PNAS RESEARCH ARTICLE | BIOPHYSICS AND COMPUTATIONAL BIOLOGY

Making invisible excited-state structures of pro-interleukin-18 visible by combining NMR and machine learning

Jeffrey P. Bonin[a,b,c,1], Jin Sub Lee[a,d], Zi Hao Liu[b,e], Philip M. Kim[a,d,f], and Lewis E. Kay[a,b,c,e,1]

Affiliations are included on p. 7.

Edited by Adriaan Bax, National Institutes of Health, Bethesda, MD; received December 18, 2025; accepted March 16, 2026

NMR spectroscopy is a powerful tool for studies of protein dynamics over timescales extending more than twelve orders of magnitude, with motion queried at many of the backbone and sidechain positions in the molecule of interest. NMR experiments can, in principle, provide atomic resolution descriptions of excited conformational states that are sparsely populated and transiently formed—i.e., invisible, and there are examples of such structures in the literature. However, in other cases, the NMR data are not sufficient for generating structural ensembles of rare and transient higher energy conformers of the protein of interest. One example is provided by the precursor form of the proinflammatory cytokine interleukin-18, pro-IL-18, studied here. Although NMR studies show that pro-IL-18 adopts two sparsely populated (<0.5%) and transiently formed (ms lifetimes) excited-state conformations in exchange with a highly populated ground state conformer and localize regions undergoing exchange to a pair of short β-strands that are preserved in at least one of the excited states, additional structural information is not forthcoming. Here, we develop a protocol whereby the NMR data are used to select alternative conformers of pro-IL-18 from ensembles predicted by the generative machine learning model AlphaFlow that are then evaluated through further NMR experiments. The method identifies distinct conformers that correspond to the pair of excited states reported by NMR, combining experiment with computation to characterize the pro-IL-18 energy landscape in ways that are not possible with either approach alone.

protein dynamics | machine learning | NMR spectroscopy | excited states

Significance

In many cases accurate models of the low energy, most populated form of a protein's three-dimensional structure can be rapidly determined using machine learning (ML) programs, such as AlphaFold. However, biomolecular function is often dependent on excursions to higher energy, invisible (excited) conformations that are not accurately predicted by current computational approaches. Here, we describe a combined solution NMR/ML-based approach focused on the proinflammatory cytokine interleukin-18 that identifies and validates a pair of excited states of the protein, going beyond the static pictures generated by AlphaFold. The study highlights the complementarity between NMR and generative ML models, suggesting a synergy between experiment and computation that promises to significantly advance current understanding of the roles of dynamics in biomolecular function.

Proteins are dynamic molecules, and their dynamics play critical roles in their function (1). Motions can occur on timescales ranging from femtoseconds (fs) to many seconds (s) or hours (h), often involving significant rearrangements of structure to accommodate specific interactions with targets or to respond to various stimuli in the environment (2). In many cases, these dynamics involve exchange between ground (lowest energy, highly populated) and excited (higher energies, sparsely populated) conformational states, the latter of which can play critical roles in molecular recognition, including binding of ligands to proteins as cancer therapeutics (3), signaling (4), enzyme catalysis (5–8), and protein folding and misfolding processes (9–11). Although ground state (GS) conformers are characterized in detail using well established biophysical techniques, excited states are sparsely populated and often transiently formed so that they are rendered invisible to many of the most used biophysical methods. While computational approaches are now widely available for predicting structures of GS conformations (12, 13), they are much less efficient in generating accurate representations of excited states for which little experimental data are available for training (14).

NMR spectroscopy is a powerful tool for studies of protein dynamics over timescales ranging from ps to h. The development of relaxation-based experiments such as Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion (15–17) and Chemical Exchange Saturation Transfer (CEST) (18) has enabled, in some cases, atomic resolution descriptions of excited conformational states that cannot be directly observed in NMR spectra, as well as quantification of the kinetics and thermodynamics involved in the underlying conformational exchange process(es). Most applications have involved excited states with significant changes to secondary structural elements relative to the GS, or with rearrangements of domains, and have focused on conformers that originate due to thermal fluctuations from the GS (9, 10, 19–23). Related relaxation experiments, termed Dark State Exchange Saturation Transfer or DEST have been used to study exchange between molecules free in solution and bound to the surface of large macromolecules to obtain structural information on the bound conformation(s) of the ligand and the kinetics and mechanism of the binding process (24). Pressure-based NMR methods offer an exciting route to

Author contributions: J.P.B., J.S.L., P.M.K., and L.E.K. designed research; J.P.B. and J.S.L. performed research; J.P.B. and Z.H.L. analyzed data; and J.P.B. and L.E.K. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: jeff.bonin@utoronto.ca or lewis.kay@utoronto.ca.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2537014123/-/DCSupplemental.

Published April 17, 2026.

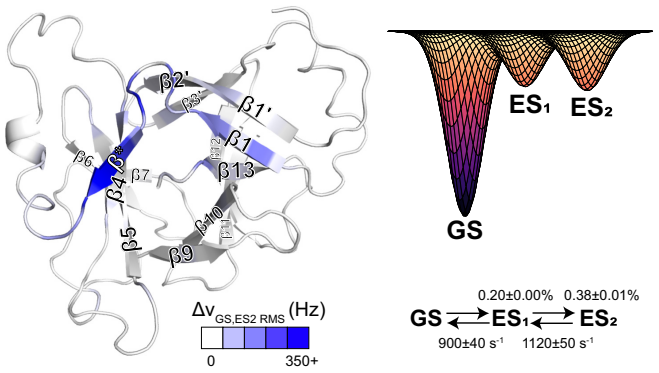

Fig. 1. Pro-IL-18 adopts invisible, excited states involving the β* strand. Pro-IL-18 accesses two excited states (ES_1 and ES_2), with the largest chemical shift changes between ground and excited states occurring in and adjacent to the β* strand (28). The root mean squared difference in chemical shifts between the WT GS and ES_2, $\Delta\nu_{GS,ES2\ RMS}$ ($\Delta\nu_{GS,ES2} = \nu_{ES2}-\nu_{GS}$; Hz), calculated based on ^{1}HN, ^{15}N, ^{13}CO, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and $^{1}H^{\alpha}$ shifts recorded at a field of 23.4 T (1 GHz), is shown on the structure as a white-blue gradient. The fitted exchange scheme, including rates of interconversion (sums of forward and backward rates connecting pairs of states) and populations, 25 °C, is shown.

characterize transient intermediates along reaction trajectories in cases where the states involved differ in volume (25–27).

