much less pronounced when $^{im}_{fd}$ RMSDs scatter plots are considered (Fig. S8), consistent with analyses using cutoffs for this metric. These results underscore that all methods struggle to correctly position FDs and IMs relative to each other in a manner consistent with available experimental structures.

As mentioned above, several of the autoinhibited proteins include PTMs or are bound to interaction partners in their experimentally resolved structures. This is especially true for systems where both active and autoinhibited conformations have been characterized, likely because PTMs or

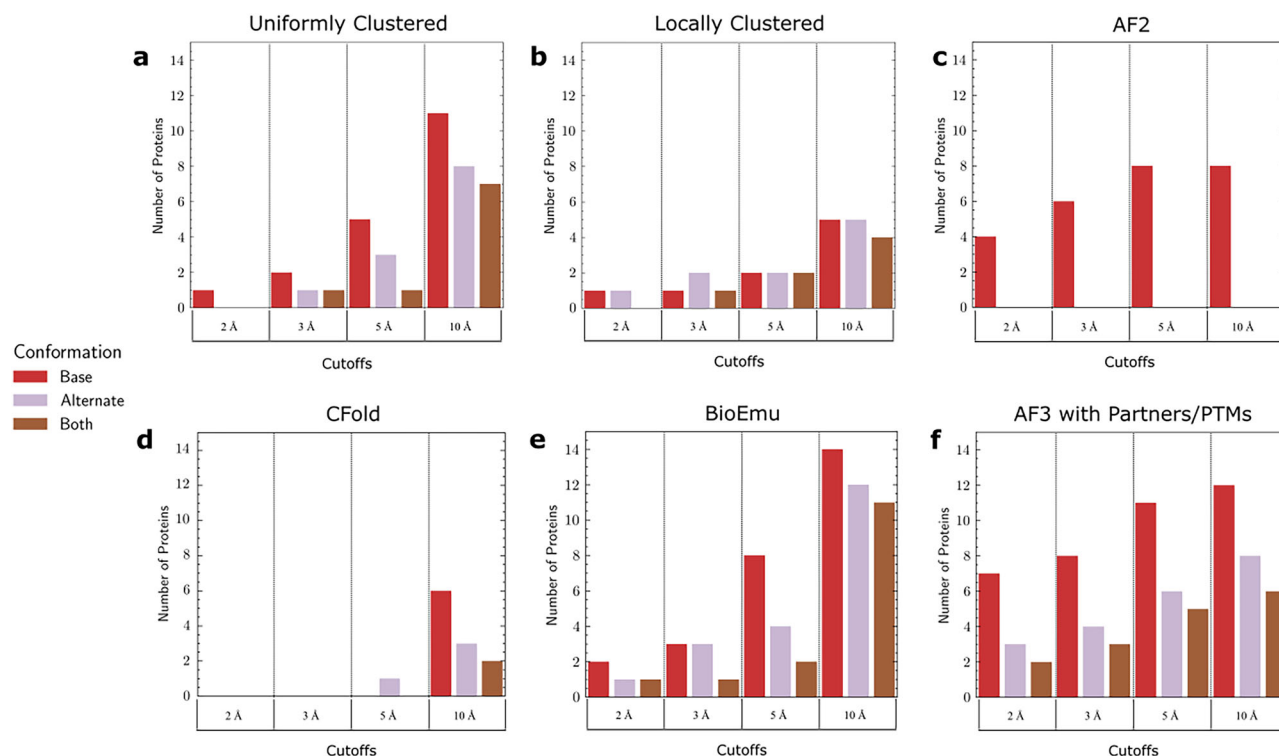
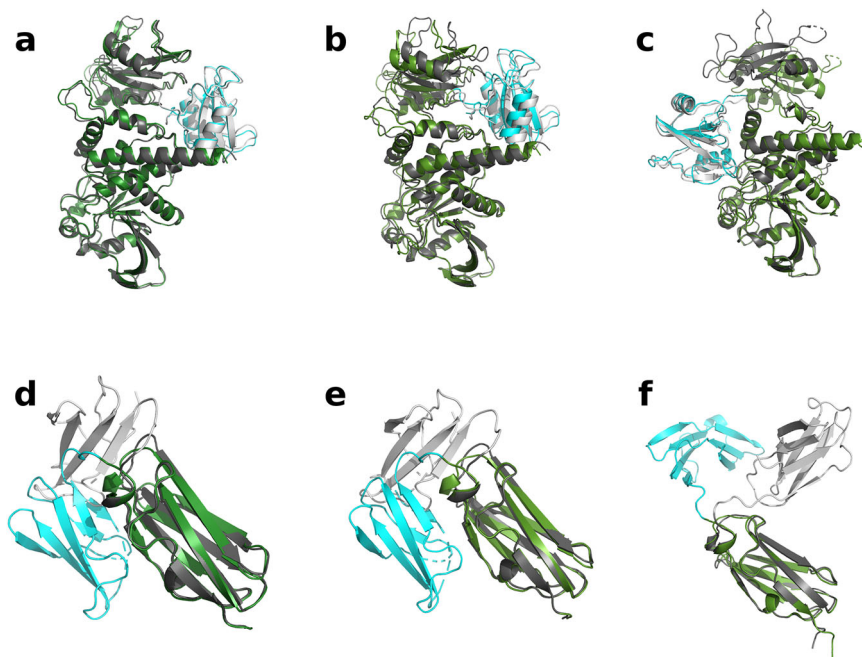


Fig. 3 | Limited accuracies in predicting conformational diversity. Number of predictions that match the base, alternate, or both conformations at the given cutoffs based on gRMSD for **a** uniformly clustered predictions; **b** locally clustered predictions; **c** AF2 predictions; **d** CFold predictions; **e** BioEmu predictions; **f** AF3 predictions with partners/PTMs. N = 15 proteins.

Fig. 4 | Predictions of tyrosine-protein phosphatase non-receptor type 6 (PTPN6) and Filamin-A (FLNA). The IMs of the experimental structures are shown in light grey and the rest of the proteins is dark grey. The predicted IMs are cyan and the rest of the predicted proteins is shown in shades of green for AlphaFold/cluster predictions. **a** PTPN6 full-depth prediction against the active state (PDB 3PS5⁷⁴). **b** PTPN6 uniformly clustered prediction (U100-007) against the active state. **c** PTPN6 uniformly clustered prediction (U100-002) against the autoinhibited state (PDB 2B3O⁷⁵). **d** Full-depth prediction compared to the autoinhibited state of Filamin-A (PDB 2J3S⁷⁶). **e** Uniformly clustered prediction (U100-008) against the autoinhibited state. **f** Uniformly clustered prediction (U10-007) against the active state (PDB 4P3W⁷⁷).



partner binding were necessary to stabilize a particular state for structural determination. AF3 has been trained to incorporate the effects of certain PTMs and binding partners on structure prediction. Since AF3 can incorporate such prior knowledge, we tested whether this feature improves agreement with experimental structures. Indeed, AF3 predictions that leveraged prior information outperformed AF2 full-depth MSA predictions in modeling base conformations across all defined RMSD ranges (Fig. 3c, f).

