JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
pubs.acs.org/JACS
Communication

Concordance of X-ray and AlphaFold2 Models of SARS-CoV-2 Main Protease with Residual Dipolar Couplings Measured in Solution

Angus J. Robertson,[†] Joseph M. Courtney,[†] Yang Shen, Jinfa Ying, and Ad Bax*

Cite This: *J. Am. Chem. Soc.* 2021, 143, 19306−19310 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The 68-kDa homodimeric 3C-like protease of SARS-CoV-2, M^{pro} (3CLpro/Nsp5), is a promising antiviral drug target. We evaluate the concordance of models generated by the newly introduced AlphaFold2 structure prediction program with residual dipolar couplings (RDCs) measured in solution for $^{15}N-^{1}H^{N}$ and $^{13}C'-^{1}H^{N}$ atom pairs. The latter were measured using a new, highly precise TROSY-AntiTROSY Encoded RDC (TATER) experiment. Three sets of AlphaFold2 models were evaluated: (1) M^{pro}_{AF}, generated using the standard AlphaFold2 input structural database; (2) M^{pro}_{AFD}, where the AlphaFold2 implementation was modified to exclude all candidate template X-ray structures deposited after Jan 1, 2020; and (3) M^{pro}_{AFS}, which excluded all structures homologous to coronaviral M^{pro}. Close agreement between all three sets of AlphaFold models and experimental RDC data is found for most of the protein. For residues in well-defined secondary structure, the agreement decreases somewhat upon Amber relaxation. For these regions, M^{pro}_{AF} agreement exceeds that of most high-resolution X-ray structures. Residues from domain 2 that comprise elements of both the active site and the homo-dimerization interface fit less well across all structures. These results indicate novel opportunities for combining experimentation with molecular dynamics simulations, where solution RDCs provide highly precise input for QM/MM simulations of substrate binding/reaction trajectories.

We demonstrate new NMR methods for precise measurement of $^{1}H^{N}-^{15}N$ and $^{13}C'-^{1}H^{N}$ residual dipolar couplings (RDCs) in the 68-kDa SARS-CoV-2 M^{pro} enzyme and compare measured RDCs with values derived from both high-resolution X-ray structures and AlphaFold2 models.

The active site of M^{pro} is highly conserved across the entire coronavirus family and quite dissimilar from any mammalian proteases, making it an attractive drug target.[1−3] Over 300 X-ray structures of SARS-CoV-2 M^{pro}, including apo and ligated structures, have been deposited in the Protein Data Bank (PDB), with all crystal structures closely reflecting the architecture and backbone coordinates of the representative substrate-free M^{pro} homodimer (PDB 5R8T).[4]

De novo structure prediction of proteins based solely on their amino acid sequence has recently undergone revolutionary advances, with some algorithms now able to accurately predict the majority of the protein structure, including side chains, at atomic resolution.[5−7] In particular, AlphaFold version 2 (AlphaFold2) integrated orientation-equivariant structural information with multi-sequence alignments (MSAs) using a bespoke transformer neural network architecture (termed the "evoformer") that uses a self-attention mechanism to iteratively improve both the structure and the processing of the MSAs.[7] This has resulted in a remarkable (<1 Å backbone C^{α} RMSD) accuracy of the predicted coordinates relative to unseen structures.[7]

X-ray structures and AlphaFold2 models are based, directly or indirectly, on data collected in the crystalline state. NMR RDCs are exquisitely sensitive reporters on bond vector orientations.[8,9] Very recently, agreement between AlphaFold2 models and RDCs was shown to rival that of high quality X-ray structures.[10] Although most certainly fully adequate for many purposes, it remains an open question to what extent such models are truly representative for pliable proteins in solution where conformational dynamics are often associated with function. One such protein is the main protease of SARS-CoV-2, M^{pro}. Here, we evaluate how well X-ray and AlphaFold2 models agree with RDCs measured in solution.

