

Assessment of Prediction Methods for Protein Structures Determined by NMR in CASP14: Impact of AlphaFold2

Yuanpeng Janet Huang¹, Ning Zhang², Beate Bersch³, Krzysztof Fidelis⁴, Masayori Inouye⁵, Yojiro Ishida⁵, Andriy Kryshchak⁴, Naohiro Kobayashi⁶, Yutaka Kuroda⁷, Gaohua Liu⁸, Andy LiWang^{2,9}, G.V.T. Swapna¹⁰, Nan Wu¹¹, Toshio Yamazaki⁶, and Gaetano T. Montelione^{1,*}

¹Dept of Chemistry and Chemical Biology, Center for Biotechnology and Interdisciplinary Sciences, Rensselaer Polytechnic Institute, Troy, New York, 12180 USA

²Department of Chemistry and Biochemistry, University of California, Merced, California, 95343 USA

³Biomolecular NMR Spectroscopy Group, Institut de Biologie Structurale, UMD-5075, CNRS-CEA-UJF, Grenoble, FRANCE

⁴Genome Center, University of California, Davis, CA 95616, USA.

⁵Dept of Biochemistry, Robert Wood Johnson Medical School, and Center for Advanced Biotechnology and Medicine, Robert Wood Johnson Medical School, Rutgers University, Piscataway, New Jersey, 08854 USA.

⁶NMR Science and Development Division. RSC, RIKEN, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama City, Kanagawa 230-0045, JAPAN

⁷Department of Biotechnology and Life Science, Graduate School of Engineering, Tokyo University of Agriculture and Technology (TUAT), 2-24-16 Nakamachi, Koganei-shi, Tokyo 184-8588, JAPAN.

⁸Nexomics Biosciences, Inc., 5 Crescent Ave, Rocky Hill, New Jersey 08553, USA.

⁹Center for Cellular and Biomolecular Machines, and Health Sciences Research Institute, University of California, Merced, California, 95343 USA

¹⁰Dept of Pharmacology, Robert Wood Johnson Medical School, Robert Wood Johnson Medical School, Rutgers University, Piscataway, New Jersey, 08854 USA.

¹¹College of Food and Bioengineering, Zhengzhou University of Light Industry, 136 Kexue Road, Zhengzhou 450001, PEOPLES REPUBLIC OF CHINA.

*Corresponding author. Email: monteg3@rpi.edu

Key words: Protein structure prediction, solution NMR, inverse structure determination,
machine learning, protein dynamics, integral membrane proteins, MipA

All authors should provide ORCID number

Yuanpeng Janet Huang: 0000-0002-3374-786X

Ning Zhang: 0000-0002-3590-3774

Beate Bersch: 0000-0003-1138-4911

Krzysztof Fidelis: 0000-0002-8061-412X

Masayori Inouye: 0000-0003-4490-600X

Yojiro Ishida: 0000-0001-6781-367X

Andriy Kryshchak: 0000-0001-5066-7178

Naohiro Kobayashi: 0000-0001-5136-2077

Yutaka Kuroda: 0000-0002-3221-4295

Gaohua Liu: 0000-0002-4343-0356

Andy LiWang: 0000-0003-4741-6946

G.V.T. Swapna: 0000-0001-9342-2525

Nan Wu: 0000-0003-2274-4900

Toshio Yamazaki: 0000-0003-4719-0828

Gaetano T. Montelione: 0000-0002-9440-3059

Abstract (250 words, currently 245)

NMR studies can provide unique information about protein conformations in solution. In CASP14, three reference structures provided by solution NMR methods were available (T1027, T1029, and T1055), as well as a fourth data set of NMR-derived contacts for an integral membrane protein (T1088). For the three targets with NMR-based structures, the best prediction results ranged from very good (GDT_TS = 0.90, for T1055) to poor (GDT_TS = 0.47, for T1029). We explored the basis of these results by comparing all CASP14 prediction models against experimental NMR data. For T1027, NMR data reveal extensive internal dynamics, presenting a unique challenge for protein structure prediction. The analysis of T1029 motivated exploration of a novel method of “inverse structure determination”, in which an AF2 model was used to guide NMR data analysis. NMR data provided to CASP predictor groups for target T1088, a 238-residue integral membrane porin, was also used to assess several NMR-assisted prediction methods. Most groups involved in this exercise generated similar beta-barrel models, with good agreement with the experimental data. However, as was also observed in CASP13, some pure prediction groups that did not use any NMR data generated models for T1088 that better fit the NMR data than the models generated using these experimental data. These results demonstrate the remarkable power of modern methods to predict structures of proteins with accuracies rivaling solution NMR structures, and that it is now possible to reliably use prediction models to guide and complement experimental NMR data analysis.

Introduction

The remarkable performance of some protein structure prediction groups in the 2020 Critical Assessment of Protein Structure Prediction experiment 14 (CASP14) has set a new standard for protein structure modeling [1]. These breakthrough technologies exploit advances in attention-based machine learning [2, 3], contact prediction based on sequence co-variance analysis using the massive data bases of genomic sequence data [4-9], and the rapidly growing database of experimental protein structures. In particular, in blind tests of protein structure prediction accuracy on 96 CASP14 targets, the performance of DeepMind AlphaFold2 (AF2) [10] had an unprecedented high accuracy, assessed by backbone atomic coordinate global distance test (GDT_TS) scores [11], of 0.88 ± 0.1 , corresponding to a root-mean-squared deviation (RMSD) between predicted and experimental protein structures of about 1.5 Å [1]. Buried sidechain conformations in these blind predictions of protein structure are also generally a remarkable good match between the predicted model and experimental structure [12]

The available machine learning methods, including AlphaFold2 [10] and RosTTAaFold [13], are trained on the extensive Protein Data Base (PDB) of protein structures with the assumption that the true structure is a single conformation. Since most of the data in the PDB, and in fact most experimental protein structures, have been provided by X-ray crystallography, this assumption is relevant. However, in their biological contexts proteins are dynamic and adopt multiple conformational states as required for their thermodynamic stability and functions. While X-ray crystallography can be used to study protein dynamics, and the various dynamic states of proteins may crystallize separately providing atomic resolution structures for alternative states, other experimental techniques have unique capabilities for characterizing the multiple

conformational states of proteins. In particular, both solution NMR and cryo electron microscopy (cryoEM) are especially powerful in identifying and characterizing multiple conformational states of proteins. To date, multiple conformational and dynamic structure prediction has not been a focus in CASP, and is not generally considered in training of machine learning and other protein structure prediction methods.

In the 2018 CASP13 experiment, we explored the concept of using incomplete “sparse” solution NMR data to assist protein structure prediction methods [14]. The aim of this earlier study was to assess if advanced structure prediction methods could be combined with the kinds of sparse NMR data that can be obtained on medium-sized (20 – 50 kDa) proteins, which are otherwise challenging for structure determination by solution NMR. NOESY data typical of that easily obtained for ^{15}N , ^{13}C -enriched, perdeuterated proteins up to about 40 kDa, were simulated for 11 CASP13 targets ranging in size from 80 to 326 residues, and used to generate tables of ambiguous contacts using simple NOESY peak assignments protocols. These Ambiguous Contact Lists were provided, together with simulated ^{15}N - ^1H residual dipolar coupling data and backbone dihedral angle restraints obtainable from chemical shift data, to the CASP prediction community for data-assisted prediction. Real NMR data collected for a *de novo* designed protein were also used to generate ambiguous contact tables and chemical-shift based backbone dihedral angle restraints, that were provided to CASP13 predictor groups, including one set of (ambiguous) NMR-based contacts in which only backbone resonance (no sidechain) assignments were available. Guided by these “sparse” experimental NMR data, some CASP13 prediction groups generated models more accurate than those produced using more traditional protein NMR modeling methods.

The best NMR-assisted models were also compared with the best “regular” predictions (i.e. pure prediction) models provided by all CASP13 groups. For 6 of 13 target data sets, the most accurate model provided by any NMR-assisted prediction group was more accurate than the most accurate model provided by any regular prediction group, as expected. *However, for the remaining 7 target data sets, one or more regular prediction method provided a more accurate model than even the best NMR-assisted model.* Here, accuracy was assessed by comparison with the reference X-ray crystal structure from which ambiguous contacts were derived, or the experimental NMR structure determined with a much larger amount of NMR data. Hence, for some of these blind structure predictions, pure prediction methods provided more accurate models than either traditional NMR structure determination or data-assisted prediction methods that used these simulated or real sparse NMR data [14]. Machine learning methods, and particularly the AlphaFold methods (the progenitor of AlphaFold2), were particularly successful in providing accurate models even without any experimental data [14].

In CASP14, three reference structures provided by solution NMR methods were available (targets T1027, T1029, and T1055), as well as a fourth data set of NMR-derived contacts for NMR data-guided structure prediction (T1088). For the three CASP14 targets with reference structures provided by solution NMR methods, the best AF2 prediction results range from very good (GDT_TS_best = 0.90, for T1055), to medium (GDT_TS_best = 0.67, for T1027), to poor (GDT_TS_best = 0.47, for T1029). We explored the basis of these results by comparing ¹H-¹H distance maps derived from these models against the experimental NOESY peak lists using recall and precision scores (RPF-DP scores) [15, 16]. Models were also compared with backbone

chemical shift data using the *TALOS_N* program [17], and residual dipolar coupling (RDC) data where available. These results demonstrate the remarkable reliability of some CASP14 prediction models, particularly AlphaFold2, and reveal different reasons for the differences between experimental and prediction models for each target.

Methods

Knowledge-based Structure Validation. Structure quality assessment included analysis of knowledge-based structure quality scores, including Ramachandran backbone analysis [18], *ProCheck* dihedral angle analysis for both backbone dihedral angles and all dihedral angles (i.e., backbone plus sidechain) [19], *ProsaII* [20], *Verify3D* [21], and *Molprobit* [22], using the Protein Structure Validation Software suite (*PSVS*) server [23]. Knowledge-based dihedral analysis was restricted to well-defined residues, defined by the method of *Cyrange* [24] as recommended by the wwPDB NMR Structure Validation task force [25]. For each of these knowledge-based structure quality assessment metrics, Z scores are reported relative to the corresponding raw scores obtained for a set of 252 X-ray crystal structures each of < 500 residues, and with resolution ≤ 1.8 Å, R factor ≤ 0.25 , and R-free ≤ 0.28 [23]; positive Z scores correspond to knowledge-based structure quality scores better than the average score in this set of reference structures. Generally speaking, acceptable NMR-based models have Z scores > -3.0 for *ProCheck* (backbone), *ProCheck* (backbone plus sidechain), *ProsaII*, and *MolProbit* [23], while *Verify3D* scores for accurate structures are more variable and dependent on the protein fold, but generally have Z scores > -5.0 .

NMR Restraint Violation Analysis. NMR distance and restraint violations were assessed consistently using experimental distance restraint lists generated by different programs and available in the Protein Data Bank using the *PDBStat* software [26]. Model agreement with backbone chemical shift data deposited in the BioMagResDatabase was assessed using the *Talos_N* program [17] .

Residual Dipolar Coupling Q Scores. The RDC Q score (or quality factor) [27] is used to quantify the extent of agreement between a structure and measured dipolar couplings. A Q score below 0.2 can be used as a rule of thumb to indicate adequate agreement. Q scores are calculated using the following equation :

$$Q \text{ score} = \sqrt{\frac{\sum_i W^2 (D_{i,obs} - D_{i,cal})^2}{N * D_a^2 (4 + 3R_h^2)/5}}$$

W is the weighting factor; N is the number of dipolar couplings; D_a is the axial component of the tensor and R_h is the rhombicity. Residual dipolar couplings and Q scores were back-calculated from models using single-value decomposition tools available on DC: Servers for Dipolar Coupling Calculations (<https://spin.niddk.nih.gov/bax/nmrserver/dc/svd.html>).

RPF-DP Scores for CASP14 NMR Structures and Prediction Models. RPF-DP scores are a set of fast and sensitive structure quality assessment measures which can be used to evaluate how well a 3D structure model fits with NOESY peak and chemical shift lists to indicate the correctness of the fold and accuracy of the structure [15, 16]. RPF-DP scores provide one type of NMR R-factor, in which models are compared against NMR NOESY data. They have been

described previously [15, 16], but as they play a key role in this work, we provide an overview of these model vs data structure quality assessment metrics here.