Despite the underlying potential of the various NMR methods there are limitations, as we recently encountered in CPMG and CEST relaxation studies of pro-IL-18 (28). Pro-IL-18 is activated via cleavage by caspases-1 and -4, removing the first 36 residues to produce the mature form of the protein (29, 30). Previously, we showed that pro-IL-18 transiently adopts two excited states (ES_1 and ES_2) on the ms timescale, with chemical shift differences between each of them and the GS conformation localized to a region of the protein containing two short β strands, β1 and β* (Fig. 1), which are involved in caspase binding. The chemical shift differences are especially large within and adjacent to β* (Fig. 1). Backbone chemical shifts indicate that these β strands do not unfold in either excited state and maintain their secondary structures in at least one of the rare conformations of the wild-type (WT) protein. Further structural insights from the NMR data could not be obtained, however, because the shifts themselves provide little information about the orientation of the strands with respect to the remaining structure, and orientational restraints derived from anisotropic interactions such as residual dipolar couplings (31) can be difficult to quantify in exchanging systems with more than a single excited state.

We wondered if new machine learning (ML)-based tools might be able to provide some insights into excited-state structures in cases where secondary structure is preserved, but where significant shift differences between ground and excited conformers indicate distinct tertiary arrangements, as for pro-IL-18. Recent advances in ML allow for rapid generative modeling of protein conformational ensembles (32–34), including for fold-switchers (35–37), but thus far only for conformations which are similarly stable to the GS. Here, we identify conformations of higher energy, excited (invisible) states by combining a large ensemble of diverse structures obtained from AlphaFlow (32), a recently developed generative framework, with knowledge of the pro-IL-18 excited states from initial CPMG and CEST NMR measurements (28). In this manner, structures within the ensemble were identified that could resemble the excited states. These potential structures were then used to suggest further NMR experiments which validated the AlphaFlow predictions, hence providing atomic-resolution models of the excited-state conformers identified through NMR relaxation experiments.

Results

AlphaFlow Predicts Two Alternative Folded Conformations of β*. To gain structural insights into the excited states of pro-IL-18 that can then be used to guide their further experimental characterization, beyond what was available through initial NMR relaxation studies, we turned to AlphaFlow. The amino acid sequence of pro-IL-18 was fed to AlphaFlow, along with its NMR structure which was provided as a template. We found that the template structure was required to prevent AlphaFlow from generating erroneous structures, similar to AlphaFold's incorrectly predicted, high confidence structure of pro-IL-18 (38) that did not incorporate the prosequence properly within the folded domain of the structure. To generate structures with greater variability within the region of interest, the coordinates of residues L45-Q60 (containing β1 and β*) were masked (i.e., withheld) in the template structure. Using this approach, 30,000 structures of pro-IL-18 were produced, showing variability within and adjacent to the β* strand. The ability to mask only the region of interest was key, as was the generation of a large ensemble, to reveal a variety of states including hidden, excited conformers. These requirements favor a rapid generative model such as AlphaFlow. Notably, attempts to use the generative deep-learning system BioEmu (33) were not successful as this tool does not allow the user to supply a template structure. The structures generated with BioEmu did not incorporate the prodomain into the β trefoil, also observed using AlphaFold, and were not consistent with the experimentally determined GS structure of pro-IL-18.

To identify distinct conformations among the 30,000 AlphaFlow structures we first used principal component analysis; however, we found that there were no clear groupings of the structures in the first two principal components, even when only the regions with conformational heterogeneity (i.e., masked regions) were considered (*SI Appendix*, Fig. S1). We next wondered whether we could use experimental chemical shifts to select members of the AlphaFlow ensemble that could be potential matches for the excited states that were obtained via analysis of CPMG and CEST experiments recorded on the WT pro-IL-18 protein. To evaluate this possibility, we compared experimentally derived chemical shifts of GS and ES_2 conformers with calculated shifts of the ensemble of GS structures predicted from AlphaFlow. Notably, the predicted GS shifts were often between the experimental values for GS and ES_2, and sometimes closer to ES_2. This indicates that chemical shift predictions of the generated models are not of sufficient accuracy to distinguish between ground and excited states in the derived ensemble, nor to link a subset of structures of the ensemble unequivocally to either ES_1 or ES_2 (*SI Appendix*, Fig. S2).

Instead, we utilized our knowledge of the excited-state structures from our initial NMR characterization to search for relevant groups of conformers within the AlphaFlow set. Using a procedure which included groupings based on i) the backbone dihedral angles of Q54 where ψ changes from ~0° (GS) to 120° (ES_2) as predicted by the program TALOS (39), which uses input backbone chemical shifts to generate the associated dihedral angles (*SI Appendix*, Text 1), and ii) the backbone dihedral angles of the residues forming β* (V55-I58), which are still in the appropriate range for a β strand for at least one of the excited states, as well as additional criteria described in (*SI Appendix*, *Materials and Methods*) and illustrated in *SI Appendix* Fig. S3, five groups of structures were identified (Fig. 2*A*). These were 1) a GS-like group, which is nearly identical to our experimentally determined structure of pro-IL-18; 2) an unfolded group (Unfolded), where residues V55-I58 do not adopt β strand-like backbone dihedral angles; 3) an alternative folded group (Alt1), where the β* strand has flipped around; 4) a second

