

Sampling alternative conformational states of transporters and receptors with AlphaFold2

Diego del Alamo^{1,2†}, Davide Sala^{3†}, Hassane S Mchaourab^{1*}, Jens Meiler^{2,3*}

¹Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, United States; ²Department of Chemistry, Vanderbilt University, Nashville, United States; ³Institute for Drug Discovery, Leipzig University, Leipzig, Germany

Abstract Equilibrium fluctuations and triggered conformational changes often underlie the functional cycles of membrane proteins. For example, transporters mediate the passage of molecules across cell membranes by alternating between inward- and outward-facing states, while receptors undergo intracellular structural rearrangements that initiate signaling cascades. Although the conformational plasticity of these proteins has historically posed a challenge for traditional *de novo* protein structure prediction pipelines, the recent success of AlphaFold2 (AF2) in CASP14 culminated in the modeling of a transporter in multiple conformations to high accuracy. Given that AF2 was designed to predict static structures of proteins, it remains unclear if this result represents an underexplored capability to accurately predict multiple conformations and/or structural heterogeneity. Here, we present an approach to drive AF2 to sample alternative conformations of topologically diverse transporters and G-protein-coupled receptors that are absent from the AF2 training set. Whereas models of most proteins generated using the default AF2 pipeline are conformationally homogeneous and nearly identical to one another, reducing the depth of the input multiple sequence alignments by stochastic subsampling led to the generation of accurate models in multiple conformations. In our benchmark, these conformations spanned the range between two experimental structures of interest, with models at the extremes of these conformational distributions observed to be among the most accurate (average template modeling score of 0.94). These results suggest a straightforward approach to identifying native-like alternative states, while also highlighting the need for the next generation of deep learning algorithms to be designed to predict ensembles of biophysically relevant states.

***For correspondence:**

hassane.mchaourab@vanderbilt.edu (HSM);

jens.meiler@vanderbilt.edu (JM)

[†]These authors contributed equally to this work

Competing interest: The authors declare that no competing interests exist.

Funding: See page 9

Preprinted: 22 November 2021

Received: 22 November 2021

Accepted: 02 March 2022

Published: 03 March 2022

Reviewing Editor: Janice L Robertson, Washington University in St Louis, United States

© Copyright del Alamo et al. This article is distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use and redistribution provided that the original author and source are credited.

Editor's evaluation

del Alamo and colleagues illustrate that restricting the depth of the input multiple sequence alignment allows AlphaFold2 to predict diverse conformational ensembles of transporters and receptors, as opposed to single static models reflecting individual states. Although they are limited to a small number of test cases of membrane proteins, the examples are of interest to members of the community. This work presents a validation of a simple approach that may be applicable to all proteins and is thus an exciting advance that is expected to be of broad interest.

Introduction

Dynamic interconversion between multiple conformations underpins the functions of integral membrane proteins in all domains of life (Campbell et al., 2016; Cournia et al., 2015; Shaw et al., 2010; Boehr et al., 2009). For example, the vectorial translocation of substrates by transporters is

mediated by movements that open and close extra- and intracellular gates (*Drew and Boudker, 2016; Kazmier et al., 2017*). For G-protein-coupled receptors (GPCRs), ligand binding on the extracellular side triggers structural rearrangements on the intracellular side that initiate downstream signaling (*Wang et al., 2020; Gusach et al., 2020*). Traditional computational prediction pipelines reliant on inter-residue distance restraints calculated from deep multiple sequence alignments (MSAs) have historically struggled to accurately predict the structures of these proteins and their movements. The resulting models are unnaturally compact and frequently distorted, preventing critical questions about ligand and/or drug binding modes from being addressed (*Ovchinnikov et al., 2015; Nicoludis and Gaudet, 2018*).

A performance breakthrough was unveiled during CASP14 by AlphaFold2 (AF2) (*Jumper et al., 2021a; Tunyasuvunakool et al., 2021; Pereira et al., 2021*), which achieved remarkably accurate *de novo* structure prediction. Upon examining the list of CASP14 targets and corresponding models, we found that AF2 modeled the multidrug transporter LmrP (target T1024) in multiple conformations, two of which were individually consistent with published experimental data (*Jumper et al., 2021b; Del Alamo et al., 2021a; Debruycker et al., 2020; Martens et al., 2016; Masureel et al., 2014*). This observation stimulated the question of whether such performance can be duplicated for other membrane proteins. At its essence, this question centers on whether AF2 can sample the conformational landscape in the minimum energy basin. Here, we investigate this hypothesis using a benchmark set of topologically diverse transporters and GPCRs. Our results demonstrate that reducing the depth of the input MSAs is often conducive to the generation of accurate models in multiple conformations by AF2, suggesting that the algorithm's outstanding predictive performance can be extended to sample alternative structures of the same target. For most proteins considered, we report a striking correlation between the breadth of structures predicted by AF2 and the corresponding cryo-EM and/or X-ray crystal structures. Finally, we propose a modeling pipeline for researchers interested in sampling alternative conformations of specific membrane proteins, which we apply to the structurally unknown GPR114/ADGRG5 adhesion GPCR as an example.

Results and discussion

The default three-stage AF2 pipeline consists of (1) querying of sequence databases and generation of an MSA, (2) inference of structure via a neural network using a randomly resampled subset of this MSA containing up to 5120 sequences, which is repeated three times (a process termed 'recycling'), and (3) resolution of steric clashes and bond geometry using a constrained all-atom molecular dynamics simulation. The neural networks used for prediction were trained on all structures deposited in the Protein Data Bank (PDB) on or before April 30, 2018 (*Jumper et al., 2021a*). Therefore, by necessity, this study is restricted to proteins whose structures were absent from the PDB before this date and have since been determined at atomic resolution in two or more conformations. We selected five transporters that not only met these criteria but also reflected a range of transport mechanisms characterized in the literature (*Drew and Boudker, 2016*), including rocking-bundle (LAT1, *Yan et al., 2019; Yan et al., 2021*; ZnT8, *Xue et al., 2020*), rocker-switch (MCT1, *Wang et al., 2021*; STP10, *Bavnhoj et al., 2021*), and elevator (ASCT2, *Garib Singh et al., 2021; Garaeva et al., 2019*). We also included three GPCRs, which were distributed across classes A (CGRPR, *Liang et al., 2020; Josephs et al., 2021*), B1 (PTH1R, *Ehrenmann et al., 2018; Zhao et al., 2019*), and F (FZD7, *Xu et al., 2021*); to serve as points of comparison, we used the active conformation of FZD7 and the inactive conformation of the nearly identical FZD4, *Yang et al., 2018*).