The RPF-DP score algorithm is outlined schematically in Figure 1. Nodes represent all protons listed in the resonance assignment table. Edges connect the nodes and represent all potential associated NOEs from the NOESY peak lists, within a chemical shift match tolerance. In constructing G_{ANOE} (shown on right side of Figure 1) each NOESY cross peak (p) may be ambiguously assigned to one or more proton pairs, as determined by chemical shift degeneracies and match tolerances. The solution network, G_{NOE} , corresponding to the true 3D structure, is a subgraph of G_{ANOE} . Given complete NOESY peak lists and resonance assignments, for each NOESY cross peak p, at least one of its possible proton pair assignments has a corresponding edge in G_{NOE} . For each structure model (shown on left side of Figure 1), a distance network G is calculated from the summation distances (sum of inverse sixth powers of individual degenerate proton-proton distances), assuming uniform effects of nuclear relaxation processes. Nodes are connected by an edge in G if the corresponding interproton summation distance in the model structure is $\leq d_{\text{NOE_max}}$, where $d_{\text{NOE_max}}$ is the (estimated) maximum distance detected in the NOESY spectrum. Summation distances are used to address the lack of stereospecific assignments of prochiral methylene proton pairs, sets of protons that are degenerate (e.g. the three hydrogens of a methyl group, degenerate methylene protons, or degenerate resonances of Tyr or Phe), or combinations of these kinds of ambiguities (e.g. for prochiral isopropyl methyl groups of Leu or Val for which stereospecific assignments are not available).

NOESY cross peaks represented in G_{ANOE} that are consistent with the short interproton distances in the network derived from the model, G , are defined as true positives (TPs), while NOESY peaks expected from the model (edges in G) that are not observed in the data, G_{ANOE} , are true negatives (TNs). As illustrated in Figure 1, particular proton pair interactions present in the atomic coordinates of a model structure, represented by the network G , may either be represented in the graphical representation of the NOESY peak list data G_{ANOE} (TP), or not (FP). Since G_{ANOE} is an ambiguous network, a FN score is assigned to the peak only if none of the several possible short proton-proton distance consistent with the possible NOESY peak assignments are observed in G . In this context, recall measures the fraction of NOE cross peaks that are consistent with the query model structures, while precision measures the fraction of proton pair interactions in the query structure that are observed in the NOESY peak list (in G_{ANOE}), weighted by interproton distance. The F-measure is the harmonic mean of the recall and precision. Equations used to calculate recall (R), precision (P), and F-measure (F, also called the performance) are shown in Figure 1.

The DP score is a normalized F-measure that accounts for lower-bound and upper-bound values of the F-measure. The lower-bound of $F(G)$ is estimated by $F(G_{\text{free}})$, where G_{free} is a distance network graph computed from interproton distances in a freely rotating polypeptide chain model, as described by Flory and co-workers [28]. The upper-bound of $F(G)$ is estimated by $F(G_{\text{ideal}})$. G_{ideal} is the graph of a hypothetical ideal structure that is perfectly consistent with G_{ANOE} . It is defined so that $\text{recall}(G_{\text{ideal}}) = 1$ and $\text{precision}(G_{\text{ideal}}) = \text{precision}(G_{\text{local}})$, where G_{local} is the network of all two and three-bond connected proton pairs; i.e. the completeness of the network G_{ANOE} is assumed to be approximately the same as the completeness of the subnetwork

of NOEs associated with these local ^1H - ^1H distances, G_{local} . With these definitions, $F(G_{\text{ideal}})$ represents the best possible performance F considering the quality of the input NOESY peak lists and resonance assignments. $F(G_{\text{ideal}})$, and particularly the precision of G_{ideal} , thus provides a measure of the combined quality of the resonance assignment and NOESY peak lists for one or more spectra. $F(G_{\text{ideal}})$ and $F(G_{\text{free}})$ describe the two bounds of the performance $F(G)$; i.e., $F(G_{\text{ideal}}) \leq F(G) \leq F(G_{\text{free}})$. With these definitions, the fold Discriminating Power (DP) for G is then estimated by scaling the F values so that $F(G_{\text{ideal}}) = \text{DP}(G_{\text{ideal}}) = 1$, and $F(G_{\text{free}}) = \text{DP}(G_{\text{free}}) = 0$. This scaling is necessary to account for the fact that the NOESY data may not be complete, and the observation that even a random coil chain model can satisfy a large part of the NOESY peak list data [16].

The default upper-bound observed distance, $d_{\text{NOE_max}}$, used in these metrics is 5 Å, but can also be calibrated from the NOESY data. In this analysis, a distance (d^{-6}) weighting of the precision metric, $\text{precision}_w(G)$, is used to reduce the otherwise dominant influence of the many weak NOEs arising from interproton distances close to the upper-bound detection limit, $d_{\text{NOE_max}}$. This weighting also makes these quality scores less sensitive to the value chosen for $d_{\text{NOE_max}}$ [16].

RPF-DP scores can be calculated for individual models, or using average distances across an ensemble. The ensemble DP score is usually higher than individual DP scores. In various studies [15, 16, 29, 30], structures within 2.0 Å RMSD of the corresponding “correct” structure have been observed to have DP scores > 0.70 for NMR ensembles, and DP scores > 0.60 for

individual conformers. Perdeuterated protein NMR data, like that obtained here for MipA, require a larger d_{\max} (7 Å), and generally provide somewhat lower scores.

ANSURR Scores. The Accuracy of NMR Structures Using RCI and Rigidity (ANSURR) method provides an independent assessment of model quality by comparing protein flexibility computed from backbone chemical shifts with protein flexibility predicted with a graph theory based measure of structural rigidity [31]. ANSURR provides two measures of similarity between these measures, a correlation score (corr) which assesses the correlation between peaks and troughs of observed and predicted structural flexibility along the sequence, and root-mean-squared deviation (RMSD) between the metrics. Both the corr and RMSD score are reported as a percentile score (ranging from 0 to 100). These scores were calculated using the ANSURR program version 1.0.2 (<https://zenodo.org/badge/latestdoi/234519929>).

Experimental NOESY Peak Lists and Model Preparation. For targets T1027, T1029, and T1055, NOESY peak lists were obtained from the experimentalists who carried out the original NMR structure analyses. For target T1088, experimental data was collected and NOESY peak lists were generated as outlined in Supplementary Methods. For assessing experimental NMR structures, the coordinates with hydrogen atoms for each conformer are used. For assessing prediction models or X-ray crystal structures, which generally do not include hydrogen atoms, the program *Reduce* [32] was used to add protons with ideal covalent geometries. DP scores were calculated by comparing individual conformers against the NOESY peak and chemical shift lists. We noticed that the program *Reduce* failed to add all protons for some of the CASP14

models, due to their unrealistic heavy atom geometry. The DP scores for these models with bad geometry tend to have very small or negative values.

Global Distance Test Scores. GDT_TS scores were computed by the CASP Prediction Center using the method of Zemla [11]. For brevity, GDT_TS scores are referred to throughout this paper as GDT scores.

Molecular Modeling. Molecular modeling was done using *PyMol* [33].

NMR Data for Integral Membrane Protein Target MipA in Detergent Micelles. MipA is an antibiotic-resistance factor, which acts to transport some drugs out of bacteria, while enhancing transport of other drugs into bacteria [34]. The expression, isotope-enrichment, and purification of MipA is outlined in the Supplementary Methods. Briefly, a synthetic codon-optimized gene (Genscript, Inc) for *Klebsiella pneumoniae* MipA was expressed using the pColdII single protein expression system [35, 36]. The resulting protein construct includes a short N-terminal 6xHis purification tag. MipA samples for NMR studies were prepared with ^2H , ^{13}C , ^{15}N , and $^{13}\text{CH}_3$ methyl-enrichment. MipA was expressed in *E. coli* BL21(DE3) $\Delta hisB$ cells harboring *pACYCmazF(ΔH)* and *pCold2-mipA*, solubilized with 8M urea, purified by Ni-NTA affinity chromatography, and then refolded by slow removal of urea by dialysis. The purified protein was prepared in 20 mM potassium phosphate buffer at pH 6.5, containing 0.2 M NaCl, 50 mM M Arg, and 0.1% d_{37} -DPC. The resulting sample was > 95% homogenous on SDS-PAGE gels. The final protein concentration for NMR studies was ~ 0.5 mM. ^2H -decoupled NMR studies were carried out using Avance 600 and 800 NMR spectrometer systems located at Rutgers

University, Princeton University, and Rensselaer Polytechnic Institute. Details of NMR data collection and processing are provided in Supplementary Methods. NMR data for MipA were provided to CASP14 predictor groups in the form of Ambiguous Contact Lists, prepared as described previously [14]. Briefly, for each NOESY cross peak we provide a list of possible assignments by analyzing NOESY peak lists together with the corresponding resonance assignment lists using the Cycle 0 module of the program *ASDP* [37], providing a simple matching between resonance frequencies of the NOESY peak and the resonance assignment list. Details of this process are outlined in Supplementary Methods.

Results

Target T1055: A20₃₀₄₋₄₂₆. The A20 protein of vaccinia virus forms a heterodimer processivity factor with the uracil-DNA glycolase, D4 protein, and binds the catalytic subunit of the DNA polymerase, E9 protein, to form the essential DNA polymerase holoenzyme E9-A20-D4 required for viral DNA synthesis. CASP14 target T1055 is the C-terminal domain of A20, corresponding to the last 123 residues. The construct used for structural studies included a C-terminal biotin acceptor protein (BPAP) tag, connected by a 10-residue linker. The solution NMR structure of A20₃₀₄₋₄₂₆ was determined by Bersch et al. [38] from triple-resonance NMR and NOESY data ($\tau_m = 100$ ms), which provided 2351 unambiguous and 566 ambiguous distance restraints, together with 218 backbone dihedral angle restraints based on analysis of chemical shift data with *Talos+*. The *UNIO10* program suite [39] was used for initial NOE assignment. The resulting NOESY peak lists generated with the *ATNOS* peak picking algorithm were then used for structure calculation using *ARIA* 2.3 [40], followed by refinement with *CNS* 1.21 [41]. Although potential NOEs were observed between the C-terminal linker-BPAP purification tag and the core

of the structure, these NOEs were excluded from the analysis because of their ambiguity in assignment [38]. The resulting well-defined structure (PDB ID 6zyc), reported as an ensemble of 20 conformers, includes 5 N-terminal α -helices, a two-stranded antiparallel β -sheet, and a long C-terminal helix. ^{15}N relaxation data indicate that A20₃₀₄₋₄₂₆-BAP has dynamic flexibility in its N-terminal ~ 10 -residue polypeptide segment, and in the C-terminal linker-BAP tag, but otherwise has a relatively static overall backbone structure.

We assessed the similarities between NMR and CASP14 prediction models, including AF2 models (Figure 2, left). In comparing the predicted AF2 structure of T1055 with the experimental structure, we excluded residue segments for which atomic positions are not well defined in either the ensemble of 20 NMR-derived conformers or the ensemble of 5 AF2 conformers. Well-defined regions of the of the NMR ensemble, residues 305 – 426 (Figure 2A) and AF2 ensemble, residues 310 – 426 (Figure 2B), were identified using the program *Cyrange*. For the residues whose backbone positions are well defined in both ensembles, i.e. for residues 310 – 426, the pairwise GDT scores between the NMR model with best DP score and the 5 AF2 conformers ranged from 0.89 to 0.90, corresponding to backbone rmsd's of about 1.3 Å (Figure 1C). Many buried sidechain conformations also have relatively good agreement between the AF2 and NMR structures (Figure 1D). This is a remarkable result considering that the AF2 prediction did not use any NMR data.

Structure quality statistics for T1055 were also analyzed with the PSVS software suite. The resulting PSVS structure quality statistics for both the NMR and AF2 model ensembles are summarized in Supplementary Tables S1-S2. Both the NMR and AF2 models generally exhibit

excellent structure quality scores and good energetics. However, the AF2 models have significantly better ProCheck G-factor (all dihedral angle) and Molprobity clash scores, attributable to more energetically consistent core sidechain packing.