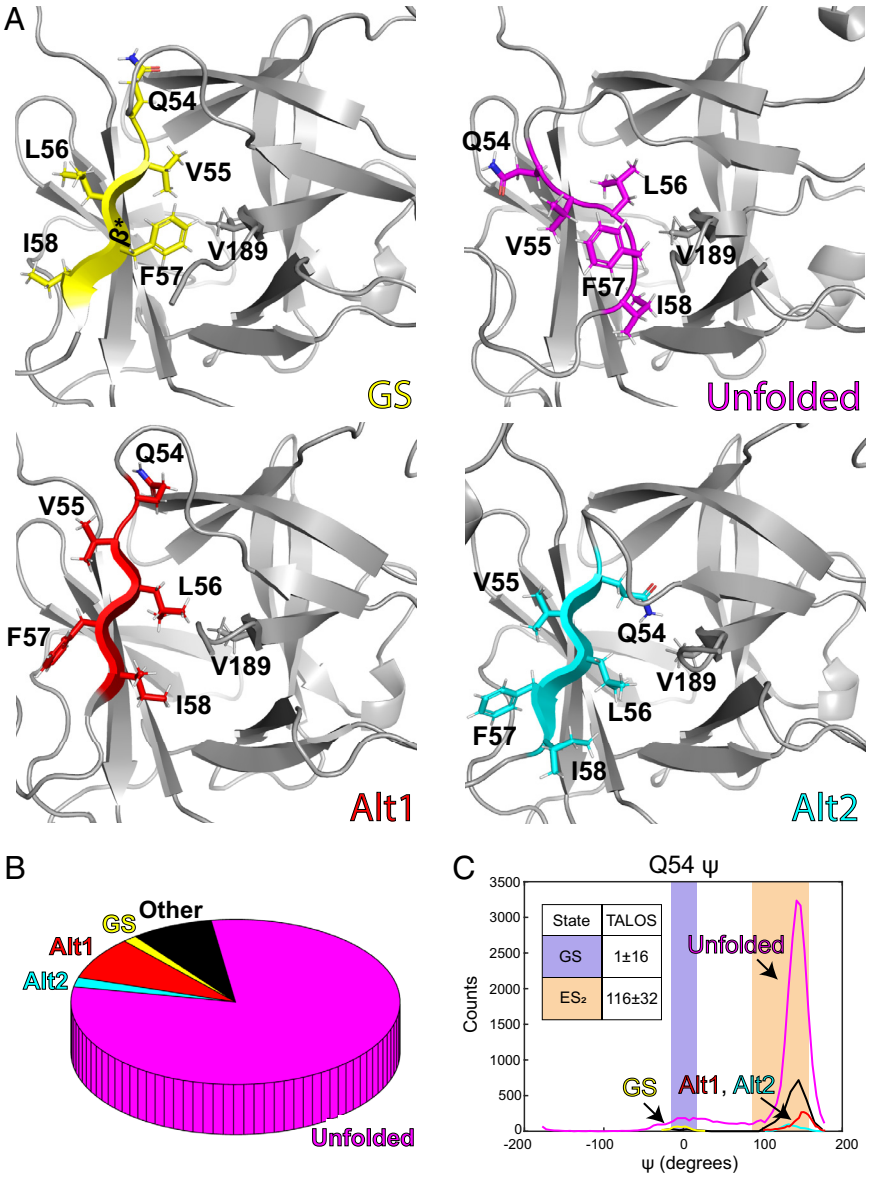

Fig. 2. AlphaFlow predicts two alternative folded conformations of β*. (*A*) Representative structures from four different classes generated by AlphaFlow: GS (similar to the experimentally determined structure), Unfolded (residues V55-I58 are disordered), Alt1 (β* strand is flipped around, shifted "up" by one residue), and Alt2 (β* strand is flipped around, shifted "down" by one residue). (*B*) Relative abundance of the structural classes across all AlphaFlow models. (*C*) Q54 ψ angle distribution from all AlphaFlow models (n = 29,950) is largely bimodal, with one mode centered about ~0° degrees (corresponding to GS), and a second mode centered on ~145° (β strand-like, corresponding to the alternative conformations, Alt1 and Alt2, as well as the "Unfolded" and "Other" classes of structures). Notably, the Q54 ψ angle prediction from TALOS for WT ES_2 (28) (*Inset*) and for ES_1 and ES_2 of the Q54V variant (present work, *SI Appendix*, Table S1) agrees with the ψ angles found in the Alt1 and Alt2 structures.

alternative folded group (Alt2), where the β* strand is again flipped, but shifted "down" by two residues relative to Alt1; 5) a final group (Other), which contains several different types of models, including those similar to GS, Alt1, and Alt2 but with incomplete interactions between the β* and β4 strands resulting from a larger distance between them, as well as models in which these strands do not interact at all (*SI Appendix*, Fig. S4). Structures from the Unfolded class were not considered further as these have β* residues that adopt conformations that are inconsistent with experiment (i.e., not β-strand). Conformers in the "Other" category were also eliminated as these lack stabilizing hydrogen bonds connecting β* with the β4 strand, and the resulting increased distance between these secondary structural elements does not support the NOEs that were observed between them in subsequent analyses (*Validation of AlphaFlow model predictions by recording methyl-methyl NOEs*).

Most of the structures generated by AlphaFlow belong to the unfolded group, and both the Alt1 and Alt2 groups contain more structures than the GS group (Fig. 2*B*), indicating that AlphaFlow does not accurately capture the relative abundances of various structures within the conformational ensemble of pro-IL-18. However, as AlphaFlow predicts two main alternative structures, Alt1 and Alt2, while our NMR experiments show that pro-IL-18 adopts two excited states, we wondered whether these predicted structures might resemble the sparse conformers identified by NMR. Alt1 and Alt2 are consistent with our initial NMR characterization both in terms of the ψ angles of Q54 for both alternative forms, which agree with the β strand-like prediction of TALOS for the Q54 ψ angle of ES_1 (*SI Appendix*, Table S1) and ES_2 (Fig. 2*C*), and with the secondary structure of V55-I58, a region which forms a β strand in both alternative structures as well as in ES_1 (*SI Appendix*, Table S1) and ES_2 (28).

Replacing Q54 with a β Strand-Favoring Residue Increases Excited-State Populations. Our previous NMR characterization of ES_2 from WT pro-IL-18 indicates that the β* strand in this sparse conformer is extended by one residue to include Q54, but not in the GS (as established by comparison of Q54 ψ values for GS and ES_2 in Fig. 2 *C*, *Inset* table). Similarly, the backbone dihedral angles of Q54 from the AlphaFlow derived Alt1 and Alt2 models are also β strand-like, and distinct from the GS for which ψ ~ 0° is obtained (Fig. 2*C*). If Alt1 and Alt2 are models for ES_1 and ES_2, then replacing Q54 with a residue that favors β strand formation would stabilize the excited states, as predicted by the

NMR data. We prepared two such mutants of Q54, Q54V and Q54I (40), both of which increase the magnitude of the dispersions observed in ^{15}N CPMG experiments, suggesting that both variants have increased excited-state populations (*SI Appendix*, Fig. S5). ^{15}N CEST data were additionally recorded on Q54V, the mutant with the largest CPMG dispersion profiles, and the combined analysis of ^{15}N CPMG and CEST data (*SI Appendix*, Tables S2 and S3) yielded fractional populations of ES_1 and ES_2 of 6.4 ± 0.2% and 7.6 ± 0.3%, respectively, an increase of roughly 30-fold for ES_1 and 20-fold for ES_2 relative to the WT protein under the same experimental conditions (Fig. 3*A* and *SI Appendix*, Table S4). The larger populations and longer lifetimes of the Q54V variant excited states (8 ms and 16 ms for ES_1 and ES_2, respectively), increase the utility of CEST experiments, and, in contrast to the WT protein where minor state dips were only observed for ES_2, the ^{1}HN, ^{15}N, ^{13}CO, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and $^{13}C^{methyl}$ CEST profiles of some of the backbone and sidechain probes now show dips for both ES_1 and ES_2 in the Q54V mutant (Fig. 3*A*). Notably, the combined analysis of the Q54V ^{15}N CPMG and CEST data required the use of a 4-state kinetic model ($ES_3 \leftrightarrow GS \leftrightarrow ES_1 \leftrightarrow ES_2$) (*SI Appendix*, Text 2 and Fig. S6), indicating that this mutation increases the population of at least one other excited state as well. In addition, the chemical shift differences between the ground and ES_2 states, $\Delta\nu_{GS,ES2} = \nu_{ES2}-\nu_{GS}$, for the WT (x-axis) and Q54V (y-axis) proteins are similar (r = 0.89), suggesting that the Q54V mutation does not significantly perturb the structure of this excited state (Fig. 3*B*). Note that a full set of backbone chemical shifts was only available for ES_2 of the WT protein (*SI Appendix*, Text 1; but available also for ES_1 in the Q54V variant, *SI Appendix*, Table S5) so that a similar comparison involving ES_1 is not possible. The correlation of $\Delta\nu_{GS,ES2}$ values is scattered, as expected, as the site of mutation is in the region where large chemical shift differences are found (Fig. 1, blue).