Notably, AF3 achieved superior accuracy at stringent cutoffs of 2 and 3 Å compared to all other tested methods. For nearly half of the 15 autoinhibited proteins, AF3 generated conformations within 10 Å of both the base and alternate experimental structures. Furthermore, AF3 also outperformed other approaches in accurately capturing the relative placement of FDs and IMs (Figs. S6f, S8). Together, these findings highlight that incorporating prior biological knowledge, such as specific regulatory PTMs or interaction

partners, can improve the prediction of alternative conformations in autoinhibited proteins. Nonetheless, a substantial number of systems remain difficult to model. For example, in the case of human Filamin-A (Fig. 4d–f), none of the tested methods produced structures with gRMSDs below 10 Å for either conformation, underscoring the continuing challenges in accurately predicting large-scale allosteric rearrangements.

Predictions based on uniformly clustered MSAs are more consistent with experimental structures

One of the unexpected findings of our analysis was that predictions based on UC MSAs are more consistent with experimental structures than those generated using LC MSAs (Fig. 3a, b; Fig. S6a, b). UC MSA-based predictions were initially intended to serve as controls, whereas LC MSAs have previously been shown to produce alternative conformations with high AlphaFold confidence scores¹⁰. To better understand this result, we examined specific autoinhibited systems in greater detail, mapping all predicted structures, derived from both LC and UC MSAs, onto the first two principal components capturing the dominant structural transitions between the experimentally determined active and autoinhibited conformations (see Methods).

Figure 5a, b show the projections of structures predicted for the human GTP-binding nuclear protein Ran using UC and LC MSAs, respectively. Ran is a GTPase that is autoinhibited by a C-terminal helix binding to and reducing the activity of the GTPase domain (the experimental structure of the autoinhibited state, PDB: 3GJ3, is shown in black on the left of Fig. 5a, b). Partner binding shifts the equilibrium toward an active conformation in which the inhibitory helix is removed from the GTPase domain (the active state, PDB: 4HAT, is shown in black on the right of Fig. 5a, b). Using LC MSAs yields a large number of predicted conformations (Fig. 5b), with some (e.g., LC_061) closely resembling the active state, where the inhibitory helix no longer interacts with the GTPase domain. However, other LC MSA-based predictions (e.g., LC_103) show the inhibitory helix interacting with the GTPase domain at a non-canonical site. Notably, the majority of these structures exhibit high averaged inter-domain PAEs (IM-FD), indicating high uncertainty regarding the placement of the inhibitory helix relative to the GTPase domain. Interestingly, even the default AF2 prediction using the full-depth MSA yields a structure with high inter-domain PAE that resembles the active conformation (Fig. 5a), despite the active experimental structure requiring an activating partner not included as input (Fig. 5b, 4HAT structure with activating partner in yellow). In contrast, predictions with low average inter-domain PAEs consistently reproduce the autoinhibited conformation, and multiple predictions emerging from UC MSAs show both low inter-domain PAEs and accurate modeling of the autoinhibited state (Fig. 5a). As a second system we chose cystathionine beta-synthase and found very similar results in the same type of analysis (Fig. S9).

Overall, these analyses suggest that predictions made with either uniformly or locally clustered MSAs can generate structures close to both the active and autoinhibited conformations. However, uniformly clustered MSAs tend to produce structures with higher confidence in the placement of the IM and, therefore, may be more likely to be accurate. To test this further, we compared the distributions of mean pLDDT and PAE scores for predictions made with UC and LC MSAs (Fig. 6a, b). These distributions clearly show that UC MSA-based predictions have significantly higher mean pLDDT scores for the entire protein as well as for the IM and FD individually, and, most notably, significantly lower mean intra- and inter-domain PAEs.

A key question, however, is whether these differences in prediction confidence translate into consistently better agreement with experimental structures. To investigate this, we compared the gRMSD distributions for the “best” predictions generated with UC and LC MSAs, where “best” was alternately defined by the lowest RMSD to the experimental structure, highest mean pLDDT, or lowest mean inter-domain PAE (Fig. 6c, d). This analysis revealed that predictions from UC MSAs selected based on PAE consistently more closely matched the experimentally determined structures than those from LC MSAs. Although the difference was not statistically significant, likely due to the limited sample size, the overall trend was clear (Fig. 6d). To strengthen the analysis, we repeated it by considering not only

the single best but also the top two or five predicted structures ranked by PAE, and in this case the differences reached statistical significance (Fig. S10). The detailed prediction analyses shown in Fig. 5 and Fig. S9 suggest that PAE-based selection tends to favor more closed conformations, those in which the IM and FD are in closer contact. Indeed, in 11 out of 15 proteins with experimentally determined active and autoinhibited states, PAE-based selection leads to more compact models. Overall, these findings suggest UC MSAs yield higher-confidence predictions and better capture conformations that align with experimental data than LC MSAs. Moreover, PAE-based selection preferentially identifies closed conformation in the set of proteins analyzed here.

Variability in sequence and evolutionary information is associated with better prediction outcomes for autoinhibited proteins

Next, we investigated the extent to which sequence and evolutionary diversity in the clustered MSAs correlate with improved prediction fidelity to experimental data. Specifically, we quantified the information content provided to AlphaFold2 by the clustered MSAs in two ways: first, by measuring sequence variability through Shannon entropy; and second, by assessing taxonomic diversity using a metric we termed the Lineage Score Δ (see below and Methods). We then compared the distributions of these measures between predictions that closely matched experimental structures (within 3 Å) and those that did not. As expected, uniform clusters exhibit higher average Shannon entropy than local clusters, consistent with the fact that DBScan clusters based on sequence similarity (Fig. S11). Roughly a third of autoinhibited UC clusters display average Shannon entropy values above 2, indicative of high sequence variability. In contrast, only a negligible fraction of LC clusters reach this level of variability. The same trend is observed in two-domain proteins. When autoinhibited protein groups were stratified based on the accuracy of their predictions (using a 3 Å cutoff relative to experimental structures), no significant difference in the average Shannon entropy across the full-length sequences is observed. However, when considering only the sequence regions corresponding to the IM and FD, the critical regions for autoinhibition, we found that accurate predictions using UC are associated with statistically significantly higher sequence variability than inaccurate predictions (Fig. 7).

As an additional metric, we assessed the taxonomic diversity of the sequences within the MSAs by calculating a lineage score for each sequence. This score represents the number of nodes within a species' taxonomic lineage, effectively, the number of branches required to reach that species from the root of the evolutionary tree (see Methods for details). The difference in lineage scores, denoted as Δ , serves as a measure of the phylogenetic diversity within each clustered MSA used for prediction. For predictions made using both UC and LC MSAs for autoinhibited proteins, we observed that accurate predictions are associated with significantly larger lineage score Δ values compared to inaccurate ones (Fig. S12a). However, since a higher number of sequences could inherently increase the likelihood of greater variance, we controlled for this potential confound. Specifically, we separated uniform clusters into two groups, those with 10 sequences and those with 100 sequences, and re-evaluated the lineage score Δ variance (Fig. S12b). In both cases, clusters that led to accurate structure predictions exhibit greater lineage score Δ variance, although the difference reaches statistical significance only for the larger clusters (100 sequences). We further analyzed the maximum lineage score Δ within each MSA and similarly find statistically significant differences between accurate and inaccurate predictions (Fig. S12c). Notably, across all analyses, clusters with the lowest average Shannon entropy or the lowest lineage score Δ consistently produce inaccurate structures. Overall, these findings demonstrate that for autoinhibited proteins, greater sequence and phylogenetic diversity are important for achieving reliable structure predictions.

