

3DSeqCheck: A Web-based Tool for Verifying Sequence Consistency Between a 3D Structure File and the Corresponding UniProt Entry ☆

Anja Conev, Suhail A. Islam, Ifigenia Tsitsa, Alessia David [†] and Michael J. E. Sternberg [†,*]

Centre for Integrative Systems Biology and Bioinformatics, *Department of Life Sciences, Imperial College London, London SW7 2AZ, UK*

Correspondence to Michael J.E. Sternberg: *m.sternberg@imperial.ac.uk (M.J.E. Sternberg)*
https://doi.org/10.1016/j.jmb.2025.169620
Edited by Rita Casadio

Abstract

UniProt is a central repository of protein sequences and annotations, with entries being updated several times a year as new sequencing evidence is collected. By contrast, protein structure resources often evolve at a different pace. The AlphaFold database remained unchanged for four years, until September 2025, during which time nearly 3% of the associated sequences underwent revisions in UniProt. In a range of bioinformatics tasks, protein structure data is paired with sequence annotations from UniProt. Mapping annotations to outdated structure files can lead to errors in downstream analysis. While this concern has been addressed for experimental structures, efforts for the modeled structures are lacking. 3DSeqCheck is a lightweight web tool that enables quick comparison of the sequence of modeled and experimental structures to the latest UniProt entries. 3DSeqCheck provides an interactive visual panel of the alignment and the comparison of the residue numbering and can be accessed freely at: https://missense3d.bc.ic.ac.uk/3dseqcheck and https://github.ic.ac.uk/ImperialCollegeLondon/check3Dseq.
© **2025 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).**

Introduction

DeepMind's AlphaFold2 [1,2] models achieved outstanding accuracy in the CASP14 evaluation of protein structure prediction. In 2022, in collaboration with the European Institute of Bioinformatics (EBI), the AlphaFoldDB (AlphaFold DataBase) was released, providing over 200 million predicted structures [2]. Structure models of an unprecedented number of sequences are now available for researchers to study for a range of applications [3] including function prediction [4], the binding of small molecules [5,6], understanding protein–protein interactions [7] and evaluating mutational effects [8]. Bioinformatics researchers are increasingly connecting their workflows to structural data such as AlphaFoldDB.

We have recently reported a discrepancy between AlphaFoldDB and UniProt that affected around 3% of all proteins in the human proteome, corresponding to 631 AlphaFoldDB entries that were found to be inconsistent with their related UniProt records in release 2025_03, published June 18, 2025 [9]. Out of the 631 inconsistent entries, 116 were made redundant by UniProt, 295 had deletions or insertions in the sequence since AlphaFoldDB deposition and 220 have the same sequence length but some variations in the sequence. The same issue persists across different species, with around 2.5% of the mouse proteome in AlphaFoldDB being inconsistent with the UniProt sequences. We highlighted the impact of sequence mismatch on the mapping of functional annotations,

☆ This article is part of a special issue entitled: 'Computation Resources 2026' published in Journal of Molecular Biology.

0022-2836/© 2025 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

such as the location of transmembrane domains and missense variants.

In recent months, the synchronization of UniProt and AlphaFoldDB has been addressed on several levels. First, the UniProt [10] and SwissMODEL [11] platforms displayed warnings when users selected outdated AlphaFoldDB models on their platforms. Similarly, AlphaFoldDB at first provided warnings on its site. Finally, in September 2025, the AlphaFoldDB announced a new release of the models, synced with UniProt version 2025_03. While the latest release AlphaFoldDB provides up to date models, it is not clear how frequent releases will be in the future. Resources that were built on top of the old AlphaFoldDB models such as AlphaFill and AlphaMissense currently remain outdated. The need to keep the modeled structures up to date with UniProt entries extends beyond the AlphaFoldDB. The 3D-Beacons [12] initiative, launched in 2022, allows users to fetch structure models corresponding to UniProt accession codes from various resources including the AlphaFoldDB [2], HegeLab.org [13], Isoform.io [14], ModelArchive [15], LevyLab [16], AlphaFill [17], SWISS-MODEL [11]. Many of these models may require checking synchronization with UniProt before use.

