Hidden Structural States of Proteins Revealed by Conformer Selection with AlphaFold-NMR

Yuanpeng J. Huang* and Gaetano T. Montelione*

Dept of Chemistry and Chemical Biology, Center for Biotechnology and Interdisciplinary Sciences, Rensselaer Polytechnic Institute, Troy, New York, 12180 USA

Abstract

Recent advances in molecular modeling using deep learning can revolutionize our understanding of dynamic protein structures. NMR is particularly well-suited for determining dynamic features of biomolecular structures. The conventional process for determining biomolecular structures from experimental NMR data involves its representation as conformation-dependent restraints, followed by generation of structural models guided by these spatial restraints. Here we describe an alternative approach: generating a distribution of realistic protein conformational models using artificial intelligence-(AI-) based methods and then selecting the sets of conformers that best explain the experimental data. We applied this *conformational selection* approach to redetermine the solution NMR structure of the enzyme Gaussia luciferase. First, we generated a diverse set of conformer models using AlphaFold2 (AF2) with an enhanced sampling protocol. The models that best-fit NOESY and chemical shift data were then selected with a Bayesian scoring metric. The resulting models include features of both the published NMR structure and the standard AF2 model generated without enhanced sampling. This "AlphaFold-NMR" protocol also generated an alternative "open" conformational state that fits nearly as well to the overall NMR data but accounts for some NOESY data that is not consistent with first "closed" conformational state; while other NOESY data consistent with this second state are not consistent with the first conformational state. The structure of this "open" structural state differs from that of the "closed" state primarily by the position of a thumb-shaped loop between α-helices H5 and H6, revealing a cryptic surface pocket. These alternative conformational states of Gluc are supported by "double recall" analysis of NOESY data and AF2 models. Additional structural states are also indicated by backbone chemical shift data indicating partially-disordered conformations for the C-terminal segment. Considered as a multistate ensemble, these multiple states of Gluc together fit the NOESY and chemical shift data better than the "restraint-based" NMR structure and provide novel insights into its structure-dynamic-function relationships. This study demonstrates the potential of AI-based modeling with enhanced sampling to generate conformational ensembles followed by conformer selection with experimental data as an alternative to conventional restraint satisfaction protocols for protein NMR structure determination.

*Corresponding authors. Email: huangy26@rpi.edu, monteg3@rpi.edu

Key words: Solution NMR, AlphaFold2, Gaussia luciferase, Bayesian scoring function, conformer selection, conformational dynamics, enhanced sampling.

Abbreviations: AF2 – AlphaFold2 multimer; AFsample – AlphaFold2 with enhanced sampling; AI – Artificial Intelligence; CI – statistical Confidence Interval; LDDT – Local Difference Distance Test score, a superimposition independent metric of structure prediction accuracy; NOE – Nuclear Overhauser Effect; Gluc – Gaussia luciferase; NOESY – NOE SpectroscopY; pLDDT – predicted per-residue LDDT from AF2; <pLDDT> - pLDDT score averaged over all residues in the protein sequence; $pLDDT_{avg}$ – average per-residue pLDDT across a set of models; PSVS – Protein Structure Validation Software suite; pTM – predicted TM score for overall protein structure from AF2; RCI – Random Coil Index for assessing conformational flexibility from chemical shift data; $RCI_{0.6}$ – Random Coil Index scaled by 1/0.6; RCI_{rmsf} – Random Coil Index scaled for RMSF comparison; RDC – Residual Dipolar Coupling; RMSD – Root Mean Squared Deviation; RMSF - Root Mean Squared atomic Fluctuation; RPF-DP score – Recall, Precision, F-measure, and Discrimination Power score; SA – surface area; SCC – Spearman's correlation coefficient; SEM – standard error of the mean; TM score – template model score, a measure of structural similarity.

Introduction

Recent advances in molecular modeling using deep learning methods of artificial intelligence (AI) have the potential to revolutionize the process of experimental protein structure determination. For example, for relatively rigid proteins the AlphaFold2 (AF2) deep learning platform (Jumper *et al.*, 2021) has been shown to generate models of protein structures in excellent agreement with experimental data, even for "blind" protein targets which were not used in the machine learning training process (Huang *et al.*, 2021; Kryshtafovych *et al.*, 2021; Li *et al.*, 2023). AF2 structures are now routinely used as phasing models for molecular replacement in X-ray crystallography (Terwilliger *et al.*, 2023; Terwilliger *et al.*, 2024), and to guide analysis and model building using cryogenic electron microscopy (cryoEM), small angle X-ray scattering (SAXS), and NMR [for examples see (Robertson *et al.*, 2021; Fontana *et al.*, 2022; Heo *et al.*, 2022; Tejero *et al.*, 2022; Chinnam *et al.*, 2023; Li *et al.*, 2023)]. Although they were not trained to model structural dynamics, AF2-based methods are also useful in identifying flexible or dynamic regions of proteins (Jumper *et al.*, 2021; Ruff and Pappu, 2021; Guo *et al.*, 2022; Saldano *et al.*, 2022; Ma *et al.*, 2023), compute alternative conformational states of proteins (Del Alamo *et al.*, 2022; Heo and Feig, 2022; Stein and McHaourab, 2022; Sala *et al.*, 2023; Jing *et al.*, 2024; Wayment-Steele *et al.*, 2024), and more recently for modeling populations of alternative protein conformational states (Sedinkin *et al.*, 2023; Monteiro da Silva *et al.*, 2024; Zheng *et al.*, 2024). However, despite their potential impact, the accuracy of these evolving methods for modeling multiple conformational states, and particularly for predicting their Boltzmann-weighted populations, has not been extensively tested by rigorous comparison with experimental data.

NMR-based methods of experimental protein structure determination in solution have tremendously enhanced our understanding of molecular recognition processes and relationships between protein dynamics and biological function. Although small protein structures determined by X-ray crystallography and NMR are generally quite similar, significant and important differences are sometimes observed. These differences are generally attributable to the plasticity of protein structures. For example, in some cases only one (or a subset) of the multiple conformational states populated by the protein in solution may be selected by the crystal lattice in the crystallization process. In other cases, because of resonance line broadening due to slow molecular tumbling and/or the spectroscopic features of conformational exchange, only incomplete NMR data is obtainable, and the resulting model based on sparse restraints derived from these data has low accuracy and/or high uncertainty.

The canonical process for determining biomolecular structures from experimental NMR data involves representing these data as conformation-dependent restraints, followed by generation of protein structure models guided by these spatial restraints (Braun *et al.*, 1981; Williamson *et al.*, 1985). This approach has many strengths, but suffers from various challenges in generating accurate distance restraints from NMR data. These challenges include difficulties in calibrating interatomic distance restraints, ambiguities arising from the lack of stereospecific resonance assignments for prochiral atom sets, and resonance degeneracies that complicate the assignment of Nuclear Overhauser Enhancement spectroscopy (NOESY) cross peaks (Tejero *et al.*, 2013; Baskaran *et al.*, 2024). In addition, transiently-populated conformational states may yield only sparse NOESY data insufficient for modeling the corresponding structures using conventional restraint-satisfaction methods. NOEs arising from multiple conformational states may lead to conflicting distance restraints, resulting in models that only partially satisfy the restraints and/or overly constrained and unrealistic structures. An alternative approach to biomolecular structure determination is to compute a collection of realistic protein conformational models and select the one (or more) conformers from the distribution that best explain the experimental data (Mondal *et al.*, 2023). Although

this approach has rarely been used historically, the recent advances in protein structure modeling, and their potential for generating accurate structures of single or multiple conformational states, suggests that *conformer selection* could be sufficiently reliable to be used routinely for protein structure determination from NMR data.

Recent results have demonstrated that AlphaFold2 (AF2) models of small, relatively rigid proteins can rival the accuracy of their experimental NMR structures (Huang *et al.*, 2021; Robertson *et al.*, 2021; Zweckstetter, 2021; Tejero *et al.*, 2022; Li *et al.*, 2023), despite the fact that NMR structures were specifically excluded from the AF2 training data (Jumper *et al.*, 2021). However, for flexible protein structures, AlphaFold models may not always match the dominant solution structures (Huang *et al.*, 2021; Saldano *et al.*, 2022; Sala *et al.*, 2023; Bonin *et al.*, 2024; Wayment-Steele *et al.*, 2024) . The accuracy of AF2 modeling was initially demonstrated in the Critical Assessment of Protein Structure Prediction 14 (CASP14), which included three NMR-determined protein targets, each classified as a difficult prediction target lacking a homologous protein structure template for comparative modeling (Huang *et al.*, 2021). For two of these CASP14 targets, AF2 models were observed to be an excellent fit to the experimental NMR data (although for one of these CASP targets, the original NMR data and NMR-based model had to be corrected by the corresponding authors subsequent to the CASP14 assessment process). However, for the third CASP14 NMR target (target T1027), Gaussia luciferase (Gluc), significant structural differences were observed between the computed AF2 model and NMR-based experimental model (Wu *et al.*, 2020; Huang *et al.*, 2021). These differences primarily involve two helices (helix H1 and helix H2) in the N-terminal segment of Gluc that are packed against the core of the protein structure in the NMR model but less ordered and packed differently in the AF2 model, and a C-terminal helical segment that is packed against the protein core in the AF2 model but is disordered in the NMR model. Interestingly, while the NMR data is largely consistent with the original NMR-based model, a significant number of NOE interactions, which arise from short interproton distances, were observed that are inconsistent with this model but consistent with the computed AF2 model (Huang *et al.*, 2021). This observation led to the conclusion that in solution Gluc is an ensemble of protein structural states in dynamic equilibrium, which can not be modeled correctly by the standard AF2 modeling protocol (Huang *et al.*, 2021).

Gluc is a 168-residue (18.2 kDa) monomeric enzyme isolated from the marine *Gaussia princeps*, which catalyzes the oxidation of coelenterazine to create a bright blue bioluminescent product (Kaskova *et al.*, 2016). It is widely used as a reporter protein, and is an important target for enzyme design (Yeh *et al.*, 2023). Gluc has five disulfide bonds and is difficult to produce in its native form, but was successfully produced in *E. coli* for NMR studies by Wu, Kobayoshi, Kuroda,Yamazaki, and co-workers (Wu *et al.*, 2020; Wu *et al.*, 2022). Their work provided extensive resonance assignments and a high-quality NMR structure. Recently, a second solution NMR structure of Gluc was reported (Dijkema *et al.*, 2024), with some significant structural differences. These NMR studies reveal extensive conformational dynamics of Gluc in solution, making it an excellent system for exploring AI-based methods for modeling multiple conformational states.

Here, we further investigate the dynamic structural features of Gluc using a novel conformational selection protocol with NMR data. We first generated a diverse set of conformer models using AlphaFold-sample (AFsample) (Wallner, 2023a), an AI-based enhanced sampling protocol that exploits multiple parameter settings of AF2 together with neural network dropouts to generate conformational diversity. Five models that best-fit the NOESY and chemical shift data were then selected to represent the dominant solution structure (state 1). In addition, AFsample generated an alternative structural state (state

2) that also fits well to the NMR data but differs from the state 1 structures primarily by the position of a thumb-shaped loop between α-helices H5 and H6. In the state 1 models, this dynamic loop forms a lid over a cryptic pocket that contains the proposed (Wu *et al.*, 2020; Wu *et al.*, 2022; Dijkema *et al.*, 2024) enzyme active site. Comparison of these AFsample models against the NMR NOESY data reveals a subset of NOEs specific to state 1 and another subset specific to state 2. Additional conformations of Gluc are also indicated by backbone chemical shift data consistent with a disordered structure for the C-terminal segment in a subpopulation of molecules. Considered as an ensemble, these multiple conformational states of Gluc together fit the combined NOESY and chemical shift data better than the published "restraint-generated" solution NMR structure. Our study demonstrates the potential of using AI-based modeling methods with enhanced sampling to generate conformational ensembles that can be fit to experimental data by conformer selection as an alternative to conventional restraint satisfaction protocols for protein NMR structure determination.

Methods

Experimental NMR data. The published NMR structure of Gluc (PDB_ID 7d2o) was determined from extensive ^{1}H,^{13}C and ^{15}N NMR assignments together with about 2500 NOE-based distance restraints. including about 580 long-range restraints (Wu *et al.*, 2020; Wu *et al.*, 2022). Chemical shift assignments for Gluc were downloaded from the BioMagResDatabase (BMRB), entry ID 36385. NOESY peak lists were obtained from the group of Wu, Kobayoshi, Kuroda,Yamazaki, and co-workers (Wu *et al.*, 2020; Wu *et al.*, 2022) who determined the original NMR structure model used for assessment of structure predictions in CASP14 (Huang *et al.*, 2021).

AlphaFold2 modeling. AF2 modeling was carried out using the *Colab-Fold* server of AlphaFold-multimer (version 1.5.3) (Mirdita *et al.*, 2022). No X-ray crystal structures of proteins homologous to Gluc were available in the PDB at the time of training AF2 (Jumper *et al.*, 2021). NMR structures of the query target proteins structures themselves and the structures of any possible homologs detected with HMMPred (Soding *et al.*, 2005) were excluded as modeling templates. Hydrogen atoms were added by the *ColabFold* prediction pipeline to the top five-scoring conformers along with estimates of prediction reliability (per-residue pLDDT), as described elsewhere (Jumper *et al.*, 2021).

AF2 modeling with AFsample. AFsample modeling calculations were carried out as described by Wallner (Johansson-Akhe and Wallner, 2022; Wallner, 2023a) using their modified version of *AlphaFold v2.2.0* (Evans *et al.*, 2022) (https://github.com/bjornwallner/alphafoldv2.2.0) installed on the AiMOSx NPL cluster in the Center for Computational Innovation at Rensselaer Polytechnic Institute. AF2 was trained using the PDB database of 2018-04-30 and did not include any NMR structures in the training data (Jumper *et al.*, 2021). AFsample calculations, which use six different settings with network dropouts to create conformational heterogeneity(Johansson-Akhe and Wallner, 2022; Wallner, 2023a), were carried out to generate ~6000 models. Hydrogen atoms were added using *run_relax_from_results_pkl.py* from the AFsample package. AFsample is quite aggressive in generating conformational diversity, and computes some models that are not physically reasonable, with incorrect amino acid chirality, non-native cis peptide bonds, missing residues, and other biophysically incorrect features, particularly in the not-well-packed residue segments of the modeled proteins. The most egregious of these physically unreasonable models were identified and removed, as described elsewhere (Spaman et al, manuscript in preparation). The

resulting 4990 relaxed models were used for further analysis. pTM scores were retrieved from the AF2 output file *result_model_*.pkl.json*. and pLDDT scores from the corresponding atomic coordinate files.