AF2 generates multiple conformations of all eight target proteins

The sequences of all targets were truncated at the N- and C-termini to remove large soluble and/or intrinsically disordered regions which represent a challenge for AF2 (see *Methods*). The structures were then predicted using the default AF2 structure prediction pipeline in the absence of templates. However, the resulting models were largely identical to one another and failed to shed light on the target protein's conformational space. To diversify the models generated by AF2, we reduced the number of recycles to one and restricted the depth of the randomly subsampled MSAs to contain as few as 16 sequences. To sample the conformational landscape more exhaustively, we generated 50 models of each protein for each MSA size, while skipping the final MD simulation to reduce the

pipeline's total computational cost. For the targets, each model's similarity to the experimental structures was quantified using template modeling (TM) score (Zhang and Skolnick, 2004; Zhang and Skolnick, 2005; Xu and Zhang, 2010), a metric ranging from 0 to 1, which indicates how well the two backbone (C_{α} atoms) structures superimpose over one another (higher values corresponding to greater similarity; Figure 1A).

Accurate models of all eight protein targets were obtained for at least one conformation (TM score ≥ 0.9), consistent with published performance statistics (Figure 1B). MSAs with hundreds or thousands of sequences were generally observed to engender tighter clustering in conformations specific to each protein. Decreasing the depth of the subsampled MSAs, by contrast, appeared to promote the generation of alternative conformations in most proteins. The increased diversity coincided with the generation of misfolded or outlier models. However, unlike the models of interest that resembled experimentally determined structures, misfolded models virtually never coclustered and could thus be identified and excluded from further analysis (example shown in Figure 1—figure supplement 1). Increasing the depth of subsampled MSAs had the desirable effect of eliminating these models, while also limiting the extent to which alternative conformations were sampled. Thus, our results revealed a delicate balance that must be achieved to generate models that are both diverse and natively folded. No general pattern was readily apparent regarding the ideal MSA depth required to achieve this balance, even when accounting for sequence length of the target (Figure 1—figure supplement 2).

One target, MCT1, was exclusively modeled by AF2 in either inward-facing (IF) or fully occluded conformations; over 99% of the models had TM scores of ≥ 0.9 and < 0.9 to the IF and outward-facing (OF) structures, respectively, regardless of MSA depth. Therefore, we investigated the effect of providing templates of homologs in exclusively OF conformations alongside MSAs of various sizes (see Methods for details on template selection). Accurate OF models were obtained only with MSAs containing 16–32 sequences and constituted a minor population in an ensemble dominated by IF models. Thus, the generation of large numbers of models appeared to be necessary to yield intermediate conformations of interest. Similar results were observed when we modeled PTH1R using either inactive or active templates, as well as LAT1 using either OF or IF templates (Figure 1—figure supplement 3), further indicating that the information content provided by the templates diminishes as the depth of the MSA increases.

Overall, these results demonstrate that both conformations of all eight protein targets could be predicted with AF2 to high accuracy (TM score ≥ 0.9) by using MSAs that are far shallower than the default. However, because the optimal MSA depth and choice of templates varied for each protein, these parameters need to be explored for conformational sampling of a particular target.

Predicted conformational fluctuations correlate with implied conformational dynamics

To further investigate the structural heterogeneity predicted by these models, we calculated each residue's C_{α} atom distance between the two superimposed experimental structures, as well as each residue's root mean square fluctuation (RMSF) among all 50 models following structure-based alignment (Figure 2). Correlation between these two measures was observed in most cases and was notable for ASCT2, LAT1, CGRPR, and MCT1 with templates ($R^2 \geq 0.75$). The exception was MCT1 without templates, which was likely due to a lack of conformational diversity among the sampled models. The inclusion of templates restored this correlation in MCT1 but contributed negligibly to those of PTH1R and LAT1 (Figure 2—figure supplement 1). The correlation demonstrates that predicted flexibility by AF2 is related to the protein's dynamics inferred from the experimental structures. In contrast with a recent preprint (Saldaño et al., 2021), the predicted flexibility values failed to correlate with their pLDDT values, which reflect the confidence of the AF2 prediction of each residue's local environment (Mariani et al., 2013).

Distributions of predicted models relative to the experimental structures

Visual examination suggested that many of the predicted models fall 'in between' the two experimentally determined conformations (example shown in Figure 3A). Furthermore, certain structural features expected to be conformationally heterogeneous, such as long loops, appeared to be nearly identical across these models. Both observations raised questions about the relationship between