In the apo state, M^{pro} is subject to auto-proteolysis at the elevated concentrations needed for solution NMR. We therefore focus on the C145A active-site mutant (M^{pro}_{C145A}). Uniformly $^{2}H/^{13}C/^{15}N$-enriched samples were prepared,[11] followed by reprotonation of the exchangeable backbone amides. Conformational exchange broadening on intermediate time scales resulted in variable intensities in the TROSY-HSQC NMR spectrum.[12] At 35 °C, ca. 260 out of 293 non-Pro residues yielded a signal-to-noise ratio greater than 20:1 in the 3D HNCO-TROSY[13] NMR spectrum, which represents a lower limit for measuring RDCs at high precision (Supporting Information (SI) Tables S1 and S2).[14]

$^{1}D_{NH}$ RDCs were measured using a 3D HNCO version of the ARTSY experiment[14] (Figure S1) that yielded a precision ≤0.7 Hz for the vast majority of backbone amides (Table S2). A comparably complete set of $^{2}D_{C'H}$ couplings were recorded using a novel combination of two 3D TROSY-HNCO spectra termed the **T**ROSY- and **A**nti**T**ROSY-**E**ncoded **R**DC

Received: October 6, 2021
Published: November 10, 2021

https://doi.org/10.1021/jacs.1c10588
J. Am. Chem. Soc. 2021, 143, 19306−19310

(TATER) method (Figure S2). Even though the orientations of one-bond $^{15}N-^{1}H^{N}$ and two-bond $^{13}C'-^{1}H^{N}$ vectors only differ by ~35°, their experimentally observed RDCs, as well as values predicted by X-ray structures, correlate only weakly (Figure S3), indicating that they provide largely independent structural information. Due to the larger internuclear separation, $^{2}D_{C'H}$ values are intrinsically ca. 3.2 times smaller than $^{1}D_{NH}$ RDCs,[15] but because the measurement precision ($\lesssim$0.5 Hz) is far smaller than the range spanned by these couplings (Table S2), they are comparably valuable in providing structural information (Figure S4). Cross sections through the 3D $^{1}D_{NH}$ ARTSY-HNCO and $^{2}D_{C'H}$ TATER spectra highlight the quality of the data (Figure 1).

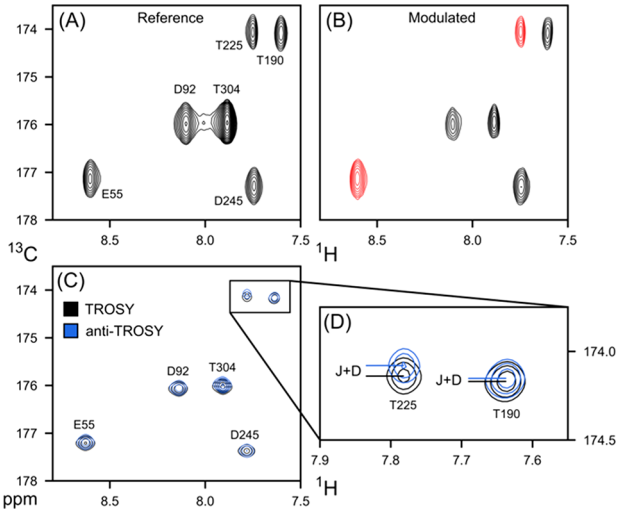

Figure 1. NMR spectra of M^{pro}_{C145A} in 11 mg/mL Pf1, from which RDCs were derived. (A, B) Small regions of cross sections through the ($^{1}J_{NH}$ + $^{1}D_{NH}$)-modulated 900-MHz 3D ARTSY-HNCO spectrum, recorded with dephasing delays of (A) 5.38 ms and (B) 10.75 ms. Positive (black) and negative (red) intensities correspond to $|^{1}J_{NH} + {}^{1}D_{NH}| < 93$ Hz and > 93 Hz, respectively.[14] (C, D) Overlay of small regions of cross sections taken through the 600-MHz TROSY- (black) and AntiTROSY-HNCO (blue) TATER spectra. The difference in $^{13}C'$ resonance frequency equals $^{2}J_{C'H}$ + $^{2}D_{C'H}$ (Table S2).