Discussion

Proteins are inherently dynamic, constantly adjusting their structures to fulfill functional demands¹⁷. This principle lies at the heart of the modern view of allostery, which holds that a protein's sequence encodes an ensemble

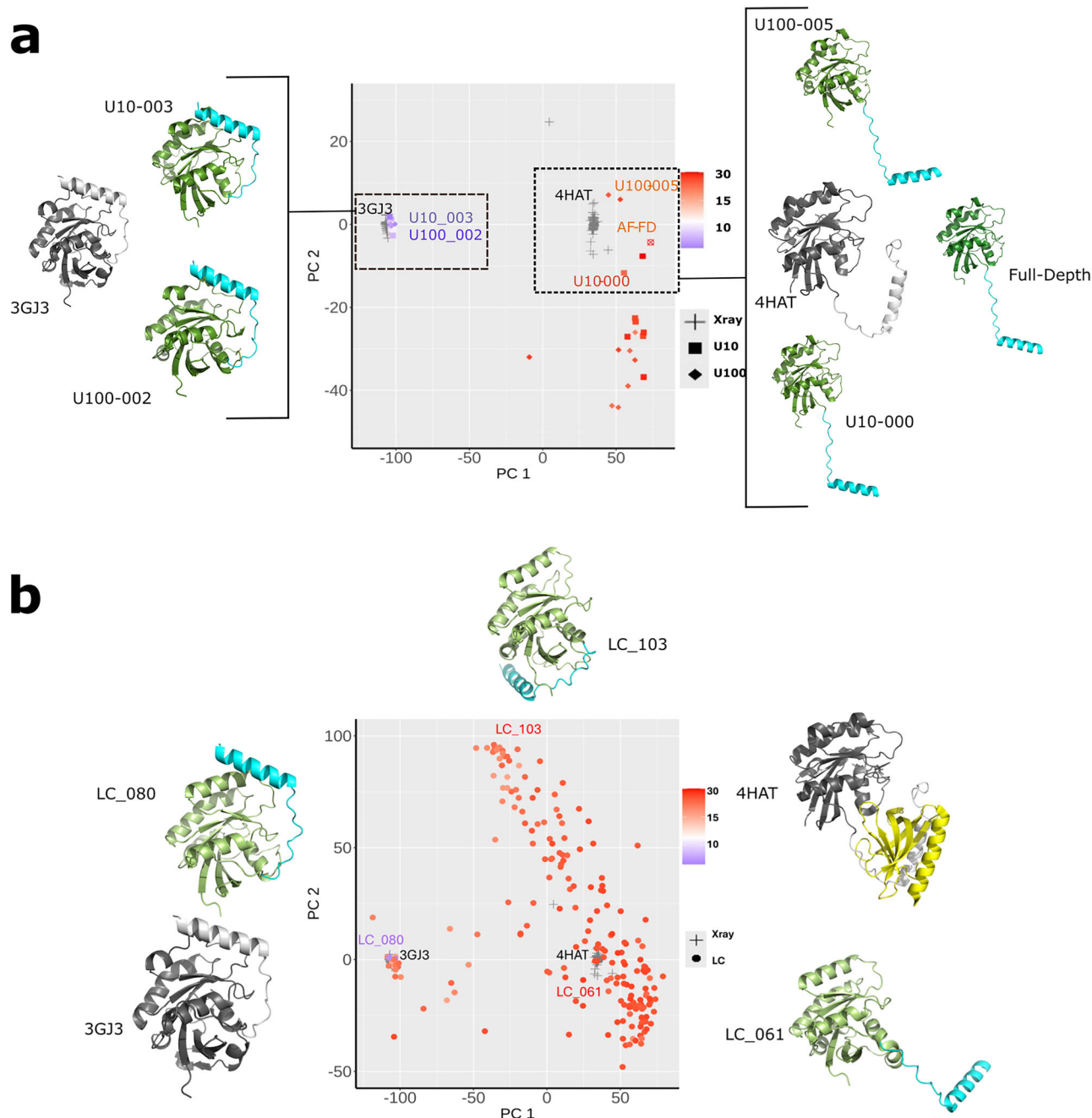


Fig. 5 | PCA analysis of cluster-based predictions of GTP-binding nuclear protein Ran (RAN). **a** Predictions using UC MSAs and **b** LC MSAs were projected onto the first and second lowest frequency eigenvectors of experimentally determined structures and color-coded according to their averaged inter-domain (IM-FD) PAEs (color scale shown on the right). Locations of selected experimental structures on this projections are also provided with PDB IDs. Experimentally determined structures representing the autoinhibited state (PDB ID 3GJ3⁷⁸) and the active state

(PDB ID 4HAT⁷⁹) are shown on the left and right of the PCA plots, respectively. IMs of experimental structures are shown in light grey and the rest of the proteins is dark grey. RAN's partner protein, Ran-specific GTPase-activating protein 1, is shown in yellow in (b). Structures of representative UC and LC predictions are also provided. The predicted IMs are cyan and the rest of the predicted proteins is shown in shades of green.

of conformations, each sampled under native conditions with distinct statistical weights. The ultimate goal of protein structure prediction is to accurately capture this entire conformational landscape, with populations that reflect specific experimental conditions. While this remains a distant target, AlphaFold and related tools have opened exciting new avenues toward realizing it. Here, we systematically assessed the capabilities of AlphaFold2, AlphaFold3, and recent method extensions to predict the structures of autoinhibited proteins, complementing important benchmarking analyses that have been carried out by others^{9–11,14}.

By comparing predictions for protein sequences encoding autoinhibited proteins, i.e., sequences that, by definition, encode at least two functionally distinct conformations, often characterized by large inter-domain rearrangements, with predictions for two-domain proteins lacking autoinhibition annotations (some of which are known to form obligate inter-domain contacts), we found that AlphaFold2 and AlphaFold3 exhibit lower confidence when modeling the autoinhibited proteins. This trend also holds for autoinhibited proteins for which only a single conformation is available in the PDB, structures that AlphaFold