Experimentally resolved structures have also been subject to the issue of integrating function annotations between the structure and the UniProt entries. For experimental structures, a substantial effort was undertaken by the community to provide such mappings. Structure Integration with Function, Taxonomy and Sequences (SIFTS) resource [18–20] has addressed this concern and has been maintained by UniProt, PDBe and EBI since 2002. SIFTS provides residue mappings such that other downstream services can renumber the residues in the structure according to the given UniProt entry. This is important because the residue identifier in a PDB file is simply a label assigned by the authors of the file and there is no requirement for it to correspond to the UniProt numbering and associated functional annotations. Other resources, such as PDBRenum [21], have been built to integrate SIFTS annotations and mappings. As part of their integrated sequence-structure viewer, the RCSB-PDB [22] shows the sequence from the structure aligned to the corresponding UniProt sequence using the SIFTS mappings. However, SIFTS infrastructure and the related tools only handle experimental structures.

In outline, 3DSeqCheck provides a simple entry. The user inputs the UniProt identifier for the protein whose sequence they wish to compare (Figure S1). They can compare models from different structure databases or upload their own structure file. The results page of 3DSeqCheck shows the alignment of the UniProt sequence to that extracted from the structure. It provides details of the alignment on two levels. First, it provides a detailed view of the sequence alignment between the UniProt sequence and structure (Figure 1). Regions that are missing or added in structure, as well as regions that are different in the two sequences, are highlighted and an overall match percentage is provided. Second, it compares the residue identifiers inside the structure to those in the UniProt sequence (Figure 2). Each aligned residue is presented in a table with columns containing UniProt identifiers as well as the residue identifiers of the structure. Users are presented with the overall percentage of residues that match in identifiers, and they can download a renumbered file as well as the results in both TSV and JSON formats.

3DSeqCheck can be accessed at https://missense3d.bc.ic.ac.uk/3dseqcheck, where it ensures reliable sequence–structure correspondence across diverse modeling resources, helping researchers avoid annotation errors and make more confident use of structural data in their analyses. In addition, the source code is available for download on GitHub at https://github.ic.ac.uk/ImperialCollegeLondon/check3Dseq.

3DSeqCheck user pipeline

Input description

The 3DSeqCheck submission form is displayed on the main page for users to interact with the app (Figure S1). Here, the user is prompted to input the data for the structure they want to validate.

UniProt ID is the first required input to 3DSeqCheck. The UniProt ID should correspond to the entry that your structure describes, it is used to fetch the most up-to-date sequence that is used as a *reference sequence* in the 3DSeqCheck. Structure Source is the second required input to 3DSeqCheck. The Structure Source should correspond to the structure the user wants to validate. The *query sequence* will be extracted from this given structure. There are currently two supported ways to supply a structure:

- *From a Database* – a structure will be fetched from the corresponding database (options include: AlphaFoldDB, AlphaFill, ModelArchive, Isoform.io, Protein Ensemble Database, HegeLab, LevyLab, PDBe, SWISS-MODEL) using the 3DBeacons API [12].
- *Custom PDB File* – by choosing a custom file as source, the uploaded file will be used as input. In addition to the PDB file, the user needs to provide the chain ID for the chain corresponding to the supplied UniProt ID.