We next assessed how well the NMR and AF2 structures fit to the experimental NMR chemical shift (bmr_b_id 34545) and NOESY peak list data using the RPF-DP score [15, 16]. Plots of DP score vs GDT for all CASP14 predictor groups have a strong correlation, and DP scores ranged from -3.06 to 0.63 (Figures 1E,F). The prediction model with highest DP score, 0.63 for AF2 model 2 (model 427_2) is higher than the highest DP score for any of the NMR conformers, 0.58 (Figure 1E); *i.e., some AF2 models fit the NOESY data better than the NMR model itself.*

RPF DP analysis also provides information about which regions of experimental and prediction models fit to, or violate, the NOESY data. This analysis for target 1055 is summarized on the left side of Figure 3. The recall analysis (NOESY peaks that cannot be explained by the model) indicates that most NOESY peaks are consistent with both the NMR and AF2 models. Overall, the NMR models ($R = 0.97$) have fewer recall violations than the AF2 models ($R = 0.95$ - -0.96). There are a small number of NOESY peak data that are consistent with the AF2 models, but not the NMR model (Figure 3A and C), and a small number of NOESY peaks that are consistent with the NMR models but not with the AF2 models (Figure 3B and C). The histogram plot (Figure 3C) indicates only 7 NOESY peaks consistent with the AF2 structure, but not the NMR structure, while 54 NOESY peaks are consistent with the NMR structure, but not the AF2 structure.

On the other hand, both the NMR ($P = 0.74 - 0.76$) and AF2 ($P = 0.78 - 0.79$) models have significant numbers of precision violations; i.e. short distances that are not supported by NOESY peaks. These are distributed throughout the structures (cf. Figure 3D and E). These precision violations arise mostly from sidechain packing that is not fully consistent with the NOESY data. Overall, the AF2 models have much fewer precision violations, consistent with the better ProCheck G-factor (all dihedrals) and Molprobit scores, cited above, which indicate more energetically-consistent core sidechain packing in the AF2 models. These differences may be related to the quality of force fields and energy refinement protocols used in the NMR and AF2 modeling processes.

Finally, we also assessed how well the NMR and AF2 models satisfy backbone dihedral restraints derived from backbone chemical shift data using *Talos_N* [17]. All of the NMR and AF2 models satisfy these chemical shift data (Figure 3 F and G). Overall, both the NMR and AF2 models of T1055 fit well to the NOE and chemical shift data, although the smaller number of precision violations (short distances that are not supported by the NMR data) and better *ProCheck* (all dihedral) and *MolProbit* clash scores for the AF2 models indicates they have more accurate core sidechain packing.

Target T1027: *Gaussia* luciferase (GLuc). Luciferases are bioluminescent enzymes. CASP14 target T1027 (GLuc) is a 168-residue luciferase isolated from the marine organism *Gaussia princeps* [42]. GLuc catalyzes the oxidation of coelenterazine generating a bright blue light, and is attracting interest as a genetically-encodable reporter protein. Recombinant GLuc for NMR

structure determination was expressed in *E. coli* [43] and despite its five disulfide bonds, it was refolded into its active form in amounts sufficient for structural analysis using a Solubility Enhancement Peptide tag [44]. The solution NMR structure of T1027 was determined by Wu, Kobayashi, Kuroda, Yamazaki and colleagues [45] using triple-resonance NMR and NOESY data ($\tau_m = 80$ ms), providing $2,573 \pm 42$ NOESY-derived distance restraints, together with 183 backbone dihedral angle restraints determined from *Talos+* analysis of chemical shift data, 25 hydrogen bond restraints indicated by amide $^1\text{H}/^2\text{H}$ exchange data, and restraints for 3 disulfide bonds. *CYANA* 3.98 [46] was used for both automated NOESY peak assignment and structure generation; no additional energy refinement was done.

The resulting structure (PDB ID 7d2o), reported as an ensemble of 19 conformers, includes 9 α -helices, and 5 disulfide bonds. The pairing of 3 disulfide bonds (C59/C120, C65/C77, and C136/C148) were determined unambiguously by the NMR structure. However due to the proximity of Cys residues along the sequence, pairing of the remaining four cysteine (C52, C56, C123, and C127) disulfide pairings were could not be unambiguously distinguished from the NMR structures. The C52/C127 pairing is, however, consistent with at least one proteolytic fragment observed in mass spectroscopy, and statistical analysis of $S\gamma$ - $S\gamma$ distances across the ensemble of NMR structures strongly suggested C52/C127 and C56/C123 as the most likely pairings.

Structural convergence, proton linewidth, ^{15}N relaxation dispersion, and ^1H - ^{15}N heteronuclear NOE data (HetNOE) data indicate that GLuc has extensive internal conformational dynamics. Residues 1 – 9 (N-terminal segment), 19 – 35 (helix α_2), 82- 95 (loop

between helices $\alpha 5$ and $\alpha 6$), and 148 – 168 (C-terminal segment) exhibit HetNOE values that indicate flexibility. However, some of these segments include strongly conserved residues, and several lines of evidence suggest that some of these regions, particularly the C-terminal segment, adopt transient structures due to conformational exchange between folded and unfolded states [45].

We assessed the similarities between these NMR models and CASP14 prediction models, including AF2 models (Figure 2, right). Again, we excluded residue segments for which atomic positions are not well defined in either the ensemble of 19 NMR-derived conformers or the ensemble of 5 AF2 conformers. Well-defined regions of the of the NMR ensemble include residue ranges 10 - 18, 36 - 81, and 96 - 145 (Figure 2G), while for the AF2 structure residues 36 - 75 and 96 - 164 are well-defined (Figure 2H), as defined by *Cyrange*. The pairwise GDT scores between the NMR model with best DP score and the 5 AF2 conformers for the residues whose backbone positions are well defined in both ensembles, i.e. for residues 36 – 75 and 96 - 145, ranged from 0.66 to 0.67, corresponding to backbone rmsd's of about 4 Å (Figure 1I). The primary differences between the NMR and AF2 structures involve the packing of helices into the core of the protein structure that is formed by the two antiparallel helical bundles ($\alpha 3$, $\alpha 4$, $\alpha 7$, and $\alpha 8$). In the NMR structure, N-terminal helix $\alpha 1$ is “grabbed” by this bundle, while in the AF2 model helix $\alpha 1$ is replaced in this core by a new helix formed by the C-terminal segment (which is at least partially disordered in the experimental NMR structure). This “switch” also reorients helix $\alpha 2$. These interactions are mutually exclusive. However it is possible that the two forms (i.e., the NMR structure with helix $\alpha 1$ in the core, and the AF2 structure with the C-terminal region forming a helix and replacing helix $\alpha 1$ in the core) could exist in dynamic

equilibrium, consistent with the observed conformational dynamics in both helices $\alpha 1$ and $\alpha 2$ and in the C-terminal region, described above.

The AF2 model also includes core sidechain conformations that are very similar to those in the solution NMR structure. Particularly notable are the positions and pairing of the 5 disulfides bonds, which are in good agreement with the experimental structure (Figure 2J), particularly for the 4 disulfide paired cysteines located in the well-defined regions of the NMR structure, including the ambiguous C52/C127 and C56/C123 disulfide pairs. There is somewhat less agreement in superimposition for the C136/C148 disulfide pair, that includes residue Cys148 located in a not-well-defined (possibly flexible) region of the NMR model. This prediction of correct disulfide pairing and sidechain conformation is quite remarkable considering that no experimental disulfide pairing information was used in the AF2 modeling.

Structure quality statistics for T1027 were analyzed with the *PSVS* software suite [23]. The resulting structure quality statistics for both the NMR and AF2 model ensembles are summarized in Supplementary Tables S3-S4. Both the NMR and AF2 models generally exhibit excellent structure quality scores. The T1027 NMR structure provides a marginally acceptable wwPDB structure validation report (Supplementary Figure S1); the *ProCheck* (backbone plus sidechain) and *MolProbity* Z scores are at the lower end of the normally acceptable range, which probably simply reflects the fact that no specific energy minimization was used in the structure refinement. As was observed for T1055, the AF2 models of T1027 have better *ProCheck* G-factor (backbone plus sidechain) and *Molprobity* clash scores, attributable to more energetically consistent core sidechain packing.

We next assessed how well the NMR and AF2 structures fit to the experimental NMR chemical shift data (bmr_id 36288) and NOESY peak list data using the RPF-DP score. Plots of DP score vs GDT for all CASP14 predictor groups have a strong correlation, with DP scores ranging from -2.02 to 0.58 (Figures 2K and L). The prediction model with highest DP score, 0.58 for AF2 model 4 (model 1027_427_4) is not as high as the DP scores of NMR conformers, 0.64 - 0.68 (Figure 1K). *In this case the NMR models fit the NOESY data better than any CASP14 model, including the AF2 models.* Although the GDT score between the AF2 models and this NMR structure is lower than for most AF2 predictions, the NMR model is clearly a better fit to the unassigned NOESY data.

A more detailed RPF DP analysis for T1027 is summarized on the right side of Figure 3. Overall, the NMR models ($R = 0.89$) have less recall violations than the AF2 models ($R = 0.85$ - 0.86). The recall analysis also documents that there are many NOESY peaks that are consistent with NMR model but not consistent with the AF2 models and (color coded in Figure 3I). Residues with NOESY peaks that are assigned to consistent interactions in the NMR model but not consistent with the AF2 models are color coded on the AF2 model according to their residue assignment in the NMR model; i.e. residues colored light blue (1 - 3 recall violations), green (4 - 5 recall violations), orange (6 - 10 recall violations), or red (> 13 recall violations) on the AF2 model have structurally consistent assignments in the NMR model. NOESY peaks that are not consistent with the AF2 model, but assigned in the NMR model, are also indicated with their residue assignments in the NMR model as blue bars in the histogram plot Figure 3J. Hence, the NMR models explain many more NOESY peaks than the AF2 model.

However, there are also a significant number of NOESY peaks that are consistent with the AF2 models but not with the NMR models. These residues are colored light blue or green in Figure 3H, and as orange histogram bars in Figure 4J, and include residues 80, 82, and 144-149 in the C-terminal segment. These NOESY peaks, though inconsistent with the NMR model, could be explained by a population of conformers similar to the AF2 structure, with a C-terminal helix interacting with the core in place of the N-terminal helix.

Both the NMR ($P = 0.78 - 0.80$) and AF2 ($P = 0.76 - 0.78$) models of T1027 have a modest numbers of precision violations. Precision violations are short distances in the model that cannot be explained by any NOESY cross peak. Figure 3K highlights precision violations of the NMR model, located primarily in helices $\alpha 5$ and $\alpha 6$ (Figure 3K). These precision violations may result from exchange broadening or resonances in or near these residues, due to conformational dynamics. In the AF2 models, the precision violations occur mostly where the C-terminal segment forms a helix that interacts with the core (Figure 3L); i.e. this packing interaction is not fully supported by the NOESY data. Interestingly, however, some of the short distances resulting from packing the C-terminal region into the core, and displacing helix $\alpha 1$, are consistent with some of the NOESY data (Figure 3J, orange bars), suggesting the potential for a small population of conformers in solution with the helical packing predicted by AF2.

Finally, we assessed how well the NMR and AF2 models of T1027 satisfy backbone dihedral restraints derived from backbone chemical shift data using *Talos_N* [17]. The NMR models satisfy most of these chemical shift data (Figure 3M). Residues identified by *Talos_N* as

“dynamically disordered” (colored in yellow color in Figures 3 M and N) are located mostly in the N- and C- terminal segments or loop regions. Residues whose dihedral angles in the model are inconsistent with backbone chemical shift data are colored red in Figures 3 M and N. These are mostly located in the loop regions, and may reflect conformational dynamics in these loops. However, both the conformations of N-terminal and C-terminal helical segments of the AF2 models are not supported by these chemical shift data (Figure 3N); if present in solution the predicted C-terminal helix is populated only to a low level, and are not reflected in the (population-weight-averaged) chemical shift data.

