Multistate modelling using LIT-AlphaFold

Chiesa L, Urvas L., Bret G., Jacquemard C., Kellenberger E.

UMR 7200 – Centre National de la Recherche Scientifique et Université de Strasbourg,

Institut du Médicament de Strasbourg, Faculté de Pharmacie, 67400 Illkirch-Graffenstaden, France.

AlphaFold [1] has brought a revolution in the field of protein modelling, allowing to predict the structure of a target with near experimental accuracy using only its amino acid sequence as input. AlphaFold however predicts only one conformational state for a given protein, limiting its use since many proteins (G-protein coupled receptors, kinases, ion channels, etc.) are characterized by their ability to adopt different conformations, each with a specific biological function.

Different conformational states can be generated using AlphaFold by subsampling or masking the multiple sequence alignment (MSA) [2]. By reducing the number of sequences used during predictions more diverse structures are generated, leading to a better sampling of the target's conformational space. Additionally, by selecting structural templates in a specific conformational state, it is possible to stir the predictions towards structures in the desired conformation [2].

In this tutorial we will present you AlphaFold2 and its use, as well as specific application aimed at modelling different conformational states of a target protein. To participate to the tutorial click on this link to the Colab Notebook: Link to the tutorial.

All calculations are performed using the LIT-AlphaFold pipeline, which simplifies input generation and customization compared to the original AlphaFold implementation.

About LIT-AlphaFold

Multiple modified pipelines have been developed using MSA subsampling and template selection, each with its own strengths and limitations. These methods however were developed independently, with each one requiring a different installation, different scripts, and different input information, making the comparison and use of multiple methods difficult. To overcome this issue, we developed LIT-AlphaFold, a modified input preparation and prediction pipeline which includes all the basic tools developed for multistate modelling (MSA masking, MSA subsampling, MSA clustering, template selection, dropout, etc.), and extends their use to AlphaFold-Multimer. LIT-AlphaFold allows to easily re-implement known multistate modelling protocols, and to develop new ones.

As a proof of concept LIT-AlphaFold has been used to generate structures of chemokine receptors in apo form (AlphaFold2) and bound to a chemokine (AlphaFold-Multimer), obtaining conformations of the receptor in both active-like and inactive-like states for most targets [3].

How to

Here we include a few general notes that might be useful during the tutorial.

Google Colab

Colab is a hosted Jupyter Notebook service that requires no setup to use and provides free access to computing resources, including GPUs and TPUs. A Google account is required to access the free computing resources of Colab. By using Colab you do not need to install anything on your machine, therefore there should be no issue regardless of the used machine.

py3Dmol

In the tutorial the 3D structure of the generated proteins will be visualized using py3Dmol.

To navigate the images, the following command might be useful:

Rotation: left-click and move the mouse

Translation: center-click or ctrl+left-click, and move the mouse

Zoom: Scrool wheel, or shift+left-click and move the mouse

Show/hide atom labels: single left-click (this feature is not universal and is specific for this tutorial)

LIT-AlphaFold

LIT-AlphaFold stores protein information (multiple sequences alignments and templates) as compressed pickle file (.pkl.bz2). These file are generally read by the LIT-AlphaFold prediction pipeline to perform calculations, or can be used in python scripts using the package litaf.

LIT-AlphaFold can be installed locally or used online via Google Colab from: https://github.com/LIT-CCM-lab/LIT-AlphaFold

Overview of the tutorial

The tutorial explores three detailed examples that highlight how AlphaFold2 impacts target validation, structural plasticity assessment, and multistate modeling. The examples are drawn from recent articles [4,5] and illustrate the practical applications of AlphaFold2 in drug discovery.

Example 1: Segment of a viral polyprotein

In the first example, AlphaFold2 is used to model the replicase encoded by the 1708 amino acid polyprotein of the human-infecting hepatitis E virus (HEV-3). The levels of confidence of the model generated by AlphaFold2 are assessed using the predicted local distance difference test (pLDDT) scores, which provide a measure of the reliability of the predicted structures.

Example 2: SARS-CoV-2 3CL-like proteinase

AlphaFold2's predictions for the 3CL-like proteinase are compared with experimentally determined structures of the protein in its free and ligand-bound forms. The goal is to assess how well AlphaFold2 can capture the dynamic nature of the protein's binding site.

Example 3: Multistate Modeling of β2 Adrenergic Receptor

The third example focuses on the β2 adrenergic receptor, a G-protein coupled receptor used as drug target. GPCRs, including the β2 adrenergic receptor, exist in multiple states, typically active and inactive, which are crucial for their function and for the design of state-specific drugs.

AlphaFold mostly generates for class A GPCRs only conformations in the inactive state. As an example we report the structure of ADRB2. To avoid any bias we removed all ADRB2 structures from the used templates.

References

[1] Jumper J. et al., *Nature 2021, **596**:583-589 [doi: 10.1038/s41586-021-03819-2]*

[2] Sala D. et al., *Current Opinion in Structural Biology 2023, **81**:102645 [doi: 10.1016/j.sbi.2023.102645]*

[3] Urvas L., Chiesa L. et al., *Journal of Chemical Information and Modelling 2024, **64**:4587-4600 [doi: 10.1021/acs.jcim.3c01835]*

[4] Chiesa L., Sick E., Kellenberger E. *Molecular Informatics 2023, **42**:e202300141 [doi: 10.1002/minf.202300141]*

[5] Borkakoti N, Thornton JM, *Current Opinion in Structural Biology, 2023, **78** :102526 [doi: 10.1016/j.sbi.2022.102526]*