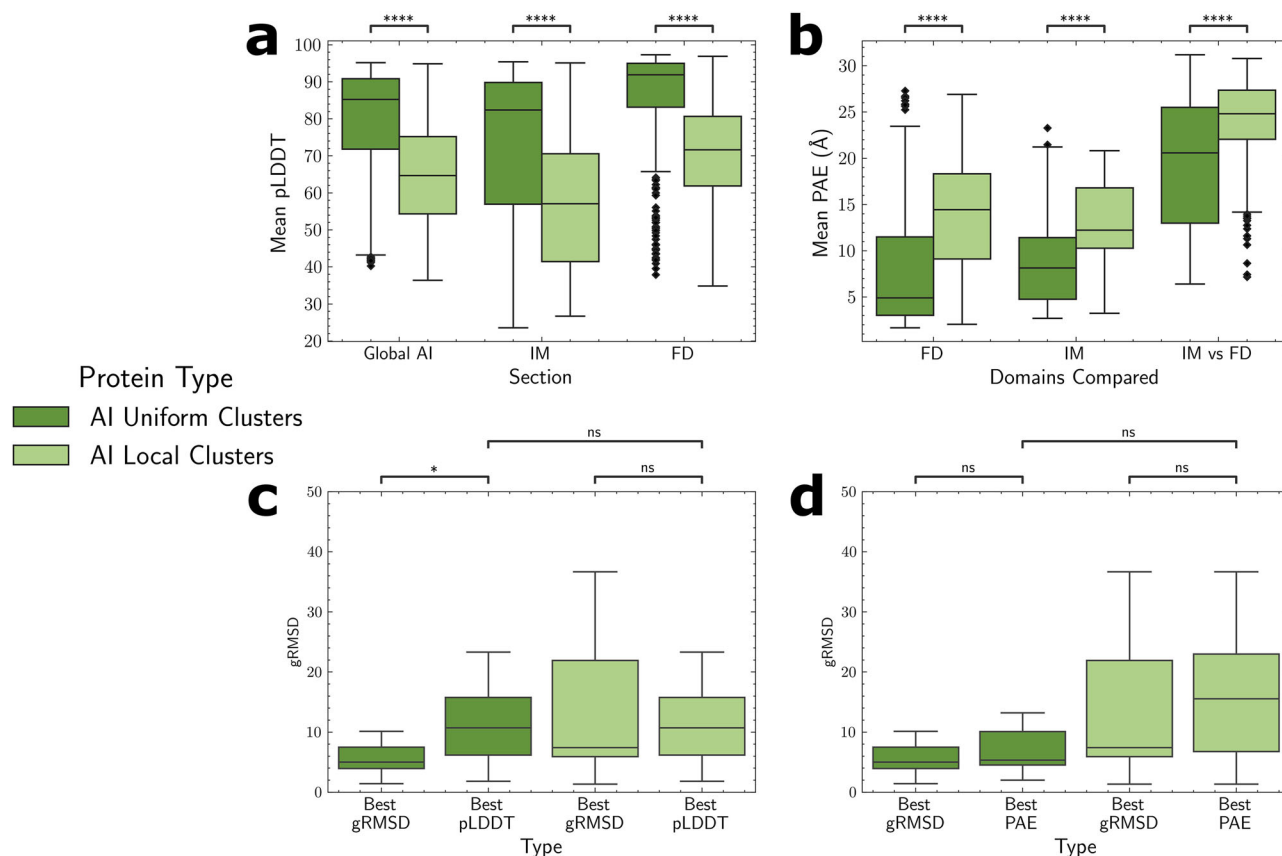
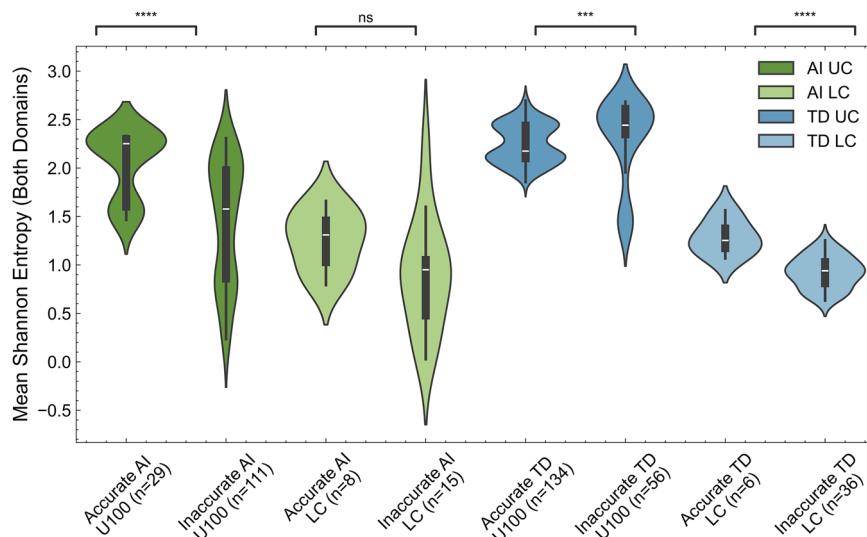


Fig. 6 | Predictions based on uniform MSA subsampling display greater confidence and accuracy than those based on local subsampling. Confidence metrics and consistency with experimental structures for predictions made with UC and LC MSAs for the 15 autoinhibited proteins with two known reference conformations. **a** Distributions of average pLDDT for UC and LC predictions. N = 300 uniform clusters, N = 1059 local clusters. **b** Distributions of average inter-domain (IM-FD)

PAE for UC and LC predictions. N = 300 uniform clusters, N = 1059 local clusters. Distributions of gRMSDs for predictions selected based on the best gRMSD, (highest) pLDDT per protein (in **c**) and best (lowest) inter-domain PAE (in **d**). N = 15 proteins. Statistical differences in all panels were assessed using the Mann-Whitney test (****: p-value ≤ 0.0001 , ***: p-value ≤ 0.001 , **: p-value ≤ 0.01 , and *: p-value ≤ 0.05).

Fig. 7 | Accurate predictions result from multiple sequence alignments with higher sequence diversity. Distributions of average Shannon entropy broken down by protein type (autoinhibited or two-domain), cluster type (uniform or local), and accuracy to experimental structures at a 3 Å cutoff (accurate or inaccurate). The average Shannon entropy was taken over the sequence ranges encompassing the IMs and FDs. Sample sizes are shown in category labels. Statistical differences were assessed using the Mann-Whitney test (****: p-value ≤ 0.0001 , ***: p-value ≤ 0.001 , **: p-value ≤ 0.01 , and *: p-value ≤ 0.05).



should, in principle, reproduce with high confidence if its predictions were primarily driven by memorization of training data. Notably, AlphaFold2 and AlphaFold3 also display reduced confidence, particularly in positioning the inhibitory module relative to the functional domain, for autoinhibited proteins lacking homologous structural

templates in the PDB. These findings, combined with the fact that the majority of the autoinhibited proteins analyzed predate AlphaFold's training set cutoff, suggest that AlphaFold may have learned aspects of the energy landscapes of these proteins^{11,36,37}, rather than merely memorizing structures^{12,14}. Nonetheless, this may be a finding specific to autoinhibited

proteins and more detailed analyses will be required to rigorously test this hypothesis.