The sensitivity of an RDC to its internuclear vector orientation, represented by polar coordinates θ and φ in a molecular frame that coincides with the alignment tensor principal axis system, is given by

$$\frac{d(D^{NH}(\theta, \phi))}{d\theta} = D_a^{NH}[-3\sin(2\theta) + {}^{3}/_{2}Rh\sin(2\theta)\cos(2\phi)] \quad (1a)$$

$$\frac{d(D^{NH}(\theta, \phi))}{d\phi} = -D_a^{NH}[3Rh\sin^2\theta\sin(2\phi)] \quad (1b)$$

where D_a^{NH} is the magnitude of the alignment tensor and Rh is its rhombicity; the same equations apply for $^{2}D_{C'H}$, but with $D_a^{C'H} \approx D_a^{NH}/3.2$.[15] As can be seen from eq 1, the derivative of an RDC with respect to orientation depends strongly on orientation and becomes zero when the vector coincides with any of the alignment tensor's principal axes, generating "blind spots". To reduce this differential sensitivity to orientation and remove these blind spots, for each amide we instead use the averaged normalized deviations between observed and predicted RDCs, $\Delta D = \sqrt{\{[(\Delta^{1}D_{NH})^2 + 3.2^2(\Delta^{2}D_{C'H})^2]/2\}}$.

We fit the alignment tensor to the observed $^{1}D_{NH}$ and $^{2}D_{C'H}$ RDCs and a reference structure (protonated by DYNAMO; Figure S5) using singular value decomposition (SVD).[16] A perfect, noise-free fit has a quality factor, Q, of 0.0, while a high-quality fit in the presence of experimental noise has a Q of ~0.2.[17]

M^{pro} contains three domains (Figure 2C) numbered sequentially from N- to C-terminus. Separately fitting the three domains of 1.3-Å "representative" X-ray structure 5R8T[4] yields different Q factors (0.25, 0.51, and 0.39, for domains 1, 2, and 3, respectively), but with nearly identical alignment tensors, pointing to the absence of significant domain reorientation in solution versus crystal.

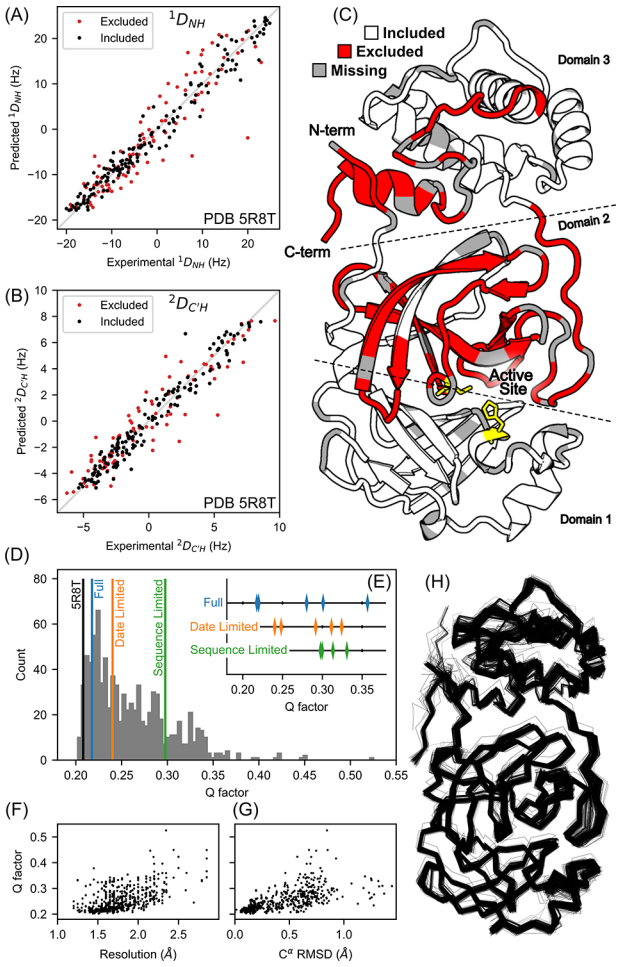

Figure 2. Agreement between measured M^{pro} RDCs and values predicted by structural models. (A) $^{1}D_{NH}$ and (B) $^{2}D_{C'H}$ experimental couplings vs those predicted from X-ray structure 5R8T. (C) Excluded residues (red) illustrated on a ribbon diagram (PDB: 5R8T; only a single chain is shown for clarity); residues with missing RDCs are shown in gray, and the catalytic dyad is shown in yellow. (D) Q-factors from SVD fits of $^{1}D_{NH}$ and $^{2}D_{C'H}$ RDCs to the *included* region of all available M^{pro} X-ray structures plotted as a histogram, with the top-ranked (Amber-relaxed) AlphaFold2 models obtained using *full*, *date-limited*, and *sequence-limited* implementations marked. (E) Q-factors of all Amber-relaxed models. (F) X-ray structure resolution versus Q-factor and (G) C^{α} RMSD (relative to 5R8T) versus Q-factor. (H) C^{α} wireframe of all 352 PDB structures.