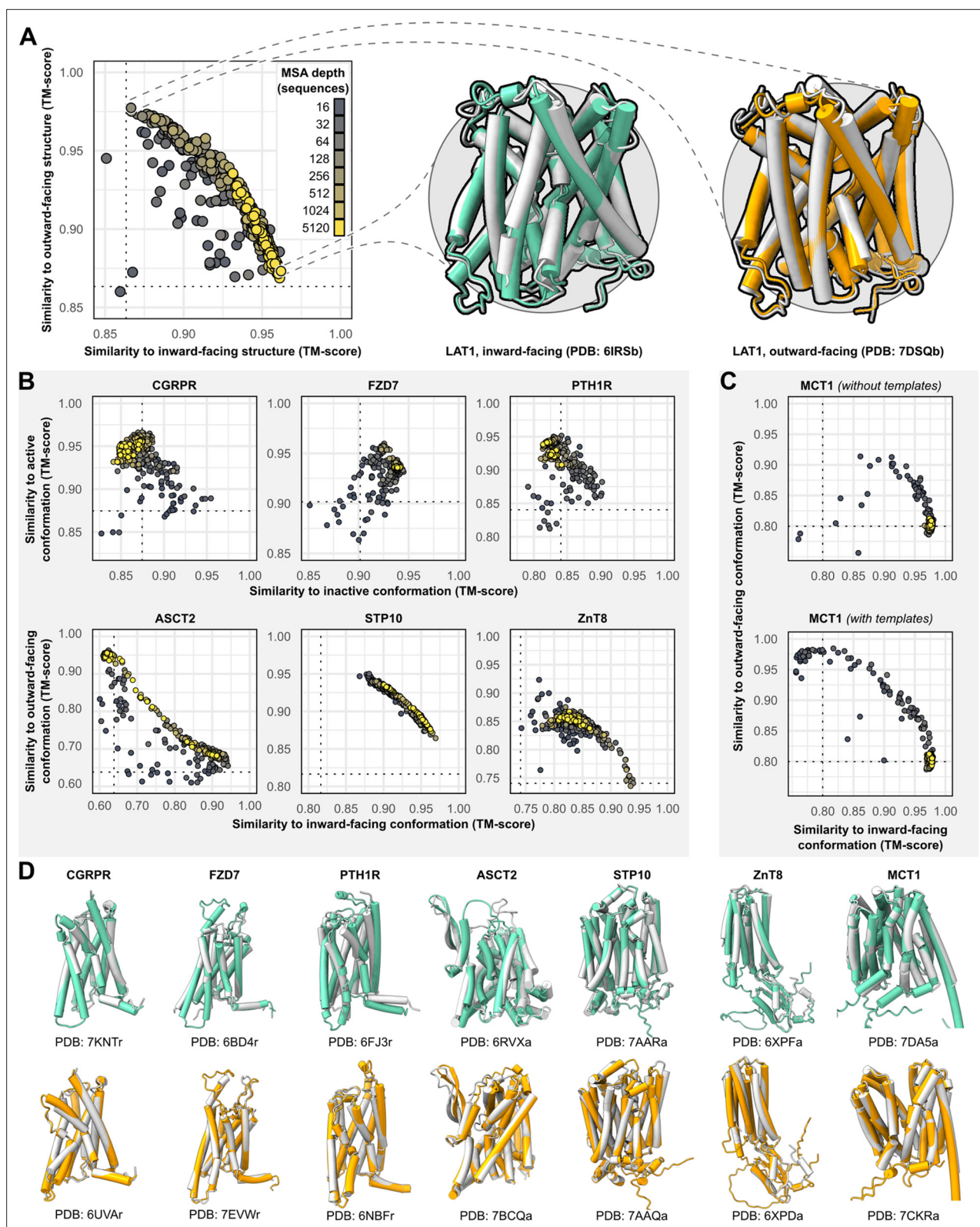


Figure 1. Alternative conformations of transporters and G-protein-coupled receptors (GPCRs) can be predicted by AlphaFold2 (AF2). **(A)** Representative models of the transporter LAT1 in inward-facing (IF) and outward-facing (OF) conformations. Experimental structures shown in gray and models shown in colors. **(B)** Comparison of AF2 models with inactive/active or IF/OF experimental structures as a function of multiple sequence alignment (MSA) depth for GPCRs (top) and transporters (bottom), respectively. All models shown here were generated without templates. Dashed lines indicate the template

Figure 1 continued on next page

Figure 1 continued

modeling (TM) score between experimental structures and are shown for reference. (C) Supplementing shallow MSAs with OF templates allows AF2 to predict the OF conformation of MCT1. (D) Experimental structures superimposed over models with the greatest TM scores. Inactive/IF and active/OF cartoons shown on the top and bottom in teal and orange, respectively.

The online version of this article includes the following figure supplement(s) for figure 1:

Figure supplement 1. Example principal component analysis (PCA) of ASCT2 models generated by AlphaFold2 (AF2) containing outlier models.

Figure supplement 2. Conformational homogeneity as a function of multiple sequence alignment (MSA) depth.

Figure supplement 3. Templates contribute to conformational sampling only when shallow multiple sequence alignments (MSAs) are provided.

Figure supplement 4. Protein targets with one conformation in the training set cannot be predicted in the alternative conformation.

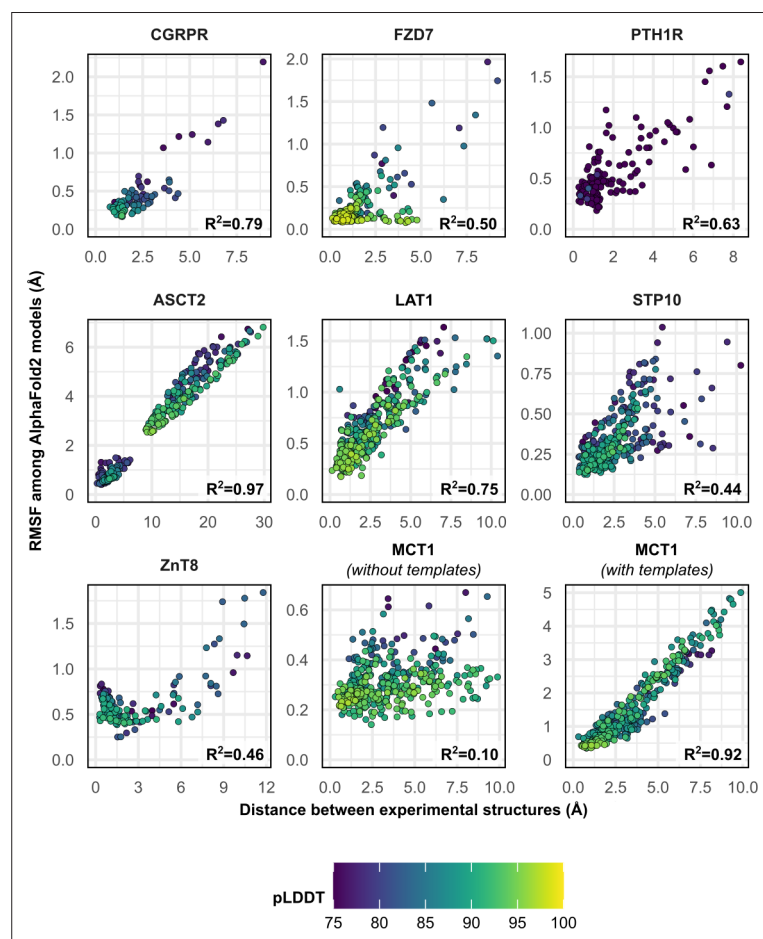


Figure 2. Comparison between the movement undergone by each C_{α} atom between the two superimposed experimental structures and their root mean square fluctuation (RMSF) values among AlphaFold2 (AF2) models. Residues with low confidence (pLDDT ≤ 75) were omitted from this plot for all proteins except PTH1R. Multiple sequence alignment (MSA) sizes of 128 sequences were used for all predictions, except for MCT1 with templates, which instead used 32 sequences to capture the outward-facing (OF) conformation. pLDDT refers to each residue's predicted accuracy, with a value of 100 indicating maximum confidence.

The online version of this article includes the following figure supplement(s) for figure 2:

Figure supplement 1. Comparison between the movement undergone by each C_{α} atom between the two superimposed experimental structures of PTH1R and LAT1 and their root mean square fluctuation (RMSF) values among AlphaFold2 (AF2) models generated with templates.