The initial assessment of AlphaFold's predictions of autoinhibited proteins clearly highlights both the challenges the method faces and its intrinsic recognition of these challenges as reflected in its lower confidence scores. To address the central goal of our benchmarking effort, we next evaluated recent AlphaFold extensions to assess their ability to capture not just one, but two functional conformations of autoinhibited proteins. In the case of CFold, we observed that its predictions for autoinhibited proteins perform poorly. However, this finding is consistent with CFold's own evaluation¹⁴, which reported that CFold struggles particularly with “hinge” motions, the type of structural changes that characterize the majority of autoinhibited proteins in our dataset. While MSA subsampling strategies lead to more favorable results, only a limited number of autoinhibited proteins show accurate overall predictions of both the active and autoinhibited conformations. These findings mirror those from the study on fold-switching proteins, where accurate prediction of multiple conformations was observed in only a minority of cases¹². Previous work that generated alternative conformations through MSA alteration^{9–11} proposed that these manipulations deconvolve evolutionary signals, enabling AlphaFold to predict alternative states. Intriguingly, for autoinhibited proteins, we observed the opposite: uniform subsampling of the MSA, and the resulting increase in sequence diversity, is associated with more accurate predictions. Our analyses suggest that greater sequence diversity may be necessary to accurately predict conformations of autoinhibited proteins. Interestingly, recent work has demonstrated that reducing subsampling to very shallow depths improves predictions of alternative conformations in fold-switching proteins³⁸.

Better results than with MSA subsampling were achieved using BioEmu, as expected given that the model was specifically optimized for conformational diversity through the inclusion of MD simulation data during training. Nevertheless, only in one analyzed system did BioEmu generate structures matching both conformations within 3 Å gRMSD. AlphaFold3 performed similarly when PTM information and interaction partner sequences were incorporated, enabling accurate modeling of both active and autoinhibited conformations that most closely resembled the experimentally observed ones. Since PTMs and partner binding are major factors that shift the structural equilibrium of autoinhibited systems, and are often leveraged experimentally to stabilize specific states, it is perhaps unsurprising that providing this information enhances AlphaFold3's ability to predict alternative conformations. However, this advantage also highlights a key limitation: the method requires prior knowledge of the factors governing conformational shifts, which substantially constrains its generalizability. Despite the encouraging results achieved with BioEmu and AlphaFold3, it is important to emphasize that all tested methods, including these two, performed poorly in reproducing the relative placement of IM and FD as observed in experimental structures, underscoring the persistent challenges of predicting autoinhibited protein structures.

A key issue when working with protein sequences that encode complex energy landscapes is the definition of ‘ground truth.’ Autoinhibited proteins may not only sample states of full inactivity or full activity, often the ones captured experimentally, but also a spectrum of partially active states, each associated with distinct conformations. Moreover, fully active states may contain various conformations. This is particularly evident when the IM loses stable interactions with the functional domain (Fig. 5). In this case, the IM likely adopts an ensemble of conformations relative to the FD, of which only a subset has been experimentally determined. Moreover, experimental conditions used for structure determination (e.g., solvation) can bias which conformations are captured or even lead to crystallographic artifacts, and PDB entries do not fully represent the conformational diversity of proteins^{7,8,39}. As a result, AlphaFold predictions for autoinhibited proteins that come close but do not exactly match an experimentally observed structure should not automatically be considered ‘incorrect’. For this reason, more lenient structural similarity cutoffs, such as those applied here, may provide a fairer basis for assessing predictions and avoid overly negative evaluations.

This raises the question of whether confidence scores provided by prediction tools can mitigate these issues. Ideally, a predictor would assign high confidence only to conformations that are functionally relevant. Yet it remains unclear whether AlphaFold's confidence metrics truly capture biological relevance¹⁰ or whether correlations with accuracy are largely coincidental^{9,14,40}. While pLDDT is often used as a measure of prediction reliability, several studies have shown that AlphaFold can assign high confidence to incorrect structures, particularly in flexible or ambiguous regions^{12,14}. Our analysis cannot determine whether AlphaFold's confidence scores correlate with the biological or functional relevance of predicted conformations of autoinhibited proteins. However, we found that, for this class of proteins, PAE is a more reliable indicator of structural similarity to experimental data than pLDDT. Based on these observations, we propose using PAE as a selection metric when employing full-depth and uniformly subsampled MSAs, although this strategy may favor the selection of closed conformations, i.e., states in which the IM and FD interact. Whether this PAE-driven bias toward compact structures generalizes to other complex systems, and the extent to which it favors alternative closely packed conformations, remains an important question for future studies.

Finally, although our analyses cover a substantial number of proteins, including over 100 proteins not previously examined by large-scale efforts^{12,14}, the generalizability of our results remains limited by sample size and diversity. While analysis of autoinhibited, two-conformation proteins outside of AlphaFold2's training set serves as an important control, such an evaluation would be restricted to only around seven proteins. Furthermore, our dataset is heavily skewed toward vertebrate proteins: of the 894 high-quality experimental structures analyzed, 90% are human, and accurate predictions from clustered MSAs were limited to humans and brown rats (*Rattus norvegicus*), highlighting a likely species bias in AlphaFold. Achieving accurate, generalizable predictions of alternative conformations will ultimately require training on a broader and more phylogenetically diverse structural dataset^{5,14}. Notably, while this study was underway, additional approaches to generate conformational diversity have emerged^{41,42}. However, whether such methods can accurately recapitulate the conformational diversity of the systems analyzed here remains to be seen.

Methods

Data sets

We compiled the set of autoinhibited proteins from proteins deposited in the autoinhibited protein database (AiPD)³¹. An autoinhibited protein is a protein that self-regulates via transient intramolecular interactions between different parts of the protein. The initial data compilation resulted in a list of 442 autoinhibited proteins. After filtering for proteins with associated PDB structures containing $\geq 60\%$ domain completeness, 128 proteins remained representing 70 homologous superfamilies. Complete information on InterPro classifications can be found in `project_pipeline/data/interpro.tsv` at the github repo (https://github.com/bperkinsj/autoinhibitory_protein_prediction). The autoinhibited protein database AiPD³¹ provided annotated sequence boundaries for both the inhibitory module (IM) and the functional domain (FD). IM boundaries are typically defined through deletion assays, where removal of specific regions results in altered activity. The FD corresponds to the domain responsible for the protein's primary function—such as partner binding or catalytic activity—and is the focus of the assay.

For autoinhibited proteins with no homologous structures in the PDB, we selected proteins from AiPD according to specific criteria. It was required that the location of the IM of the candidate proteins had been experimentally determined. We checked for the absence of experimental structural information for the protein (no PDBs), as well as for any of its known homologs. Homologs were obtained from the Homologous Gene Database (HGD)⁴³.

We also assembled a control set consisting of non-autoinhibited two-domain proteins. We created the list of two-domain proteins by selecting all proteins annotated as “Multi-domain proteins (alpha and beta) (56572)” under SCOP-e on the RCSB and then filtering for proteins with two

domains annotated in UniProt⁴⁴ using custom Python code. This resulted in a list of 61 proteins. Filtering for proteins with associated PDB structures containing $\geq 60\%$ domain completeness left 40 proteins. Sequence boundaries of the domains were taken from UniProt annotations. Obligate proteins are two-domain proteins with permanent inter-domain interactions. A list of PDB files with permanent interactions was taken from ref. 45 and used to find associated proteins in the list of two-domain proteins. Ultimately, seven proteins in the two-domain list were labeled as obligate.