Target T1029: Se0862. Biofilms are communities of microorganisms that are enclosed in extracellular polymeric matrices. They provide protection from environmental stresses, and can confer antibiotic resistance. The cyanobacterium *S. elongatus* encodes a conserved protein Se0862, CASP14 target T1029, that is required for biofilm development [47]. Isotope-enriched samples of Se0862 were produced by Zhang, LiWang, and coworkers as a SUMO fusion, which was processed by Ulp1 SUMO protease cleavage to provide the native 125-residue protein with no non-native residues. In this work, a chemical-shift based CS-Rosetta model was used to guide the NOESY peak assignments, and NOESY peak assignments were restricted to only cross peaks with low assignment ambiguity. The solution structure was determined from 2,045 distance restraints, 192 dihedral angle restraints derived from backbone chemical shift data using *Talos-N*, and 175 residual dipolar couplings for H^N-N , $H^\alpha-C^\alpha$, and $C^\alpha-C'$ bond vectors [47] using *XPLOR-*NIH** [48]. The resulting structure is a well-converged $\alpha + \beta$ structure with $\alpha\alpha\beta\beta\beta\alpha\alpha$ topology. This NMR structure satisfies the NOE-based distance restraints, has an acceptable RDC Q-score of 0.173, and provided an acceptable wwPDB NMR Structure Validation Report (Supplementary

Figure S1). TALOS chemical-shift-based dynamic order parameters S^2 indicate localized conformational dynamics in surface loops between helices $\alpha 1$ and $\alpha 2$, strands $\beta 1$ and $\beta 2$, and strands $\beta 3$ and $\beta 4$ [47].

Structure quality statistics for T1029 (PDB 6uf2) were analyzed with the PSVS software suite [23], and the resulting structure quality statistics for both the NMR and AF2 model ensembles are summarized in Supplementary Tables S5-S6. The NMR structure exhibits acceptable knowledge-based structure quality scores. Notably, the *ProCheck* backbone, *ProCheck* (backbone and sidechain), and *MolProbity* Z scores are all > -1.0 , typical of good structures [23]. The wwPDB Structure Validation Report (Supplementary Figure S1) also does not flag any serious problems with this NMR structure. Consistent with the observations for the other NMR targets, the AF2 models have even better *ProCheck* backbone, *ProCheck* all dihedral and *MolProbity* Z scores, reflecting the good quality of the energy refinement protocols used by AF2. It should be noted, however, that acceptable values for these metrics are necessary, but not sufficient, for validating the accuracy of a structure, and even models with poor accuracy may have good knowledge-based structure quality scores [29].

We assessed the similarities between NMR and all CASP14 prediction models (Figure 4). Well-defined regions of the NMR ensemble, residue ranges 2-19, and 29 - 122 (Figure 4A) were identified using *Cyrange*. For the AF2 models, residue ranges 2-46, and 53 - 123 are well-defined based on *Cyrange* (Figure 4B), and the pairwise GDT scores between the NMR model with the best DP score and 5 AF2 conformers for residues 3 - 19, 29 - 46, 53 - 122 (i.e., well-defined in the NMR and AF2 ensemble, and revised NMR ensemble described below), range

from 0.46 to 0.47 (Figure 4C), corresponding to a backbone RMSD of ~ 7 Å. Considering only the common secondary structure elements, the GDT is 0.54 – 0.55 and backbone RMSD is ~ 5 Å. The best GDT score for all prediction models is also quite low, GDT = 0.50, for model 1 of prediction group 071 (model 071_1, for residues 3-19, 29-46, 53-122). T1029 is a significant outlier for AF2 and other CASP14 predictions.

We next assessed how well the NMR and CASP14 structures fit to the experimental NMR data, including chemical shift (bmr_id 30676) and NOESY peak list data, using the RPF-DP score [15, 16]. For T1029, the plot of DP score vs GDT for all CASP14 predictor groups has a poor correlation (Figure 4D), and DP scores range from -1.62 to 0.57 (Figures 1K and L). The highest DP score for all prediction models, 0.57 for model 4 of predictor group 323 (model 323_4), is significantly higher than the range of DP scores obtained for the NMR conformers, $0.19 - 0.27$ (Figure 4D). Indeed, more than 50% of the CASP14 prediction models have DP scores > 0.27 , and are a better fit to these NMR data than the NMR structure itself.

Inverse Structure Determination of T1029. The low DP score for the T1029 NMR model (DP_best = 0.27) is attributable primarily to poor precision scores (P_best = 0.57); i.e. there are many short distances in the model that are not explained by the NOESY data. Although a low precision score can result from conformational exchange broadening [15, 16], T1029 does not exhibit extensive internal dynamics that cause exchange broadening [47]. This led us to assess the quality of the ^{13}C - and ^{15}N -edited 3D NOESY peak lists. Because of the strategy of focusing on unambiguously-assigned NOESY peaks used in the original structure determination process, many peaks present in the NOESY spectra were not included in the original NOESY peak lists,

particularly for the 3D ^{13}C -edited NOESY. Accordingly, we carried out a careful repicking of the 3D ^{15}N - and ^{13}C -edited NOESY data. Automatic peak picking resulted in far too many peaks, particularly for the ^{13}C -edited NOESY. In order to guide this pick picking, we then used the recall violations provided by the RPF webserver [15] to further edit these NOESY peak lists. The resulting improved NOESY peak lists provided better DP scores for the original NMR structure PDB_id 6uf2, of 0.49 – 0.51, and for many of the CASP prediction models.

Considering these observations, we (N.Z., A.L., Y.J.H., and G.T.M.) next undertook a refinement of the solution NMR structure of T1029, guided by the AF2 prediction model. This process is outlined on the left side of Figure 5. The resonance assignments, dihedral restraints from *TALOS_N*, and RDC restraints, together with the manually-refined NOESY peak lists, were used as input for NOESY peak assignment with the program ASDP. However, rather than initializing the ASDP NOESY peak assignment process with an extended or random conformation, the program was initiated with the coordinates of the five AF2 prediction models. Backbone dihedral angle restraints for residues 40, 41, 61, 63 and 123 that were strongly violated by the AF2 models were also removed from the dihedral restraint list. In this way, the NOESY peak assignment process was intentionally guided by the AF2 prediction models.

In the course of analyzing NOESY peak assignments, *ASDP* uses a structure generation program to produce structural models; in this case the *Cyana* program was used with the NOESY peak assignments and restraints provided as input to *Cyana* by *ASDP*. The output of *ASDP* also includes assigned NOESY peak lists, distance restraints, and a RFP recall / precision analysis. The recall violation list (NOESY peaks not consistent with resulting models) was then used to

further guide manual refinement of the NOESY peak list, and the process was reiterated. The resulting restraints (distance, dihedral, and RDC) were then used as input to *XPLOR-NIH*, using the same protocols used to generate the original NMR structure PDB_ID 6uf2. Two structure determination protocols were used with *XPLOR-NIH*: (i) refinement of the *ASDP - Cyana* models and (ii) generation of revised NMR models starting from extended conformations. Although both protocols provided acceptable structures, only the results of the second protocol (starting from extended conformations) was selected for deposition and release in the PDB (PDB_ID 7n82).

The revised NMR models were analyzed for restraint satisfaction and knowledge-based structure quality statistics using PSVS program. The knowledge-based Z scores of *ProCheck* (backbone), *ProCheck* (backbone and sidechain), *ProsaII*, and *MolProbity* for the revised T1029 structure (Supplementary Table S7) are all significantly better than for the original NMR structure (Supplementary Table S5), though still a bit lower than those for the AF2 structure (Supplementary Table S6). The revised NMR models are also a better fit to the RDC data (right side of Figure 5 and Table 1); the Q-scores for N-H^N, C^α-C', and C^α-H^α are all significantly lower (better). In this analysis, we also assessed ANSURR scores [31]. These are significantly higher (better) for both the *ASDP-Cyana* NMR models and for the revised NMR structure of T1029 (PDB 7n82) than for the original NMR structure (PDB 6uf2) (Table 1). The revised NMR structures also have DP scores that are much higher (better) than the original NMR structure, ranging from 0.66 to 0.69, with improved recall and precision statistics (R = 0.86 - 0.87, P = 0.75 - 0.77). Accordingly, the AF2 model was successfully used to guide the analysis of NMR data to produce a revised NMR model with excellent energetics, restraint satisfaction,

and a better fit to the NOESY and RDC data than the original NMR structure. Even though the re-analysis of the T1029 NMR data was guided by the AF2 models, the resulting structures are not identical to the AF2 models, and in fact the DP scores of the revised NMR models are a bit higher than the AF2 models; i.e. the revised NMR models are a better fit to the NOESY data than the AF2 models.

The revised NMR models (Figure 4F) were then used to reanalyze the DP vs GDT score plot for all CASP14 predictions (Figure 4E). Using the revised NMR model with highest DP score as a reference, the DP vs GDT plot is much more monotonic and linear, as expected for a good quality NOESY peak list and reference model. The prediction models with highest GDT and DP scores were all AF2 models (GDT = 0.89 to 0.90, DP = 0.66 to 0.67). These AF2 models also have very good C $^{\alpha}$ -C' RDC scores (Table 1). (N-H N and C $^{\alpha}$ -H $^{\alpha}$ RDC scores depend on the details of H atom placement, which are not provided in the AF2 model coordinates). The core sidechains in AF2 models also superimpose remarkably well with sidechain conformations in the revised NMR models (Figure 4H).

NMR Guided Prediction of an Integral Membrane Protein Structure in CASP14. A

preliminary solution NMR structure of 238-residue [^2H , ^{13}C , ^{15}N -enriched, $^{13}\text{CH}_3$ labeled]-MipA in detergent micelles has been determined using *ASDP* with *Cyana*, followed by refinement with *Rosetta*. The structure is a 10-12 stranded beta-barrel. The solution NMR structure analysis of MipA is challenging due to extensive exchange broadening in polypeptide segment 43 - 67, which appears to involve multiple conformations for two strands of the beta-barrel. The current “best” experimental NMR model has a DP score of 0.54; it is not considered a final structure.

Ongoing studies are aimed at properly characterizing these multiple conformational states of MipA, and their relationship to MipA's function.

Since the experimental dynamic solution NMR structure analysis of MipA is still in progress, CASP14 prediction models were assessed only against the NMR NOESY and chemical shift data, using the DP score and *TALOS_N*, rather than against atomic coordinates. In CASP14, eight prediction methods submitted results for “NMR-assisted prediction” of MipA, in which prediction was assisted by the NOESY-based Ambiguous Contact List. A ninth group deposited results in this category, but later informed us that their result did not actually use the Ambiguous Contact Lists derived from NMR data. The relative performance of these eight groups was assessed by DP score for both the top model selected by the submitting group (DP_first) and the best scoring model (DP_best) (Figure 6A). CASP14 prediction groups 018 (UNRES_template), 360 (UNRES), 71 (Kihara Lab) and 96 (UNRES-contact) all submitted similar beta – barrel structures with DP scores > 0.50; the best-scoring model (N1088TS018_2) has a DP score of 0.55. Analysis of this top-scoring model against chemical shift data using *TALOS_N* (Figure 6B) showed generally good agreement over most of the structure, except in the polypeptide segment 43 – 67 for which chemical shift and other NMR data indicate multiple conformations local structure. Some residues in the two short α -helices predicted to form in segment 162-169 also violate the chemical shift data (Figure 6B, residues colored red and yellow). Overall, the top scoring NMR-assisted prediction models are consistent with one another and in good agreement with the NMR data, except for regions 43 – 67 (predicted to form two strands of the β -barrel) and 162-169 (predicted to form two small helices) for which experimental data indicate conformational dynamics.

Next we assessed all “pure” predictions (i.e. predictions that did not use the NMR – derived Ambiguous Contact List data) of MipA, using the DP score. Thirty eight top-scoring prediction groups submitted models with $DP_first \geq 0.54$ (Figure 6C) that fit these NMR data better than or equal to the best NMR-assisted models, and 64 groups with $DP_best \geq 0.54$. The best-scoring models include T1088TS226_5 with $DP = 0.62$ (Zhang-TBM), T1088TS024_5 with $DP = 0.61$ (DeepPotential), T1088TS031_5 with $DP=0.61$ (Zhang-CEthreader), T1088TS328_2 with $DP=0.60$ (FoldXpro), T1088TS013_2 with $DP = 0.60$ (FEIG-S), T1088TS067_3 with $DP=0.60$ (ProQ2), and T1088TS498_3 with $DP = 0.60$ (VoroMQA-select). Interestingly, these best-performing pure prediction groups include (but is not lead by) the DeepMind AF2 group ($DP_first = 0.54$, and $DP_best = 0.55$, highlighted by the red histogram bars in Figures 6C). Hence, as was observed in the NMR-data-assisted component of CASP13, *some advanced pure prediction methods used in CASP14 provided models that fit the NMR data better than traditional or data-assisted prediction methods that utilize the NMR data itself.*

We also tried the inverse structure determination method with MipA, using AF2 models to guide the NOESY assignment process. However, unlike what was observed for target T1029, we did not obtain a complete 12-stranded β -barrel structure with this protocol, as the proton resonances that form the key inter strand NOEs needed to form the two missing β -strands are exchange-broadened and these NOESY peaks are not present in peak list. The success of the inverse structure determination method is mainly driven by assignment of experimental NOESY cross peaks, rather than being defined directly by the input prediction models.