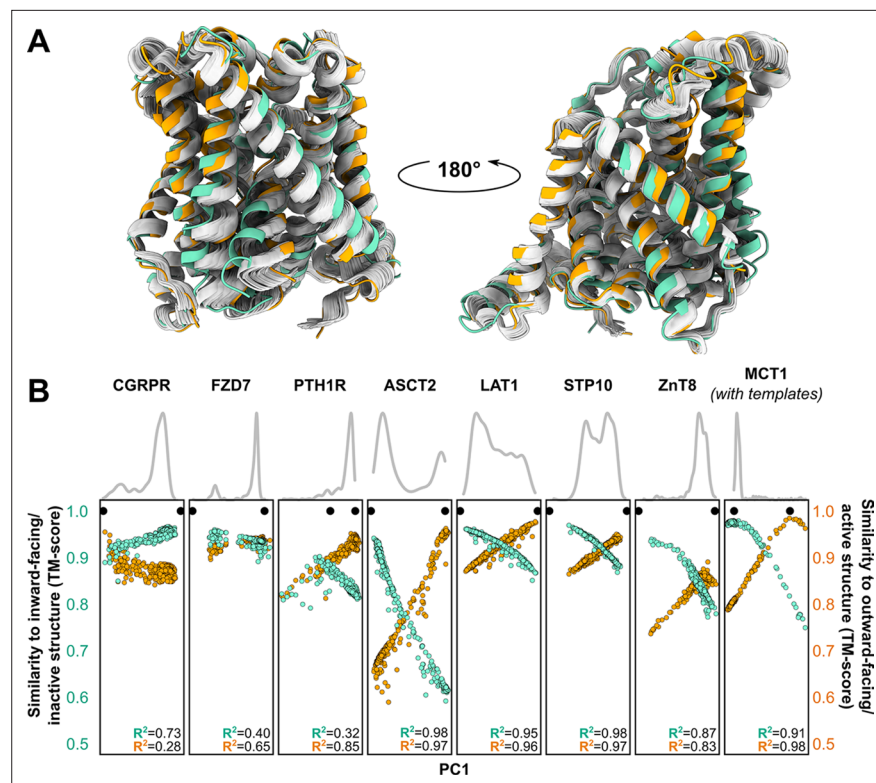


Figure 3. Distinct conformations can be delineated using principal component analysis (PCA). **(A)** Conformational heterogeneity in AlphaFold2 (AF2) models of LAT1. Experimental inward-facing (IF) and outward-facing (OF) conformations shown in teal and orange, respectively, while the gallery of AF2 models generated using 128 sequences are shown in gray. **(B)** Distribution of AF2 models generated using multiple sequence alignments (MSAs) with 32 or more sequences across the first principal component (PC1) following PCA (gray traces). Scatter plots comparing each model's position along PC1 and its structural similarity to experimentally determined structures. Teal: similarity to IF (transporters) or inactive (G-protein-coupled receptors, GPCRs) conformation. Orange: similarity to OF (transporters) or active (GPCRs) conformation. Each model is shown twice, once in teal and once in orange. Native structures are shown as black dots.

The online version of this article includes the following figure supplement(s) for figure 3:

Figure supplement 1. Models sampled by AlphaFold2 (AF2) in multiple conformations cannot be fully explained by linear interpolation of two end structures.

Figure supplement 2. Example predictions of the adhesion G-protein-coupled receptor (aGPCR) GPR114/ADGRG5.

the diversity of the predicted models and the breadth of the conformational ensembles bracketed by the experimental structures. To quantitatively place the predicted conformational variance in the context of the experimentally determined structures, we used principal component analysis (PCA), which reduces the multidimensional space to a smaller space representative of the main conformational motions. In our benchmark set, the first principal component (PC1) captured $64.9\% \pm 16.1\%$ of the structural variations among the models generated using MSAs with 32 or more sequences (**Figure 3B**), while comparison of PC1/PC2 values suggested that the predicted dynamics deviate from simple interpolation of two end states (**Figure 3—figure supplement 1**). The experimental structures virtually always occupied well-defined extreme positions. In every case, a correlation was evident between each model's PC1 values and their TM scores. Indeed, the models with the most extreme PC1 values were also among the most accurate: average TM scores were 0.94 for the top 1, top 3, and top 10 PC1 models, and Pearson's correlation coefficients between PC1 and TM scores of the ensemble of models exceeded 0.8 for all transporters in this dataset. Moreover, the experimental structures virtually always flanked the AF2 models along PC1. The exception, PTH1R, was determined in a partially inactive and active conformation (**Zhao et al., 2019**), suggesting that models extending

beyond the former state along PC1 may represent the fully inactive conformation. Therefore, these results indicate that accurate representative models of conformations of interest can be selected from the extreme positions along PC1.

Limited conformational sampling is observed for proteins with structures in the AF2 training set

A follow-up question centers on whether this strategy can yield similar results for proteins with one conformation present in the AF2 training set. We investigated this question using four membrane proteins with two experimentally determined conformations, at least one of which was included in the AF2 training set: the class A GPCR CCR5 (Zheng *et al.*, 2017; Zhang *et al.*, 2021), the serotonin transporter SERT (Coleman *et al.*, 2016; Coleman *et al.*, 2019), the multidrug transporter PfMATE (Tanaka *et al.*, 2013; Zakrzewska *et al.*, 2019), and the lipid flippase MurJ (Kuk *et al.*, 2017; Kuk *et al.*, 2019). Using the template-free prediction pipeline outlined above, we determined the resultant models' similarity to the structures included in and absent from the training set. Unlike the results presented above, the majority of the transporter models generated this way were more similar to the conformation present in the training set than the conformation absent from the training set (i.e., their TM scores were greater; Figure 1—figure supplement 4). The conformational diversity of these models, including those generated using shallow MSAs, was far more limited than what was generally observed for the proteins discussed above, with the exception of MCT1 (Figure 1—figure supplement 2). Although conformational diversity was demonstrated to a limited extent by the generation of occluded models of MurJ and PfMATE, none of the models observed adopted the alternative conformer. By contrast, while models of CCR5 were less biased toward the training set conformation, deep MSAs reduced conformational diversity. This divergence in performance may stem from the composition of the AF2 training set, which featured the structures of many active GPCRs but no structures, for example, of IF MATEs (Claxton *et al.*, 2021).

Concluding remarks: proposed workflow and future directions

Our results indicate that the state-of-the-art *de novo* structural modeling algorithm AF2 can be manipulated to accurately model alternative conformations of transporters and GPCRs whose structures were not available in the training set. The use of shallow MSAs was instrumental to obtaining structurally diverse models in most proteins, and in one case (MCT1) accurate modeling of alternative conformations also required the manual curation of template structures. Thus, while the results presented here provide a blueprint for obtaining AF2 models of alternative conformations, they also argue against an optimal one-size-fits-all approach for sampling the conformational space of every protein with high accuracy. Indeed, whereas the DeepMind team reportedly required templates to obtain models of LmrP in an OF conformation (Jumper *et al.*, 2021b), we found that this procedure was usually unnecessary. Accurate representatives of distinct conformers were generally obtainable with exhaustive sampling and could be identified by performing PCA and selecting models at the extreme positions of PC1. Nevertheless, prediction pipelines will likely require a combination of iterative fine-tuning specific to each target of interest followed by experimental verification to identify proposed conformers. Moreover, this approach showed limited success when applied to transporters whose structures were used to train AF2, hinting at the possibility that traditional methods may still be required to capture alternative conformers (Crawley *et al.*, 2011; Ollikainen *et al.*, 2013).