We selected fifteen autoinhibited proteins for predictions of multiple conformations. All 15 were selected on the basis of having at least two experimentally observed conformations (see 4.2). In addition, we selected nineteen two-domain proteins randomly from the list of 40 for control structure predictions via MSA subsampling. For any predictions using trimmed sequences, we extracted protein segments from UniProt⁴⁴, including only residues corresponding to the IM, FD, and any intervening regions. When multiple isoforms were available, the canonical sequence annotated in UniProt was used.

Protein clustering

We used hierarchical clustering to determine the structural heterogeneity of the experimental structures and identify proteins with distinct conformations. Specifically, we performed the clustering using SciPy's⁴⁶ linkage function with method='single' and metric='euclidian'. Proteins were considered having distinct conformations if at least two cluster emerged when using a cutoff of 3 Å.

Predictions using full-depth MSAs

We downloaded AF2 predictions for full-length sequences of the proteins in the two data sets (128 autoinhibited and 40 two-domain proteins) and autoinhibited proteins with no homologous templates in the PDB from AlphaFold's Google Cloud repository. AlphaFold3 (AF3) predictions were generated using the AlphaFold Server (<https://alphafoldserver.com/welcome>) for all 128 autoinhibited proteins, without providing binding partners or post-translational modifications (PTMs). Additionally, for a subset of 15 proteins with at least two known experimental conformations, we performed AF3 predictions with partner sequences and PTMs specified based on experimental data. The full list of experimental structures, along with their associated partners and PTMs, is available in the file `af3_partner_prediction_structures.xlsx` on GitHub (https://github.com/bperkinsj/autoinhibitory_protein_prediction).

Predictions using MSA subsampling, CFold and BioEmu

To generate predictions by MSA subsampling, we followed the steps laid out by Wayment-Steele et al.¹⁰ in their AF-Cluster approach. First, an initial multiple MSA was created using the ColabFold notebook (<https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb>), which utilizes MMSeqs2 and HHsearch^{47,48}. AF-Cluster was run with these MSAs to generate a number of clustered MSAs equal to e^{max} , the maximum number of clusters. AF-Cluster also generates twenty MSAs that are uniformly sampled across the full sequence alignment, ten of which have ten sequences and the other ten having one hundred sequences. We referred to these MSAs as “uniformly clustered” (UC) and the former as “locally clustered” (LC). We referred to structures generated with the full MSA depth (a.k.a. the original AlphaFold2 pipeline) as “full depth” predictions. We fed clustered MSAs generated by AF-Cluster to an instance of ColabFold (v. 1.5.2) running on Compute Canada's Graham server cluster. The configurations were set to default, i.e. `num_recycles` is set to 3 and `num_seeds` is set to 1. We collected ten predictions from each MSA.

For CFold predictions, we first generated MSAs using HHblits⁴⁹, based on the protein sequences from the RCSB, following the approach described in ref. 14. These MSAs were then used as input in the CFold Google Colab notebook (<https://colab.research.google.com/github/patrickbryant1/Cfold/blob/master/Cfold.ipynb>) to generate structure predictions.

We submitted protein sequences to the BioEmulator Google Colab notebook (<https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/BioEmu.ipynb>) on the Compute Canada's Fir server cluster to generate 10,000 samples per protein¹⁵. For two proteins (P21333 and P26358) technical errors allowed for only 100 samples to be generated.

Confidence scores

To assess differences in pLDDT scores across protein categories, we extracted scores from prediction CIF files and calculated average pLDDT values for the full structure, the inhibitory module (domain 1), and the functional domain (domain 2).

PAE) values were extracted from the JSON output files provided by AlphaFold and ColabFold. For each comparison—either within a single region or between two regions—we isolated the relevant submatrix (e.g., residues 1–102 vs. 1–102 or 1–102 vs. 344–485) and calculated the mean PAE.

RMSD calculations

There are multiple ways to compare protein structures, the most common being the root mean square deviation, which is given in Å and calculated as

$$RMSD = \sqrt{\frac{1}{n} \sum_{i=1}^n d_i^2}$$

where n is the number of equivalent atom pairs and d_i is the distance between atom i in a given conformation and its counterpart in the reference structure⁵⁰. Although various atom sets can be used for alignment, we used the standard approach of aligning C_α atoms. We calculated four types of RMSDs: the global RMSD of a protein when aligning the entire sequences for which coordinates are available in all PDB structures (gRMSD); the RMSD of the inhibitory modules when aligned to each other ($_{im}$ RMSD); the RMSD of the functional domains when aligned to each other ($_{fd}$ RMSD); and the RMSD of the inhibitory modules when aligned on the functional domain ($_{fd}^{im}$ RMSDs).

DockQ

We also compared experimental structures and full-depth predictions using DockQ³². Predictions with the best ($_{fd}^{im}$ RMSDs) were chosen for this analysis. Structures were split into three chains—chain A for the functional domain (domain 2), chain B for the inhibitory module (domain 1), and chain C for the rest of the protein—using custom Python scripts and Snakemake (Snakemake_DockQ). We then ran the DockQ scoring script from the official GitHub repository (<https://github.com/bjornwallner/DockQ>) to compute interface quality scores.

Conformational ensembles mapping

We used principal component analysis (PCA) to capture the major conformational differences between structures. Cartesian coordinates of aligned C_α atoms of all experimental structures of a given protein were used to calculate the covariance matrix. The covariance matrix was then diagonalized to derive principal components with their associated variances using Bio3D package⁵¹, a family of related R packages containing utilities for the analysis of biomolecular structure, sequence and trajectory data^{51,52}. Over 90% of the total mean-square displacement (or variance) of atom positional fluctuations was captured in two dimensions for the two systems analyzed (P62826 and P35520). Thus, the first two principal components provide a useful description of the conformational space of the systems. We projected structures predicted with LC, UC and full-depth MSAs as well as experimental structures representing active and autoinhibited states on the two first principal components.

Shannon Entropy

We calculated Shannon entropy⁵³ of cluster MSAs using a modified version of Joe Healey's Shannon.py script⁵⁴. We calculated the Shannon entropies over both the entire protein sequence and the sequence ranges representing

the IM and FD and then the average entropy was taken for both. We calculated average Shannon entropy for cluster MSAs with ≥ 100 sequences.

Lineage score

We calculated the lineage scores as follows. Each MSA contains a number of sequences pulled from UniRef100⁵⁵. UniRef100 combines identical sequences and fragments into a single entry with a representative sequence. For each sequence, we queried UniRef100 for the UniRef ID denoted in the MSA and the species from which the representative sequence was selected. The number of nodes on the species' NCBI taxonomic lineage⁵⁶ was counted, giving the lineage score. To calculate the LS Δ for each sequence, we subtracted the LS of the sequence from the LS of the reference sequence and the absolute value was taken.