Knowledge-based protein structure model validation. All structure quality statistical analyses were performed using the *Protein Structure Validation Software* (Bhattacharya *et al.*, 2007) (*PSVS* ver 2.0) (https://montelionelab.chem.rpi.edu/PSVS/PSVS2/) and the protein model assessment program *PDBStat* (Tejero *et al.*, 2013). *PSVS* runs a suite of knowledge-based software tools including *ProCheck* (ver 3.5.4) (Laskowski *et al.*, 1993) and *MolProbity* (mage ver 6.35.040409) (Chen *et al.*, 2010). The structure validation scores of these programs are used to calculate a normalized Z score relative to mean values and standard deviations obtained for a set of 252 reference X-ray crystal structures of < 500 residues, resolution < 1.80 Å, R-factor < 0.25 and R-free < 0.28; for all metrics assessed, positive Z scores indicate 'better' scores (Bhattacharya *et al.*, 2007).

NMR NOESY RPF-DP scores. RPF-DP scores are a set of fast and sensitive structure quality assessment measures used to evaluate how well a 3D structure model fits with NOESY peak and resonance assignment lists, and hence to assess the accuracy of the structure (Huang *et al.*, 2005; Huang *et al.*, 2012; Huang *et al.*, 2021). RPF-DP scores were computed using the *RPF server* (ASDP ver 1.0) https://montelionelab.chem.rpi.edu/rpf/. *RPF* returns several statistics, including a *recall* metric corresponding to the fraction of NOESY peaks that can be explained by the model (or collection of models), *precision*, the fraction of short distances in the model(s) that can be explained by NOESY peaks, a *performance score* (F measure) which is a geometric mean of recall and precision, and *discriminating power* (*DP)*, a normalized F-measure accounting for the completeness of the NOESY data and the baseline F-measure expected for a random coil chain (Huang *et al.*, 2005; Huang *et al.*, 2012; Huang *et al.*, 2021). RPF-DP scores provide a type of NMR "R-factor", in which models are compared against unassigned NMR NOESY peak list data. Output folders (zip files) were downloaded from the *RPF server* for further Double Recall analysis (as described below).

NMR NOESY Double Recall analysis. Double recall analysis was carried out using *DoubleRecall* version 1.0 in a Jupyter notebook. This method compares two ensembles of conformers, A and B, and identifies 3D ^{15}N/^{13}C-resolved NOESY peaks (from provided peak lists) explained by short ^{1}H-^{1}H distances present in one conformer ensemble (e.g. in the A ensemble) but not by the other (e.g. in the B ensemble). This analysis uses *RPF*'s recall violation reports from the zipped RPF output folder. NOEs in 3D ^{15}N and ^{13}C-resolved NOESY spectra consistent with short mid-range ^{1}H-^{1}H distances unique to ensemble A (blue) or unique to ensemble B (orange) are plotted on a 2D plot at the coordinates of the residue number containing the first H(-N/C) proton (x-axis, "^{1}H(-N/C)-donor Match") versus the residue number of the second H proton (y-axis, "^{1}H-acceptor Match"). This contact map-like plot allows for visual evaluation of the agreement of each conformer ensemble with the NOESY data by highlighting the differences that are specific to each of the two ensembles being compared. Multiple NOEs between the same two residues appear as only a single dot on the per residue plot. A 1D histogram plot at the top of the 2D double-recall plot shows the per-residue count of the matching NOESY recall violation peaks explained uniquely by each ensemble for the ^{1}H(-N/C)-donor proton ("# Resolved H(-N/C)"), which are identified more confidently due to the additional frequency resolution in the 3rd dimension from the attached ^{15}N or ^{13}C nucleus. The general Double Recall analysis allows for comparison of any two pairs of ensembles, using the midrange interproton distances within each ensemble for the NOESY comparison. The DoubleRecall method was also extended to compare merged ensembles of multiple structural states against a reference ensemble, using the shortest midrange distances among the multiple structural states.

In either case, the number of models in each conformer ensemble does not need to be identical to perform the comparison.

Well-defined residue ranges and TM scores to reference structures. For NMR and AF2 structures, the ranges of "well-defined" residues were determined by the program Cyrange separately for each ensemble (Kirchner and Guntert, 2011). These residue ranges were used in superimpositions, structure quality assessment, and TM score calculations (Zhang and Skolnick, 2004), as described elsewhere (Tejero *et al.*, 2022; Li *et al.*, 2023). For TM-score comparisons to the NMR structure, the representative "medoid" conformer (Snyder and Montelione, 2005; Montelione *et al.*, 2013; Tejero *et al.*, 2013) from PDB deposition 7d2o, referred to here as NMR_{7d2o}, was used, with the well-defined residue range defined by Cyrange (Kirchner and Guntert, 2011). Similarly, TM comparison scores to the AF2 ensemble were calculated using the rank1 AF2 conformer, with only the well-defined residues for that ensemble.

Per-residue Random Coil Index (RCI) and RCI_{rmsf}). RCI values were calculated from backbone chemical shifts by the method of Wishart and co-workers (Berjanskii and Wishart, 2007) implemented on their server (http://www.randomcoilindex.ca/cgi-bin/rci_cgi_current.py) with the default options:

$$RCI = (\langle A|\Delta\delta C\alpha|+B|\Delta\delta CO|+C|\Delta\delta C\beta|+D|\Delta\delta N|+E|\Delta\delta NH|+F|\Delta\delta H\alpha|\rangle)^{-1} \qquad \text{Eqn. 1}$$

where $|\Delta\delta C\alpha|$, $|\Delta\delta CO|$, $|\Delta\delta C\beta|$, $|\Delta\delta N|$, $|\Delta\delta NH|$ and $|\Delta\delta H\alpha|$ are the absolute values of the conformation-dependent chemical shifts $\Delta\delta X = \delta X_{observed} - \delta X_{random_coil}$ (in ppm) of Cα, CO, Cβ, N, NH and Hα resonance, respectively, and *A*, *B*, *C*, *D*, *E* and *F* are weighting coefficients. Left angle and right angle brackets (<·>) indicate that the RCI per residue is the average value across the several resonances (^{1}H, ^{13}C, and ^{15}N) of that residue. If the per-residue RCI values are above a 'ceiling limit' of 0.6, they are replaced with this ceiling value of 0.6. The final RCI values given by the server are obtained after a second smoothing along the sequence by three-point averaging. For comparison with pLDDT values, the RCI values are divided by 0.6, resulting in a range of 0.0 to 1.0 (i.e. $RCI_{0.6}$) scaled for comparison with pLDDT scores. RCI values are an experimental estimate of conformational flexibility, and can also be reported as estimated root-mean-squared-fluctuations (i.e. RCI_{rmsf}) for comparison against atomic coordinate RMSFs, as described by Wishart and co-workers (Berjanskii and Wishart, 2007),.

$$RCI_{rmsf} = RCI * 12.7 \text{ Å} \qquad \text{Eqn. 2}$$

pLDDT vs RCI and RMSF vs RCI agreement plots. For pLDDT vs RCI agreement plots, per-residue AF2 pLDDT values and $RCI_{0.6}$ values were plotted along the protein sequence to provide a visual analysis of their correlation. For ensembles of conformers, the ensemble-average of the per-residue pLDDT values, $pLDDT_{avg}$, was calculated. In most cases we have examined, RCI and pLDDT (or $pLDDT_{avg}$) have an inverse correlation. The agreement plot between RCI and pLDDT was constructed by reporting the sum of $RCI_{0.6}$ and $pLDDT_{avg}$ for each residue along the sequence. Global mean and standard deviation (SD) values were calculated accordingly. In order to determine confidence interval (CI) values, the unbiased Standard Error of the Mean (SEM) was first calculated from the standard deviation value using $SEM = SD / (N -1)^{1/2}$, where N is the number of residues used for determining the mean and SD. CI values were then calculated by multiplying the SEM by the critical value (Zp) for the CI t-test with one degree of freedom using Equation 3:

$$CIp = SEM \cdot Zp \qquad \text{Eqn. 3}$$

where CIp denotes the CI at percent threshold p. Dashed lines indicating the mean value with the 95% CIs (Zp=12.71 for p=95%) are shown on the plot to emphasize regions of the protein sequencea with large deviations in the agreement plot. Residues with pLDDT ≤ 50 were excluded from these mean, SD, and CI calculations.

Similarly, for RMSF vs RCI agreement plots, atomic coordinate RMSF values for model ensembles and RCI_{rmsf} (Eqn. 2) values were plotted along the protein sequence to provide a visual analysis of their correlation. The agreement plot was constructed by reporting the difference between RCI_{rmsf} and RMSF of ensembles for each residue along the sequence.

Principal Component and RMSF analysis. Bio3D (Grant *et al.*, 2021) was used for the Cα distance matrix calculations and principal component analysis (PCA) for AFsample models. Bio3D was also used for Cα root-mean-squared fluctuation (RMSF) analysis. In RMSF calculations, residues were included in the superimpositions if they were "well-defined" (i.e. well-converged) in the experimental NMR_{7d2o} ensemble.

Alphafold-NMR Bayesian model selection metric. For each cluster of conformations generated by AFsample, the model(s) with highest recall score(s) were identified, a recall_cutoff near to this value was defined for that cluster, and selection was done using a Bayesian metric that estimates the likelihood of the model given NMR data [P(model|NOE, RCI)]. This metric accounts for both the agreement of the model with the available NMR data, and the probability of the model, P(model) given by <pLDDT> (for residue-specific information) or pTM (for overall fold information), where <pLDDT> is the average pLDDT value along the sequence and pTM is the estimated likelihood of the overall tertiary structure, as defined by AF2 (Jumper *et al.*, 2021).

We introduce here a per model score assessing the agreement between pLDDT, a per residue estimate of the reliability of AF2 structure models, and the NMR RCI value, an experimental estimate of conformational flexibility.

For each model, we define

$$P(RCI|model) \sim SCC(pLDDT, RCI) \qquad \text{Eqn. 4}$$

where SCC(x,y) is the Spearman Correlation Coefficient between the values of x and y. Hence, by Bayes theorem

$$P(model|RCI) \sim P(RCI|model) * P(model) \sim SCC(pLDDT, RCI) * \langle pLDDT \rangle \qquad \text{Eqn. 5}$$

where both |SCC(pLDDT, RCI)| and <pLDDT> are scaled to have positive values between 0.0 and 1.0

For the purpose of scoring models, we take

$$P(model|RCI) = [|SCC(pLDDT, RCI)| * \langle pLDDT \rangle]^{1/2} \qquad \text{Eqn. 6}$$

and define RCI_pLDDT score ≡ P(model|RCI) Eqn. 7

This new RCI_pLDDT score (Eqns. 6 and 7) is a *global* score measuring how well the model and associated estimates of model uncertainty fit with conformational flexibility encoded in the per residue chemical shift data. A high RCI_pLDDT score requires a model with overall high reliability (high <pLDDT>), and also a good correlation between local per-residue reliability (pLDDT) and experimental flexibility (RCI), demonstrating a well-modeled structure.

For scoring each model using NOESY data, we focus only on the set of models with maximum (or near maximum) NOESY recall scores within each cluster.

$$P(NOE|model) = (recall(model) - recall_min) / (recall_max - recall_min)$$

$$\text{if } recall(model) \geq recall_cutoff; \text{ otherwise } 0 \qquad \text{Eqn. 8}$$

where recall_max (e.g. 0.88) and recall_min (e.g. 0.84) are the maximum and minimum values of recall scores for all models generated by enhanced sampling with AF2.

Since recall is a global (rather than per-residue) score, the pTM score (rather than <pLDDT> is used here to estimate the overall probability of the model.

Hence by Bayes theorem:

$$P(model|NOE) \sim P(NOE|model) * P(model) \sim P(NOE|model) * pTM \qquad \text{Eqn. 9}$$

and, for the purpose of scoring models, we take

$$P(model|NOE) = [P(NOE|model) * pTM]^{1/2} \qquad \text{Eqn. 10}$$

Based on these assumptions, each model in each cluster is scored as:

$$P(model|NOE, RCI) = (P(model|NOE) + P(model|RCI))/2 \qquad \text{Eqn. 11}$$

which we define as the AlphaFold-NMR Bayesian model selection metric.

State combination probability scores. All of the models generated by enhanced sampling are clustered based on structural similarly defined by their Cα distance matrices using principal component analysis (PCA). Alternative structural clustering methods could also be used for this purpose. For each cluster of conformers, the five models with the highest Bayesian model selection scores are identified to represent a "single conformational state". These "single conformational states" or "classes" (class a, b, c, …) are then selected and combined in order to identify the combined set of states that together fit the data as well (or better) than a single state.

We assign "state 1" conformer to the class with highest average P(model|NOE, RCI), and assess the agreement between the root-mean-squared fluctuations of Cα coordinates (RMSF) and the RCI using the Spearman Correlation Coefficient

$$P(state\ 1 \mid RCI) \sim SCC\ (RMSF(state\ 1), RCI) \qquad \text{Eqn. 12a}$$

We next ask if adding additional classes improves (or only minimally reduces) this probability by assessing

$$P(\text{state 1 + class } x \mid \text{RCI}) = \text{SCC}(\text{RMSF}(\text{state 1 + class } x), \text{RCI})$$
$$\text{if SCC}(\text{RMSF}(\text{state 1 + class } x), \text{RCI}) > P(\text{state 1} \mid \text{RCI}) * (1 - \delta); \text{ otherwise } 0 \qquad \text{Eqn. 12b}$$

δ is an error tolerance (5% for this work) that allows for small decrease in the SCC estimation in adding an additional conformational state to the structural ensemble. Class x is any of the remaining single state classes suitable for state combination selection. The class x with the best P(state 1 + class x | RCI) is then identified as state 2 conformer, with score P(state 1 + state 2 | RCI). In case there are more than one classes with the same or similar best P(state 1 + class x | RCI) scores, then P(model|NOE, RCI) will be used as the tie breaker. Classes are added so long as either SCC increases or decreases no more than δ. Any classes which decrease SCC more than δ are considered unsuitable for state combination selection. This process is then iterated by adding in additional classes until there are no more classes suitable for state combination selection. Atomic coordinates for each of these states were used for analysis and structure quality assessment, and deposited in the PDB-dev section of the Protein Data Bank.