As a final verification of this proposed pipeline, we tested it on GPR114/ADGRG5, a class B2 adhesion GPCR whose structure has not been experimentally determined. The structural model deposited in the AF2 database, which likely depicts an active conformation that diverges from the structure of the homolog GPR97 (Ping *et al.*, 2021), could be recapitulated by using deep MSAs. The use of shallow MSAs (≤ 64 sequences), by contrast, yielded a range of intermediate conformations distributed across three well-separated clusters (Figure 3—figure supplement 2). One of these clusters contains models with an orientation of TM6 and TM7 that fully occludes the orthosteric site and partially blocks the cytosolic pocket where G proteins bind. The physiological relevance of these proposed structural movements nonetheless requires experimental validation.

While these results reinforce the notion that AF2 can provide models to guide biophysical studies of conformationally heterogeneous membrane proteins, they represent a methodological 'hack', rather than an explicit objective built into the algorithm's architecture. Several preprints have provided

evidence that AF2, despite its accuracy, likely does not learn the energy landscapes underpinning protein folding and function (Saldaño *et al.*, 2021; Pak *et al.*, 2021; Akdel *et al.*, 2021). Moreover, AF2 does not directly account for the lipid environment, which has been experimentally shown to bias the conformational equilibria of membrane proteins (Martens *et al.*, 2018; Muller *et al.*, 2019; Immadisetty *et al.*, 2019). As our results show, the exploration of the conformational space is in part a byproduct of low sequence information provided for inference. Ultimately, they highlight the need for further development of artificial intelligence methods capable of learning the conformational flexibility intrinsic to protein structures.

Materials and methods

Overview of the prediction pipeline

Prediction runs were executed using AlphaFold v2.0.1 and a modified version of ColabFold (Mirdita *et al.*, 2021) that is available at https://github.com/delalamo/af2_conformations, (Del Alamo, 2021b copy archived at [swh:1:rev:d60db86886186e80622deaa91045cacca4103d3](https://www.ncbi.nlm.nih.gov/bioRx/2021/11/11/2021.11.11.458886)). The pipeline used in this study differs from the default AF2 pipeline in several aspects. First, all MSAs were obtained using the MMSeqs2 server (Steinegger and Söding, 2017), rather than the default databases. Second, template search was disabled, except when explicitly performed with specific templates of interest (see below). Third, the number of recycles was set to one, rather than three by default. Finally, models were not refined following their prediction. This study utilized all 5 neural networks when predicting structures without templates, with 10 predictions per neural network per MSA size. The following residues were omitted from modeling: 1–131 and 401–461 of CGRPR, 1–247 of FZD7, 1–175 and 492–593 of PTH1R, and 1–49 of LAT1.

MSA subsampling

MSA subsampling was carried out randomly by AF2, and depth values were controlled by modifying 'max_msa_clusters' and 'max_extra_msa' parameters prior to execution. The former parameter determines the number of randomly chosen sequence clusters provided to the AF2 neural network. The latter parameter determines the number of extra sequences used to compute additional summary statistics. Throughout this manuscript, we refer to the latter when describing the depth of the MSA used for prediction and set the former to half this value in all cases except when 5120 sequences were used, in which case we set the former to 512. No manual intervention was carried out to fine-tune the composition of these alignments.

Template-based predictions

Templates were fetched using the MMSeqs2 server used by ColabFold. All templates were manually inspected, and those with structures similar to the desired conformation of interest were retained. The following templates were used: OF MCT1 with FucB (PDB 3O7Pa and 3O7Qa, 18% sequence identity calculated using Needleman Wunsch Needleman and Wunsch, 1970; Madeira *et al.*, 2019); OF LAT1 with AdiC (3OB6a and 5J4Ib, 22%); IF LAT1 with b(0,+)AT1 (6LI9, 45%), KCC3 (6Y5Ra, 12%), GkApcT (6F34a, 24%), NKCC1 (6NPLa, 6PZTa, and 6NPHa, 12%), BasC (6F2Ga and 6F2Wa, 29%), and GadC (4DJLa and 4DJKb, 21%); active PTH1R with GCGR (6LMLr and 6VCBr, 30%) PAC1 (6M1I, 34%), and CRF1 (6P9X, 34%); inactive PTH1R with GLP1 (6LN2, 30%) and GCGR (5YQZ, 5XF1, and 5EE7, 28%). Template processing then proceeded as previously described, except that the parameter 'subsample_templates' was set to True and the template similarity cutoff was reduced from 10% to 1%. Additionally, as only 2 of the 5 AF2 neural networks were parametrized to use templates, each of these 2 neural networks generated 25 models in order to arrive at 50 total models per MSA depth.

Structural analysis

TM scores were calculated using TM align (Zhang and Skolnick, 2005). PCA and RMSF calculations were carried out in CPPTRAJ (Roe and Cheatham, 2013). Loop residues were omitted from PCA.

Acknowledgements

This study was funded by the National Institutes of Health (HSM: GM 128087) and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC 1423, project number

421152132, subproject Z04. The authors would like to thank Dr. John Jumper for explaining how the DeepMind team predicted the structure of LmrP in CASP14 and Mr. Taylor Jones for helpful discussions on modeling MCT1 without templates.

Additional information

Funding

Funder	Grant reference number	Author
National Institutes of Health	GM 128087	Hassane S Mchaourab
Deutsche Forschungsgemeinschaft	CRC 1423, project number 421152132, subproject Z04	Jens Meiler

The funders had no role in study design, data collection, and interpretation, or the decision to submit the work for publication.

Author contributions

Diego del Alamo, Conceptualization, Data curation, Formal analysis, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review and editing; Davide Sala, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review and editing; Hassane S Mchaourab, Jens Meiler, Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review and editing

Author ORCIDs

Diego del Alamo  <http://orcid.org/0000-0003-1757-9971>
 Davide Sala  <http://orcid.org/0000-0002-3900-0011>
 Hassane S Mchaourab  <http://orcid.org/0000-0002-5673-0980>
 Jens Meiler  <http://orcid.org/0000-0001-8945-193X>

Decision letter and Author response

Decision letter <https://doi.org/10.7554/eLife.75751.sa1>
 Author response <https://doi.org/10.7554/eLife.75751.sa2>

Additional files

Supplementary files

- Transparent reporting form

Data availability

All scripts and data presented in this study are made available for download at https://github.com/delalamo/af2_conformations, (copy archived at [swh:1:rev:d60db86886186e80622deaa91045caccaf4103d3](https://www.swh.io/rev/d60db86886186e80622deaa91045caccaf4103d3)).