Validation of AlphaFlow Model Predictions by Recording Methyl-Methyl NOEs. As the Q54V mutant increases the populations of ES_1 and ES_2 by over an order of magnitude relative to the WT protein, additional experiments can be recorded using it to validate the AlphaFlow-predicted conformers. The β* strand and its surroundings contain numerous Ile, Leu, and Val residues, and consequently there are several pairs of methyl groups which are only close together in Alt1 (red arrows), Alt2 (blue arrows), or in both (black arrows) where the β* strand is flipped around. This is shown schematically in Fig. 4*A*, top and in the distograms below, highlighting the expected carbon–carbon distances for some of the interactions based on analysis of the AlphaFlow predicted GS, Alt1, and Alt2 ensembles. Thus, observing Nuclear Overhauser Effect crosspeaks (NOEs) connecting methyl groups that are predicted to be proximal in Alt1/Alt2 but not in the GS in the Q54V mutant would provide evidence that Alt1 and Alt2 are structural models of the pair of excited states observed experimentally. This would then be further corroborated by the absence of such NOEs, or their significant reduction, in spectra recorded of the WT protein where the populations of ES_1/ES_2 are much smaller. Fig. 4*B* shows a schematic of the 3D CCH Nuclear Overhauser Effect Spectroscopy (NOESY) experiment that was used in these studies, focusing on the magnetization transfer pathway that is most relevant for Alt1/Alt2 validation. During all three chemical shift encoding periods, the signal of interest derives from the GS. However, during the mixing time conversion from the ground to the excited states and subsequent build-up of NOEs occurs within these lowly populated conformers, as has been reported previously by Vallurupalli and coworkers (41). The cross peaks of interest, thus, report on interactions in the excited state, but are observed through GS chemical shifts, providing a facile route to validate the alternative structures predicted by AlphaFlow.

CCH NOESY datasets were recorded on a 1 mM, highly deuterated, ILV-methyl labeled sample of Q54V pro-IL-18 (*SI Appendix*, *Materials and Methods*), as well as on an equivalently labeled and concentration matched WT sample as a control. There are five pairs of methyl groups which are expected to be close together in both Alt1 and Alt2 (Alt1 + Alt2), but distant in the GS, including both δ methyls of L56 and both γ methyls of V189, giving rise to four combinations of NOEs, and V55γ2-I82δ1 (Fig. 4*A* and *SI Appendix*, Table S6). In the CCH NOESY experiment recorded on the Q54V variant, cross peaks (weak relative to those originating from the GS) are observed between all five of these methyl pairs, and three of these five were sufficiently resolved to be assigned (Fig. 4*C*). Importantly, there is comparatively very little to no signal observed between these methyl pairs in the WT dataset (Fig. 4*C*, yellow trace; *SI Appendix*, Fig. S7). This indicates that the NOESY cross peaks in the spectrum of Q54V originate from magnetization transfer in ES_1 and/or ES_2 and are not simply the result of spin diffusion within the GS. Thus, L56δ1/2-V189γ1/2 and V55γ2-I82δ1 are close together in at least one of the excited

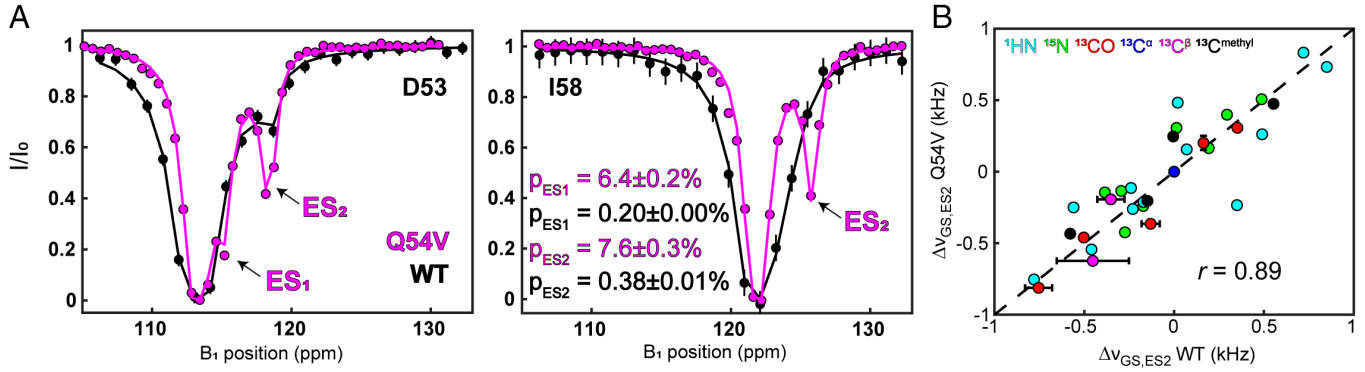

Fig. 3. Replacing Q54 with a strand-favoring residue increases excited-state populations. (*A*) ^{15}N CEST profiles for D53 and I58, 25 °C. Profiles from the Q54V mutant (20 Hz spin lock, magenta) are overlaid with those from the WT (40 Hz spin lock, black), showing significantly more pronounced excited-state dips for the mutant despite the lower spin lock power. The CEST profiles have been shifted so that the GS dips overlap, and each point has been divided by $\exp(-R_1T_{CEST})$ such that each baseline is at I/I_0 = 1 for ease of comparison. The Q54V data (4-state, ^{15}N CPMG recorded at 1 GHz and 800 MHz and one ^{15}N CEST dataset, *SI Appendix*, Tables S2 and S3) are fit with populations for ES_1 and ES_2 that are more than an order of magnitude larger than those of the WT. (*B*) Chemical shift differences between ground and ES_2 states of Q54V and WT (^{1}HN, ^{15}N, ^{13}CO, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{13}C^{methyl}$) are in good agreement (r = 0.89), indicating that the mutation has not significantly changed the conformation of this excited state; corresponding shifts for ES_1 are not available for the WT protein as the exchange kinetics preclude observation of CEST dips for this conformer (28). The $\Delta\nu_{GS,ES2}$ chemical shifts are given in kHz, calculated assuming a 1 GHz magnetic field. Chemical shifts of D53, Q/V54, and V55 are omitted from this comparison due to proximity to the site of mutation.