Additional predictions, plotting and statistical Analysis

To identify structured IMs, we predicted their disorder with Espritz⁵⁷ using the Disprot prediction type and the Best Sw decision threshold. We analyzed all our data in Jupyter notebooks⁵⁸ with custom Python scripts utilizing Biopython⁵⁹, NumPy⁶⁰, Pandas⁶¹, and SciPy⁶². K-means clustering was performed using sklearn⁶². Protein structures visualizations were created with PyMOL⁶³. Plots were constructed using a mixture of matplotlib⁶⁴ and seaborn⁶⁵ and stylized with scienceplots⁶⁶. Two-sided Mann-Whitney U-tests were carried out and annotated using statannotations⁶⁷. Final figures were created using Inkscape⁶⁸.

Data availability

Data used to generate plots, as well as the original lists of proteins, can be found at Github⁶⁹. Raw structure data for AlphaFold and ColabFold predictions, as well as generated MSA clusters, are publicly available⁷⁰ at the Open Science Framework⁷¹.

Code availability

Code used for data collection, analysis and visualization can be found at Github⁶⁹. Software versions and requirements can be found in the README.

Received: 14 May 2025; Accepted: 20 October 2025;

Published online: 26 November 2025

References

- Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
- Baek, M. et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).
- Zhang, X. et al. Bending and binding: Predicting protein flexibility upon ligand interaction using diffusion models (2023-10-27). <https://openreview.net/forum?id=PQa3giMLZp>.
- Wallner, B. AFsample: improving multimer prediction with AlphaFold using massive sampling. *Bioinformatics* **39**, btad573 (2023).
- Ahdritz, G. et al. OpenFold: retraining AlphaFold2 yields new insights into its learning mechanisms and capacity for generalization. *Nat. Methods* **21**, 1514–1524 (2024).
- Mirdita, M. et al. ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
- Boehr, D. D., Nussinov, R. & Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* **5**, 789–796 (2009).
- Motlagh, H. N., Wrabl, J. O., Li, J. & Hilser, V. J. The ensemble nature of allostery. *Nature* **508**, 331–339 (2014).
- del Alamo, D., Sala, D., Mchaourab, H. S. & Meiler, J. Sampling alternative conformational states of transporters and receptors with AlphaFold2. *eLife* **11**, e75751 (2022).
- Wayment-Steele, H. K. et al. Predicting multiple conformations via sequence clustering and AlphaFold2. *Nature* **625**, 832–839 (2024).
- Stein, R. A. & Mchaourab, H. S. SPEACH_af: Sampling protein ensembles and conformational heterogeneity with alphafold2. *PLOS Comput. Biol.* **18**, e1010483 (2022).
- Chakravarty, D. et al. AlphaFold predictions of fold-switched conformations are driven by structure memorization. *Nat. Commun.* **15**, 7296 (2024).
- Porter, L. L. & Looger, L. L. Extant fold-switching proteins are widespread. *Proc. Natl. Acad. Sci. USA* **115**, 5968–5973 (2018).
- Bryant, P. & Noé, F. Structure prediction of alternative protein conformations. *Nat. Commun.* **15**, 7328 (2024).
- Lewis, S. et al. Scalable emulation of protein equilibrium ensembles with generative deep learning. *Science* **389**, ead9817 (2025).
- Li, P., Martins, I. R. S., Amarasinghe, G. K. & Rosen, M. K. Internal dynamics control activation and activity of the autoinhibited vav DH domain. *Nat. Struct. Mol. Biol.* **15**, 613–618 (2008).
- Smock, R. G. & Gierasch, L. M. Sending signals dynamically. *Science* **324**, 198 (2009).
- Pufall, M. A. & Graves, B. J. Autoinhibitory domains: modular effectors of cellular regulation. *Annu. Rev. Cell Develop. Biol.* **18**, 421–462 (2002).
- Fenton, M., Gregory, E. & Daughdrill, G. Protein disorder and autoinhibition: The role of multivalency and effective concentration. *Curr. Opin. Struct. Biol.* **83**, 102705 (2023).
- Trudeau, T. et al. Structure and intrinsic disorder in protein autoinhibition. *Structure* **21**, 332–341 (2013).
- Parikh, C. et al. Disruption of PH-kinase domain interactions leads to oncogenic activation of AKT in human cancers. *Proc. Natl. Acad. Sci. USA* **109**, 19368–19373 (2012).
- Ruiz-Saenz, A. et al. Proteomic analysis of Src family kinase phosphorylation states in cancer cells suggests deregulation of the unique domain. *Mol. Cancer Res. MCR* **19**, 957–967 (2021).
- Niault, T. et al. From autoinhibition to inhibition in trans: the Raf-1 regulatory domain inhibits Rho-alpha kinase activity. *J. Cell Biol.* **187**, 335–342 (2009).
- Holguin-Cruz, J. A., Bui, J. M., Jha, A., Na, D. & Gsponer, J. Widespread alteration of protein autoinhibition in human cancers. *Cell Syst.* **15**, 246–263 (2024).
- Schaufele, F. et al. The structural basis of androgen receptor activation: Intramolecular and intermolecular amino-carboxy interactions. *Proc. Natl. Acad. Sci. USA* **102**, 9802–9807 (2005).
- Underbakke, E. S., Iavarone, A. T. & Marletta, M. A. Higher-order interactions bridge the nitric oxide receptor and catalytic domains of soluble guanylate cyclase. *Proc. Natl. Acad. Sci. USA* **110**, 6777–6782 (2013).
- Zhou, G. et al. Role of AMP-activated protein kinase in mechanism of metformin action. *J. Clin. Invest.* **108**, 1167 (2001).
- Onuchic, J. N., Luthey-Schulten, Z. & Wolynes, P. G. THEORY OF PROTEIN FOLDING: The energy landscape perspective. *Annu. Rev. Phys. Chem.* **48**, 545–600 (1997).
- Brini, E., Simmerling, C. & Dill, K. Protein storytelling through physics. *Science* **370**, eaaz3041 (2020).
- Sippl, M. J. Calculation of conformational ensembles from potentials of mean force. An approach to the knowledge-based prediction of local structures in globular proteins. *J. Mol. Biol.* **213**, 859–883 (1990).
- Cho, D. et al. Autoinhibited protein database: a curated database of autoinhibitory domains and their autoinhibition mechanisms. *Database* **2024**, baae085 (2024).
- Basu, S. & Wallner, B. DockQ: a quality measure for protein-protein docking models. *PLoS ONE* **11**, e0161879 (2016).
- Omidi, A., Möller, M. H., Malhis, N., Bui, J. M. & Gsponer, J. AlphaFold-multimer accurately captures interactions and dynamics of intrinsically disordered protein regions. *Proc. Natl. Acad. Sci. USA* **121**, e2406407121 (2024).
- Guo, H.-B. et al. AlphaFold2 models indicate that protein sequence determines both structure and dynamics. *Sci. Rep.* **12**, 10696 (2022).