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Del Alamo D	2022	Sampling alternative conformational states of transporters and receptors with AlphaFold2	https://github.com/delalamo/af2_conformations	GitHub, conformations

References

Akdel M, Pires DEV, Porta Pardo E, Jänes J, Zalevsky AO, Mészáros B, Bryant P, Good LL, Laskowski RA, Pozzati G, Shenoy A, Zhu W, Kundrotas P, Ruiz Serra V, Rodrigues CHM, Dunham AS, Burke D, Borkakoti N,

- Velankar S, Frost A, et al. 2021. A structural biology community assessment of AlphaFold 2 applications. *bioRxiv*. DOI: <https://doi.org/10.1101/2021.09.26.461876>
- Bavnhoj L, Paulsen PA, Flores-Canales JC, Schiøtt B, Pedersen BP. 2021. Molecular mechanism of sugar transport in plants unveiled by structures of glucose/H symporter STP10. *Nature Plants* 7:1409–1419. DOI: <https://doi.org/10.1038/s41477-021-00992-0>, PMID: 34556835
- Boehr DD, Nussinov R, Wright PE. 2009. The role of dynamic conformational ensembles in biomolecular recognition. *Nature Chemical Biology* 5:789–796. DOI: <https://doi.org/10.1038/nchembio.232>, PMID: 19841628
- Campbell E, Kaltenbach M, Correy GJ, Carr PD, Porebski BT, Livingstone EK, Afriat-Jurnou L, Buckle AM, Weik M, Hollfelder F, Tokuriki N, Jackson CJ. 2016. The role of protein dynamics in the evolution of new enzyme function. *Nature Chemical Biology* 12:944–950. DOI: <https://doi.org/10.1038/nchembio.2175>, PMID: 27618189
- Claxton DP, Jagessar KL, Mchaourab HS. 2021. Principles of Alternating Access in Multidrug and Toxin Extrusion (MATE) Transporters. *Journal of Molecular Biology* 433:166959. DOI: <https://doi.org/10.1016/j.jmb.2021.166959>, PMID: 33774036
- Coleman JA, Green EM, Gouaux E. 2016. X-ray structures and mechanism of the human serotonin transporter. *Nature* 532:334–339. DOI: <https://doi.org/10.1038/nature17629>, PMID: 27049939
- Coleman JA, Yang D, Zhao Z, Wen PC, Yoshioka C, Tajkhorshid E, Gouaux E. 2019. Serotonin transporter-ibogaine complexes illuminate mechanisms of inhibition and transport. *Nature* 569:141–145. DOI: <https://doi.org/10.1038/s41586-019-1135-1>, PMID: 31019304
- Cournia Z, Allen TW, Andricioaei I, Antonny B, Baum D, Brannigan G, Buchete N-V, Deckman JT, Delemotte L, Del Val C, Friedman R, Gkeka P, Hege H-C, Hénin J, Kasimova MA, Kolocouris A, Klein ML, Khalid S, Lemieux MJ, Lindow N, et al. 2015. Membrane Protein Structure, Function, and Dynamics: a Perspective from Experiments and Theory. *The Journal of Membrane Biology* 248:611–640. DOI: <https://doi.org/10.1007/s00232-015-9802-0>, PMID: 26063070
- Crawley SW, Gharaei MS, Ye Q, Yang Y, Raveh B, London N, Schueler-Furman O, Jia Z, Côté GP. 2011. Autophosphorylation activates *Dictyostelium* myosin II heavy chain kinase A by providing A ligand for an allosteric binding site in the alpha-kinase domain. *The Journal of Biological Chemistry* 286:2607–2616. DOI: <https://doi.org/10.1074/jbc.M110.177014>, PMID: 21071445
- Debruycker V, Hutchin A, Masureel M, Ficici E, Martens C, Legrand P, Stein RA, Mchaourab HS, Faraldo-Gómez JD, Remaut H, Govaerts C. 2020. An embedded lipid in the multidrug transporter LmrP suggests a mechanism for polyspecificity. *Nature Structural & Molecular Biology* 27:829–835. DOI: <https://doi.org/10.1038/s41594-020-0464-y>, PMID: 32719456
- Del Alamo D, Govaerts C, Mchaourab HS. 2021a. AlphaFold2 predicts the inward-facing conformation of the multidrug transporter LmrP. *Proteins* 89:1226–1228. DOI: <https://doi.org/10.1002/prot.26138>, PMID: 33973689
- Del Alamo D. 2021b. Prediction of alternative conformations using AlphaFold 2. swb:1:rev:d60db86886186e80622deaa91045caccaf4103d3. Software Heritage. https://archive.softwareheritage.org/swb:1:dir:732ab369a1b141030fbc5728d9ab3f50e222b3ea;origin=https://github.com/delalamo/af2_conformations;visit=swb:1:snp:cb0ae8b45df2c5f548867364bbfe52debb482bda;anchor=swb:1:rev:d60db86886186e80622deaa91045caccaf4103d3
- Drew D, Boudker O. 2016. Shared Molecular Mechanisms of Membrane Transporters. *Annual Review of Biochemistry* 85:543–572. DOI: <https://doi.org/10.1146/annurev-biochem-060815-014520>, PMID: 27023848
- Ehrenmann J, Schöppe J, Klenk C, Rappas M, Kummer L, Doré AS, Plückthun A. 2018. High-resolution crystal structure of parathyroid hormone 1 receptor in complex with a peptide agonist. *Nature Structural & Molecular Biology* 25:1086–1092. DOI: <https://doi.org/10.1038/s41594-018-0151-4>, PMID: 30455434
- Garaeva AA, Guskov A, Slotboom DJ, Paulino C. 2019. A one-gate elevator mechanism for the human neutral amino acid transporter ASCT2. *Nature Communications* 10:1–8. DOI: <https://doi.org/10.1038/s41467-019-11363-x>, PMID: 31366933
- Garib Singh RAA, Ndaru E, Garaeva AA, Shi Y, Zielewicz L, Zakrepine P, Bonomi M, Slotboom DJ, Paulino C, Grewer C, Schlessinger A. 2021. Rational design of ASCT2 inhibitors using an integrated experimental-computational approach. *PNAS* 118:e2104093118. DOI: <https://doi.org/10.1073/pnas.2104093118>, PMID: 34507995
- Gusach A, Maslov I, Luginina A, Borshchevskiy V, Mishin A, Cherezov V. 2020. Beyond structure: emerging approaches to study GPCR dynamics. *Current Opinion in Structural Biology* 63:18–25. DOI: <https://doi.org/10.1016/j.sbi.2020.03.004>, PMID: 32305785
- Immadisetty K, Hettige J, Moradi M. 2019. Lipid-Dependent Alternating Access Mechanism of a Bacterial Multidrug ABC Exporter. *ACS Central Science* 5:43–56. DOI: <https://doi.org/10.1021/acscentsci.8b00480>, PMID: 30693324
- Josephs TM, Belousoff MJ, Liang Y-L, Piper SJ, Cao J, Garama DJ, Leach K, Gregory KJ, Christopoulos A, Hay DL, Danev R, Wootten D, Sexton PM. 2021. Structure and dynamics of the CGRP receptor in apo and peptide-bound forms. *Science (New York, N.Y.)* 372:eabf7258. DOI: <https://doi.org/10.1126/science.abf7258>, PMID: 33602864
- 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 SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, et al. 2021a. Highly accurate protein structure prediction with AlphaFold. *Nature* 596:583–589. DOI: <https://doi.org/10.1038/s41586-021-03819-2>, PMID: 34265844