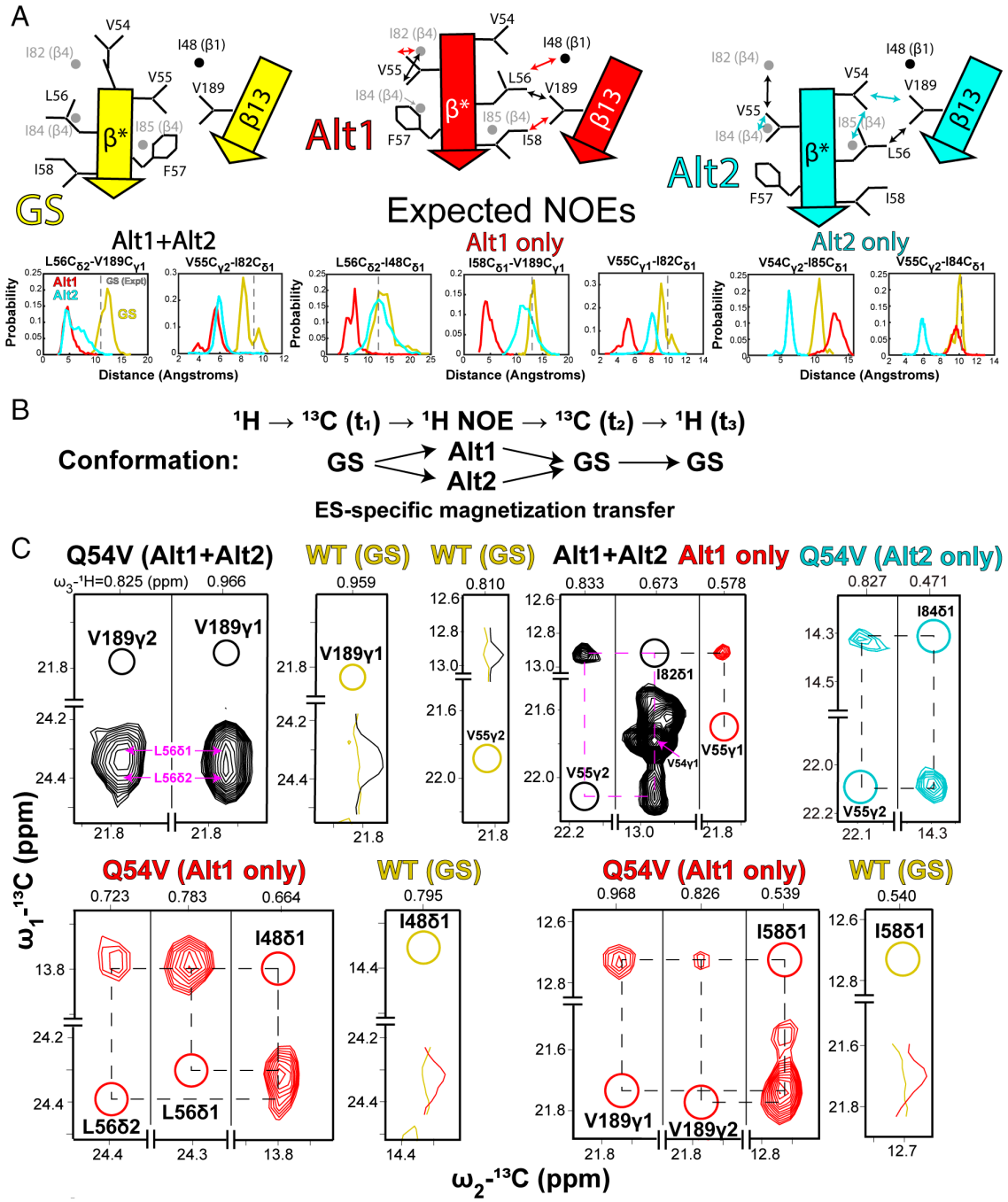

Fig. 4. Validation of AlphaFlow models via methyl-methyl NOEs. (*A*) (*Top*) Orientation of β* is shown for GS, Alt1, and Alt2 structures, with proximal methyl groups in either Alt1 (red), Alt2 (cyan), or both (black), but distant in GS, marked with arrows. For simplicity, only a single arrow connects proximal sidechains, even if NOEs are observed between all possible methyl groups in each sidechain. Sidechains for residues from β* and β13 are indicated explicitly, while those from β1 and β4 are denoted by circles. (*Bottom*) Distance distributions between selected methyl carbons among all Q54V AlphaFlow models from the GS, Alt1, and Alt2 classes are highlighted. The dashed line indicates the distance determined from the GS, WT NMR structure (PDB 8URV) (42). (*B*) Schematic of the methyl-methyl CCH NOESY experiment, focusing on magnetization transfer steps that give rise to NOEs connecting proximal methyl groups within the excited states, observed through the GS. (*C*) CCH NOESY slices for Q54V and WT pro-IL-18. NOEs derived from methyl groups close together in both Alt1 and Alt2 (but distant in GS), including L56δ1,2 → V189γ1,2 (reciprocal cross peaks are observed but are overlapped with other signals) and V55γ2-I82δ1 (both reciprocal cross peaks identified), give rise to cross peaks in the Q54V dataset (black) but not in the WT spectrum, or are much weaker (gold; compare black vs gold 1D traces in the adjacent WT slice). NOEs between proximal methyl groups only in Alt1, including I82δ1 → V55γ1, I48δ1-L56δ1,2 and I58δ1-V189γ1,2, lead to cross peaks in spectra recorded on samples of Q54V (red) but not of the WT (gold; compare red vs gold 1D traces in the proximal WT slice). 1D traces from the Q54V dataset are shifted to align with the corresponding traces from the spectrum recorded on the WT protein. Also shown is a pair of NOESY peaks for Alt2 connecting V55γ2 with I84δ1(plotted at a 30% lower level than all other slices). Diagonal peaks are not shown because of dynamic range issues but are indicated with circles and labeled to highlight the destination of magnetization transfer in the experiment.

states, validating the flip of the β* strand predicted by AlphaFlow. Notably, very weak cross peaks between L56δ1/2 and V189γ1 are observed in spectra recorded on the WT protein as well (*SI Appendix*, Fig. S8), further indicating that the Q54V excited states are structurally similar to those of the WT.