Discussion

In CASP14, the AF2 prediction approach performed remarkably well in predicting 3D structures relative to reference experimental structures determined by X-ray crystallography and cryoEM, generally providing GDT scores > 0.85 [1, 10]. For two of the three targets for which the reference structures were determined by NMR, these scores were generally lower. We initially asked the question whether these lower GDT scores for T1027 and T1029 are due to inaccuracies in the NMR models. For the three NMR structures, we plotted the DP score of best scoring model in the NMR ensemble, a measure of the quality of the NMR structure, against the GDT score for the best-scoring AF2 model (Figure 7). This analysis suggested that the low GDT scores for AF2 models of targets T1055 and T1027 are not attributable to serious problems in the accuracy of these models. However, the low DP_{best} score for target T1029 suggested possible inaccuracies in the NMR structure. We investigated this carefully, and refined the resonance assignments and peak list data. Using the improved NOESY peak list, the DP score for the original T1029 structure PDB_id 6uf2 increased significantly, from a range of 0.19 – 0.27 to 0.49 – 0.51. Using the inverse structure determination protocols, the individual conformers of the T1029_{revised} NMR structure have much higher DP scores (0.66 to 0.69), indicating they are more accurate, than those of the original model. The AF2 models also have much higher GDT scores, 0.89 - 0.90, relative to the revised experimental models T1029_{revised} (Figure 7).

Our analysis revealed alternate bases for the differences between experimental and prediction models for each CASP14 target. T1055 is a well-defined, relatively static structure, for which the NOESY and chemical shift data are fit well by either the NMR models deposited in the PDB, or the AF2 models. Interestingly, the AF2 models fit the NMR data a bit better than

the experimental structure. These differences are attributable to differences in structure refinement protocols; overall the AF2 and NMR models are nearly identical, with backbone GDT score of 0.90 (corresponding to a backbone RMSD of about 1.3 Å). T1027, on the other hand, is a dynamic structure, for which the experimental data, particularly ^{15}N relaxation dispersion data, indicate interconversion between multiple conformational states [45]. The experimental NMR models are a much better fit to the NOESY and chemical shift data than the AF2 structures; however some NOEs are not explained by the experimental structure. Features of the distribution of recall and precision violations, and the dynamic NMR data, suggest the potential for a small dynamic population of the AF2-predicted structure in solution. This interesting hypothesis could be pursued with further data collection and analysis.

For T1029, our analysis revealed that the AF2 model, and other CASP14 prediction models, are a significantly better fit to the NOESY data than the reported NMR structure itself. This observation motivated exploration of a novel method of “inverse structure determination”, in which the predicted AF2 model was used to guide a more complete and accurate analysis of the NMR data. The resulting experimental restraints were then used to generate a revised NMR model ensemble which better fits the NOE, chemical shift, and residual dipolar coupling data. This NMR structure, PDB ID 7n82, has excellent structure validation scores, including *RPF-DP* and *ANSURR* scores that are significantly better than the original NMR structure.

In computing GDT scores or other superimposition-based metrics, it is critical to properly exclude those regions of the structure models that are not consistently determined / predicted. In

this study we used the software *Cyrange* [24] to distinguish well-defined from not-well-defined regions of the structure, as recommended by the wwPDB Task Force on Protein Structure Validation. However, other methods for defining this convention are also useful for this purpose [49].

A novel result of this work is the sensitivity of the DP vs GDT plots to the correct choice of reference structure for the GDT score calculation. By improving the accuracy of atomic coordinates for target T1029, and using this revised structure as a reference for the GDT calculation (along with improved NOESY peak lists), the correlation between DP and GDT across CASP14 prediction models become much more monotonic and linear. This correlation coefficient is an interesting metric for assessing the correctness of a NMR structural model, a concept which merits further investigation.

In this study, we focus structure validation on RPF-DP scores [15, 16], which compare models against unassigned NOESY peak lists, as well as knowledge-based Z scores [23], RDC Q scores [27], and dihedral angle ranges indicated by chemical shift data [17]. Generally speaking, distance restraint validation is also an essential metric for NMR structure assessment. The NMR structures deposited in the PDB for the three original targets, T1027, T1029, and T1055, have no significant (> 0.5 Å) restraint violations relative to the deposited restraint lists. The T1029_revised structure also satisfies the restraints used to generate the structure (Supplementary Table S7). However, distance restraints used in NMR structure determination are derived during an iterative process of NOESY peak assignment, structure generation, and restraint assessment; in some cases NOESY cross peaks may be misassigned, resulting in

incorrect restraints, and some restraints may be modified or culled in the process of structure analysis by automated NOESY peak assignment programs. For this reason, our assessment of CASP14 prediction models did not include an extensive analysis of restraint violations relative to the corresponding deposited distance restraint lists.

Another important observation involves the sensitivity of existing structure validation metrics to model inaccuracies. The original T1029 NMR structure has very good structure quality scores, which by standard criteria are acceptable. However, the DP and ANSURR scores suggest some inaccuracies in this ensemble of structures. These problems were not detected by the PSVS structure quality score analysis, the wwPDB NMR Structure Validation Report, the RDC Q score analysis, the TALOS_N analysis, or even by the NOE-derived restraint violation analysis. These results highlight the weaknesses of these standard NMR structure validation scores for assessing NMR-derived model accuracy, and the need for more general use of structure vs data scores, like the DP and ANSURR scores, for assessing NMR structure quality.

In CASP14, NMR data was also provided for target T1088, a beta-type integral membrane porin protein, used by several CASP13 predictor groups to generate NMR-guided prediction models. Most groups involved in this exercise generated similar beta-barrel models, with good agreement with the experimental data. However, as was also observed in CASP13 (Sala et al, 2019), some regular prediction groups, which did not use the NMR data, generated models for T1088 which better fit the NMR data than the NMR-guided methods. In all cases, the most severe discrepancies between the predicted models and NMR data are in the

segment 43-67 for which NMR data indicate intermediate-exchange conformational dynamics. These results demonstrate the power of the most advanced current modeling methods to predict structures of small proteins with accuracies rivaling solution NMR structures.

The CASP14 blind protein structure prediction results have also provided the opportunity to assess the potential for using predicted protein structures to guide experimental NMR data analysis. This goal appears to have been successfully achieved using current best methods of protein structure prediction, for proteins of up to about 200 residues. The best prediction results (e.g. AF2 models) generally fit to the experimental NMR data as well (or better) than experimental structures generated using conventional approaches. Considering the results with more than 90 protein targets, the accuracy of structures predicted by AlphaFold2 should be generally sufficient to provide reliable guidance to NMR data analysis. AlphaFold2 provided models which, in two of the three cases studied here, match the experimental data as well or better than structures generated by conventional NMR structure determination methods. Several other structure prediction methods applied in CASP14 also achieved nearly this level of model accuracy for relatively static structures.

The availability of source code for AlphaFold2[10], RosTTAaFold[13], and other successful co-variance and machine learning methods strongly motivates future efforts to explore using these methods to guide NMR data analysis. There is potential to use these structures not only to guide structure analysis, as was done here, but to provide a complete analysis of both resonance assignments and 3D structures. Accurate models provided by methods like AlphaFold2 [10] and RosTTAaFold [13] open the potential of complete structure determination

of small, relatively rigid protein structures from a single NOESY spectrum; e.g. from a single simultaneous ^{13}C , ^{15}N -resolved NOESY spectrum. However, care must be exercised in using prediction models to interpret these experimental data, as was observed for T1029 using a CS-Rosetta structure to guide the analysis of the original T1029 structure [47]. For example, when there are significant conformational dynamics of the target protein structure, like target T1027, the prediction methods used in CASP14 cannot yet accurately describe these conformational distributions. For such dynamic structures, current prediction methods have limited value in guiding the data analysis, and might in fact misguide the structure analysis process. Methods for predicting chemical shift assignments from models are also not yet sufficiently accurate. Moreover, while structure prediction for less dynamic structures is more reliable, efforts in “inverse structure determination” are susceptible to any errors in the predicted structure that are not contraindicated by the data, and must be carefully cross validated by multiple model vs. data structure quality assessment scores including the RPF-DP score, RDC Q score, and ANSURR score.

Acknowledgements

We thank Drs. K. Bafna, N. Fowler, J. Moult, T. Ramelot, R. Tejero, and M. Williamson for helpful discussions and comments on the manuscript. This research was supported by grants from the National Institutes of Health (R01-GM120574 and R35-GM141818 to G.T.M.; R01-GM100482 to K.F.). Air Force Office of Scientific Research (FA9550-17-1-0447 to A.L.), and NSF-CREST: Center for Cellular and Biomolecular Machines at the University of California, Merced (NSF-HRD-1547848).

Declaration of Interests

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

List of Supplementary Material

1. Supplementary Tables S1 – S7. PSVS Structure Quality Reports for experimental NMR and predicted AF2 models.
2. Supplementary Table S8. Top scoring regular predictor groups for T1088, with DP_best ≥ 0.54
3. Supplementary Figure S1. wwPDB NMR Structure Quality Summaries for experimental NMR and predicted AF2 models.

References

1. Kryshchuk A, Schwede T, Topf M, Fidelis K, Moult J. Critical Assessment of Methods of Protein Structure Prediction (CASP) – Round XIV,. PROTEINS. 2021; THIS ISSUE.
2. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, California, USA: Curran Associates Inc.; 2017. p. 6000–6010.
3. Huang Z, Wang X, Huang L, Huang C, Wei Y, W. L. CCNet: Criss-Cross Attention for Semantic Segmentation. Proceedings of the IEEE/CVF International Conference on Computer Vision 2019.
4. Morcos F, Pagnani A, Lunt B, Bertolino A, Marks DS, Sander C, Zecchina R, Onuchic JN, Hwa T, Weigt M. Direct-coupling analysis of residue coevolution captures native contacts across many protein families. Proc Natl Acad Sci U S A. 2011; 108:E1293-1301. <https://doi.org/10.1073/pnas.1111471108>
5. Marks DS, Colwell LJ, Sheridan R, Hopf TA, Pagnani A, Zecchina R, Sander C. Protein 3D structure computed from evolutionary sequence variation. PLoS One. 2011; 6:e28766. <https://doi.org/10.1371/journal.pone.0028766>
6. Marks DS, Hopf TA, Sander C. Protein structure prediction from sequence variation. Nat Biotechnol. 2012; 30:1072-1080. <https://doi.org/10.1038/nbt.2419>