Note that 3DSeqCheck imposes limits to the input size to prevent excessive computational overhead. Namely, custom PDB files are restricted to a maximum size of 5 MB. In addition, AlphaFoldDB entries containing more than 2700 amino acids are not supported for comparison as they are split

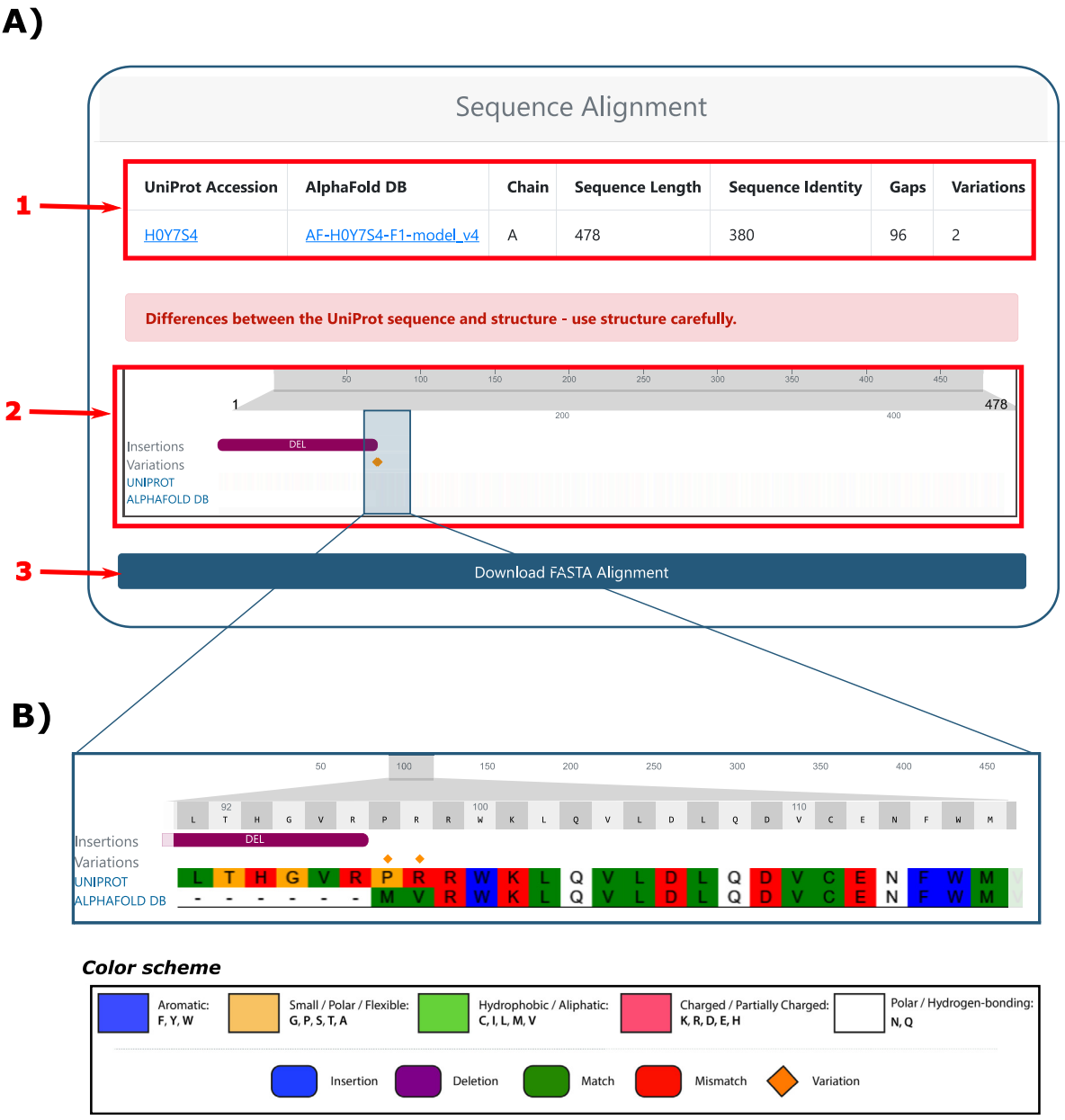


Figure 1. Check3DSeq Results Page – Sequence Alignment. (A) Alignment results between the structure and UniProt sequences. (1) Summary table; (2) alignment panel; (3) button for downloading the alignment in a FASTA format. (B) Zoomed in perspective of the alignment segment between residue position 91 and residue position 116 in the UniProt sequence. The deleted region is shown with a purple round rectangle track above the sequences, while the variations are depicted in the orange diamond track. The color scheme for the sequence and the tracks is shown below the alignment.

across multiple files and are not unavailable through the AlphaFoldDB API. Users are presented with a warning message if their inputs exceed these limits.