In addition, there are five pairs of methyl groups which are expected to be close only in Alt1, and distant in the GS and Alt2. These include both δ methyls of L56 that are predicted to be proximal to I48δ1, I58δ1 that is proximal to both γ methyls of V189, and a V55γ1–I82δ1 contact (Fig. 4*A*, red arrows and

SI Appendix, Table S6). In the CCH NOESY experiment recorded on the Q54V sample, cross peaks are observed between all five of these methyl pairs (Fig. 4*C*, red), with the corresponding cross peaks absent in the WT dataset (Fig. 4*C*, yellow trace; *SI Appendix*, Fig. S7). The presence of these cross peaks, along with those connecting L56δ1/2-V189γ1/2 and V55γ2-I82δ1, indicate that the Alt1 structure closely resembles one of the two experimentally observed excited states.

Several methyl pairs are also predicted to be proximal only in Alt2 and distant in both the GS and Alt1, Fig. 4*A*. Although spectral overlap prevents a definitive assignment for many of these to distinct interactions, NOE contacts between V54γ2-I85δ1 (*SI Appendix*, Fig. S7 and Table S6) and V55γ2-I84δ1 (Fig. 4*C* and *SI Appendix*, Table S6) are observed that are telltale reporters of Alt2.

The presence of Alt1/Alt2 NOEs in spectra of the Q54V pro-IL-18 variant and their absence or significant attenuation in the dataset recorded on the WT protein provide strong evidence that the observed NOEs report faithfully on the proximity of residues in the Alt structures. It is highly unlikely that the observed NOEs assigned to the Alt models are the result of spin diffusion within the GS due to the introduction of additional methyl groups at residue 54, as established through analysis of members of the GS structural ensemble of the Q54V variant (*SI Appendix*, Fig. S9).

In addition to the NOE connectivities, other experimental data confirm important features of the Alt1/2 structures. For example, a TALOS analysis of ^{1}HN, ^{15}N, ^{13}CO, ^{13}C$^{\alpha}$, and ^{13}C$^{\beta}$ CEST-derived chemical shifts of ES_1 and ES_2 of the Q54V variant establishes that residues V54-I58 are in β-sheet conformations, with the ψ of V54 transitioning from ~0° in the GS to ~130° in both ES_1 and ES_2 (*SI Appendix*, Table S1). This change in ψ is a critical feature in the flip of the β* strand, providing strong evidence that both excited-state conformations have similar flipped orientations and further validating the AlphaFlow models. Additional evidence in support of the robustness of the Alt1 and Alt2 structures derives from a qualitative evaluation of *relative* intensities of NOEs. For example, NOEs connecting L56δ1/2 with V189γ1/2 are much stronger than the corresponding I58δ1-Val189γ1/2 correlations, as well as other Alt1/2 specific NOEs (Fig. 4*C* and *SI Appendix*, Fig. S10). This can be explained, at least qualitatively, by the AlphaFlow models where residues 56 and 189 are close in both Alt1 and Alt2, while I58 and V189 are in similarly close proximity in Alt1 but distal in Alt2. Thus, both Alt conformers contribute to the L56-V189 NOE intensity, while only Alt1 is relevant for the I58-V189 NOE. Finally, as described above in this section, our Alt models show contacts between β* and β4 strands, as observed in the NOE dataset recorded on the Q54V variant; such contacts are not present in the "Other" class of computed structures (*AlphaFlow predicts two alternative folded conformations of β**), justifying the grouping of structures used in the present analysis.

Discussion

The development of computational approaches for predicting protein structures has transformed the field of structural biology and it is now possible to generate three-dimensional models of biomolecules and their complexes with a high degree of confidence. However, computational approaches are not at the stage where they can predict biomolecular dynamics and energy landscapes of biomolecules with the same level of accuracy as GS structures. Here, we present a protocol which combines strengths of both NMR spin relaxation experiments and AlphaFlow to obtain validated structural models for sparsely populated and transiently formed pro-IL-18 conformers that are invisible to traditional biophysical experiments.

The protocol begins with a combined CPMG and CEST NMR study which establishes that pro-IL-18 interconverts between a GS and a pair of excited states, enables quantification of the kinetics and thermodynamics of the exchange process, localizes the region of interconversion, and informs on the secondary structure of the exchanging regions in the sparse states. However, tertiary information is not forthcoming. Although the Q54V mutation, as suggested by the relaxation experiments and further by subsequent AlphaFlow modeling, results in an approximately 20-fold increase in populations of the sparse states (from ~0.3 to 7%) leading, therefore, to significantly more robust experimental data, the chemical shifts of the excited states by themselves are not sufficient to obtain structural models. Indeed, the chemical shifts do not readily distinguish ES_1 and ES_2 (*SI Appendix*, Fig. S2*B*), as might be expected from the fact that both excited states maintain the same secondary structure. In this regard our AlphaFlow strategy is complementary to the NMR approach because it produces a distribution of structures that can both be analyzed using existing NMR data and used for developing additional experiments to generate models of the excited states. In this manner a pair of AlphaFlow models are identified, Alt1 and Alt2, as potential candidates. Both Alt structures show that Q54 becomes part of a β-sheet, distinct from the GS and consistent with the NMR results, supporting the use of a Q54V variant to stabilize the excited states experimentally, as valine favors formation of β-strands more than glutamine (40). The complementarity is further established by the fact that Alt1 and Alt2 models highlight the utility of methyl groups of Ile, Leu, and Val sidechains in and around the β* strand as ideal probes to characterize these conformers experimentally. The AlphaFlow models also provide a framework with which to analyze the Q54V methyl-NOE dataset, which contains cross peaks from both excited states, enabling validation of Alt1 and Alt2 as structural models of the pro-IL-18 excited states. The importance of NMR relaxation data to this process is underscored by the fact that the region undergoing exchange between ground and excited states needs to be masked in the generation of AlphaFlow conformers, since when provided the complete (i.e., unmasked) structure as a template AlphaFlow was unable to produce structures other than the GS (*SI Appendix*, Fig. S11), and if a structural template is not provided correct structures of pro-IL-18 cannot be generated at all. The NMR data further provide an avenue to experimentally validate the utility of AlphaFlow, and more generally any generative computational model, for predicting invisible higher energy states on a protein's energy landscape.

AlphaFlow generated models of the excited states of pro-IL-18 indicate that the β* strand is flipped relative to the GS and that the primary difference in the excited-state structures lies in a register shift of two amino acids in β*. β-strand flipping and/or slipping has been observed in a variety of proteins, including in ubiquitin (27, 43), in mutants of the β-sheet OspA protein (44), and in protein–peptide complexes that are linked via intermolecular β-sheets (45). In the case of pro-IL-18 these less stable configurations may play an important role in promoting unfolding of β* to prime binding of pro-IL-18 to caspases-1 and -4 for cleavage to the mature state of the protein.