Molecular graphics*, *secondary structure analysis, and pocket/cavity analysis. Molecular visualization and graphical representations of molecular structures for figures were done using *ChimeraX* (Meng *et al.*, 2023) and PyMol (Schrödinger and DeLano, 2020). The *Define Secondary Structure of Proteins (DSSP)* (Kabsch and Sander, 1983) server (https://2struc.cryst.bbk.ac.uk/twostruc) was used for identifying regular secondary structures. The *CASTp 3.0* (Tian *et al.*, 2018) server (http://sts.bioe.uic.edu/castp/calculation.html) was used for surface pocket/cavity analysis of the relaxed top-scoring model for each conformational state.

Results

Comparison of conventional NMR and standard AF2 models using NOESY Double Recall. The ColabFold (Mirdita *et al.*, 2022) AF2 models of Gluc generated for this study are essentially identical to the AF2 models reported by Deep Mind, Inc. in CASP14 (backbone RMSD 0.45 Å and TMscore = 0.98 between the Rank1 conformers for ordered regions 42-78 and 93-162). We used these ColabFold models (referred to as "standard" AF2 models) for comparison with the NMR structure models reported by Wu et al. (Wu *et al.*, 2020) using conventional NMR methods. As illustrated in **Figure 1** (and discussed previously (Huang *et al.*, 2021)), the principal differences between these models involve two helices (α-helix H1 and α-helix H2) in the N-terminal segment that are packed against the protein core in the NMR-based model but are less-ordered and have different interactions in the standard AF2 model, and two C-terminal helices, H10 and H11, that are packed against the protein core in the AF2 model but are disordered in the conventional NMR model (**Figure 1a**). The structurally variable regions that differ between the NMR and AF2 models are color-coded in **Figure 1**; N-terminal disordered "tail" (dark blue), α-helix H1 (cyan), α-helix H2 (light blue), a surface loop between α-helix H5 and α-helix H6 (green), the C-terminal segment (red) that is disordered the NMR model but forms a broken helix, H10 and H11, in the AF2 model, and the protein core (gray) of α-helices H3, H4, H5, H6, H7, H8, and H9. The NMR and AF2 models have the same set of five disulfide bonds. While most of the NMR data are well-fit by the conventional NMR model, a significant number of NOESY peaks arising from short interproton distances were observed that are inconsistent with this model but consistent with the standard AF2 model (Huang *et al.*, 2021). This is illustrated in the double-recall analysis of the conformers of the AF2 and NMR-based models (**Figure 1b**). Specifically, there are a large number of NOEs (540) consistent with the NMR-based model, but not consistent with the AF2 model (blue dots in the double-recall plot of **Figure 1b**),

particularly NOEs consistent with N-terminal helices H1 interacting with the core of the protein structure. Conversely, there are also a significant number of NOEs (137) that cannot be explained by the NMR model, indicating helices H10/H11 packing against the core helices as modeled by AF2 (orange dots inside red stripe, Figure 1b). However, as both RCI and Talos (Shen and Bax, 2015) analysis of backbone chemical shifts indicate a partially disordered C-terminal segment (Wu *et al.*, 2020; Huang *et al.*, 2021), the NOEs supporting H10/H11 helix formation must arise from a subpopulation of Gluc with an ordered C-terminal segment present under NMR conditions.

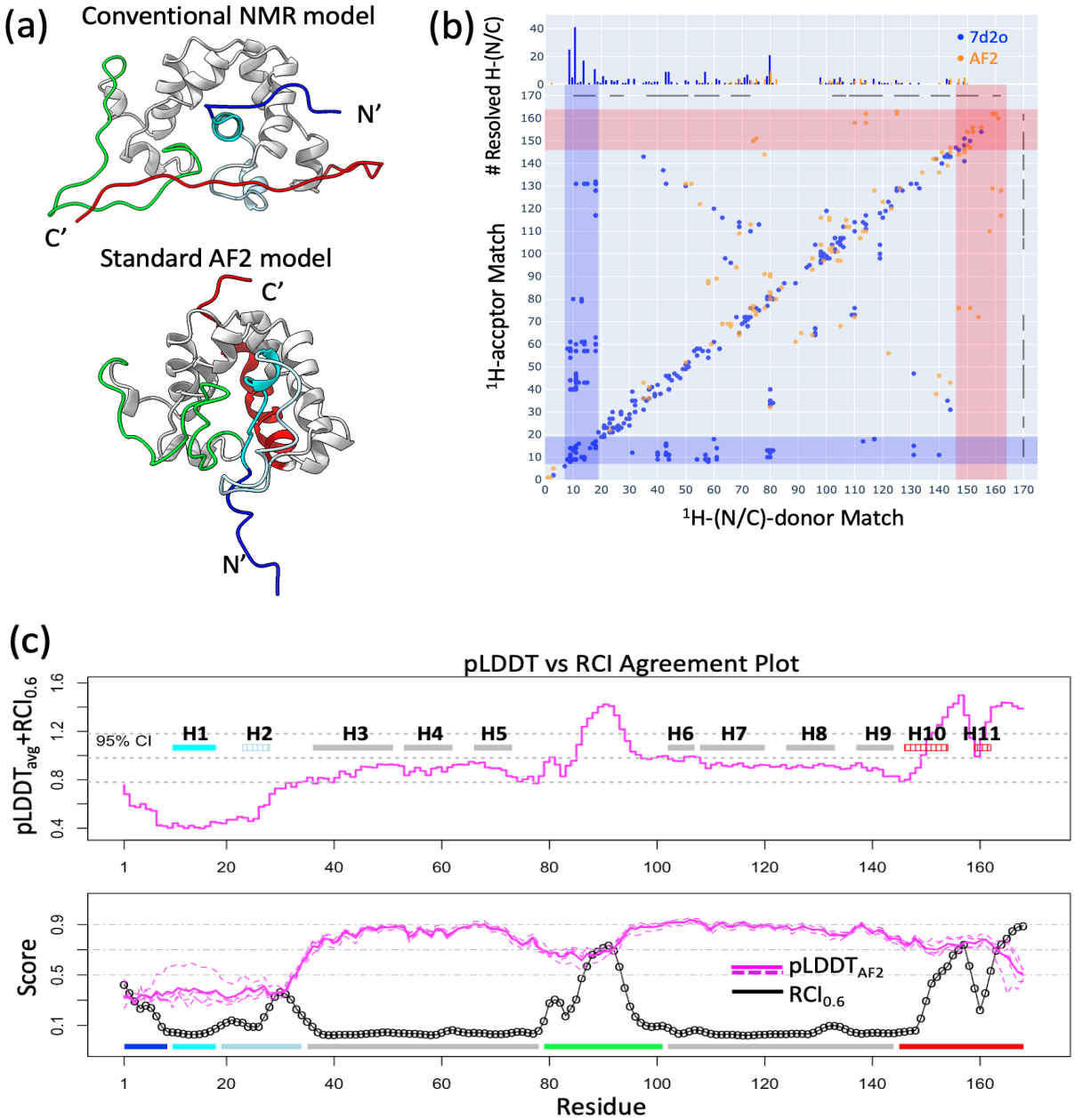


Fig 1. Comparison of conventional NMR and AF2 models of Gluc. (a) Ribbon diagrams for the published conventional NMR structure (PDB ID 7d2o) (top) and standard AF2 model (bottom). Structural elements are represented using the following color scheme: blue: the N-terminal disordered region (residues 1-9); cyan: the ordered helix H1 (residues 10-18); light blue: the flexible helix H2 and loop (residues 19-34); green: the disordered loop between helices H5 and H6 (residues 79-101); red: the C-terminal disordered region (residues 146-168). In the C-terminal region (red) of AF2 models, residue segments

146-154 and 159-162 adopt helical conformations, with a kink between them resulting in a "broken helix". (b) Double Recall plot for NMR_{7d2o} vs. AF2 models of Gluc. Interresidue interactions indicated by one or more NOEs from the experimental NOESY spectra that are specific to the NMR models (blue) or specific to the AF2 models (orange) are plotted as dots on this contact map. Brown dots are overlapping blue and orange dots. Regions corresponding to contacts between the N-terminal helix H1 (residues 7-19) and core residues are indicated by blue highlighted stripes. Regions corresponding to contacts between the C-terminal broken helices H10/H11 (residues 146-164) and core residues are highlighted as red stripes. (c) Comparison between pLDDT and experimental RCI data. (top) The agreement plot (see Methods) between $pLDDT_{avg}$ and $RCI_{0.6}$. The average and 95% confidence intervals are shown as gray dashed lines. The locations of nine helices (Wu *et al.*, 2020), together with helices H10/H11 computed by AF2, are indicated by horizontal rectangles. The three most flexible helices (based on RCI), H2, H10 and H11, are marked with striped lines. (bottom) Per-residue pLDDT (dashed line for each model, and solid line for the average per-residue pLDDT) and $RCI_{0.6}$ scores along the sequence. The Spearman correlation coefficient (SCC) between RCI and average pLDDT is -0.56 (ranging from -0.59 to -0.51).

Comparison of conventional NMR and standard AF2 models using RCI chemical shift data. We also compared the conformational variability between NMR and AF2 models using an experimental measure of conformational flexibility, the random coil backbone chemical shift index (RCI) (Berjanskii and Wishart, 2007), and the pLDDT scores of AF2, which measure local uncertainty in the structural model (Jumper *et al.*, 2021) (**Figure 1c**). pLDDT scores have been observed to correlate reasonably well (though not perfectly) with regions predicted or observed to have conformational flexibility (Jumper *et al.*, 2021; Ruff and Pappu, 2021; Guo *et al.*, 2022; Saldano *et al.*, 2022; Ma *et al.*, 2023). While there is a generally good agreement between these experimental and AI-based measures of flexibility / uncertainty in Gluc, some notable divergence is observed, as illustrated by the agreement plot in the top panel of **Figure 1c**. Specifically, the entire N-terminal region preceding helix $\alpha 3$ (colored blue-cyan-light blue) including both helices H1 and H2, the loop between helix $\alpha 5$ and $\alpha 6$ (green), and most of the C-terminal region (red) do not have good agreement between uncertainty indicated by pLDDTs and flexibility indicated by RCIs. These are the same regions showing significant structural differences in the AF2 and NMR-based models. These results demonstrate that regions of the standard AF2 models with high per-residue pLDDT scores (i.e. the structural core formed by helices H3 - H5 and H6 - H9) are not particularly flexible, as indicated by low RCI values, and exhibit good correlation between RCI and pLDDT methods, while regions of the AF2 models with low pLDDT scores generally have conformational flexibility, indicated by high RCI values, and generally exhibit poor correlation between RCI and pLDDT metrics. One notable exception is helix H1, which has low RCI (low flexibility) but also low pLDDT scores using standard AF2 modeling.

AFsample generates conformational diversity. We next asked the question of whether AI-based modeling with enhanced sampling methods can provide alternative models that fit the NMR data better than the original AF2 models. Although several methods of enhanced sampling with AF2 have been recently described, here we focus on the AFsample method of Wallner and colleagues (Johansson-Akhe and Wallner, 2022; Wallner, 2023a,b). Gluc has a relatively shallow MSA (~ 229 sequences), and neither AF_alt (Del Alamo *et al.*, 2022) nor AF_cluster (Wayment-Steele *et al.*, 2024) methods worked well for this system. AFsample models were compared by TM score to the conventional NMR structure (medoid model of NMR_{7d2o}) and to the rank1 model of the standard AF2 ensembles (**Figure 2**). These plots demonstrate the diversity achieved by AFsample over AF2. While most AFsample models have lower NOESY recall scores, many models have scores (recall 0.88) higher than the standard AF2 model (recall 0.86), and comparable to the conventional NMR models (recall 0.89).

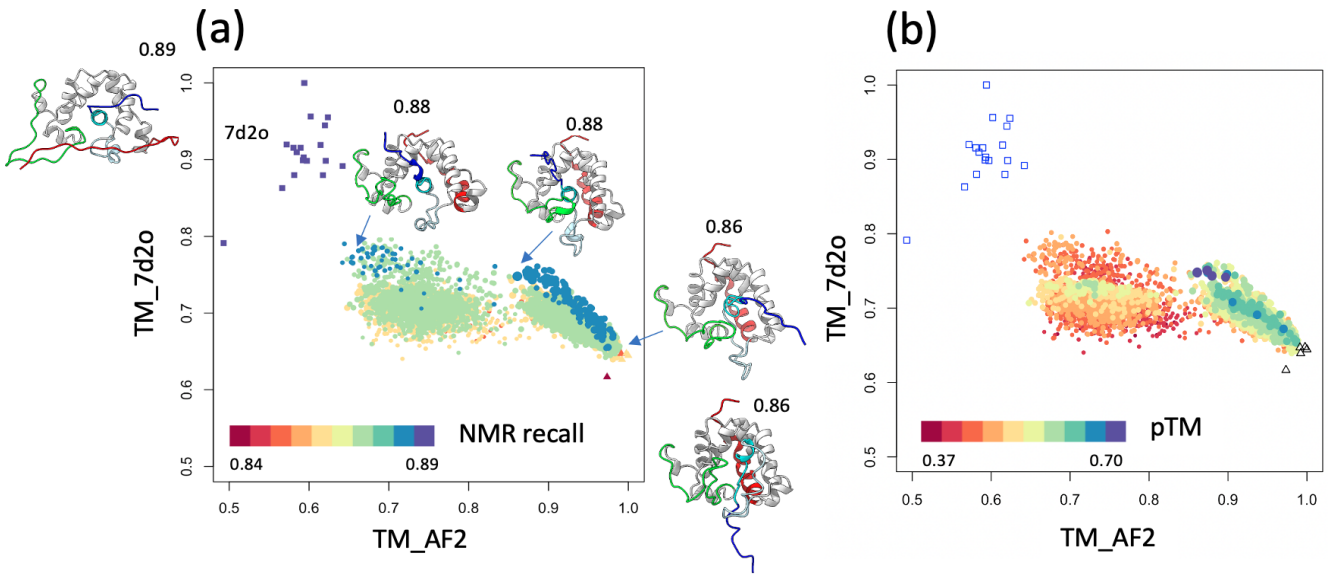


Fig 2. TM plots of AFsample vsNMR$_{7d2o}$ and AF2 models. TM_AF2 scores are calculated by comparing each conformer with the AF2 rank1 model (for ordered regions 42-78 and 93-162). TM_NMR$_{7d2o}$ scores are calculated by comparing each conformer with the medoid NMR$_{7d2o}$ model (model #3, for ordered regions 7-19, 34-81 and 95-143). The plotted dots for each conformer is colored by its corresponding (a) NOESY recall or (b) pTM (b) scores for AFsample (circles), conventional NMR$_{7d2o}$ (squares), and standard AF2 (triangles) models. The size of the circular points for AFsample models are also coded by pTM score; the larger-sized dots have higher pTM values.