Extracting the sequences

To obtain the up-to-date *reference sequence*, 3DSeqCheck queries the UniProt REST API [23] using the input UniProt ID. The *sequence of the query structure* is obtained by extracting it from the structure source by identifying the residues that are in the 'ATOM' section of the structure file (with the matching chain ID, in the case of custom file input).

Alignment

The pairwise sequence alignment of the reference and query sequences is performed using the implementation of the Needleman-

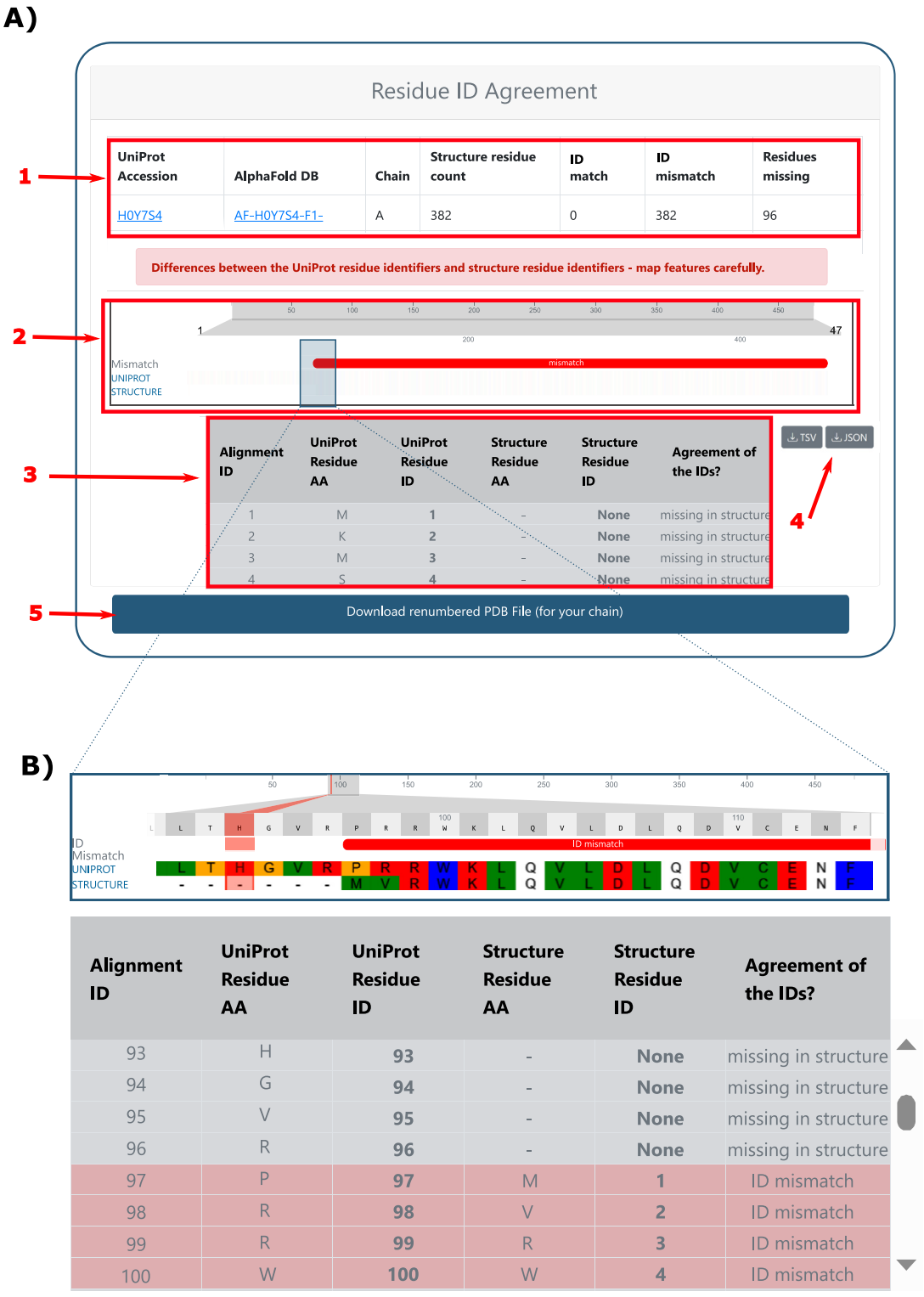


Figure 2. Check3DSeq Results Page – Residue Identifier Agreement. (A) Residue ID comparison between the structure and UniProt sequences. (1) Summary table; (2) alignment panel; (3) detailed residue table; (4) buttons for downloading the results in both TSV and JSON formats. (5) Button for downloading the renumbered structure. (B) Zoomed in perspective of the alignment segment between residue position 91 and residue position 114 in the UniProt sequence. Residues where the structure and UniProt IDs are different are marked in the red rectangle track above the sequences. Detailed residue table for residues 93–100 is displayed below the alignment and highlights residues where the IDs in UniProt and in the structure do not match.

Wunsch algorithm provided in BioPython's Align module [24] with the default parameters.

Output visualization

Upon submission of the input form, the user is presented with a Results page which displays: (1) a concise overview of the alignment between the reference sequence and the amino acid sequence derived from the structure file (Figure 1), and (2) a detailed residue-level analysis, tabulating the type and extent of discrepancies identified (Figure 2).

The results of the concise analysis are displayed at the top of the Results page in a summary table (Figure 1A) which presents:

- length of aligned sequences
- sequence identity between the query sequence (structure) and the reference sequence (UniProt)
- number of gaps (insertions/deletions to the reference UniProt sequence as compared to the query sequence) in the alignment