7. Ovchinnikov S, Kinch L, Park H, Liao Y, Pei J, Kim DE, Kamisetty H, Grishin NV, Baker D. Large-scale determination of previously unsolved protein structures using evolutionary information. *Elife*. 2015; 4:e09248. <https://doi.org/10.7554/eLife.09248>
8. Ovchinnikov S, Park H, Kim DE, Liu Y, Wang RY, Baker D. Structure prediction using sparse simulated NOE restraints with Rosetta in CASP11. *Proteins*. 2016; 84 Suppl 1:181-188. <https://doi.org/10.1002/prot.25006>
9. Buchan DWA, Jones DT. Improved protein contact predictions with the MetaPSICOV2 server in CASP12. *Proteins*. 2018; 86 Suppl 1:78-83. <https://doi.org/10.1002/prot.25379>
10. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Židek A, Potapenko A, Bridgland A, Meyer C, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021. <https://doi.org/10.1038/s41586-021-03819-2>
11. Zemla A. LGA: A method for finding 3D similarities in protein structures. *Nucleic Acids Res*. 2003; 31:3370-3374. <https://doi.org/10.1093/nar/gkg571>
12. Pereira J, Simpkin AJ, Hartmann MD, Rigden DJ, Keegan RM, Lupas AN. High-accuracy protein structure prediction in CASP14. *Proteins*. 2021. <https://doi.org/10.1002/prot.26171>
13. Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, Wang J, Cong Q, Kinch LN, Schaeffer RD, Millan C, Park H, et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science*. 2021. <https://doi.org/10.1126/science.abj8754>
14. Sala D, Huang YJ, Cole CA, Snyder DA, Liu G, Ishida Y, Swapna GVT, Brock KP, Sander C, Fidelis K, Kryshtafovych A, Inouye M, et al. Protein structure prediction assisted with sparse NMR data in CASP13. *Proteins*. 2019; 87:1315-1332. <https://doi.org/10.1002/prot.25837>
15. Huang YJ, Rosato A, Singh G, Montelione GT. RPF: a quality assessment tool for protein NMR structures. *Nucleic Acids Res*. 2012; 40:W542-546. <https://doi.org/10.1093/nar/gks373>
16. Huang YJ, Powers R, Montelione GT. Protein NMR recall, precision, and F-measure scores (RPF scores): structure quality assessment measures based on information retrieval statistics. *J Am Chem Soc*. 2005; 127:1665-1674. <https://doi.org/10.1021/ja047109h>
17. Shen Y, Bax A. Protein structural information derived from NMR chemical shift with the neural network program TALOS-N. *Methods in molecular biology*. 2015; 1260:17-32. https://doi.org/10.1007/978-1-4939-2239-0_2
18. Lovell SC, Davis IW, Arendall WB, 3rd, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC. Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins*. 2003; 50:437-450. <https://doi.org/10.1002/prot.10286>
19. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK - a Program to Check the Stereochemical Quality of Protein Structures. *J Appl Crystallogr*. 1993; 26:283-291.
20. Sippl MJ. Recognition of errors in three-dimensional structures of proteins. *Proteins*. 1993; 17:355-362.
21. Luthy R, Bowie JU, Eisenberg D. Assessment of protein models with three-dimensional profiles. *Nature*. 1992; 356:83-85. <https://doi.org/10.1038/356083a0>
22. Chen VB, Arendall WB, 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. MolProbity: all-atom structure validation for

- macromolecular crystallography. *Acta crystallographica Section D, Biological crystallography*. 2010; 66:12-21. <https://doi.org/10.1107/S0907444909042073>
23. Bhattacharya A, Tejero R, Montelione GT. Evaluating protein structures determined by structural genomics consortia. *Proteins*. 2007; 66:778-795. <https://doi.org/10.1002/prot.21165>
 24. Kirchner DK, Guntert P. Objective identification of residue ranges for the superposition of protein structures. *BMC Bioinformatics*. 2011; 12:170. <https://doi.org/10.1186/1471-2105-12-170>
 25. Montelione GT, Nilges M, Bax A, Guntert P, Herrmann T, Richardson JS, Schwieters CD, Vranken WF, Vuister GW, Wishart DS, Berman HM, Kleywegt GJ, et al. Recommendations of the wwPDB NMR Validation Task Force. *Structure*. 2013; 21:1563-1570. <https://doi.org/10.1016/j.str.2013.07.021>
 26. Tejero R, Snyder D, Mao B, Aramini JM, Montelione GT. PDBStat: a universal restraint converter and restraint analysis software package for protein NMR. *Journal of biomolecular NMR*. 2013; 56:337-351. <https://doi.org/10.1007/s10858-013-9753-7>
 27. Clore GM, Schwieters CD. Amplitudes of protein backbone dynamics and correlated motions in a small alpha/beta protein: correspondence of dipolar coupling and heteronuclear relaxation measurements. *Biochemistry*. 2004; 43:10678-10691. <https://doi.org/10.1021/bi049357w>
 28. Flory PJ. *Statistical Mechanics of Chain Molecules*. New York: Interscience Publishers; 1969.
 29. Rosato A, Tejero R, Montelione GT. Quality assessment of protein NMR structures. *Curr Opin Struct Biol*. 2013; 23:715-724. <https://doi.org/10.1016/j.sbi.2013.08.005>
 30. Rosato A, Vranken W, Fogh RH, Ragan TJ, Tejero R, Pederson K, Lee HW, Prestegard JH, Yee A, Wu B, Lemak A, Houlston S, et al. The second round of Critical Assessment of Automated Structure Determination of Proteins by NMR: CASD-NMR-2013. *Journal of biomolecular NMR*. 2015; 62:413-424. <https://doi.org/10.1007/s10858-015-9953-4>
 31. Fowler NJ, Sljoka A, Williamson MP. A method for validating the accuracy of NMR protein structures. *Nat Commun*. 2020; 11:6321. <https://doi.org/10.1038/s41467-020-20177-1>
 32. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB, 3rd, Snoeyink J, Richardson JS, Richardson DC. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res*. 2007; 35:W375-383.
 33. DeLano WL. *The PyMOL Molecular Graphics System, Version 1.5.0.4* Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 1504 Schrödinger, LLC. 2002.
 34. Li H, Zhang DF, Lin XM, Peng XX. Outer membrane proteomics of kanamycin-resistant *Escherichia coli* identified MipA as a novel antibiotic resistance-related protein. *FEMS Microbiol Lett*. 2015; 362. <https://doi.org/10.1093/femsle/fnv074>
 35. Suzuki M, Zhang J, Liu M, Woychik NA, Inouye M. Single protein production in living cells facilitated by an mRNA interferase. *Molecular cell*. 2005; 18:253-261.
 36. Schneider WM, Inouye M, Montelione GT, Roth MJ. Independently inducible system of gene expression for condensed single protein production (cSPP) suitable for high efficiency isotope enrichment. *J Struct Funct Genomics*. 2009; 10:219-225. <https://doi.org/10.1007/s10969-009-9067-x>

37. Huang YJ, Tejero R, Powers R, Montelione GT. A topology-constrained distance network algorithm for protein structure determination from NOESY data. *Proteins*. 2006; 62:587-603. <https://doi.org/10.1002/prot.20820>
38. Bersch B, Tarbouriech N, Burmeister WP, Iseni F. Solution Structure of the C-terminal Domain of A20, the Missing Brick for the Characterization of the Interface between Vaccinia Virus DNA Polymerase and its Processivity Factor. *J Mol Biol*. 2021; 433:167009. <https://doi.org/10.1016/j.jmb.2021.167009>
39. Guerry P, Herrmann T. Comprehensive automation for NMR structure determination of proteins. *Methods in molecular biology*. 2012; 831:429-451. https://doi.org/10.1007/978-1-61779-480-3_22
40. Rieping W, Habeck M, Bardiaux B, Bernard A, Malliavin TE, Nilges M. ARIA2: automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics*. 2007; 23:381-382. <https://doi.org/10.1093/bioinformatics/btl589>
41. Brunger AT. Version 1.2 of the Crystallography and NMR system. *Nat Protoc*. 2007; 2:2728-2733. <https://doi.org/10.1038/nprot.2007.406>
42. Shimomura O. *Bioluminescence: Chemical Principles and Methods*, 4th Edition ed. Singapore: World Scientific; 2012.
43. Rathnayaka T, Tawa M, Nakamura T, Sohya S, Kuwajima K, Yohda M, Kuroda Y. Solubilization and folding of a fully active recombinant Gaussia luciferase with native disulfide bonds by using a SEP-Tag. *Biochim Biophys Acta*. 2011; 1814:1775-1778. <https://doi.org/10.1016/j.bbapap.2011.09.001>
44. Islam MM, Khan MA, Kuroda Y. Analysis of amino acid contributions to protein solubility using short peptide tags fused to a simplified BPTI variant. *Biochim Biophys Acta*. 2012; 1824:1144-1150. <https://doi.org/10.1016/j.bbapap.2012.06.005>
45. Wu N, Kobayashi N, Tsuda K, Unzai S, Saotome T, Kuroda Y, Yamazaki T. Solution structure of Gaussia Luciferase with five disulfide bonds and identification of a putative coelenterazine binding cavity by heteronuclear NMR. *Sci Rep*. 2020; 10:20069. <https://doi.org/10.1038/s41598-020-76486-4>
46. Guntert P, Buchner L. Combined automated NOE assignment and structure calculation with CYANA. *Journal of biomolecular NMR*. 2015; 62:453-471. <https://doi.org/10.1007/s10858-015-9924-9>
47. Zhang N, Chang YG, Tseng R, Ovchinnikov S, Schwarz R, LiWang A. Solution NMR structure of Se0862, a highly conserved cyanobacterial protein involved in biofilm formation. *Protein Sci*. 2020; 29:2274-2280. <https://doi.org/10.1002/pro.3952>
48. Schwieters CD, Kuszewski JJ, Clore GM. Using Xplor-NIH for NMR molecular structure determination. *Prog Nucl Mag Res Sp*. 2006; 48:47-62. <https://doi.org/10.1016/J.Pnmrs.2005.10.001>
49. Snyder DA, Grullon J, Huang YJ, Tejero R, Montelione GT. The expanded FindCore method for identification of a core atom set for assessment of protein structure prediction. *Proteins*. 2014; 82 Suppl 2:219-230. <https://doi.org/10.1002/prot.24490>

Table 1. RDC Data Fits and ANSURR Scores for Target T1029

Model Ensemble	<Q Scores>			<ANSUUR Scores>	
	N-H ^N	C ^α -C'	C ^α -H ^α	rmsd	corr.
Original NMR (6uf2)	0.154 ±	0.021 ±	0.159 ±	48.2 ± 7.7	30.43 ± 5.7
(10 conformers)	0.0084	0.00047	0.018		
AF2	n.d.	0.013 ±	n.d.	87.6 ± 2.0	55.2 ± 7.5
(5 conformers)		0.00055			
ASDP-Cyana NMR	0.115 ±	0.016 ±	0.135 ±	75.5 ± 7.17	52.9 ± 12.0
(18 conformers)	0.0078	0.00059	0.013		
Xplor Revised NMR (7n82)	0.106 ±	0.014 ±	0.128 ±		53.8 ± 9.2
(20 conformers)	0.0028	0.00057	0.015	67.9 ± 6.05	

Figure Legends

Fig. 1. DP Scores in CASP14. Schematic description of RPF-DP scores. In this analysis, the graph G with nodes corresponding to all ^1H 's and edges representing all short ($< 5 \text{ \AA}$) ^1H - ^1H distances in a structure model (left), is compared with a graph G_{ANOE} (right), in which nodes again correspond to all ^1H 's and edges describe all possible assignments for each NOESY cross peak. True positives (TPs) are edges common to both G and G_{ANOE} , false positives (FPs) are edges present in G but not in G_{ANOE} , and false negatives (FNs) are the set of edges in G_{ANOE} representing the multiple possible assignments of a NOESY cross peak, none of which are present in G . These metrics are used to compute recall (R), precision (P), and F-measure as shown in the figure and outlined in the Methods Section. The F-measure is the harmonic mean of the recall and precision. The Discriminating Power (DP) is a normalized F-measure corrected to account for the F-measure expected for a random-coil chain ($\text{DP} = 0$) and the best F-measure possible considering the completeness of the NMR data ($\text{DP} = 1.0$) [16]. Accurate structures generally have DP for individual models > 0.60 .

Fig. 2. Structural Analysis for CASP14 Targets 1055 and 1027. (left) Superimposed ensembles for (A) NMR structure (pdb id 6zyc) (green) and (B) AF2 structures (blue) of T1055, illustrating the not-well-defined segments (brown) as defined by *Cyrange* [24]. For the NMR structure, residues 3 - 124 are well-defined, while for the AF2 structure residues 8 – 126 are well-defined. (C, D) Comparison of AF2 conformer with highest GDT score (blue) with the representative conformer from the NMR structure ensemble with best DP score, for residues 8 – 124 of T1055. The well-defined backbone (N, C^α , C') atoms are superimposed and both the

backbone superimposition and associated core sidechains are illustrated. DP vs. GDT scores (E) and DP scores vs. predictor group (F) for target T1055. (right) Superimposed ensembles for (G) NMR structure (pdb id 7d2o) (green) and (H) AF2 structure (blue) of T1027, illustrating the not-well-defined segments (brown). For the NMR structure, residues ranges 10 - 18, 36 - 81, and 96 - 145 are well-defined [45], while for the AF2 structure residues 36 - 75 and 96 - 164 are well-defined, as defined by *Cyrange*. (I, J) Comparison of AF2 conformer with highest GDT score (blue) with the conformer from the NMR structure ensemble with highest DP score, for T1027. The well-defined backbone (N, C α , C') atoms are superimposed for residue ranges 36 – 75 and 96 – 145. In the NMR structure, the N-terminal helix sits in a pocket in the core of the protein, while the C-terminal region is disordered; while in the AF2 structure, the N-terminal region is disordered, and the C-terminal region forms a C-terminal helix that packs into the core of the protein structure. The five disulfide bonds are illustrated in panel J. DP vs. GDT scores (K) and DP scores vs. predictor group (L). The red horizontal lines in (E) and (K) are drawn at the DP scores of the best scoring conformation from the ensemble of experimental structures. For both targets, only residues that are well-defined in both the NMR or AF2 structures were included in superimposition and GDT score calculations.