The present study provides an illustration of how a machine learning/NMR strategy can be used to discover excited conformational states of the pro-IL-18 protein, but it is, of course, only one such example. In an effort to establish whether this approach is more general we attempted to examine other systems, such as L99A T4 lysozyme and the FF domain, for which excited-state

structures were determined de novo using NMR relaxation methods (10, 19), or Abl kinase, where sparse conformers identified by NMR have been shown to play a significant role in resistance to drugs against leukemia used in the clinic (3). Notably, AlphaFlow has been trained on structures of both ground and excited states of these proteins making it difficult to properly assess the efficacy of our approach. However, using the ψ angle of F114 to group L99A T4 lysozyme structures as either "ground" or "excited" in the structural ensemble of this protein produced using AlphaFlow, where only the region identified as undergoing conformational exchange by NMR was masked, we were able to assign computed structures to an excited state that agreed with the de novo determined conformation by NMR. In this case F114, for which a large change in ψ is observed between ground and excited states, essentially mimics Q54 of pro-IL-18. While certainly not establishing generality, these two cases, involving three sparse conformers, suggest that where changes in conformation involve significant rearrangements of one or more dihedral angles it may be possible to use this information to select appropriate models for excited states observed experimentally, assuming that such models can be predicted in the first place.

Although the role of ML in structural biology will continue to advance rapidly (46), applications to studies of excited conformers will remain challenging, at least in the near future, due to the sparse data that are presently available for training. For example, while for pro-IL-18 it is possible to generate structural ensembles using AlphaFlow that include the pair of excited states that are observed by NMR, the populations of members of the ensemble are not consistent with experiment. In the general case, therefore, it will be difficult to rely exclusively on computational approaches to faithfully select low-lying excited-state conformers without recourse to experimental data. In this regard, a combined NMR/ML strategy for studies of biomolecular dynamics (47) provides a powerful approach for exploring energy landscapes of increasing complexity in ways that cannot be achieved with either method alone.

Materials and Methods

Cloning, Expression, Purification, and Sample Preparation for NMR Studies. Isotopically labeled samples of pro-IL-18 were expressed and purified from *Escherichia coli* using a previously established protocol (42). For a detailed description of the protocol, see (*SI Appendix*, *Materials and Methods*). Samples were concentrated to ~0.5 to 1 mM in 20 mM MES, 50 mM KCl, and 10 mM DTT at pH 6.5, 3% D_2O, with 0.5 mM EDTA added to each sample.

NMR Spectroscopy. NMR experiments were collected at 25 °C on Bruker AVANCE NEO 23.5 T (1.0 GHz), AVANCE III HD 18.8 T (800 MHz), and AVANCE III HD 14.1 T (600 MHz) NMR spectrometers equipped with 5-mm TCI triple-axis gradient cryoprobes. For a detailed description of the experimental setup and data analysis, see (*SI Appendix*, *Materials and Methods*).

Generation of Structural Ensembles Using AlphaFlow. 30,000 conformations of pro-IL 18 (both WT and Q54V) were generated using the 48-layer MD + template base version of AlphaFlow in which the NMR structure of the protein is provided to the model as a template. Residues L45–Q60 were masked by setting the atomic coordinates to zeros in the template. For a detailed description of the structure generation and analysis, see (*SI Appendix*, *Materials and Methods*).

Data, Materials, and Software Availability. Structures, NMR spectra, NMR relaxation data and chemical shifts; WT pro-IL-18 GS chemical shifts have been deposited in Zenodo (48) and Biological Magnetic Resonance Data Bank (BMRB) (49). All other data are included in the manuscript and/or *SI Appendix*.

ACKNOWLEDGMENTS. This work was supported through Grants from the Canadian Institutes of Health Research (CIHR, FND-503573) and the Natural Sciences and Engineering Council of Canada (024-03872) to L.E.K. J.P.B. acknowledges salary funding in the form of a postdoctoral fellowship from the CIHR.

Author affiliations: [a]Department of Molecular Genetics, University of Toronto, Toronto, ON M5S 1A8, Canada; [b]Department of Biochemistry, University of Toronto, Toronto, ON M5S 1A8, Canada; [c]Department of Chemistry, University of Toronto, Toronto, ON M5S 3H6, Canada; [d]Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, ON M5S 3E1, Canada; [e]Program in Molecular Medicine, The Hospital for Sick Children Research Institute, Toronto, ON M5G 0A4, Canada; and [f]Department of Computer Science, University of Toronto, Toronto, ON M5S 3E1, Canada