https://doi.org/10.1021/jacs.1c10588
J. Am. Chem. Soc. 2021, 143, 19306−19310

We avoided deleterious effects of strong outliers on the alignment tensor accuracy by iteratively restricting the fit (SI text) to two contiguous regions, K5-L115 and T199-L272, that yielded good agreement with M^{pro} X-ray crystal structures and AlphaFold2 models.

For these two regions, the fit to 5R8T yields $Q = 0.208$ (Figure 2A,B), indicating close agreement between RDCs and the X-ray structure. The full set of X-ray structures shows Q-factors ranging from below 0.2 to ~0.53 (Figure 2E), showing a positive correlation with crystallographic resolution (Figure 2F). The excluded regions fit considerably worse to the RDC data, reflecting their reduced secondary structure content (Figure 2, Table 1).

Table 1. Assessment of X-ray and AlphaFold2 Structure Quality Using 254 $^1D_{NH}$ and 257 $^2D_{C'H}$ RDCs

model	C^α RMSD[a]	pLDDT	Q_{incl}[b]	Q_{excl}[b]
PDB: 5R8T	–	–	0.208	0.356
M^{pro}_{AF1} unrelaxed[c]	0.408	95.5	0.213	0.360
M^{pro}_{AF1} relaxed[c]	0.406		0.220	0.341
M^{pro}_{AFD1} unrelaxed[c]	0.644	95.0	0.226	0.348
M^{pro}_{AFD1} relaxed[c]	0.644		0.248	0.337
M^{pro}_{AFS1} unrelaxed[c]	1.015	91.7	0.282	0.616
M^{pro}_{AFS1} relaxed[c]	1.023		0.300	0.577

[a]C^α RMSD values determined by alignment of residues 5−115 and 199−272 of each model relative to PDB 5R8T. [b]Q_{incl} refers to residues 5−115 and 199−272, used to determine the alignment tensor. Q_{excl} pertains to residues 2−5, 117−198, and 273−301, using the same alignment tensor. [c]"Relaxed" and "unrelaxed" refer to AlphaFold2 models with and without the final Amber relaxation step.

Application of the released[7] AlphaFold2 implementation to predict the SARS-CoV-2 M^{pro} structure generated five quality-ranked structures (M^{pro}_{AF1-5}), each with per-residue predicted local-distance difference test (pLDDT) confidence scores (Figure 3G, Table 1, and Table S3). The current implementation of AlphaFold2 only predicts monomeric structures, whereas in solution M^{pro} forms a tight C_2-symmetric homodimer, which could adversely impact the accuracy of the prediction. Despite this limitation, the model with the highest pLDDT score (M^{pro}_{AF1}) has a low backbone C^α RMSD of 0.47 Å relative to PDB 5R8T (Table 1). AlphaFold2 employs a final Amber-based[18] constrained-relaxation step to remedy backbone and side-chain clashes (Figure 3D). For M^{pro}_{AF1}, the Amber relaxed (Amber unrelaxed) model agrees with solution RDC data better than 82% (92%) of all X-ray structures. In fact, when focusing on RDCs of residues involved in secondary structure, M^{pro}_{AF1} (Amber unrelaxed) shows a modest improvement over the X-ray structure (Table S3).

M^{pro} and its close homologues are well represented in the template database (pdb70) used by AlphaFold2. Consequently, a large number of M^{pro} homologues are chosen as templates by both the *full* (M^{pro}_{AF1-5}; 16 out of 20) and *date-limited* (M^{pro}_{AFD1-5}; 15 out of 20) implementations of AlphaFold2 (Table S4). To recreate a CASP-like scenario,[19] we modified the AlphaFold2 code to reject structures homologous to SARS-CoV-2 M^{pro} (M^{pro}_{AFS1-5}; 0 out of 14; termed *sequence-limited*; see SI for details). The resulting M^{pro}_{AFS} models remain remarkably similar to the output of the "unblinded" implementation, with close agreement of backbone C^α RMSDs ≤ 1.5 Å relative to PDB: 5R8T (Table 1 and Tables S3−S5). The Amber-relaxed (Amber-unrelaxed) M^{pro}_{AFD1} and M^{pro}_{AFS1} models with the highest pLDDT scores agreed better with RDCs than 51% (69%) and 18% (28%) of all available X-ray structures, respectively (Figure 2D).