As we wanted to develop a protocol that does not require reference structures (e.g. conventional NMR structures) for implementation, we also analyzed the conformational diversity of AFsample models using principal component analysis (PCA). These results are plotted as 2D projections of the 3 principal components in **Figure 3,** with color coding corresponding to NMR NOESY recall (**Figure 3 - top panels**), pTM (**Figure 3 - middle panels**), and RCI_pLDDT (Eqns. 6 and 7) (**Figure 3 - bottom panels**) scores. PCA can be sensitive to even small changes in the protein structure, and groups the AFsample models into four clusters. PCA clusters a and b include many models with high recall scores (i.e. 163 from PCA cluster a and 49 models from PCA cluster b have NOESY recall ≥ 0.88). PCA clusters c and d have generally lower NOESY recall and pTM scores. Structurally, clusters a vs c (and clusters b vs d) mainly differ by the location of helix H1, while clusters a vs b (and clusters c vs d) that separate along PC2 differ primarily in the position of the flexible H5/H6 thumb-shaped loop. No additional clusters were found when comparing PC3 to PC2 or PC1.

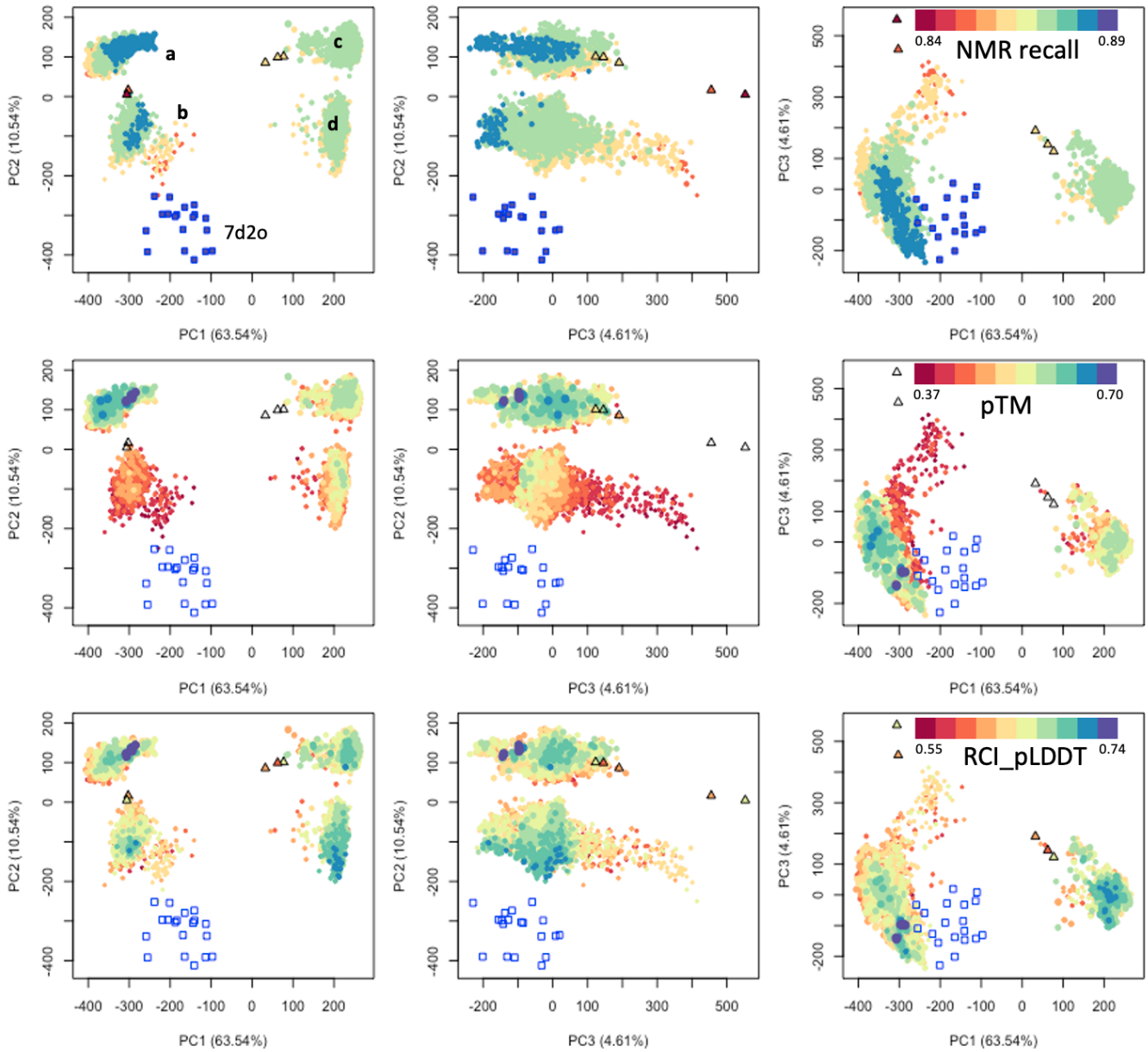


Fig. 3. PCA of AFsample models. 2D projections of Cα coordinate principal component analysis (PCA). Models are colored by NOESY recall (top), pTM (middle), and RCI_pLDDT (bottom) scores. Conventional NMR models (PDB_ID 7d2o) are indicated as blue open squares (marked as "NMR") and standard AF2 models are indicated as open triangles. In each panel, the size of each point is coded by the pTM score of the model, the bigger the point the higher the pTM. The four PCA clusters are labeled a, b, c, and d in the top left panel.

Conformer selection and identification of states using state-combination probability scores. We next asked if we could use the conformational sampling generated by *AFsample* together with experimental NMR data to *select* the one or more conformational states that best explain the NOESY peak list and chemical shift data. These conformational ensembles considered as alternative states in this process are referred to here as *classes*, class a ensemble from cluster a (Bayesian Model Selection Score: 0.770-0.787), class b ensemble from cluster b (Bayesian Model Selection Score: 0.690-0.697), class c ensemble

from cluster c (Bayesian Model Selection Score: 0.676-0.682), class d ensemble from cluster d (Bayesian Model Selection Score: 0.660-0.664).

Following the protocol outlined in the Methods for combining conformational classes (or states) to best fit the NMR data, we first defined the top-five scoring models in the class a ensemble as state 1 - the first AlphaFold-NMR state. NMR RCI data, which monitor conformational flexibility (Berjanskii and Wishart, 2007), were compared with structural variability (root-mean-squared Cα fluctuations, RMSF) of state 1 using Eqn 12a; i.e. SCC (RMSF(state 1), RCI) = 0.75. We then calculated SCC for an equal mixture of state 1 with each of the three other classes using Eqn 12b; i.e. SCC(RMSF(state 1 + class b), RCI) = 0.74 , SCC(RMSF(state 1 + class c), RCI) = 0.54 , SCC(RMSF(state 1 + class d), RCI) = 0.65. Both class c and class d have much lower SCC when combined with state 1 indicating that class c and class d have regions that are very different from state 1 and do not agree with the experimental RCI flexibility measure. The largest RMSF difference is located around the helix H1 region (**Figure 4**) in classes c and d. Since experimental RCI (**Figure 4**) and ^{15}N-^{1}H NOE data (Wu *et al.*, 2020) indicate that helix H1 is not particularly flexible, resulting in a significant reduction in SCC(RMSF,RCI) when including these classes, we ruled out the possibility that class c and d are significantly populated alternative conformations under these NMR sample conditions. Accordingly, we assigned class b ensemble as state 2 ensemble - the second AlphaFold-NMR conformational state.

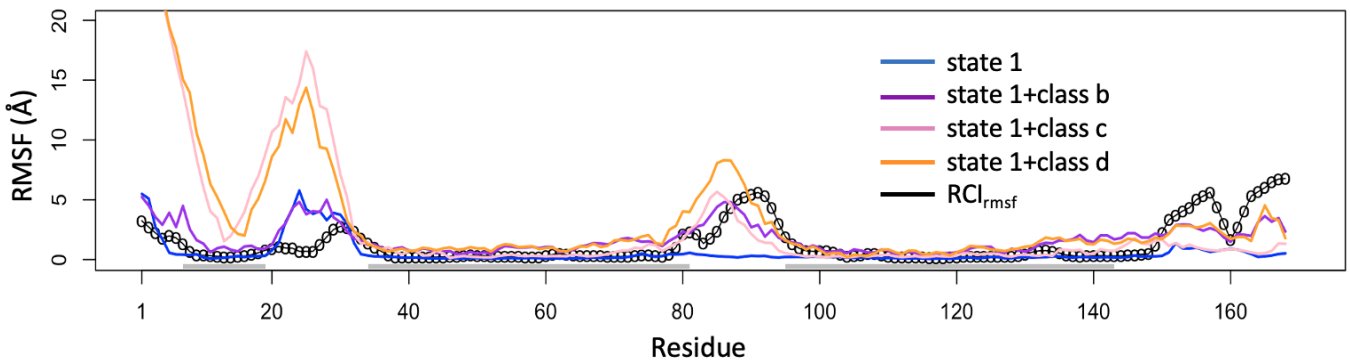


Fig. 4. RMSF plot. The RCI-based RMSF (RCI$_{rmsf}$) derived from chemical shifts (black circled line) is compared to the standard coordinated-based RMSF from the selected ensembles. The "well defined" regions of NMR$_{7d2o}$ are shown as gray bars at the bottom. RMSFs of conformers are state 1, and mixtures of state 1 + class b, state 1 + class c and state 1 + class d for state combination selection.

Based on these state combination probability scores (Eqns 12), we conclude that the state 1 and state 2 represent two highly populated alternative conformational states of Gluc. In order to confirm that these structures are physically-reasonable, these two "NMR-selected" conformational states were further assessed with knowledge-based structural quality statistics using the Protein Structure Validation Server (PSVS) (Bhattacharya *et al.*, 2007). Key statistics for sets of 5 conformers representing each of these two states and the 19-conformer ensemble of the conventional restraint-based NMR model, are summarized in Table 1, with complete PSVS reports in Supplementary Tables S1 - S3. In the PSVS structure quality assessment statistics, Z scores relative to values obtained for high-resolution X-ray crystal structures are reported for various structure quality metrics (Bhattacharya *et al.*, 2007); more positive Z scores indicate better structure quality, good NMR structures generally have Z scores > -3.0. This analysis demonstrates that AFsample models selected with AlphaFold-NMR scores, have superior knowledge-based structure quality statistics compared to the conventional restraint-based NMR models. In particular, ProCheck(all)

and MolProbity scores, which are sensitive to sidechain conformational distributions and backbone and sidechain atomic packing, respectively, are significantly better for the AFsample models.

Table 1. Structure quality statistics for conformer-selected and restraint-based NMR models

	ProCheck[a] (bb)	ProCheck[a] (all)	MolProbity[a]	Ramachandran Statistics[b] (percent most favored/ allowed/disallowed)	RPF[c] ($Recall_{avg}$ / $Precision_{avg}$ / F-$measure_{avg}$ / DP_{avg})
state 1	0.28	1.01	1.26	96.4 / 3.6 / 0.0	0.88 / 0.79 / 0.82 / 0.60
state 2	-0.51	-0.89	> 3.00[d]	92.4 / 6.7 / 1.0	0.88 / 0.79 / 0.82 / 0.60
7d2o	-0.63	-3.73	-2.53	91.3 / 6.6 / 2.1	0.89 / 0.79 / 0.84 / 0.67

[a] Z scores from the PSVS server (Bhattacharya *et al.*, 2007) using ordered residues selected by Cyrange . [b]Percent of backbone residues in percent most favored / allowed / disallowed of Richardson Ramachandran map (Chen *et al.*, 2010) using ordered residues selected by Cyrange . [c] RPF scores (Huang *et al.*, 2005) are calculated for each model of the ensemble and then averaged. [d]MolProbity returns NA for models with no clash violations, and the corresponding Z score is reported as > 3.00 (Tejero *et al.*, 2022; Li *et al.*, 2023).

The two AFsample states were also compared with experimental NOESY peak list data using *RPF*. NOESY RPF scores (Huang *et al.*, 2005) were computed for each model of each state and then averaged (Table 1). The averaged recall ($recall_{avg}$) for NMR_{7d2o} is higher than the $recall_{avg}$ for either state 1 and state 2 alone, probably because the restraints used to calculate the NMR_{7d2o} ensembles are generated from a mixture of NOEs arising from states 1 and 2; the restraint-based method forces satisfaction of the NOESY data arising from multiple states to generate a single model. However, as shown below by Double Recall analysis, a combination of state 1 and state 2 models can explain more NOESY peaks than either state alone (Table 2); i.e., more NOEs can be explained by structural features that were not previously identified by the restraint-based NMR_{7d2o} structure.

The final five state 1 and five state 2 models were also assessed against the NMR RCI data (**Figure 5**). Per residue $pLDDT_{avg}$ values for both states correlate with experimental RCI data, often better than the initial AF2 models, and agree primarily within the 95% confidence interval (**Figure 5(a)-top**). Considering both pLDDT and RCI, in both state 1 and state 2 helices H1, H3 - H5, and H6 - H9 have low flexibility, while helix H2, as well as the loop between H5 and H6, are more conformationally dynamic (**Figure 5(a)-bottom**). The largest deviations between RCI and pLDDT are in the most flexible regions, particularly helix H2, the loop between helices H5 and H6, and the C-terminal helices H10 and H11. Helices H10 / H11 identified by both AF2 and AFsample are predicted to have higher flexibility (lower pLDDT) in the state 1 and less flexibility (higher pLDDT) in state 2; while helix H1 is predicted to have higher flexibility (lower pLDDT) in the state 2 and less flexibility (higher pLDDT) in state 1.