1. M. Karplus, J. Kuriyan, Molecular dynamics and protein function. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 6679–6685 (2005).
2. M. Kurplus, J. A. McCammon, Dynamics of proteins: Elements and function. *Annu. Rev. Biochem.* **52**, 263–300 (1983).
3. T. Xie, T. Saleh, P. Rossi, C. G. Kalodimos, Conformational states dynamically populated by a kinase determine its function. *Science* **370**, eabc2754 (2020).
4. A. L. Hansen, X. Xiang, C. Yuan, L. Bruschweiler-Li, R. Brüschweiler, Excited-state observation of active K-Ras reveals differential structural dynamics of wild-type versus oncogenic G12D and G12C mutants. *Nat. Struct. Mol. Biol.* **30**, 1446–1455 (2023).
5. D. D. Boehr, D. Mcelheny, H. J. Dyson, P. E. Wright, The dynamic energy landscape of dihydrofolate reductase catalysis. *Science* **313**, 1638–1642 (2006).
6. K. A. Henzler-Wildman *et al.*, A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* **450**, 913–916 (2007).
7. S. K. Whittier, A. C. Hengge, J. P. Loria, Conformational motions regulate phosphoryl transfer in related protein tyrosine phosphatases. *Science* **341**, 899–903 (2013).
8. J. S. Fraser *et al.*, Hidden alternative structures of proline isomerase essential for catalysis. *Nature.* **462**, 669–673 (2009).
9. P. Neudecker *et al.*, Structure of an intermediate state in protein folding and aggregation. *Science* **336**, 362–366 (2012).
10. D. M. Korzhnev, T. L. Religa, W. Banachewicz, A. R. Fersht, L. E. Kay, A transient and low-populated protein-folding intermediate at atomic resolution. *Science* **329**, 1312–1316 (2010).
11. A. Ceccon, V. Tugarinov, R. Ghirlando, G. M. Clore, Abrogation of prenucleation, transient oligomerization of the Huntingtin exon 1 protein by human profilin I. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 5844–5852 (2020).
12. M. Baek *et al.*, Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).
13. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature.* **596**, 583–589 (2021).
14. A. M. Ille, E. Anas, M. B. Mathews, S. K. Burley, From sequence to protein structure and conformational dynamics with artificial intelligence/machine learning. *Struct. Dynam.* **12**, 030902 (2025).
15. A. G. Palmer, C. D. Kroenke, Loria J. Patrick, Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol.* **339**, 204–238.
16. H. Y. Carr, E. M. Purcell, Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Phys. Rev.* **94**, 630–638 (1954).
17. S. Meiboom, D. Gill, Modified spin-echo method for measuring nuclear relaxation times. *Rev. Sci. Instrum.* **29**, 688–691 (1958).
18. P. Vallurupalli, G. Bouvignies, L. E. Kay, Studying "invisible" excited protein states in slow exchange with a major state conformation. *J. Am. Chem. Soc.* **134**, 8148–8161 (2012).
19. G. Bouvignies *et al.*, Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. *Nature* **477**, 111–114 (2011).
20. D. M. Korzhnev *et al.*, Nonnative interactions in the FF domain folding pathway from an atomic resolution structure of a sparsely populated intermediate: An NMR relaxation dispersion study. *J. Am. Chem. Soc.* **133**, 10974–10982 (2011).
21. K. Madhurima, B. Nandi, S. Munshi, A. N. Naganathan, A. Sekhar, Functional regulation of an intrinsically disordered protein via a conformationally excited state. *Sci. Adv.* **9**, eadh4591 (2023).
22. J. B. Stiller *et al.*, Structure determination of high-energy states in a dynamic protein ensemble. *Nature* **603**, 528–535 (2022).
23. P. Vallurupalli, D. F. Hansen, L. E. Kay, Structures of invisible, excited protein states by relaxation dispersion NMR spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 11766–11771 (2008).
24. N. L. Fawzi, J. Ying, D. A. Torchia, G. M. Clore, Probing exchange kinetics and atomic resolution dynamics in high-molecular-weight complexes using dark-state exchange saturation transfer NMR spectroscopy. *Nat. Protoc.* **7**, 1523–1533 (2012).
25. C. Charlier, J. M. Courtney, P. Anfinrud, A. Bax, Interrupted pressure-jump nmr experiments reveal resonances of on-pathway protein folding intermediate. *J. Phys. Chem. B* **122**, 11792–11799 (2018).
26. R. Kitahara, S. Yokoyama, K. Akasaka, NMR snapshots of a fluctuating protein structure: Ubiquitin at 30 bar to 3 kbar. *J. Mol. Biol.* **347**, 277–285 (2005).
27. E. Masoumzadeh *et al.*, Structure of a transient protein-folding intermediate by pressure-jump NMR spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.* **122**, e2519493122 (2025).
28. J. P. Bonin, J. M. Aramini, L. E. Kay, Structural plasticity as a driver of the maturation of pro-interleukin-18. *J. Am. Chem. Soc.* **146**, 30281–30293 (2024).
29. P. Devant *et al.*, Structural insights into cytokine cleavage by inflammatory caspase-4. *Nature* **624**, 451–459 (2023).
30. G. Kaplanski, Interleukin-18: Biological properties and role in disease pathogenesis. *Immunol. Rev.* **281**, 138–153 (2018).
31. P. Vallurupalli, D. F. Hansen, E. Stollar, E. Meirovitch, L. E. Kay, Measurement of bond vector orientations in invisible excited states of proteins. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 18473–18477 (2007).
32. B. Jing, B. Berger, T. Jaakkola, AlphaFold meets flow matching for generating protein ensembles. arXiv [Preprint] (2024), https://doi.org/10.48550/arXiv.2402.04845 (Accessed 2 April 2026).

33. S. Lewis *et al.*, Scalable emulation of protein equilibrium ensembles with generative deep learning. *Science* **389**, eadv9817 (2025).
34. Y. Kalakoti, B. Wallner, AFsample2 predicts multiple conformations and ensembles with AlphaFold2. *Commun Biol.* **8**, 373 (2025).
35. H. K. Wayment-Steele *et al.*, Predicting multiple conformations via sequence clustering and AlphaFold2. *Nature.* **625**, 832–839 (2024).
36. M. Lee *et al.*, Large-scale predictions of alternative protein conformations by AlphaFold2-based sequence association. *Nat. Commun.* **16**, 5622 (2025).
37. Y. J. Huang, T. A. Ramelot, L. E. Spaman, N. Kobayashi, G. T. Montelione, Hidden structural states of proteins revealed by conformer selection with AlphaFold-NMR. bioRxiv [Preprint] (2024), 10.1101/2024.06.26.600902 (Accessed 2 April 2026).
38. J. P. Bonin, J. M. Aramini, Y. Dong, H. Wu, L. E. Kay, AlphaFold2 as a replacement for solution NMR structure determination of small proteins: Not so fast! *J. Magn. Reson.* **364**, 107725 (2024).
39. Y. Shen, F. Delaglio, G. Cornilescu, A. Bax, TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR.* **44**, 213–223 (2009).
40. L. Regan, Protein structure: Born to be beta. *Curr. Biol.* **4**, 656–658 (1994).
41. D. De *et al.*, Mapping the FF domain folding pathway via structures of transiently populated folding intermediates. *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2416682121 (2024).
42. Y. Dong *et al.*, Structural transitions enable interleukin-18 maturation and signaling. *Immunity* **57**, 1533–1548 (2024).
43. C. Gladkova *et al.*, An invisible ubiquitin conformation is required for efficient phosphorylation by PINK1. *EMBO J.* **36**, 3555–3572 (2017).
44. K. Makabe, S. Yan, V. Tereshko, G. Gawlak, S. Koide, β-Strand flipping and slipping triggered by turn replacement reveal the opportunistic nature of β-strand pairing. *J. Am. Chem. Soc.* **129**, 14661–14669 (2007).
45. M. T. Panteva, R. Salari, M. Bhattacharjee, L. T. Chong, Direct observations of shifts in the β-Sheet register of a protein-peptide complex using explicit solvent simulations. *Biophys. J.* **100**, L50–L52 (2011).
46. G. Janson, M. Feig, Generation of protein dynamics by machine learning. *Curr. Opin. Struct. Biol.* **93**, 103115 (2025).
47. H. K. Wayment-Steele *et al.*, Learning millisecond protein dynamics from what is missing in NMR spectra. bioRxiv [Preprint] (2025), https://doi.org/10.1101/2025.03.19.642801 (Accessed 2 April 2026).
48. J. P. Bonin, J. S. Lee, Z. H. Liu, P. M. Kim, L. E. Kay, Making invisible excited state protein structures visible by combining NMR and machine learning. Zenodo. https://zenodo.org/records/19186893. Deposited 23 March 2026.
49. J. P. Bonin, J. M. Aramini, L. E. Kay, Solution NMR structure of pro-IL-18. Biological Magnetic Resonance Data Bank. https://bmrb.io/data_library/summary/index.php?bmrbId=31122. Deposited 26 October 2023.