- Jumper J**, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Židek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, et al. 2021b. Applying and improving AlphaFold at CASP14. *Proteins* **89**:1711–1721. DOI: <https://doi.org/10.1002/prot.26257>, PMID: 34599769
- Kazmier K**, Claxton DP, Mchaourab HS. 2017. Alternating access mechanisms of LeuT-fold transporters: trailblazing towards the promised energy landscapes. *Current Opinion in Structural Biology* **45**:100–108. DOI: <https://doi.org/10.1016/j.sbi.2016.12.006>, PMID: 28040635
- Kuk ACY**, Mashalidis EH, Lee SY. 2017. Crystal structure of the MOP flippase MurJ in an inward-facing conformation. *Nature Structural & Molecular Biology* **24**:171–176. DOI: <https://doi.org/10.1038/nsmb.3346>, PMID: 28024149
- Kuk ACY**, Hao A, Guan Z, Lee SY. 2019. Visualizing conformation transitions of the Lipid II flippase MurJ. *Nature Communications* **10**:1736. DOI: <https://doi.org/10.1038/s41467-019-09658-0>, PMID: 30988294
- Liang YL**, Belousoff MJ, Fletcher MM, Zhang X, Khoshouei M, Deganutti G, Koole C, Furness SGB, Miller LJ, Hay DL, Christopoulos A, Reynolds CA, Danev R, Wooten D, Sexton PM. 2020. Structure and Dynamics of Adrenomedullin Receptors AM and AM Reveal Key Mechanisms in the Control of Receptor Phenotype by Receptor Activity-Modifying Proteins. *ACS Pharmacology & Translational Science* **3**:263–284. DOI: <https://doi.org/10.1021/acscptsci.9b00080>, PMID: 32296767
- Madeira F**, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, Lopez R. 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research* **47**:W636–W641. DOI: <https://doi.org/10.1093/nar/gkz268>, PMID: 30976793
- Mariani V**, Biasini M, Barbato A, Schwede T. 2013. IDDT: A local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics (Oxford, England)* **29**:2722–2728. DOI: <https://doi.org/10.1093/bioinformatics/btt473>, PMID: 23986568
- Martens C**, Stein RA, Masureel M, Roth A, Mishra S, Dawaliby R, Konijnenberg A, Sobott F, Govaerts C, Mchaourab HS. 2016. Lipids modulate the conformational dynamics of a secondary multidrug transporter. *Nature Structural & Molecular Biology* **23**:744–751. DOI: <https://doi.org/10.1038/nsmb.3262>, PMID: 27399258
- Martens C**, Shekhar M, Borysik AJ, Lau AM, Reading E, Tajkhorshid E, Booth PJ, Politis A. 2018. Direct protein-lipid interactions shape the conformational landscape of secondary transporters. *Nature Communications* **9**:1–12. DOI: <https://doi.org/10.1038/s41467-018-06704-1>, PMID: 30297844
- Masureel M**, Martens C, Stein RA, Mishra S, Ruyschaert J-M, Mchaourab HS, Govaerts C. 2014. Protonation drives the conformational switch in the multidrug transporter LmrP. *Nature Chemical Biology* **10**:149–155. DOI: <https://doi.org/10.1038/nchembio.1408>, PMID: 24316739
- Mirdita M**, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. 2021. ColabFold - Making protein folding accessible to all. *bioRxiv*. DOI: <https://doi.org/10.1101/2021.08.15.456425>
- Muller MP**, Jiang T, Sun C, Lihan M, Pant S, Mahinthichaichan P, Trifan A, Tajkhorshid E. 2019. Characterization of Lipid-Protein Interactions and Lipid-Mediated Modulation of Membrane Protein Function through Molecular Simulation. *Chemical Reviews* **119**:6086–6161. DOI: <https://doi.org/10.1021/acs.chemrev.8b00608>, PMID: 30978005
- Needleman SB**, Wunsch CD. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology* **48**:443–453. DOI: [https://doi.org/10.1016/0022-2836\(70\)90057-4](https://doi.org/10.1016/0022-2836(70)90057-4), PMID: 5420325
- Nicoludis JM**, Gaudet R. 2018. Applications of sequence coevolution in membrane protein biochemistry. *Biochimica et Biophysica Acta. Biomembranes* **1860**:895–908. DOI: <https://doi.org/10.1016/j.bbmem.2017.10.004>, PMID: 28993150
- Ollikainen N**, Smith CA, Fraser JS, Kortemme T. 2013. Flexible backbone sampling methods to model and design protein alternative conformations. *Methods in Enzymology* **523**:61–85. DOI: <https://doi.org/10.1016/B978-0-12-394292-0.00004-7>, PMID: 23422426
- Ovchinnikov S**, Kinch L, Park H, Liao Y, Pei J, Kim DE, Kamisetty H, Grishin NV, Baker D. 2015. Large-scale determination of previously unsolved protein structures using evolutionary information. *eLife* **4**:e09248. DOI: <https://doi.org/10.7554/eLife.09248>, PMID: 26335199
- Pak MA**, Markhieva KA, Novikova MS, Petrov DS, Vorobyev IS, Maksimova ES, Kondrashov FA, Ivankov DN. 2021. Using AlphaFold to predict the impact of single mutations on protein stability and function. *bioRxiv*. DOI: <https://doi.org/10.1101/2021.09.19.460937>
- Pereira J**, Simpkin AJ, Hartmann MD, Rigden DJ, Keegan RM, Lupas AN. 2021. High-accuracy protein structure prediction in CASP14. *Proteins* **89**:1687–1699. DOI: <https://doi.org/10.1002/prot.26171>, PMID: 34218458
- Ping Y-Q**, Mao C, Xiao P, Zhao R-J, Jiang Y, Yang Z, An W-T, Shen D-D, Yang F, Zhang H, Qu C, Shen Q, Tian C, Li Z, Li S, Wang G-Y, Tao X, Wen X, Zhong Y-N, Yang J, et al. 2021. Structures of the glucocorticoid-bound adhesion receptor GPR97–Go complex. *Nature* **589**:620–626. DOI: <https://doi.org/10.1038/s41586-020-03083-w>, PMID: 33408414
- Roe DR**, Cheatham TE. 2013. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of Chemical Theory and Computation* **9**:3084–3095. DOI: <https://doi.org/10.1021/ct400341p>, PMID: 26583988
- Saldaña T**, Escobedo N, Marchetti J, Zea DJ, Mac Donagh J, Velez Rueda AJ, Gonik E, Melani AG, Novomisky Nechcoff J, Salas MN, Peters T, Demitroff N, Fernandez Alberti S, Palopoli N, Fornasari MS, Parisi G. 2021. Impact of protein conformational diversity on AlphaFold predictions. *bioRxiv*. DOI: <https://doi.org/10.1101/2021.10.27.466189>