While the input PDB sets for the *full*, and *date-limited*, AlphaFold2 implementations differed by only a single template—PDB 6YB7—removal of this SARS-CoV-2 M^{pro} structure noticeably reduces the agreement between predicted

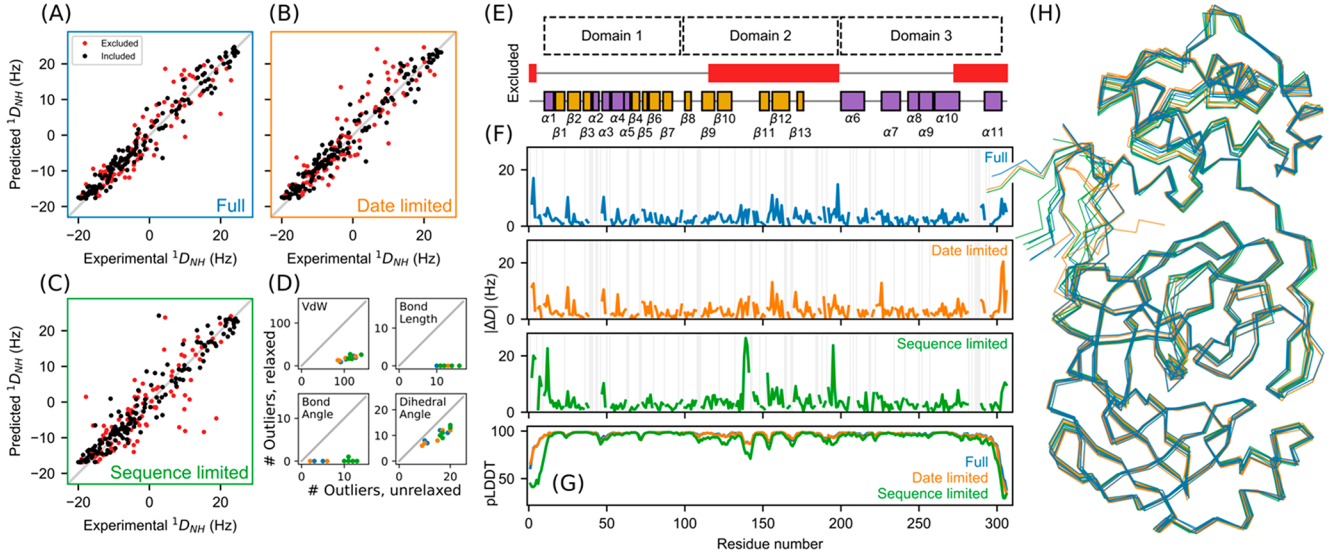

Figure 3. Agreement between measured M^{pro} RDCs and values predicted by AlphaFold2 models. $^1D_{NH}$ vs values predicted by the top-ranked model generated by (A) the *full* AlphaFold2, (B) the *date-limited*, and (C) the *sequence-limited* versions. (D) Comparison of the number of outliers in PSVS[22] structure validation of unrelaxed and relaxed AlphaFold2 models. Colors in D match those in (F). (E) Domain partitioning, excluded regions (red), and secondary structure (α-helix, purple; β-sheet, yellow). (F) Combined $^1D_{NH}$ and $^2D_{C'H}$ absolute deviation, ΔD, of predicted $^1D_{NH}$ and $^2D_{C'H}$ from experimental values for the full (blue), date-limited (orange), and sequence-limited (green) AlphaFold2 models, with missing data marked by gray stripes. (G) AlphaFold2 pLDDT for each of the top-scoring AlphaFold2 models. (H) C^α wireframe overlay of all AlphaFold2 models, color-matched to A−C.

https://doi.org/10.1021/jacs.1c10588
J. Am. Chem. Soc. 2021, 143, 19306−19310

structures and RDCs. A further decrease is observed between M^{pro}_{AFD1-5} and M^{pro}_{AFS1-5}, reflecting reduced accuracy of the predicted models, despite the comparable C^{α} RMSD (Figure 3A−C,H and Figures S6−S9), with a surprisingly large range in RDC fit quality for *full* and *date-limited* models (Figure 2E). Importantly, even for models predicted by the *sequence-limited* AlphaFold2 implementation, RDC data agree very well in regions with defined secondary structure (Q = 0.287; Table S3).