35. Ester, M., Kriegel, H.-P., Sander, J. & Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, KDD'96, 226–231 (AAAI Press, 1996).
36. Roney, J. P. & Ovchinnikov, S. State-of-the-art estimation of protein model accuracy using AlphaFold. *Phys. Rev. Lett.* **129**, 238101 (2022).
37. Monteiro da Silva, G., Cui, J. Y., Dalgarno, D. C., Lisi, G. P. & Rubenstein, B. M. High-throughput prediction of protein conformational distributions with subsampled AlphaFold2. *Nat. Commun.* **15**, 2464 (2024).
38. Lee, M. et al. Large-scale predictions of alternative protein conformations by AlphaFold2-based sequence association. *Nat. Commun.* **16**, 5622 (2025).
39. Marino-Buslje, C., Monzon, A. M., Zea, D. J., Fornasari, M. S. & Parisi, G. On the dynamical incompleteness of the protein data bank. *Brief. Bioinforma.* **20**, 356–359 (2019).
40. Schafer, J. W. et al. Sequence clustering confounds AlphaFold2. *Nature* **638**, E8–E12 (2025).
41. Watson, J. L. et al. De novo design of protein structure and function with RFdiffusion. *Nature* **620**, 1089–1100 (2023).
42. Zheng, S. et al. Predicting equilibrium distributions for molecular systems with deep learning. *Nat. Mach. Intell.* **6**, 558–567 (2024).
43. Duan, G. et al. HGD: an integrated homologous gene database across multiple species. *Nucleic Acids Res.* **51**, D994–D1002 (2023).
44. The UniProt Consortium. UniProt: the universal protein knowledgebase in 2023. *Nucleic Acids Res.* **51**, D523–D531 (2023).
45. Sidhanta, S. P. D., Sowdhamini, R. & Srinivasan, N. Comparative analysis of permanent and transient domain-domain interactions in multi-domain proteins. *Proteins* **93**, 197–208 (2023).
46. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
47. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017).
48. Söding, J. Protein homology detection by HMM-HMM comparison. *Bioinformatics* **21**, 951–960 (2005).
49. Steinegger, M. et al. HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinforma.* **20**, 473 (2019).
50. Kufareva, I. & Abagyan, R. Methods of protein structure comparison. *Methods Mol. Biol.* **857**, 231 (2012).
51. Grant, B. J., Rodrigues, A. P. C., ElSawy, K. M., McCammon, J. A. & Caves, L. S. D. Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* **22**, 2695–2696 (2006).
52. Skjærven, L., Yao, X.-Q., Scarabelli, G. & Grant, B. J. Integrating protein structural dynamics and evolutionary analysis with bio3d. *BMC Bioinforma.* **15**, 399 (2014).
53. Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423 (1948).
54. bioinfo-tools/shannon.py at master · jrjhealey/bioinfo-tools. <https://github.com/jrjhealey/bioinfo-tools/blob/master/Shannon.py> (visited on 05/01/2025).
55. Suzek, B. E., Huang, H., McGarvey, P., Mazumder, R. & Wu, C. H. UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics* **23**, 1282–1288 (2007).
56. O'Leary, N. A. et al. Exploring and retrieving sequence and metadata for species across the tree of life with NCBI datasets. *Sci. Data* **11**, 732 (2024).
57. Walsh, I., Martin, A. J. M., Di Domenico, T. & Tosatto, S. C. E. ESpritz: accurate and fast prediction of protein disorder. *Bioinformatics* **28**, 503–509 (2012).
58. Kluyver, T. et al. *Jupyter Notebooks—a publishing format for reproducible computational workflows* (Harvard, 2016).
59. Cock, P. J. A. et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422–1423 (2009).
60. Harris, C. R. et al. Array programming with NumPy. *Nature* **585**, 357–362 (2020).
61. team, T. p. d. pandas-dev/pandas: pandas <https://zenodo.org/records/10957263> (2024-04-10).
62. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
63. Schrödinger, L. & DeLano, W. The PyMOL molecular graphics system, version 2.0 schrödinger, LLC (2017). *Google Scholar There is no corresponding record for this reference* (2020).
64. Hunter, J. D. Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).
65. Waskom, M. Seaborn: statistical data visualization. *J. Open Source Softw.* **6**, 3021 (2021).
66. Garrett, J. et al. garrettj403/SciencePlots: 2.1.1. Zenodo <https://ui.adsabs.harvard.edu/abs/2021zndo...4106649G> (2023-11-25).
67. Charlier, F. et al. trevismd/statannnotations: v0.6 <https://zenodo.org/records/8396665> (2023-10-01).
68. Inkscape. <https://inkscape.org/>.
69. Perkins-Jechow, B. bperkinsj/autoinhibitory_protein_prediction: Release for publishing <https://zenodo.org/records/17307832> (2025-10-09).
70. Perkins-Jechow, B. & Ahualli, J. autoinhibition_and_alphafold2 (2025-05-01). <https://osf.io/hkpgz/>. Publisher: OSF.
71. Foster, E. D. & Deardorff, A. Open science framework (OSF). *J. Med. Libr. Assoc. JMLA* **105**, 203–206 (2017).
72. DiNitto, J. P. et al. Structural basis and mechanism of autoregulation in 3-phosphoinositide-dependent GRP1 family ARF GTPase exchange factors. *Mol. Cell* **28**, 569–583 (2007).
73. Malaby, A. W. et al. Structural dynamics control allosteric activation of cytohesin family arf GTPase exchange factors. *Structure* **26**, 106–117 (2018).
74. Wang, W. et al. Crystal structure of human protein tyrosine phosphatase SHP-1 in the open conformation. *J. Cell. Biochem.* **112**, 2062–2071 (2011).
75. Yang, J. et al. Crystal structure of human protein-tyrosine phosphatase SHP-1 *. *J. Biol. Chem.* **278**, 6516–6520 (2003).
76. Lad, Y. et al. Structure of three tandem filamin domains reveals auto-inhibition of ligand binding. *EMBO J.* **26**, 3993–4004 (2007).
77. Bank, R. P. D. RCSB PDB - 4p3w: crystal structure of the human filamin A Ig-like domains 20-21 in complex with migfilin peptide. <https://www.rcsb.org/structure/4P3W> (visited on 04/29/2025).
78. Partridge, J. R. & Schwartz, T. U. Crystallographic and biochemical analysis of the ran-binding zinc finger domain. *J. Mol. Biol.* **391**, 375–389 (2009).
79. Sun, Q. et al. Nuclear export inhibition through covalent conjugation and hydrolysis of leptomycin B by CRM1. *Proc. Natl. Acad. Sci. USA* **110**, 1303–1308 (2013).

Acknowledgements

This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada. This research was enabled in part by support provided by the Digital Research Alliance of Canada (alliancecan.ca).

Author contributions

Conceptualization: B.H.P.-J. and J.G. Methodology: B.H.P.-J.; Software: A.O. and N.M.; Investigation: B.H.P.-J., J.P.I.A., T.H.N., A.O., C.L., and J.M.B.; Data Curation: B.H.P.-J., D.C., D.N., and J.A.H.-C.; Visualization: B.H.P.-J., J.M.B., and T.H.N.; Writing - original draft: B.H.P.-J. and J.G.; Writing - review & editing: B.H.P.-J. and J.G.; Supervision: J.G.; Project administration: J.G.; Funding acquisition: J.G.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s42004-025-01763-0>.

Correspondence and requests for materials should be addressed to Jörg Gsponer.

Peer review information *Communications Chemistry* thanks Domenico Scaramozzino and the other, anonymous, reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025