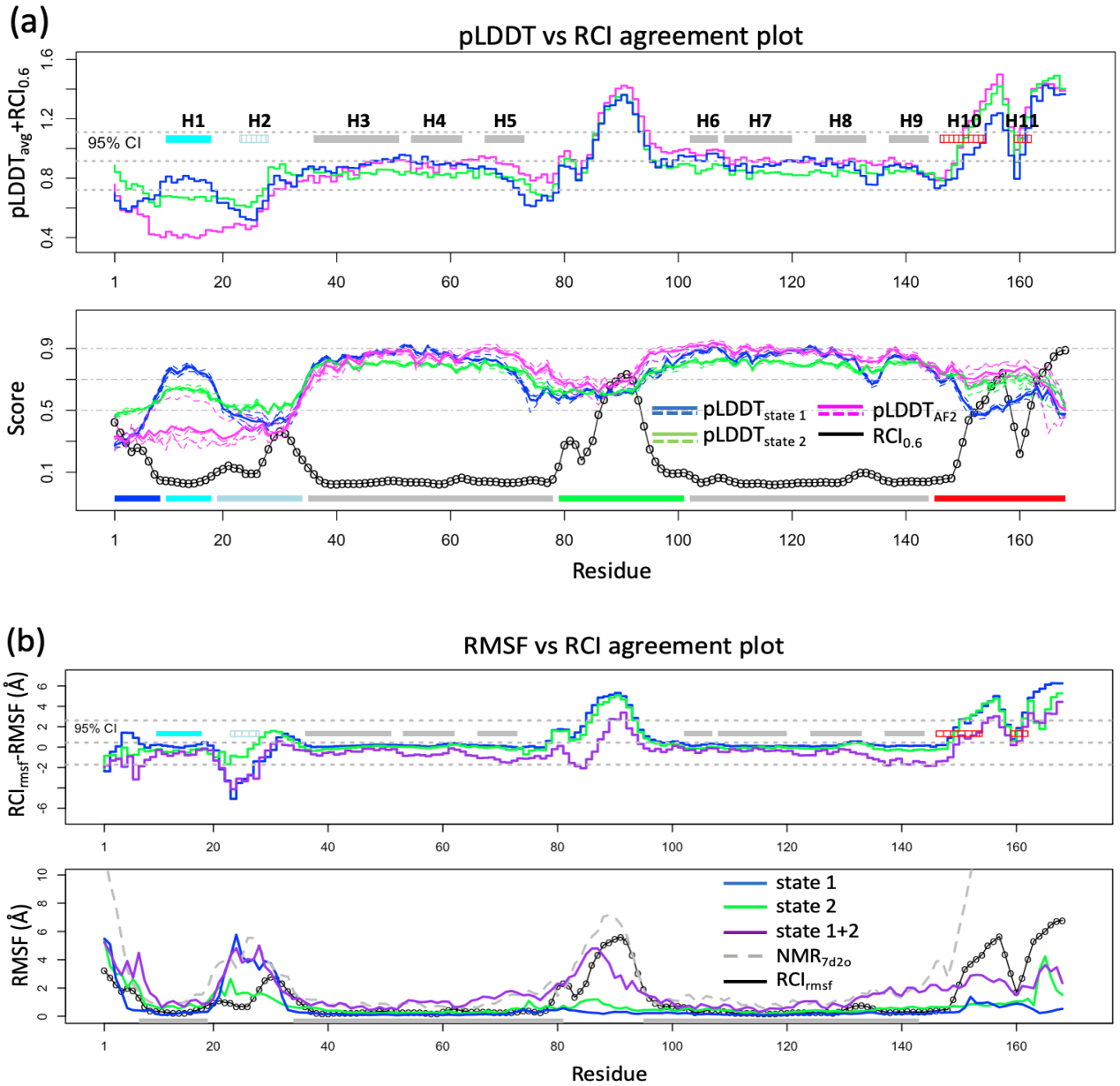
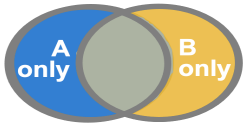


Fig. 5. Comparison of per-residue model uncertainty (pLDDT) of state 1, state 2 and AF2 ensembles with experimental chemical shift RCI data. (a) (top) Agreement plot (see Methods) between $RCI_{0.6}$ and $pLDDT_{avg}$. The locations of nine helices (H1 - H9) identified in the solution NMR structure (Wu *et al.*, 2020) along with two additional helices identified by AF2 (H10 and H11) are indicated. The three most flexible helices H2, H10 and H11 are shown in striped lines. The mean and 95% confidence interval (CI) is between the gray dashed lines. (bottom) Per-residue pLDDT (dashed line for each model, and solid line for the average per-residue $pLDDT_{avg}$) and $RCI_{0.6}$ scores along the sequence. pLDDT values of 0.5, 0.7, 0.9 are shown as gray dot dashed lines for reference. The residue segment color scheme of Fig. 1 is shown along the bottom. The Spearman correlation coefficients (SCCs) between $RCI_{0.6}$ and $pLDDT_{avg}$ for AFsample state 1 are -0.75 (-0.77 to -0.72); for AFsample state 2, -0.67 (-0.71 to -0.64); for AF2, -0.56 (-0.59 to -0.51). **(b)** RMSF vs RCI_{rmsf} agreement plots. The RCI_{rmsf} derived from chemical shifts (black circled line) is compared to the coordinate-based RMSF from the selected ensembles. (top) The agreement plots between RMSF and RCI_{rmsf} of state 1, state 2 and mixture of state 1+2. (bottom) RMSFs for state 1 and state 2 and mixture of state 1+2 compared with RCI_{rmsf}. Spearman correlation coefficients (SCCs) between RMSF and RCI_{rmsf} for state 1 are 0.75; for state 2, 0.71; for state 1 + state 2, 0.74. The "well defined" regions of NMR_{7d2o} are shown as gray bars at the bottom.

The coordinate-based RMSFs for ensembles corresponding to individual state 1 and state 2 and the mixture of both states correlate well with the RCI_{rmsf}'s based on the experimental chemical shift data (**Figure 5(b)**). The RMSF vs RCI_{rmsf} agreement plots (**Figure 5(b)** top) shows the regions between H5/H6 and the C-terminal region have much better agreement within 95% confidence interval for the mixture of the two states. In particular, the coordinate RMSF fluctuations for the H5/H6 thumb-shaped loop, or "lid", segment match the RCI_{rmsf} data much better for the mixture of state 1 and state 2 models than for either of the individual model ensembles. Both of these RMSF estimates for the mixture of state 1 and state 2 models also correlate well with regions of the conventional NMR structure that are "well defined", and with backbone conformational flexibility indicated by ^{15}N-^{1}H NOE data previously published for Gluc (see Figure 2 of Wu et al (Wu *et al*., 2020)).

NOESY Double Recall analysis of AFsample-NMR models. We also carried out NOESY Double Recall analysis for the conventional NMR_{7d2o} models vs AFsample models (**Table 2**). Many NOEs (184 total, 100 long-range NOEs) not consistent with the conventional NMR models are explained by the AFsample state 1 models; however many NMR_{7d2o} specific NOEs remain (273 total, 70 long-range NOEs). We next asked if the NMR NOESY data might be better explained by the mixture of AFsample state 1 and state 2 conformational states. The addition of state 2 explains 37 more long-range NOEs [(70 state 1) - 33 (state 1+2) = 37]; only ~ 33 NOEs assigned as long-range interactions in NMR_{7d2o} are not explained by this mixture of states. These 33 unaccounted long-range NOEs mostly involve atoms of residues Thr79, Tyr80, Glu81 interacting with atoms of residues Ala31, Lys33, Lys34, and Leu35, suggesting an additional population of "open state 2" conformers involving the H5/H6 loop interacting with the flexible helix H2.

Table 2. Double Recall analysis of *AFsample* state 1 and state 2 ensembles

		# of ensemble A-only NOEs (represented by blue dots in Fig. 6)		# of ensemble B-only NOEs (represented by orange dots in Fig. 6)	
A	B	total	long-range / resolved[a]	total	long range / resolved
NMR_{7d2o}	state 1	273	70 / 60	184	100 / 66
	state 2	245	87 / 70	162	70 / 41
	state 1+ state 2	126	33 / 27	184	100 / 66
state 2	state 1	173	54 / 39	157	89 / 58

[a] resolved - NOESY peaks matched to long-range (≥ 5 residues apart in sequence) interactions that are not overlapped with other peaks in the ^{1}H-(X) dimension of the X-resolved NOESY spectrum, where X is ^{15}N or ^{13}C.

These Double Recall analyses are also illustrated graphically in **Figure 6**, where the blue dots represent interresidue NOEs consistent with the NMR_{7d2o} ensemble but not explained by the AFsample ensemble(s). In these Double Recall plots, the AFsample state 1 comparison with NMR_{7d2o} resulted in fewer blue dots than standard AF2 models (*cf.* **Figures 6a and 1b**); i.e., the state 1 models account for 267 more NOEs than the original AF2 models. However, using the combination of state 1 + state 2 provides a Double Recall plot (**Figure 6b)** with even fewer blue dots, corresponding to 147 additional NOEs explained by the combination of state 1 and state 2 compared to state 1 alone. In particular,

AFsample state 1 accounts for many NOESY peaks between helix H1 and the rest of the protein (*cf.* blue highlighted region of **Figure 6a** with the corresponding regions of **Figure 1b**), while adding state 2 explains many additional NOESY peaks assigned to interactions between the core helices that were not explained by state 1 alone; *i.e.* there are less antidiagonal blue dots in **Figure 6b** within the corresponding region of **Figure 6a**. These antidiagonal blue dots in **Figure 6a** represent NOEs between the N- and C-terminal halves of the core Gluc structure, which can be explained by adding a second state (open state 2) with some different interactions between these regions. In addition, the combined AFsample state 1 and state 2 (**Figure 6b)** accounts for many NOESY peaks not explained by the NMR_{7d2o} model. These NOESY peaks explained by the combination of states are assignable to interactions between (i) the C-terminal helices H10 /H11 with the H1, H2, H3, H7 and H8 helices (orange dots in red highlighted regions in **Figure 6**), (ii) interactions between the H5/H6 thumb-shape loop and the H1 and H7 (orange dots in green highlighted regions in **Figure 6**), and (iii) interactions within between helices of the core (orange dots on the antidiagonal in **Figure 6**).

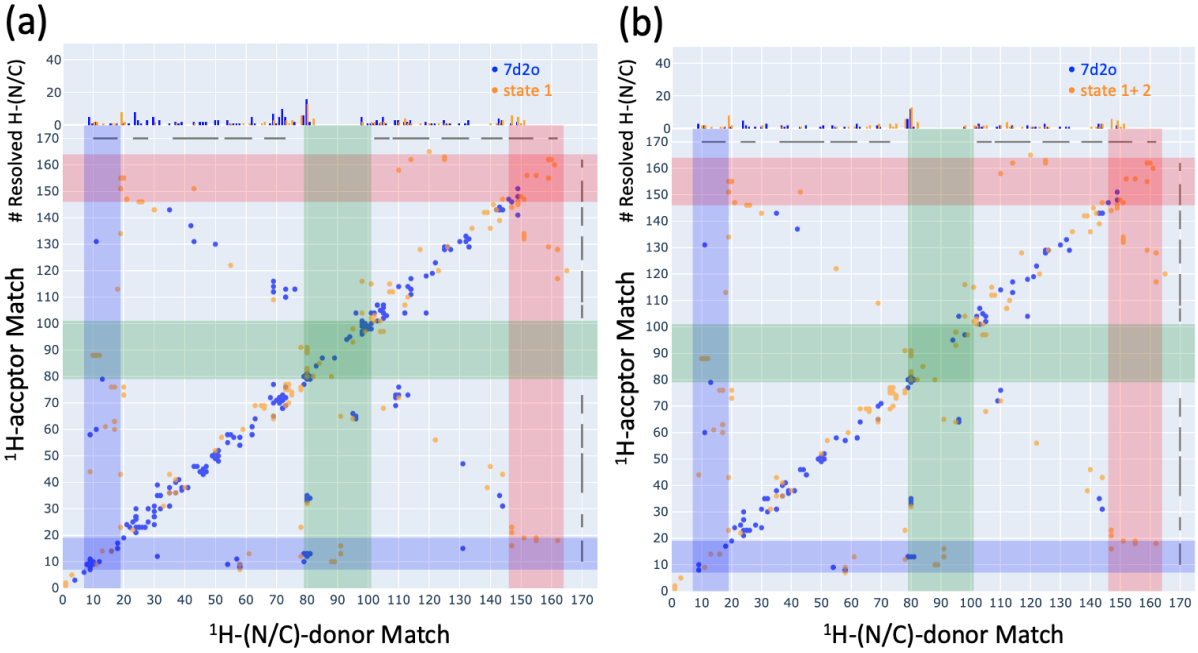


Fig. 6. Double Recall Analysis for Gluc comparing conventional NMR models against AFsample models. (a) NMR_{7d2o} (ensemble A) vs AFsample state 1 models (ensemble B) and (b) NMR_{7d2o} (ensemble A) vs AFsample state 1 + 2 models (ensemble B). NOEs in the experimental NOESY spectra that can be explained only by ensemble A are plotted as blue dots between residue number of the ^{1}H(-N/C)-donor Match (x-axis) and that of the ^{1}H-acceptor Match (y-axis). NOEs s explained only by ensemble B are indicated by orange dots. Brown dots - overlapping blue and orange dots. The upper 1D bar chart shows the number of peaks with resolved ^{1}H-(N/C) donor matches for each residue in either model A-only (blue) or model B-only (orange). Scaling is the same as for Figure 1b. On these 2D Double Recall contact maps, NOEs corresponding to interactions between helix H1 and core residues are in the blue highlighted stripes, interactions between helices H10/H11 and core residues are in the red highlighted stripes, and interactions involving the H5/H6 loop are in the green highlighted stripes.

Comparing the two-state model provided by AlphaFold-NMR to the deposited NMR_{7d20} model, there are ~7 NOEs explained by contacts in the AFsample state 1 + state 2 models between atoms in or near Tyr 80 and in or near residues Gly90, which define the thumb-shaped feature of the H5/H6 loop. However, no such thumb-shaped feature is defined in the reported NMR_{7d2o} ensemble. While the *RPF* server does not currently support analysis of mixtures of conformational states (e.g. how well a weighted combination of state 1 and state 2 fit the NOESY peak list data), there are many more NOESY peaks attributable to long-range NOEs consistent with the combined AFsample models, but not the NMR_{7d2o} models than vice versa

(**Table 2**). Specifically, ~33 NOEs are explained by NMR_{7d2o} but not the mixed-state model, while ~100 NOEs are explained by the mixed-state model but not by the NMR_{7d2o} model. In this analysis, we noticed that the flexible regions of the H5/H6 loop and H11 tend to have ambiguous matches in the ^{1}H(-N/C) dimension due to resonance degeneracies. It is very challenging to assign those NOEs using conventional methods, but they can be accounted for based on interproton distances of AlphaFold-NMR state 1 + state 2 models.