Fig. 3. RPF and Talos_N Analysis for CASP14 Targets 1055 and 1027. (left) Ensemble Recall analysis for the NMR structure (A) and AF2 model (B) of T1055. Residues with a few NOEs that are assigned and satisfied in the NMR model, but with recall violations for the AF2 models, are colored in light blue in the AF2 model. (C) Plot of number of NOEs that are satisfied in NMR structures but not in AF2 models (blue), or satisfied in AF2 models but not in NMR structures (orange), are plot along the sequence; most NOEs can be explained by both

structures. (D, E) Precision analysis for the NMR structure ensemble and AF2 model ensemble of T1055. Residues with modest numbers of Precision violations are colored light blue or green, and those with significant numbers of precision violations are colored yellow, orange and red. (F, G) Talos_N analysis for the ensembles of NMR structures and AF2 models of T1055. No significant violations of dihedral angle restraints derived from backbone chemical shift data are observed in any of the NMR structures or AF2 models. (right) Ensemble recall analysis for the NMR structure (H) and AF2 model (I) of T1027. Residues with a NOEs that are assigned and satisfied in the NMR model, but with recall violations for the AF2 models, are colored as outlined in the text on the AF2 model, and visa versa. (J) Plot of number of NOEs that are satisfied in NMR structures but not in AF2 models (blue), or satisfied in AF2 models but not in NMR structures (orange), are plot along the sequence; many NOEs can be explained only by the NMR models. (K, L). Precision analysis for the NMR structure ensemble and AF2 model ensemble of T1027. Residues with modest numbers of Precision violations are colored light blue or green, and those with significant numbers of precision violations are colored yellow, orange and red. (M, N) Talos_N analysis for the ensembles of NMR structures and AF2 models of T1027. Residues colored yellow are indicated by chemical shift data to be flexible; residues colored red have backbone conformations in well-defined regions of the models that are inconsistent with the chemical shift data. In all images, the dark blue color indicates little or no metric violation. In mapping precision violations on the models (e.g. panels K and L) the regions of the structure that are not converged are not shown because precision violations in these regions can arise simply from the conformational variability.

Fig. 4. Structural Analysis for CASP14 Target 1029. Superimposed ensembles for (A) NMR structure (pdb_id 6uf2) (green) and (B) AF2 structure (blue) of T1029. For the NMR structure, residues 2-19 and 29-122 are well-defined, while for the AF2 structure residues 2-46 and 53-123 are well-defined. (C) Comparison of AF2 conformer with highest GDT score (blue) with the representative conformer from the original NMR structure ensembles with best DP score, for residues 3-19, 29-46 and 53-122. (D, E) DP vs. GDT scores and DP scores vs. predictor group for original NMR structure (D) and revised NMR structure (E). The red horizontal lines in (D) and (E) are drawn at the DP scores of the best scoring conformation from the ensemble of experimental structures. (F) Revised NMR structure (pdb id 7n82) (green), illustrating the not-well-defined segments (brown) as defined by *Cyrange*. For the revised NMR structure, residues 3-20 and 26-123 are well-defined. (G) Comparison of AF2 conformer with highest GDT score (blue) with the representative conformer from the original NMR structure ensembles with best DP score, for residues 3-19, 29-46 and 53-122. The well-defined backbone (N, C α , C') atoms are superimposed and both the backbone superimposition and associated core sidechains are illustrated (H). Only residues that are well-defined in both the original NMR, revised NMR and AF2 structures were included in superimposition and GDT score calculations.

Fig. 5. Inverse Structure Determination. (left) Flow chart of inverse structure determination of T1029 using AF2 model as input. The AF2 models, resonance assignments, *Talos-N* dihedral restraints, and RDC restraints were combined with the manually-refined NOESY peak lists and used as input for NOESY peak assignment with the program *ASDP*. The Recall violation list (NOESY peaks not consistent with resulting models) was then used to further guide manual refinement of the NOESY peak list, and the process was reiterated. Blue and red arrows indicate

program input and output, respectively. (right) Plots of calculated *vs.* observed RDCs for H^N -N, H^α - C^α , and C^α - C' bond vectors for original and revised NMR structures, and RDCs for C^α - C' bond vectors for AF2 models.

Fig. 6. NMR-Assisted Prediction of Integral Membrane Protein. Ranking of (A) NMR assisted and (B) unassisted CASP models based on the DP score of the predictor-defined first model (DP_first) or the best scoring model submitted (DP_best). Scores for the AF2 predictor group highlighted in red among the unassisted prediction groups. (C) NMR-assisted model and (D) regular prediction (unassisted) models with the best DP scores. The models are colored with information from *TALOS_N*: blue, residues for which backbone conformation is consistent with chemical shift data; red, residues for which backbone conformation is not consistent with chemical shift data; orange, residues with no consensus dihedral angles predicted by *Talos_N*; yellow, residues that chemical shift data indicate to be dynamic. Residues 59, 66, 164, and 169 (red) are labeled as reference points. Residues in segments 52-67 and 162-169, which have backbone conformations that are identified by *Talos_N* as dynamic (yellow), inconsistent (red), or no consensus (orange), but also located in predicted regular secondary structures are considered to be inconsistent with the backbone chemical shift data, and may involve multiple conformations.

Fig. 7. DP and GDT Scores for NMR Structures in CASP14. Plot of DP score for best-scoring experimental model *vs.* GDT of best scoring CASP model relative to coordinates of PDB_IDs 7d20, 6uf2, 7n82 and 6zyc, for targets T1027, T1029, T1029_revised and T1055,

respectively. The horizontal dashed line is an empirical cutoff for an accurate NMR structure model [15].