The highest pLDDT-scoring model of the *full* AlphaFold2 implementation fits RDCs better than 92% of all X-ray structures. However, a trend is observed that the Amber-based[18] constrained-relaxation step reduces the quality of RDC predictions (Table 1 and Table S3), while somewhat improving the predictions for non-secondary structure residues (Table S3). This minimization step perturbs bond vector orientations by an average of 3.9° across all structures, with 95% under 9.2°, but the distributions are strongly tailed (max = 53.1°; Figures S10 and S11).

Relative to the best X-ray structures, AlphaFold2 M^{pro} models agree more closely with solution RDCs for residues that are part of regular secondary structure than the remainder (Figure 3E,F, Table 1, and Table S3). This result indicates that, while there remains some room for further improvement in structure prediction, catalytic scaffolds are very well defined by AlphaFold2. Our findings therefore suggest new opportunities for combining experimentation with molecular dynamics simulations,[20] where solution RDCs provide highly precise input for QM/MM simulations of substrate binding/reaction trajectories.[21]

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.1c10588.

Pulse sequence diagrams; details on sample preparation, AlphaFold implementation, and data fitting procedure; Figures S1−S11 depicting the effect of Amber refinement; and Tables S1−S5 with experimental data and proteins used by Alphafold (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Ad Bax − *Laboratory of Chemical Physics, NIDDK, National Institutes of Health, Bethesda, Maryland 20892-0520, United States;* orcid.org/0000-0002-9809-5700; Email: bax@nih.gov

Authors

Angus J. Robertson − *Laboratory of Chemical Physics, NIDDK, National Institutes of Health, Bethesda, Maryland 20892-0520, United States*

Joseph M. Courtney − *Laboratory of Chemical Physics, NIDDK, National Institutes of Health, Bethesda, Maryland 20892-0520, United States*

Yang Shen − *Laboratory of Chemical Physics, NIDDK, National Institutes of Health, Bethesda, Maryland 20892-0520, United States*

Jinfa Ying − *Laboratory of Chemical Physics, NIDDK, National Institutes of Health, Bethesda, Maryland 20892-0520, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/jacs.1c10588

Author Contributions

†A.J.R. and J.M.C. contributed equally.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank J. L. Baber for technical support and D. A. Torchia for useful discussions. This work was supported by the Intramural Research Program of the National Institute of Diabetes and Digestive and Kidney Diseases.

■ REFERENCES

(1) Yang, H. T.; Xie, W. Q.; Xue, X. Y.; Yang, K. L.; Ma, J.; Liang, W. X.; Zhao, Q.; Zhou, Z.; Pei, D. Q.; Ziebuhr, J.; Hilgenfeld, R.; Yuen, K. Y.; Wong, L.; Gao, G. X.; Chen, S. J.; Chen, Z.; Ma, D. W.; Bartlam, M.; Rao, Z. Design of wide-spectrum inhibitors targeting coronavirus main proteases. *PLoS Biol.* **2005**, *3*, 1742−1752.

(2) Hilgenfeld, R. From SARS to MERS: crystallographic studies on coronaviral proteases enable antiviral drug design. *FEBS J.* **2014**, *281*, 4085−4096.

(3) Zhang, L.; Lin, D.; Sun, X.; Curth, U.; Drosten, C.; Sauerhering, L.; Becker, S.; Rox, K.; Hilgenfeld, R. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science* **2020**, *368*, 409−412.

(4) Douangamath, A.; Fearon, D.; Gehrtz, P.; Krojer, T.; Lukacik, P.; Owen, C. D.; Resnick, E.; Strain-Damerell, C.; Aimon, A.; Abranyi-Balogh, P.; Brandao-Neto, J.; Carbery, A.; Davison, G.; Dias, A.; Downes, T. D.; Dunnett, L.; Fairhead, M.; Firth, J. D.; Jones, S. P.; Keeley, A.; Keseru, G. M.; Klein, H. F.; Martin, M. P.; Noble, M. E. M.; O'Brien, P.; Powell, A.; Reddi, R. N.; Skyner, R.; Snee, M.; Waring, M. J.; Wild, C.; London, N.; von Delft, F.; Walsh, M. A. Crystallographic and electrophilic fragment screening of the SARS-CoV-2 main protease. *Nat. Commun.* **2020**, *11*, 5047.