Having established that the mixture of state 1 and state 2 AFsample models selected with NMR data fit these data better than the single-state NMR_{7d2o} structure, we next investigated the structural features of these two conformational states of Gluc. Representative conformers from each of these states are shown in **Figure 7**, together with a summary of the experimental NOE data consistent with key structural features. State 1 and state 2 models are generally similar to one another, and include features of both the original NMR structure and AF2 models described previously (Wu *et al.*, 2020; Huang *et al.*, 2021). One major difference between state 1 and state 2 structures is the conformation of the thumb-shapes loop (green) between helices H5 and H6. In the "closed" AFsample state 1 models, this thumb-shape-like loop forms a "lid" that caps a large surface binding pocket (**Figure 7a-d**). This model is supported by ~8 NOEs between atoms of residue segment 88-91 in the lid and atoms of residue segment 10-16 in helix H1 (orange dots in the orange box located in the intersections of blue and green shading in **Figure 7e,** and orange dashed lines in **Figure 7h**). In the "open" state 2 models, this lid moves away from helix H1, and helix H5 also moves away from helix H1 and interacts instead more closely with helix H7 (**Figure 7b**). This "open" state 2 is supported by ~15 NOEs between atoms of residue segment 68-76 covering helix H5 and atoms of residue segment 108-116 of helix H7 (circled blue dots in **Figure 7e,** and blue dashed-lines in **Figure 7g**). In transitioning from "closed" to "open" states, the thumb-shape loop itself also becomes less packed as ~5 NOEs defining the thumb shape of state 1 can no longer be explained by state 2 (orange dots between 80-90 in **Figure 7e** and orange dashed lines in **Figure 7h**). Another significant difference between state 1 and state 2 involves the interhelical angles between core helices H3 to H4, and H7 to H8, which are smaller in state 1 models than in the state 2 models, resulting in an opening of the "pincher" structure formed by core helices H4/H7 and H3/H8|9 in the transition from "closed" to "open" structures (*cf.* **Figures 7a** and **7b**). Specifically, as the lid moves away from the central helix H1, other core helices also move away from this central helix. Furthermore, the short flexible helix H2 is an α-helix in the state 1 "closed" state, but 3_{10} helix in the state 2 "open" state. Interesting, residue Met110 seems to have the largest number of unique NOEs for the "closed" form (**Figure 7e**), suggesting that appropriate mutation of this residue could potentially shift the equilibrium between these two states.

Among the five state 1 models with highest Bayesian Model Selection score, three have a small surface pocket adjacent to the C-terminal helical region (with volumes of 27, 54, and 106 A^3, respectively) (**Figure 7c)**. We denote this surface pocket as pocket #1. However, two of these five models have no binding pockets in the region, suggesting that pocket #1 is quite dynamic. More significantly, the conformational transition from "closed" to "open" states creates a new large surface pocket with volume ranging from ~600-800 A^3 (i.e. largest pocket - state 2, rank1 model has volume 786 A^3) in the five state 2 models (rendered in blue in **Figure 7d**). These pockets - pocket #1, which is present in both states and cryptic pocket #2, which is present only in the "open" state 2 - are potential substrate binding sites.

The "closed" state 1 and "open" state 2 models also have other distinguishing structural features. Each of these state 1 and state 2 conformational states include two helices H10/H11. This C-terminal broken

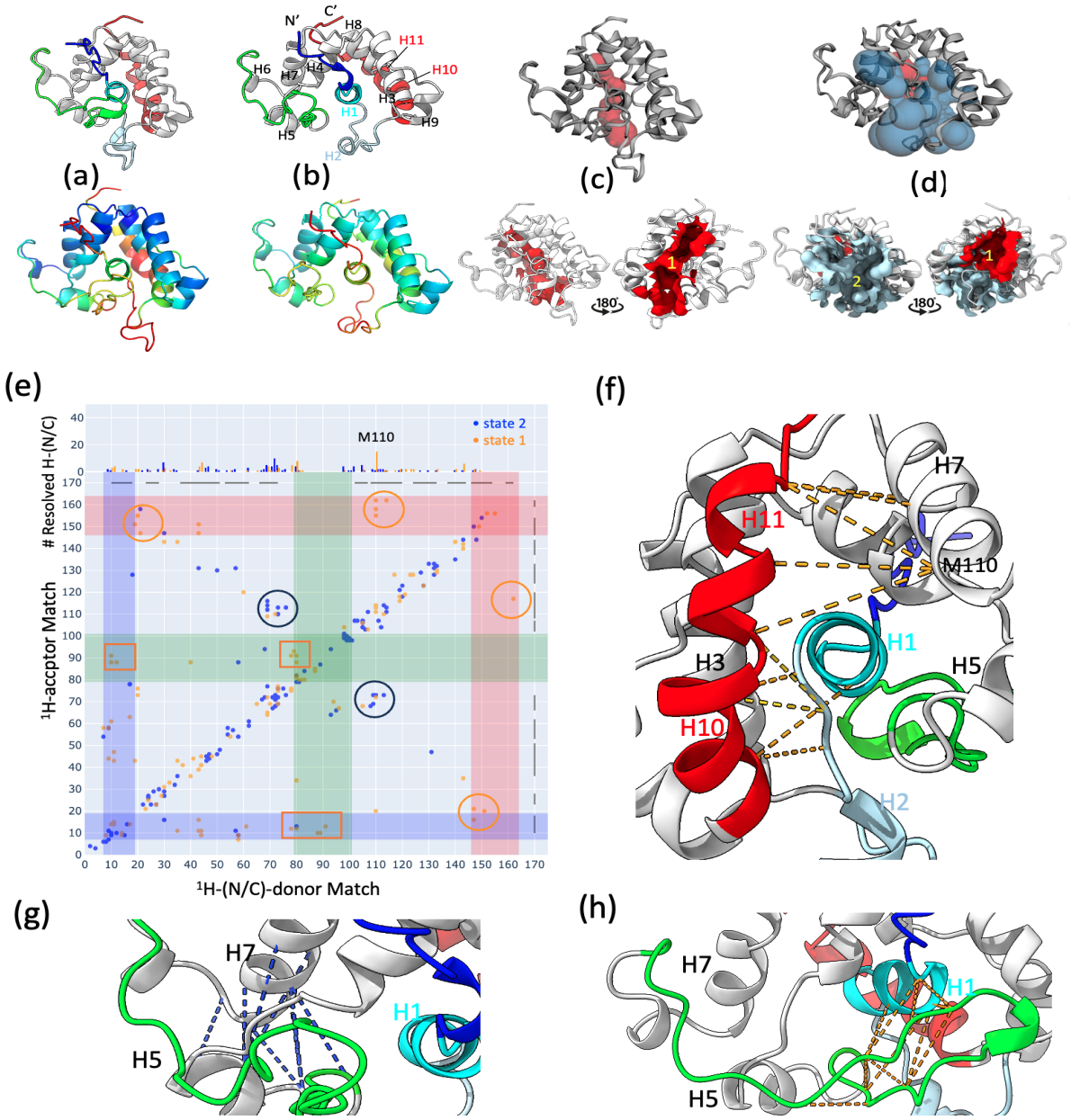


Fig. 7. Conformer-selected AlphaFold-NMR structure of Gluc. (a) Conformer-selected state 1 (state 1, rank3 model). (b) Conformer-selected state 2 (state 2, rank1 model). In panels (a) and (b) the models are colored by (top) structural element as in Fig. 1, or (bottom) based on per-residue pLDDT scores. (c) The surface pocket of "closed" state 1 is red pocket #1 on the C-terminal side (state 1, rank3 model with the highest pTM score in cluster 1: SA 274 Å^2, volume 106 Å^3) (d) Surface pockets of "open" state 2 include not only a wider red pocket #1 (state 2 rank1 model with the highest pTM score in cluster 2: SA 222 Å^2, volume 173Å^3) at the C-terminal side, but also a new large cryptic blue pocket #2 (SA 1098 Å^2, volume 786 Å^3) at the N-terminal side. (e) Double Recall plot of AFsample state 1 vs state 2, illustrating key interresidue contacts distinguishing these states; orange dots support "closed" state 1, and blue dots support "open" state 2, brown dots are overlapped blue and orange dots. Scaling is the same as for Figure 1b. (f) "Closed" state 1 with key contacts indicated as orange dashed lines (corresponding to orange dots in the red shaded region of panel e). (g) "Open" state 2 with key contacts indicated as blue dashed lines (corresponding to blue dots in the blue circles of panel e). (h) "Closed" state 1 with additional key contacts indicated as orange dashed lines (corresponding to orange dots within the orange boxes of panel e). In the "closed" state 1, the H5/H6 loop contacts helix H1 closing binding pocket #2 (h), while in the "open" state 2, this loop moves close to helices H5 and H7, opening binding pocket #2 (g). In panels f – h, the dashed lines connect the Cα-Cα atoms of the corresponding contacts.

helix is closer to the helices H1 and H7 in state 1 than in state 2. Specifically, in state 1 helices H10/11 have ~12 NOEs interacting with helix H7 (including 5 to Met110) and the end of helix H1 (**Figure 7e** and **Figure 7f**). The conformations of helices H10 and H11 for both state 1 and state 2 are a mix of α- and 3_{10}-helices, connected by β-turns. Although no Hα-HN(*i*,*i*+4) NOEs expected for an α-helix were identified, characteristic *i* -> *i*+2 NOEs (including Cys148 Hα -> Thr150 HN, Ala152 Hα ->Lys154 HN, and Asp160 Hα -> Ile162 HN) were identified in the NOESY peak lists indicating a significant population of 3_{10} helices and/or β-turns in C-terminal polypeptide segment 146-162. The transient helical structure is further stabilized in its interactions with the core structure by the disulfide bond Cys136 - Cys148. However, as the NMR chemical shift data indicate that this C-terminal region is flexible (Huang *et al.*, 2021), with extensive line broadening indicative of conformational exchange (Wu *et al.*, 2020), helices H10 and H11 appear to be in equilibrium with unfolded states. Asp160 has a much lower RCI value from its sequence neighbors suggesting that the end of H11 is less flexible than the rest of H10/11. Interestingly, helix H10 is part of the binding pocket #1 in some of the "close" state models, but becomes part of the binding pocket #2 of the "open" state. While the movement of thumb-shaped loop between H1 and H7 may regulate the formation of the cryptic binding pocket #2, the transition between disorder and helix of H10/H11 and the movement of helix H2 may regulate the sizes of both binding pockets #1 and #2.

Discussion

We describe a novel approach for protein structure determination using AlphaFold modeling with enhanced sampling (Wallner, 2023a,b) and conformational selection by NMR data. We refer to this conformational selection approach as "AlphaFold-NMR". The dynamic structural features of Gluc characterized by AlphaFold-NMR are generally consistent with structural and dynamic features described previously using conventional NMR structure determination methods (Wu *et al.*, 2020; Wu *et al.*, 2022; Dijkema *et al.*, 2024). However, additional details consistent with the NMR data but not easily derived using conventional restraint-based model generation are provided by the AlphaFold-NMR models.

These key features are illustrated in cartoon form in the left panel of **Figure 8**. In particular, while the relatively rigid architecture of the core structure formed by helices H3 - H4 and H7 - H9, which forms a clutch that holds helix H1 (**Figure 8**), is essentially identical in the restraint-based and conformer-selected models, the conformer-selection method provides novel insights regarding the transient structure of broken helix H10/H11, which packs against this core structure. More significantly, the AFsample-derived NMR landscape reveals conformational dynamics of the thumb-shaped loop between helices H5 and H6 (H5/H6 loop), creating an open/closed lid capping the cryptic pocket that includes key functional residues (Dijkema *et al.*, 2024), suggesting a functional role of these dynamics in the catalytic mechanism of its oxidation. The C-terminal segment of Gluc is in dynamic equilibrium between a structure with two helices H10 and H11 bound to the core and a disordered structure; this "broken helix" structure is more stable in the "open" state 2 where it packs with extensive interactions into both pocket #1 and pocket #2. This conclusion is supported by pLDDT data (**Figure 5**) that show that the C-terminal helices H10 and H11 helices have relatively high pLDDT scores in the state 2, but lower pLDDT in state 1. Hence, as illustrated in **Figure 8** state 1 consists of two sub-states (states 1A and 1B) corresponding folded and disordered conformations of helices H10/H11. This AlphaFold-NMR characterization is consistent with the less detailed structural and dynamic analysis of Wu et al (Wu *et al.*, 2020; Wu *et al.*, 2022) who reported broadening of resonances assigned to residues in helix H1, in the loop between helices H5 and H6, and in the C-terminal segment due to intermediate exchange between multiple conformational states,

and a potential functional dynamic role of the H5/H6 loop, based on their solution NMR structure and ^{15}N nuclear relaxation data.