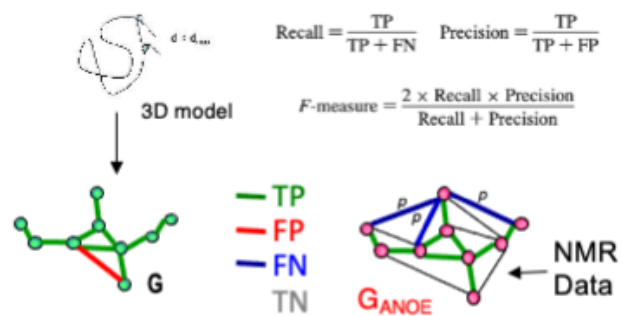


Fig. 1

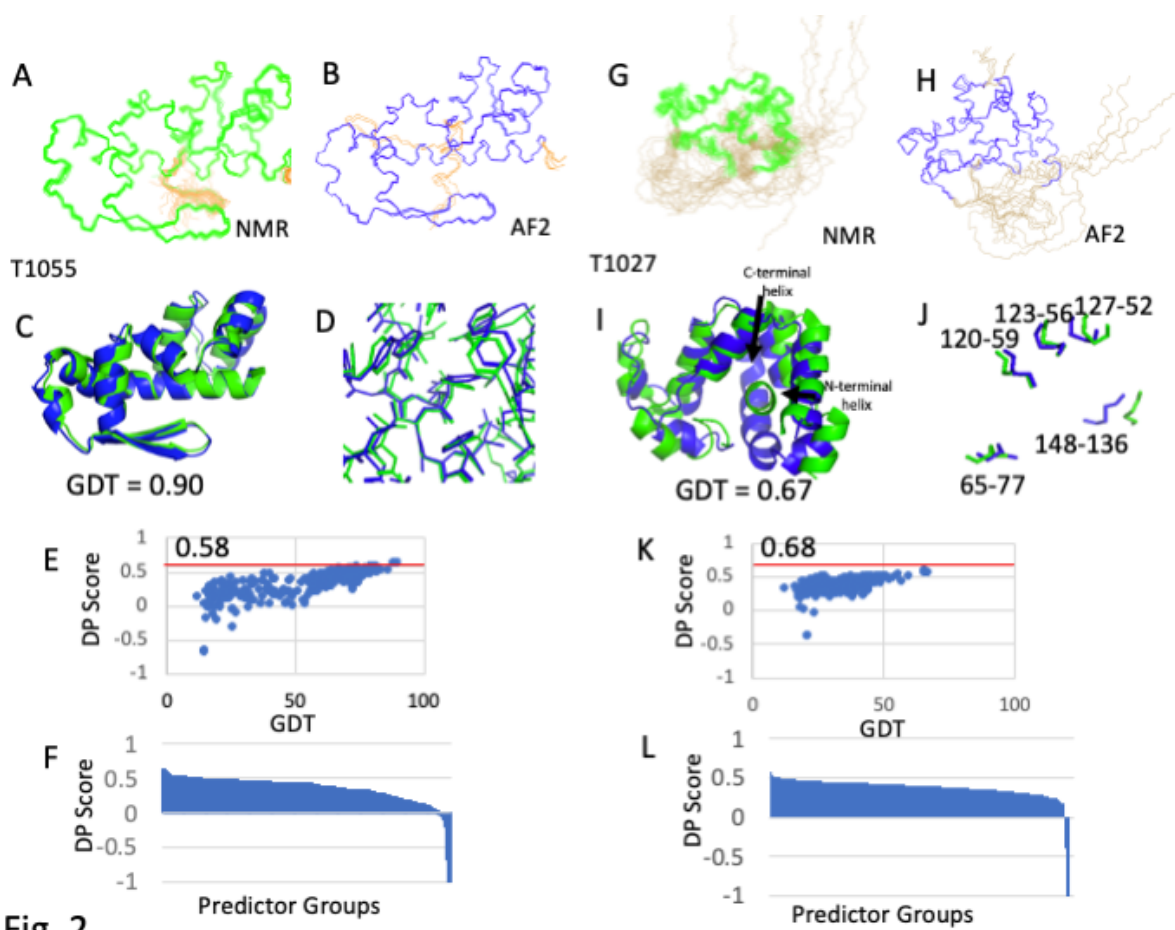


Fig. 2

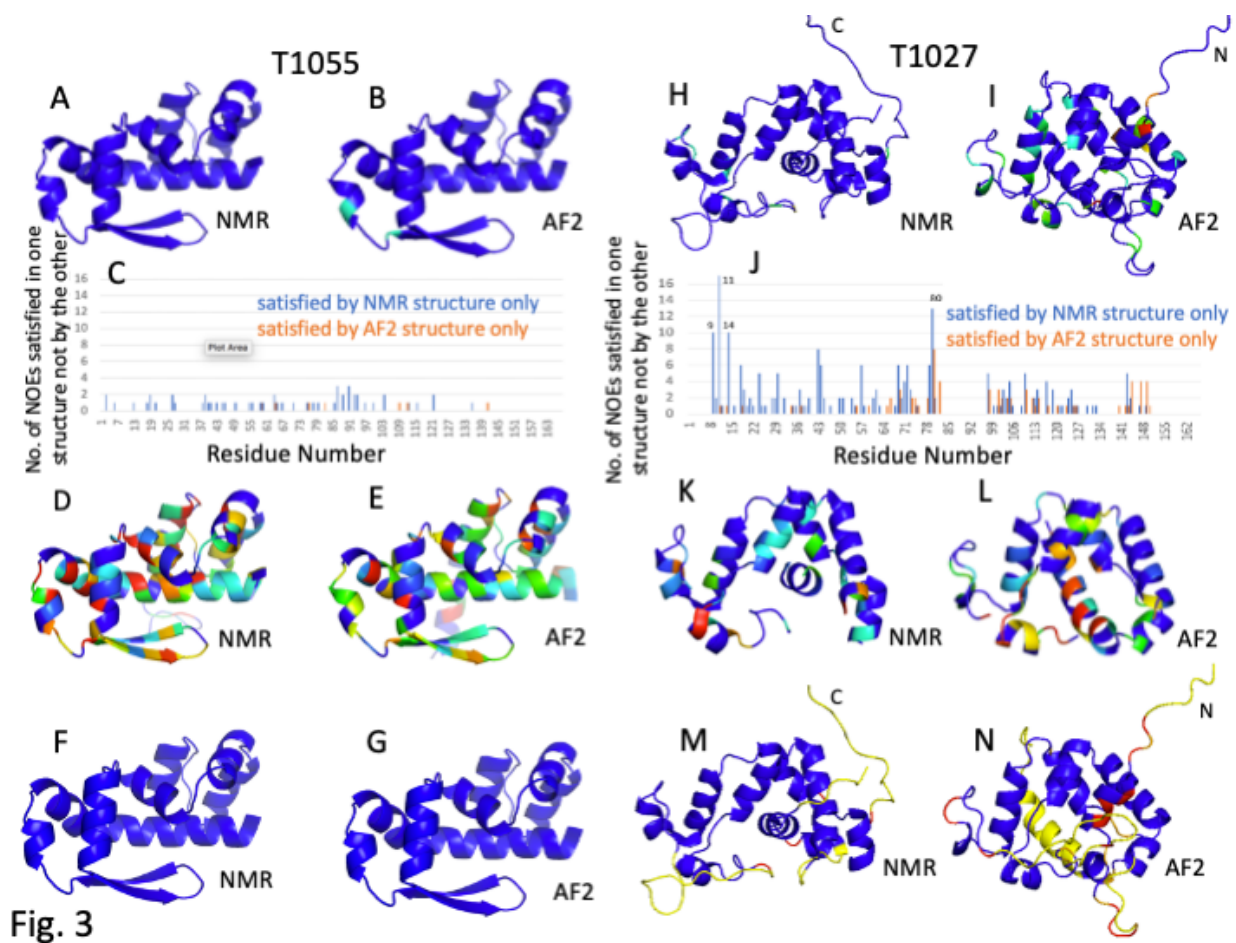
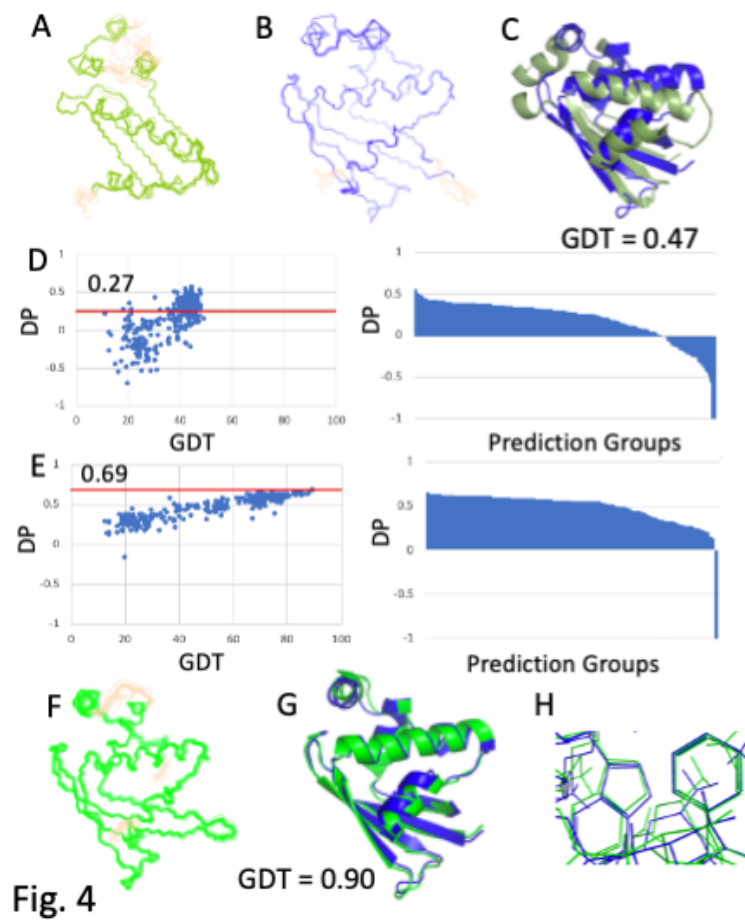


Fig. 3



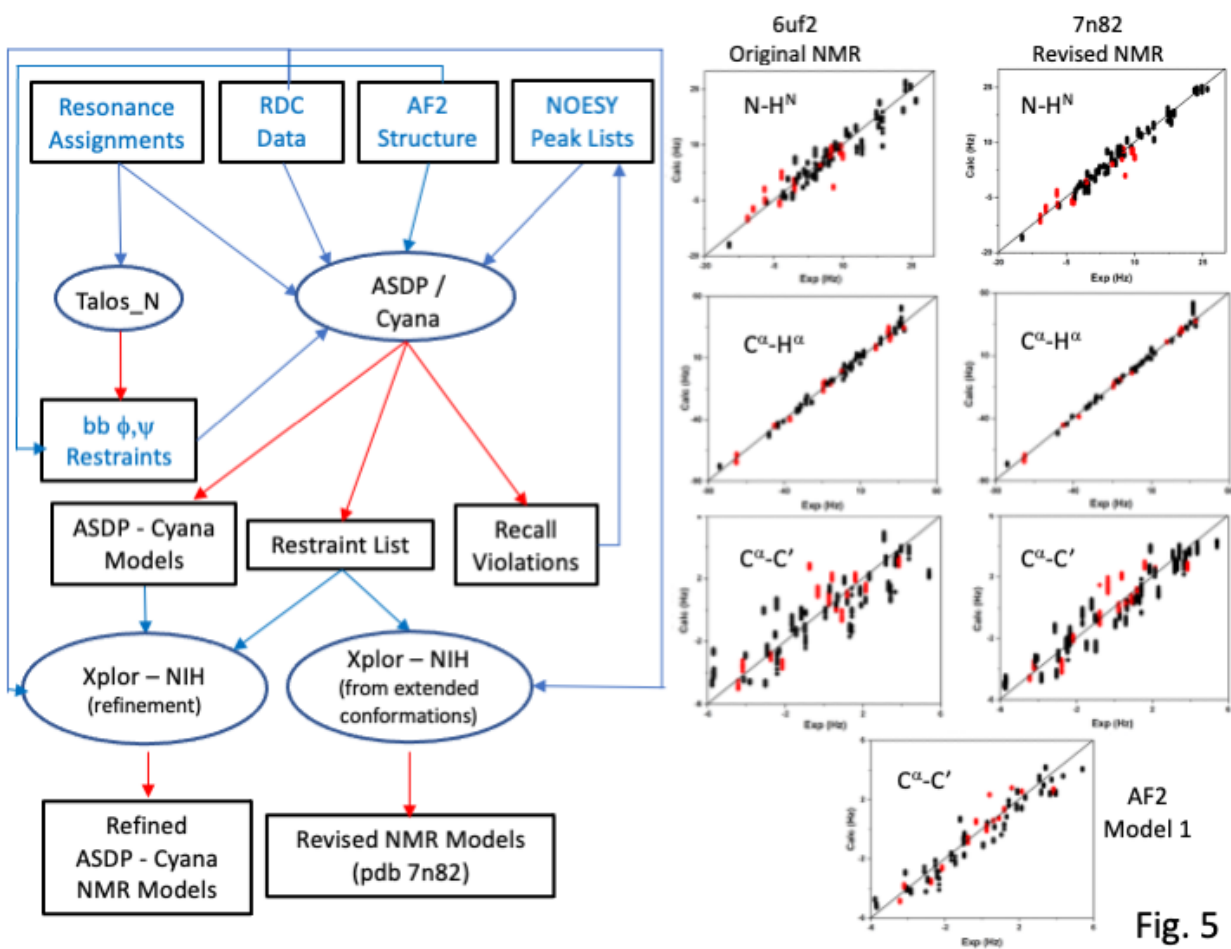
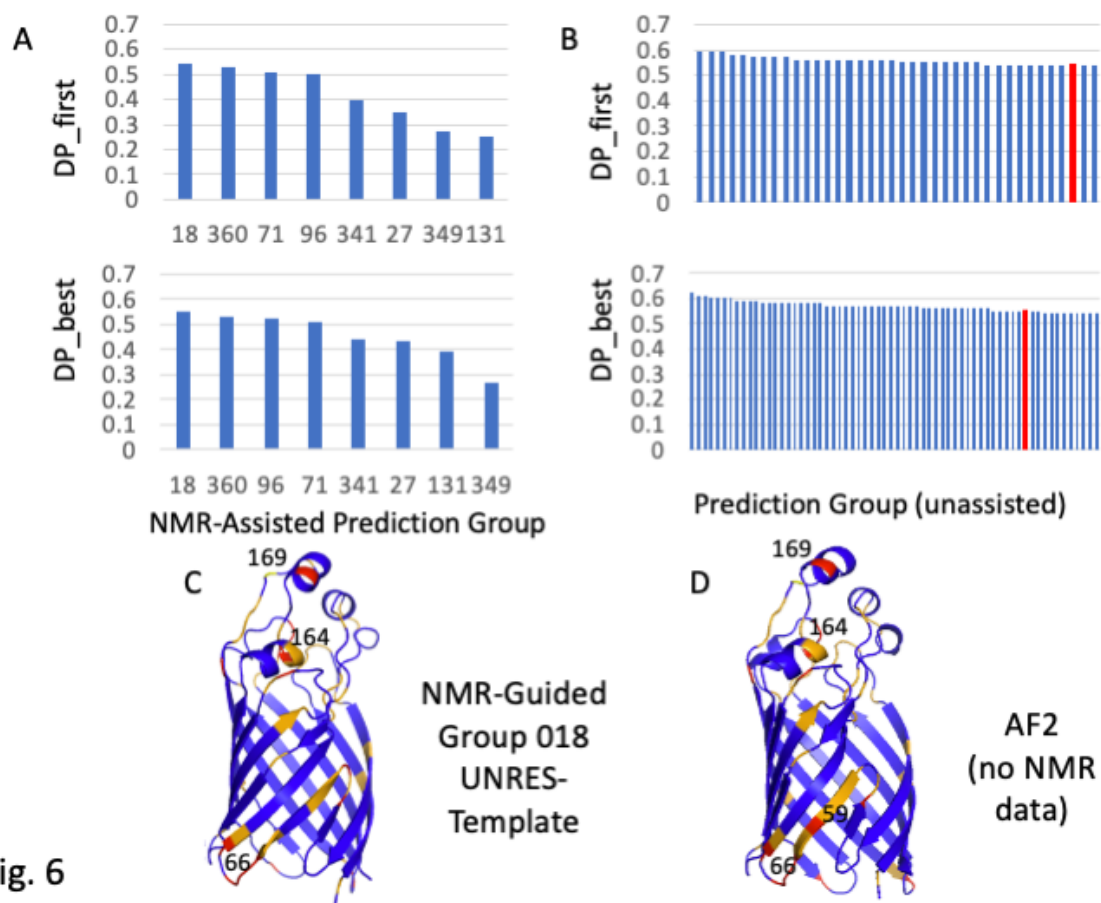


Fig. 5



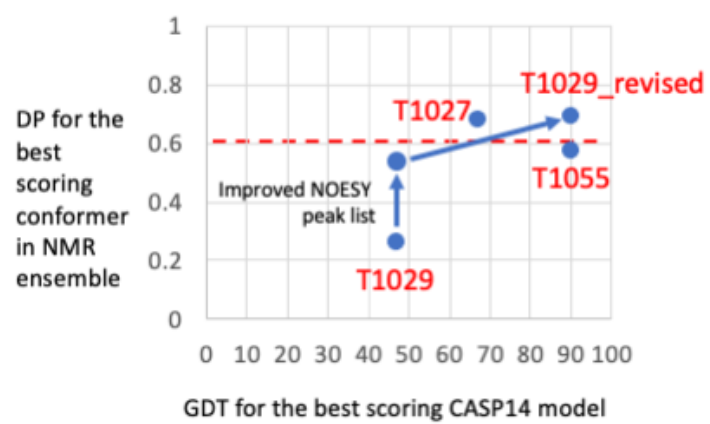


Fig. 7