(5) Senior, A. W.; Evans, R.; Jumper, J.; Kirkpatrick, J.; Sifre, L.; Green, T.; Qin, C.; Žídek, A.; Nelson, A. W. R.; Bridgland, A.; Penedones, H.; Petersen, S.; Simonyan, K.; Crossan, S.; Kohli, P.; Jones, D. T.; Silver, D.; Kavukcuoglu, K.; Hassabis, D. Improved protein structure prediction using potentials from deep learning. *Nature* **2020**, *577*, 706−710.

(6) Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G. R.; Wang, J.; Cong, Q.; Kinch, L. N.; Schaeffer, R. D.; Millán, C.; Park, H.; Adams, C.; Glassman, C. R.; DeGiovanni, A.; Pereira, J. H.; Rodrigues, A. V.; van Dijk, A. A.; Ebrecht, A. C.; Opperman, D. J.; Sagmeister, T.; Buhlheller, C.; Pavkov-Keller, T.; Rathinaswamy, M. K.; Dalwadi, U.; Yip, C. K.; Burke, J. E.; Garcia, K. C.; Grishin, N. V.; Adams, P. D.; Read, R. J.; Baker, D. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **2021**, *373*, No. eabj8754.

(7) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583−589.

(8) Tjandra, N.; Bax, A. Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* **1997**, *278*, 1111−4.

(9) Prestegard, J. H.; Bougault, C. M.; Kishore, A. I. Residual dipolar couplings in structure determination of biomolecules. *Chem. Rev.* **2004**, *104*, 3519−3540.

(10) Zweckstetter, M. NMR hawk-eyed view of AlphaFold2 structures. *Protein Sci.* **2021**, *302333*.

(11) Altincekic, N.; Korn, S. M.; Qureshi, N. S.; Dujardin, M.; Ninot-Pedrosa, M.; Abele, R.; Saad, M. J. A.; Alfano, C.; Almeida, F. C. L.; Alshamleh, I.; de Amorim, G. C.; Anderson, T. K.; Anobom, C. D.; Anorma, C.; Bains, J. K.; Bax, A.; Blackledge, M.; Blechar, J.; Bockmann, A.; Brigandat, L.; Bula, A.; Butikofer, M.; Camacho-Zarco, A. R.; Carlomagno, T.; Caruso, I. P.; Ceylan, B.; Chaikuad, A.; Chu, F. X.; Cole, L.; Crosby, M. G.; de Jesus, V.; Dhamotharan, K.; Felli, I. C.; Ferner, J.; Fleischmann, Y.; Fogeron, M. L.; Fourkiotis, N. K.; Fuks, C.; Furtig, B.; Gallo, A.; Gande, S. L.; Gerez, J. A.; Ghosh, D.; Gomes-Neto, F.; Gorbatyuk, O.; Guseva, S.; Hacker, C.; Hafner, S.; Hao, B.; Hargittay, B.; Henzler-Wildman, K.; Hoch, J. C.; Hohmann, K. F.; Hutchison, M. T.; Jaudzems, K.; Jovic, K.; Kaderli, J.; Kalnins, G.; Kanepe, I.; Kirchdoerfer, R. N.; Kirkpatrick, J.; Knapp, S.; Krishnathas, R.; Kutz, F.; zur Lage, S.; Lambertz, R.; Lang, A.; Laurents, D.; Lecoq, L.; Linhard, V.; Lohr, F.; Malki, A.; Bessa, L. M.; Martin, R. W.; Matzel, T.; Maurin, D.; McNutt, S. W.; Mebus-Antunes, N. C.; Meier, B. H.; Meiser, N.; Mompean, M.; Monaca, E.; Montserret, R.; Perez, L. M.; Moser, C.; Muhle-Goll, C.; Neves-Martins, T. C.; Ni, X. M. N.; Norton-Baker, B.; Pierattelli, R.; Pontoriero, L.; Pustovalova, Y.; Ohlenschlager, O.; Orts, J.; Da Poian, A. T.; Pyper, D. J.; Richter, C.; Riek, R.; Rienstra, C. M.; Robertson, A.; Pinheiro, A. S.; Sabbatella, R.; Salvi, N.; Saxena, K.; Schulte, L.; Schiavina, M.; Schwalbe, H.; Silber, M.; Almeida, M. D.; Sprague-Piercy, M. A.; Spyroulias, G. A.; Sreeramulu, S.; Tants, J. N.; Tars, K.; Torres, F.; Tows, S.; Trevino, M. A.; Trucks, S.; Tsika, A. C.; Varga, K.; Wang, Y.; Weber, M. E.; Weigand, J. E.; Wiedemann, C.; Wirmer-Bartoschek, J.; Martin, M. A. W.; Zehnder, J.; Hengesbach, M.; Schlundt, A. Large-Scale Recombinant Production of the SARS-CoV-2 Proteome for High-Throughput and Structural Biology Applications. *Frontiers in Molecular Biosciences* **2021**, *8*, 653148.