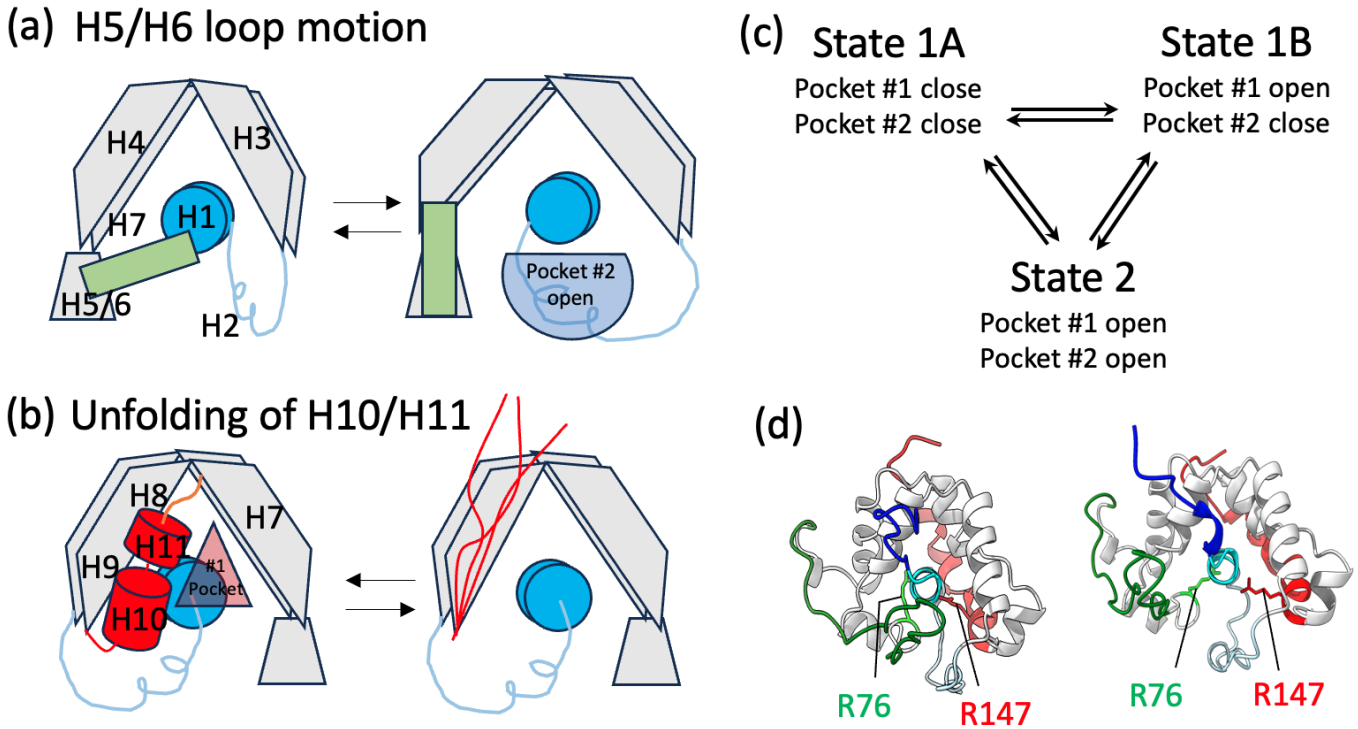


Fig. 8. Alternative conformations of Gluc indicated by the AFsample-NMR study. (a) On the N-terminal side of Gluc,, the thumb-shape loop between helices H5 and H6 (green) forms a lid that moves between close contacts with helix H1 (cyan), closing access to the hydrophobic cavity, and a more open state. (b) On the C-terminal side, the C-terminal segment exchanges between two helices H10 and H11 (red) and a partially disordered state. Chemical shift data indicate that this C-terminal segment is partially disordered, and the NOESY Double Recall analysis indicates a population of ordered structure that packs against the protein core structure. The segments labeled are: ordered helix H1 (residues 10-18), cyan; flexible helix H2 and loop (residues 19-34), light blue; H5/H6 thumb-shaped loop between helices H5 and H6 (residues 79-101), green; C-terminal segment (residues 146-168), red. (c) A conformational exchange schematic illustrating the dynamic equilibrium in solution between these three alternative states: state 1A, state 1B and state 2. State 1A and 1B together correspond to state 1, where dynamics of pocket #1 is indicated by the low pLDDT scores for H10 and H11(Fig. 5). (d) Side chains of functionally-important residues Arg76 (lime green) and Arg147 (red) are shown in state 1(left) and state 2 (right). Residue Arg76 is in binding pocket #1 while Arg147 is in binding pocket #2.

These AlphaFold-NMR models also suggest some interesting features of the catalytic mechanism of bioluminescence. Coelenterazine substrate binding has been reported to show positive cooperativity (Tzertzinis *et al.*, 2012; Larionova *et al.*, 2018), with Hill coefficient of 1.8, consistent with two binding sites. Accordingly, the two surface pockets of Gluc identified in these AlphaFold-NMR models may function as distinct binding sites, and binding of substrate in one pocket may cooperatively affect binding into the second pocket. Recently, Dijkema et al. (Dijkema *et al.*, 2024) have reported a second restraint-based NMR structure of Gluc, PDB_ID 8oqb. As part of this study, alanine scanning identified two residues, Arg76 (located near the H5/H6 loop which acts as a lid) and Arg147 (in helix H10), required for bioluminescence (Dijkema *et al.*, 2024). These key residues are separated by 15 – 35 Å in both restraint-based NMR structures 7d2o and 8oqb. In the "closed" state 1, residue Arg 76 is located within pocket #1. In the "open" state 2, residues Arg76 and Arg147 are located within pocket #1 and #2, respectively, created by the crescent-shaped core structure (**Figure 8d**). Their relative positions change in the

structural transition between state 1 and state 2 conformational states as these binding pockets expand and contract. Wu et al (Wu *et al.*, 2020; Wu *et al.*, 2022) reported that residues Phe72 and Ile73, Trp142, Leu144, and Phe151 also play a role in Gluc activities. In our AlphaFold-NMR structures, residues Phe72 and Ile73 are located together with Arg76 in binding pocket #1 of the closed state 1, and residues Trp142, Leu144, and Phe151 are located together with Arg147 in the larger cryptic pocket #2 of the open state 2. This analysis suggests that the structures selected by AlphaFold-NMR (**Figure 8**), which are more compact than either 7d2o or 8oqb, may be required for completing the functional active sites of Gluc, and that the internal dynamics of Gluc visualized by these models may impact its enzymatic function and bioluminescence properties.

AlphaFold-NMR is a general approach for combining AI-based modeling with experimental NMR data, exploiting the accuracy of these models to enable conformational selection rather than restraint-based modeling. Other kinds of NMR data, including residual dipolar couplings, back-calculated chemical shift values, nuclear relaxation data, paramagnetic relaxation enhancements, and/or paramagnetic pseudocontact shifts, could also be used for conformer selection, state selection, and/or cross validation. In addition, other AI-based modeling methods, such as AlphaFold3 (Abramson *et al.*, 2024), RosettaFold (Baek *et al.*, 2021), OpenFold (Ahdritz *et al.*, 2024), or Evolutionary Scale Modeling (Lin *et al.*, 2023) and/or alternative methods of enhanced sampling (Sala *et al.*, 2023) may ultimately be more suitable for this purpose. Hence, the AlphaFold-NMR protocol should be viewed as an evolving method for protein structure determination from NMR data.

Conclusions

In this study, we generated alternative conformational states of Gluc using AFsample, an enhanced conformational sampling method developed to improve accuracy of model predictions for protein complexes (Johansson-Akhe and Wallner, 2022; Wallner, 2023b; a). Models that best fit to NMR NOESY recall and chemical shift RCI data were then selected from these diverse conformer collections using a Bayesian selection metric. *These results demonstrate the power of enhanced sampling methods to provide information about protein conformational landscapes*.

It is remarkable and significant that AI-based modeling provides conformational diversity and indications of flexibility in the same regions of Gluc characterized as highly dynamic by NMR RCI and nuclear relaxation data (Wu *et al.*, 2020; Wu *et al.*, 2022), as well as indications of conformational rigidity in the same regions of Gluc characterized experimentally as more conformationally rigid. Importantly, the structural models of alternative conformational states revealed by enhanced sampling with AF2 are largely consistent with experimental data, including NOESY data interpreted by Double Recall analysis and chemical shift data reflected in the RCI. Hence, these AlphaFold-NMR studies of Gluc successfully exploit conformational selection of diverse AF2 models with NMR data to provide details of structural information and dynamic surface pockets which could not be obtained by conventional restraint-guided modeling.

Acknowledgements

We thank T. A. Ramelot for extensive discussions and detailed comments on the manuscript. We also thank T. Benevides, N. Dube, K. Fraga, B. A. Shurina, L. E. Spaman, G.V.T. Swapna, and R. Tejero for helpful discussions, and S. Collen for his efforts in installing and maintaining AFsample on the NPL CPU / GPU cluster of the RPI Center for Computing Innovation. The NMR data used in this study was provided by Drs. Naohiro Kobayashi, Yutaka Kuroda, Toshio Yamazaki. This research was supported by

grant R35-GM141818 (to GTM) from the National Institutes of Health, National Institute of General Medical Sciences.

Author Contributions
YJH and GTM conceptualized the study and analyzed data. YJH developed computer codes. Both authors contributed in writing and editing the manuscript.

Declaration of Interests
GTM is a founder of Nexomics Biosciences, Inc. This does not represent a conflict of interest for this study.

Supplementary Material
Table S1 - Summary of Gluc structure quality factors for AlphaFold-NMR State 1.
Table S2 - Summary of Gluc structure quality factors for AlphaFold-NMR State 2.
Table S3 - Summary of Gluc structure quality factors for conventional solution NMR structure PDB ID 7d2o.

References
Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. **Nature** *630*, 493-500. https://doi.org/10.1038/s41586-024-07487-w.

Ahdritz, G., Bouatta, N., Floristean, C., Kadyan, S., Xia, Q., Gerecke, W., O'Donnell, T.J., Berenberg, D., Fisk, I., Zanichelli, N., et al. (2024). OpenFold: retraining AlphaFold2 yields new insights into its learning mechanisms and capacity for generalization. **Nature Methods**. https://doi.org/10.1038/s41592-024-02272-z.

Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G.R., Wang, J., Cong, Q., Kinch, L.N., Schaeffer, R.D., et al. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. **Science**. https://doi.org/10.1126/science.abj8754.

Baskaran, K., Ploskon, E., Tejero, R., Yokochi, M., Harrus, D., Liang, Y., Peisach, E., Persikova, I., Ramelot, T.A., Sekharan, M., et al. (2024). Restraint validation of biomolecular structures determined by NMR in the Protein Data Bank. **Structure**. https://doi.org/10.1016/j.str.2024.02.011.

Berjanskii, M.V., and Wishart, D.S. (2007). The RCI server: rapid and accurate calculation of protein flexibility using chemical shifts. **Nucleic Acids Res** *35*, W531-537. https://doi.org/10.1093/nar/gkm328.

Bhattacharya, A., Tejero, R., and Montelione, G.T. (2007). Evaluating protein structures determined by structural genomics consortia. **Proteins** *66*, 778-795. https://doi.org/10.1002/prot.21165.

Bonin, J.P., Aramini, J.M., Dong, Y., Wu, H., and Kay, L.E. (2024). AlphaFold2 as a replacement for solution NMR structure determination of small proteins: Not so fast! **J. Magn. Reson.** https://doi.org/10.1016/j.jmr.2024.107725.

Braun, W., Bosch, C., Brown, L.R., Go, N., and Wuthrich, K. (1981). Combined use of proton-proton Overhauser enhancements and a distance geometry algorithm for determination of polypeptide

conformations. Application to micelle-bound glucagon. **Biochim Biophys Acta** *667*, 377-396. https://doi.org/10.1016/0005-2795(81)90205-1.

Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. **Acta crystallographica. Section D, Biological crystallography** *66*, 12-21. https://doi.org/10.1107/S0907444909042073.

Chinnam, N.B., Syed, A., Hura, G.L., Hammel, M., Tainer, J.A., and Tsutakawa, S.E. (2023). Combining small angle X-ray scattering (SAXS) with protein structure predictions to characterize conformations in solution. **Methods Enzymol** *678*, 351-376. https://doi.org/10.1016/bs.mie.2022.09.023.

Del Alamo, D., Sala, D., McHaourab, H.S., and Meiler, J. (2022). Sampling alternative conformational states of transporters and receptors with AlphaFold2. **Elife** *11*. https://doi.org/10.7554/eLife.75751.

Dijkema, F.M., Escarpizo-Lorenzana, M.I., Nordentoft, M.K., Rabe, H.C., Sahin, C., Landreh, M., Branca, R.M., Sørensen, E.S., Christensen, B.S., Prestel, A., et al. (2024). A suicidal and extensively disordered luciferase with a bright luminescence. **bioRxiv**, 2023.2012.2021.572513. https://doi.org/10.1101/2023.12.21.572513.

Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Žídek, A., Bates, R., Blackwell, S., Yim, J., et al. (2022). Protein complex prediction with AlphaFold-Multimer. **bioRxiv**, 2021.2010.2004.463034. https://doi.org/10.1101/2021.10.04.463034.

Fontana, P., Dong, Y., Pi, X., Tong, A.B., Hecksel, C.W., Wang, L., Fu, T.-M., Bustamante, C., and Wu, H. (2022). Structure of cytoplasmic ring of nuclear pore complex by integrative cryo-EM and AlphaFold. **Science** *376*, eabm9326.

Grant, B.J., Skjaerven, L., and Yao, X.Q. (2021). The Bio3D packages for structural bioinformatics. **Protein Sci** *30*, 20-30. https://doi.org/10.1002/pro.3923.

Guo, H.B., Perminov, A., Bekele, S., Kedziora, G., Farajollahi, S., Varaljay, V., Hinkle, K., Molinero, V., Meister, K., Hung, C., et al. (2022). AlphaFold2 models indicate that protein sequence determines both structure and dynamics. **Sci Rep** *12*, 10696. https://doi.org/10.1038/s41598-022-14382-9.

Heo, L., and Feig, M. (2022). Multi-state modeling of G-protein coupled receptors at experimental accuracy. **Proteins** *90*, 1873-1885. https://doi.org/10.1002/prot.26382.

Heo, Y., Yoon, E., Jeon, Y.-E., Yun, J.-H., Ishimoto, N., Woo, H., Park, S.-Y., Song, J.-J., and Lee, W. (2022). Cryo-EM structure of the human somatostatin receptor 2 complex with its agonist somatostatin delineates the ligand-binding specificity. **Elife** *11*, e76823.

Huang, Y.J., Powers, R., and Montelione, G.T. (2005). Protein NMR recall, precision, and F-measure scores (RPF scores): structure quality assessment measures based on information retrieval statistics. **J Am Chem Soc** *127*, 1665-1674. https://doi.org/10.1021/ja047109h.

Huang, Y.J., Rosato, A., Singh, G., and Montelione, G.T. (2012). RPF: a quality assessment tool for protein NMR structures. **Nucleic Acids Research** *40*, W542-546. https://doi.org/10.1093/nar/gks373.

Huang, Y.J., Zhang, N., Bersch, B., Fidelis, K., Inouye, M., Ishida, Y., Kryshtafovych, A., Kobayashi, N., Kuroda, Y., Liu, G., et al. (2021). Assessment of prediction methods for protein structures determined by NMR in CASP14: Impact of AlphaFold2. **Proteins** *89*, 1959-1976. https://doi.org/10.1002/prot.26246.

Jing, B., Berger, B., and Jaakkola, T. (2024). AlphaFold meets flow matching for generating protein ensembles. **arXiv:2402.04845**.

Johansson-Akhe, I., and Wallner, B. (2022). Improving peptide-protein docking with AlphaFold-Multimer using forced sampling. **Front Bioinform** *2*, 959160. https://doi.org/10.3389/fbinf.2022.959160.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. **Nature** *596*, 583-589. https://doi.org/10.1038/s41586-021-03819-2.

Kabsch, W., and Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. **Biopolymers** *22*, 2577-2637. https://doi.org/10.1002/bip.360221211.

Kaskova, Z.M., Tsarkova, A.S., and Yampolsky, I.V. (2016). 1001 lights: luciferins, luciferases, their mechanisms of action and applications in chemical analysis, biology and medicine. **Chem Soc Rev** *45*, 6048-6077. https://doi.org/10.1039/c6cs00296j.

Kirchner, D.K., and Guntert, P. (2011). Objective identification of residue ranges for the superposition of protein structures. **BMC Bioinformatics** *12*, 170. https://doi.org/10.1186/1471-2105-12-170.