- Shaw DE**, Maragakis P, Lindorff-Larsen K, Piana S, Dror RO, Eastwood MP, Bank JA, Jumper JM, Salmon JK, Shan Y, Wrighers W. 2010. Atomic-level characterization of the structural dynamics of proteins. *Science (New York, N.Y.)* **330**:341–346. DOI: <https://doi.org/10.1126/science.1187409>, PMID: 20947758
- Steinberger M**, Söding J. 2017. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology* **35**:1026–1028. DOI: <https://doi.org/10.1038/nbt.3988>, PMID: 29035372
- Tanaka Y**, Hipolito CJ, Maturana AD, Ito K, Kuroda T, Higuchi T, Katoh T, Kato HE, Hattori M, Kumazaki K, Tsukazaki T, Ishitani R, Suga H, Nureki O. 2013. Structural basis for the drug extrusion mechanism by a MATE multidrug transporter. *Nature* **496**:247–251. DOI: <https://doi.org/10.1038/nature12014>, PMID: 23535598
- Tunyasuvunakool K**, Adler J, Wu Z, Green T, Zielinski M, Židek A, Bridgland A, Cowie A, Meyer C, Laydon A, Velankar S, Kleywegt GJ, Bateman A, Evans R, Pritzel A, Figurnov M, Ronneberger O, Bates R, Kohl SAA, Potapenko A, et al. 2021. Highly accurate protein structure prediction for the human proteome. *Nature* **596**:590–596. DOI: <https://doi.org/10.1038/s41586-021-03828-1>, PMID: 34293799
- Wang J**, Hua T, Liu ZJ. 2020. Structural features of activated GPCR signaling complexes. *Current Opinion in Structural Biology* **63**:82–89. DOI: <https://doi.org/10.1016/j.sbi.2020.04.008>, PMID: 32485565
- Wang N**, Jiang X, Zhang S, Zhu A, Yuan Y, Xu H, Lei J, Yan C. 2021. Structural basis of human monocarboxylate transporter 1 inhibition by anti-cancer drug candidates. *Cell* **184**:370–383. DOI: <https://doi.org/10.1016/j.cell.2020.11.043>, PMID: 33333023
- Xu J**, Zhang Y. 2010. How significant is a protein structure similarity with TM-score = 0.5? *Bioinformatics (Oxford, England)* **26**:889–895. DOI: <https://doi.org/10.1093/bioinformatics/btq066>, PMID: 20164152
- Xu L**, Chen B, Schihada H, Wright SC, Turku A, Wu Y, Han G-W, Kowalski-Jahn M, Kozielowicz P, Bowin C-F, Zhang X, Li C, Bouvier M, Schulte G, Xu F. 2021. Cryo-EM structure of constitutively active human Frizzled 7 in complex with heterotrimeric G_s. *Cell Research* **31**:1311–1314. DOI: <https://doi.org/10.1038/s41422-021-00525-6>, PMID: 34239071
- Xue J**, Xie T, Zeng W, Jiang Y, Bai XC. 2020. Cryo-EM structures of human ZnT8 in both outward- and inward-facing conformations. *eLife* **9**:23. DOI: <https://doi.org/10.7554/eLife.58823>, PMID: 32723473
- Yan R.**, Zhao X, Lei J, Zhou Q. 2019. Structure of the human LAT1-4F2hc heteromeric amino acid transporter complex. *Nature* **568**:127–130. DOI: <https://doi.org/10.1038/s41586-019-1011-z>, PMID: 30867591
- Yan R**, Li Y, Müller J, Zhang Y, Singer S, Xia L, Zhong X, Gertsch J, Altmann KH, Zhou Q. 2021. Mechanism of substrate transport and inhibition of the human LAT1-4F2hc amino acid transporter. *Cell Discovery* **7**:16. DOI: <https://doi.org/10.1038/s41421-021-00247-4>, PMID: 33758168
- Yang S**, Wu Y, Xu T-H, de Waal PW, He Y, Pu M, Chen Y, DeBruine ZJ, Zhang B, Zaidi SA, Popov P, Guo Y, Han GW, Lu Y, Suino-Powell K, Dong S, Harikumar KG, Miller LJ, Katritch V, Xu HE, et al. 2018. Crystal structure of the Frizzled 4 receptor in a ligand-free state. *Nature* **560**:666–670. DOI: <https://doi.org/10.1038/s41586-018-0447-x>, PMID: 30135577
- Zakrzewska S**, Mehdipour AR, Malviya VN, Nonaka T, Koepke J, Muenke C, Hausner W, Hummer G, Safarian S, Michel H. 2019. Inward-facing conformation of a multidrug resistance MATE family transporter. *PNAS* **116**:12275–12284. DOI: <https://doi.org/10.1073/pnas.1904210116>, PMID: 31160466
- Zhang Y**, Skolnick J. 2004. Scoring function for automated assessment of protein structure template quality. *Proteins* **57**:702–710. DOI: <https://doi.org/10.1002/prot.20264>, PMID: 15476259
- Zhang Y**, Skolnick J. 2005. TM-align: A protein structure alignment algorithm based on the TM-score. *Nucleic Acids Research* **33**:2302–2309. DOI: <https://doi.org/10.1093/nar/gki524>, PMID: 15849316
- Zhang H**, Chen K, Tan Q, Shao Q, Han S, Zhang C, Yi C, Chu X, Zhu Y, Xu Y, Zhao Q, Wu B. 2021. Structural basis for chemokine recognition and receptor activation of chemokine receptor CCR5. *Nature Communications* **12**:1–12. DOI: <https://doi.org/10.1038/s41467-021-24438-5>, PMID: 34230484
- Zhao L-H**, Ma S, Sutkeviciute I, Shen D-D, Zhou XE, de Waal PW, Li C-Y, Kang Y, Clark LJ, Jean-Alphonse FG, White AD, Yang D, Dai A, Cai X, Chen J, Li C, Jiang Y, Watanabe T, Gardella TJ, Melcher K, et al. 2019. Structure and dynamics of the active human parathyroid hormone receptor-1. *Science (New York, N.Y.)* **364**:148–153. DOI: <https://doi.org/10.1126/science.aav7942>, PMID: 30975883
- Zheng Y**, Han GW, Abagyan R, Wu B, Stevens RC, Cherezov V, Kufareva I, Handel TM. 2017. Structure of CC Chemokine Receptor 5 with a Potent Chemokine Antagonist Reveals Mechanisms of Chemokine Recognition and Molecular Mimicry by HIV. *Immunity* **46**:1005–1017. DOI: <https://doi.org/10.1016/j.immuni.2017.05.002>, PMID: 28636951