(12) Robertson, A. J.; Ying, J.; Bax, A. Four-dimensional NOE-NOE spectroscopy of SARS-CoV-2 Main Protease to facilitate resonance assignment and structural analysis. *Magn. Reson.* **2021**, *2*, 129−138.

(13) Salzmann, M.; Wider, G.; Pervushin, K.; Senn, H.; Wuthrich, K. TROSY-type triple-resonance experiments for sequential NMR assignments of large proteins. *J. Am. Chem. Soc.* **1999**, *121*, 844−848.

(14) Fitzkee, N. C.; Bax, A. Facile measurement of H-1-N-15 residual dipolar couplings in larger perdeuterated proteins. *J. Biomol. NMR* **2010**, *48*, 65−70.

(15) Ottiger, M.; Bax, A. Determination of relative N-H-N N-C ', C-alpha-C ', and C(alpha)-H-alpha effective bond lengths in a protein by NMR in a dilute liquid crystalline phase. *J. Am. Chem. Soc.* **1998**, *120*, 12334−12341.

(16) Losonczi, J. A.; Andrec, M.; Fischer, M. W. F.; Prestegard, J. H. Order matrix analysis of residual dipolar couplings using singular value decomposition. *J. Magn. Reson.* **1999**, *138*, 334−342.

(17) Bax, A.; Grishaev, A. Weak alignment NMR: a hawk-eyed view of biomolecular structure. *Curr. Opin. Struct. Biol.* **2005**, *15*, 563−570.

(18) Case, D. A.; Cheatham, T. E., Iii; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26*, 1668−1688.

(19) Kryshtafovych, A.; Schwede, T.; Topf, M.; Fidelis, K.; Moult, J. Critical assessment of methods of protein structure prediction (CASP)-Round XIII. *Proteins: Struct., Funct., Genet.* **2019**, *87*, 1011−1020.

(20) Bottaro, S.; Lindorff-Larsen, K. Biophysical experiments and biomolecular simulations: A perfect match? *Science* **2018**, *361*, 355−360.

(21) Chan, H. T. H.; Moesser, M. A.; Walters, R. K.; Malla, T. R.; Twidale, R. M.; John, T.; Deeks, H. M.; Johnston-Wood, T.; Mikhailov, V.; Sessions, R. B.; Dawson, W.; Salah, E.; Lukacik, P.; Strain-Damerell, C.; Owen, C. D.; Nakajima, T.; Świderek, K.; Lodola, A.; Moliner, V.; Glowacki, D. R.; Spencer, J.; Walsh, M. A.; Schofield, C. J.; Genovese, L.; Shoemark, D. K.; Mulholland, A. J.; Duarte, F.; Morris, G. M. Discovery of SARS-CoV-2 Mpro peptide inhibitors from modelling substrate and ligand binding. *Chemical Science* **2021**, *12*, 13686−13703.

(22) Bhattacharya, A.; Tejero, R.; Montelione, G. T. Evaluating protein structures determined by structural genomics consortia. *Proteins: Struct., Funct., Genet.* **2007**, *66*, 778−795.

https://doi.org/10.1021/jacs.1c10588
J. Am. Chem. Soc. 2021, 143, 19306−19310