Kryshtafovych, A., Schwede, T., Topf, M., Fidelis, K., and Moult, J. (2021). Critical assessment of methods of protein structure prediction (CASP)-Round XIV. **Proteins** *89*, 1607-1617. https://doi.org/10.1002/prot.26237.

Larionova, M.D., Markova, S.V., and Vysotski, E.S. (2018). Bioluminescent and structural features of native folded Gaussia luciferase. **J Photochem Photobiol B** *183*, 309-317. https://doi.org/10.1016/j.jphotobiol.2018.04.050.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK - a program to check the stereochemical quality of protein structures. **J Appl Crystallogr** *26*, 283-291.

Li, E.H., Spaman, L.E., Tejero, R., Janet Huang, Y., Ramelot, T.A., Fraga, K.J., Prestegard, J.H., Kennedy, M.A., and Montelione, G.T. (2023). Blind assessment of monomeric AlphaFold2 protein structure models with experimental NMR data. **J Magn Reson** *352*, 107481. https://doi.org/10.1016/j.jmr.2023.107481.

Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., Kabeli, O., Shmueli, Y., et al. (2023). Evolutionary-scale prediction of atomic-level protein structure with a language model. **Science** *379*, 1123-1130. https://doi.org/10.1126/science.ade2574.

Ma, P., Li, D.W., and Bruschweiler, R. (2023). Predicting protein flexibility with AlphaFold. **Proteins** *91*, 847-855. https://doi.org/10.1002/prot.26471.

Meng, E.C., Goddard, T.D., Pettersen, E.F., Couch, G.S., Pearson, Z.J., Morris, J.H., and Ferrin, T.E. (2023). UCSF ChimeraX: Tools for structure building and analysis. **Protein Sci** *32*, e4792. https://doi.org/10.1002/pro.4792.

Mirdita, M., Schutze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. **Nat Methods** *19*, 679-682. https://doi.org/10.1038/s41592-022-01488-1.

Mondal, A., Lenz, S., MacCallum, J.L., and Perez, A. (2023). Hybrid computational methods combining experimental information with molecular dynamics. **Curr Opin Struct Biol** *81*, 102609. https://doi.org/10.1016/j.sbi.2023.102609.

Monteiro da Silva, G., Cui, J.Y., Dalgarno, D.C., Lisi, G.P., and Rubenstein, B.M. (2024). High-throughput prediction of protein conformational distributions with subsampled AlphaFold2. **Nat Commun** *15*, 2464. https://doi.org/10.1038/s41467-024-46715-9.

Montelione, G.T., Nilges, M., Bax, A., Guntert, P., Herrmann, T., Richardson, J.S., Schwieters, C.D., Vranken, W.F., Vuister, G.W., Wishart, D.S., et al. (2013). Recommendations of the wwPDB NMR Validation Task Force. **Structure** *21*, 1563-1570. https://doi.org/10.1016/j.str.2013.07.021.

Robertson, A.J., Courtney, J.M., Shen, Y., Ying, J., and Bax, A. (2021). Concordance of X-ray and AlphaFold2 models of SARS-CoV-2 main protease with residual dipolar couplings measured in solution. **J Am Chem Soc** *143*, 19306-19310. https://doi.org/10.1021/jacs.1c10588.

Ruff, K.M., and Pappu, R.V. (2021). AlphaFold and implications for intrinsically disordered proteins. **J Mol Biol** *433*, 167208. https://doi.org/10.1016/j.jmb.2021.167208.

Sala, D., Engelberger, F., McHaourab, H.S., and Meiler, J. (2023). Modeling conformational states of proteins with AlphaFold. **Curr Opin Struct Biol** *81*, 102645. https://doi.org/10.1016/j.sbi.2023.102645.

Saldano, T., Escobedo, N., Marchetti, J., Zea, D.J., Mac Donagh, J., Velez Rueda, A.J., Gonik, E., Garcia Melani, A., Novomisky Nechcoff, J., Salas, M.N., et al. (2022). Impact of protein conformational diversity on AlphaFold predictions. **Bioinformatics** *38*, 2742-2748. https://doi.org/10.1093/bioinformatics/btac202.

Schrödinger, L., and DeLano, W. (2020). PyMOL.

Sedinkin, S.L., Burns, D., Shukla, D., Potoyan, D.A., and Venditti, V. (2023). Solution structure ensembles of the open and closed forms of the approximately 130 kDa ezyme I via AlphaFold modeling, coarse grained simulations, and NMR. **J Am Chem Soc** *145*, 13347-13356. https://doi.org/10.1021/jacs.3c03425.

Shen, Y., and Bax, A. (2015). Protein structural information derived from NMR chemical shift with the neural network program TALOS-N. **Methods Mol Biol** *1260*, 17-32. https://doi.org/10.1007/978-1-4939-2239-0_2.

Snyder, D.A., and Montelione, G.T. (2005). Clustering algorithms for identifying core atom sets and for assessing the precision of protein structure ensembles. **Proteins** *59*, 673-686. https://doi.org/10.1002/prot.20402.

Soding, J., Biegert, A., and Lupas, A.N. (2005). The HHpred interactive server for protein homology detection and structure prediction. **Nucleic Acids Res** *33*, W244-248. https://doi.org/10.1093/nar/gki408.

Stein, R.A., and McHaourab, H.S. (2022). SPEACH_AF: Sampling protein ensembles and conformational heterogeneity with Alphafold2. **PLOS Computational Biology** *18*, e1010483. https://doi.org/10.1371/journal.pcbi.1010483.

Tejero, R., Huang, Y.J., Ramelot, T.A., and Montelione, G.T. (2022). AlphaFold Models of small proteins rival the accuracy of solution NMR structures. **Front Mol Biosci** *9*, 877000. https://doi.org/10.3389/fmolb.2022.877000.

Tejero, R., Snyder, D., Mao, B., Aramini, J.M., and Montelione, G.T. (2013). PDBStat: a universal restraint converter and restraint analysis software package for protein NMR. **J Biomol NMR** *56*, 337-351. https://doi.org/10.1007/s10858-013-9753-7.

Terwilliger, T.C., Afonine, P.V., Liebschner, D., Croll, T.I., McCoy, A.J., Oeffner, R.D., Williams, C.J., Poon, B.K., Richardson, J.S., Read, R.J., and Adams, P.D. (2023). Accelerating crystal structure determination with iterative AlphaFold prediction. **Acta Crystallogr D Struct Biol** *79*, 234-244. https://doi.org/10.1107/S205979832300102X.

Terwilliger, T.C., Liebschner, D., Croll, T.I., Williams, C.J., McCoy, A.J., Poon, B.K., Afonine, P.V., Oeffner, R.D., Richardson, J.S., Read, R.J., and Adams, P.D. (2024). AlphaFold predictions are valuable hypotheses and accelerate but do not replace experimental structure determination. **Nat Methods** *21*, 110-116. https://doi.org/10.1038/s41592-023-02087-4.

Tian, W., Chen, C., Lei, X., Zhao, J., and Liang, J. (2018). CASTp 3.0: computed atlas of surface topography of proteins. **Nucleic Acids Res** *46*, W363-W367. https://doi.org/10.1093/nar/gky473.

Tzertzinis, G., Schildkraut, E., and Schildkraut, I. (2012). Substrate cooperativity in marine luciferases. **PLoS One** *7*, e40099. https://doi.org/10.1371/journal.pone.0040099.

Wallner, B. (2023a). AFsample: improving multimer prediction with AlphaFold using massive sampling. **Bioinformatics** *39*. https://doi.org/10.1093/bioinformatics/btad573.

Wallner, B. (2023b). Improved multimer prediction using massive sampling with AlphaFold in CASP15. **Proteins** *91*, 1734-1746. https://doi.org/10.1002/prot.26562.

Wayment-Steele, H.K., Ojoawo, A., Otten, R., Apitz, J.M., Pitsawong, W., Homberger, M., Ovchinnikov, S., Colwell, L., and Kern, D. (2024). Predicting multiple conformations via sequence clustering and AlphaFold2. **Nature** *625*, 832-839. https://doi.org/10.1038/s41586-023-06832-9.

Williamson, M.P., Havel, T.F., and Wuthrich, K. (1985). Solution conformation of proteinase inhibitor IIA from bull seminal plasma by 1H nuclear magnetic resonance and distance geometry. **J Mol Biol** *182*, 295-315. https://doi.org/10.1016/0022-2836(85)90347-x.

Wu, N., Kobayashi, N., Kuroda, Y., and Yamazaki, T. (2022). Reflecting on mutational and biophysical analysis of Gaussia princeps luciferase from a structural perspective: a unique bioluminescent enzyme. **Biophys Rev** *14*, 1513-1520. https://doi.org/10.1007/s12551-022-01025-6.

Wu, N., Kobayashi, N., Tsuda, K., Unzai, S., Saotome, T., Kuroda, Y., and Yamazaki, T. (2020). Solution structure of Gaussia luciferase with five disulfide bonds and identification of a putative coelenterazine binding cavity by heteronuclear NMR. **Sci Rep** *10*, 20069. https://doi.org/10.1038/s41598-020-76486-4.

Yeh, A.H., Norn, C., Kipnis, Y., Tischer, D., Pellock, S.J., Evans, D., Ma, P., Lee, G.R., Zhang, J.Z., Anishchenko, I., et al. (2023). De novo design of luciferases using deep learning. **Nature** *614*, 774-780. https://doi.org/10.1038/s41586-023-05696-3.

Zhang, Y., and Skolnick, J. (2004). Scoring function for automated assessment of protein structure template quality. **Proteins: Structure, Function, and Bioinformatics** *57*, 702-710.

Zheng, S., He, J., Liu, C., Shi, Y., Lu, Z., Feng, W., Ju, F., Wang, J., Zhu, J., Min, Y., et al. (2024). Predicting equilibrium distributions for molecular systems with deep learning. **Nature Machine Intelligence**. https://doi.org/10.1038/s42256-024-00837-3.

Zweckstetter, M. (2021). NMR hawk-eyed view of AlphaFold2 structures. **Protein Sci** *30*, 2333-2337. https://doi.org/10.1002/pro.4175.

Table S1 - Summary of Gluc structure quality factors for AlphaFold-NMR State 1

Total structures computed	currently unknown		
Number of structures used	5		
RMSD Values			
	all	ordered[e]	Selected[f]
All backbone atoms	1.0 Å	0.3 Å	0.2 Å
All heavy atoms	1.3 Å	0.4 Å	0.4 Å
Structure Quality Factors - overall statistics			
	Mean score	SD	Z-score [g]
Procheck G-factor [e] (phi / psi only)	-0.01	N/A	0.28
Procheck G-factor [e] (all dihedral angles)	0.17	N/A	1.01
Verify3D	0.16	0.0045	-4.82
ProsaII (-ve)	0.82	0.0071	0.70
MolProbity clashscore	1.55	3.4659	1.26
Ramachandran Plot Summary from Procheck [f]			
Most favoured regions	89.7%		
Additionally allowed regions	10.3%		
Generously allowed regions	0.0%		
Disallowed regions	0.0%		
Ramachandran Plot Statistics from Richardson's lab			
Most favoured regions	96.4%		
Allowed regions	3.6%		
Disallowed regions	0%		

[e] Residues with sum of phi and psi order parameters >= 1.8: 5A-19A, 33A-73A, 76A-165A

[f] Residues selected (Core residues by CYRANGE): 6..19, 34..150, 164..168

[g] With respect to mean and standard deviation for a set of 252 X-ray structures < 500 residues, of resolution <= 1.80 Å, R-factor <= 0.25 and R-free <= 0.28; a positive value indicates a 'better' score

Table S2 - Summary of Gluc structure quality factors for AlphaFold-NMR State 2

Total structures computed	currently unknown		
Number of structures used	5		
RMSD Values			
	all	ordered[e]	Selected[f]
All backbone atoms	0.8 Å	0.7 Å	0.5 Å
All heavy atoms	1.2 Å	0.9 Å	0.7 Å
Structure Quality Factors - overall statistics			
	Mean score	SD	Z-score [g]
Procheck G-factor [e] (phi / psi only)	-0.21	N/A	-0.51
Procheck G-factor [e] (all dihedral angles)	-0.15	N/A	-0.89
Verify3D	0.13	0.0055	-5.30
ProsaII (-ve)	0.86	0.0167	0.87
MolProbity clashscore	No clash	No clash	>3.00
Ramachandran Plot Summary from Procheck [f]			
Most favoured regions	84.2%		
Additionally allowed regions	15.7%		
Generously allowed regions	0.2%		
Disallowed regions	0.0%		
Ramachandran Plot Statistics from Richardson's lab			
Most favoured regions	92.4%		
Allowed regions	6.7%		
Disallowed regions	1%		

[e] Residues with sum of phi and psi order parameters >= 1.8: 2A-20A, 24A-74A, 76A-80A, 83A-85A, 88A-161A

[f] Residues selected (Core residues by CYRANGE): 8..18, 31..163

[g] With respect to mean and standard deviation for a set of 252 X-ray structures < 500 residues, of resolution <= 1.80 Å, R-factor <= 0.25 and R-free <= 0.28; a positive value indicates a 'better' score

Table S3 - Summary of Gluc structure quality factors for conventional solution NMR structure PDB ID 7d2o.

Total structures computed	currently unknown		
Number of structures used	19		
RMSD Values			
	all	ordered[e]	Selected[f]
All backbone atoms	7.4 Å	1.4 Å	1.1 Å
All heavy atoms	7.2 Å	1.7 Å	1.4 Å
Structure Quality Factors - overall statistics			
	Mean score	SD	Z-score [g]
Procheck G-factor [e] (phi / psi only)	-0.24	N/A	-0.63
Procheck G-factor [e] (all dihedral angles)	-0.63	N/A	-3.73
Verify3D	0.11	0.0130	-5.62
ProsaII (-ve)	0.53	0.0497	-0.50
MolProbity clashscore	23.66	1.6608	-2.53
Ramachandran Plot Summary from Procheck [f]			
Most favoured regions	86.3%		
Additionally allowed regions	10.9%		
Generously allowed regions	1.2%		
Disallowed regions	1.5%		
Ramachandran Plot Statistics from Richardson's lab			
Most favoured regions	91.3%		
Allowed regions	6.6%		
Disallowed regions	2.1%		

[e] Residues with sum of phi and psi order parameters >= 1.8: 8A-18A, 23A-27A, 35A-78A, 97A-134A, 137A-144A

[f] Residues selected (Core residues by CYRANGE): 7..19, 34..81, 95..143

[g] With respect to mean and standard deviation for a set of 252 X-ray structures < 500 residues, of resolution <= 1.80 Å, R-factor <= 0.25 and R-free <= 0.28; a positive value indicates a 'better' score