- number of variations (different amino acids between the reference UniProt sequence and the query sequence) in the alignment

The match assessment message is displayed below the summary table and contains two possible outcomes:

- *Your structure matches the UniProt sequence, safe to use!* is displayed when the sequence identity is 100% and the UniProt sequence fully matches the sequence in the structure. This corresponds to a perfect alignment, and the recommendation is to freely use the UniProt annotations for the supplied UniProt ID on this structure.
- *Differences between the UniProt sequence and structure – use structure carefully.* is displayed in case the sequence identity is less than 100%. This corresponds to an alignment of varying quality due, for example, to the presence of gaps, insertions or residue mismatch. The recommendation is to use the UniProt annotations carefully for this structure.

The alignment is presented in an interactive dashboard using the Nightingale web components [25], and the insertions/deletions/variations are displayed in separate tracks on top of the alignment (Figure 1B). FASTA file contents are available for download.

The second section of the Results page presents the detailed analysis of the agreement between the identifiers of residues in the structure and the UniProt identifiers (Figure 2A). Structure residue identifiers are contained in columns 23–26 of the ATOM lines of the PDB file. These identifiers, along with the chain identifiers are used to identify parts of the structure that the user is interested in. UniProt residue identifiers are associated with the position of the residue in the UniProt sequence (e.g., the first residue has the identifier 1, the tenth residue identifier 10, etc.). In the second section of the Results page, the agreement between these identifiers is inspected in detail. The following information is included in a summary table:

- number of residues in the structure
- number of residues in the structure for which the identifiers match the UniProt sequence identifiers
- number of residues in the structure for which the identifiers do not match the UniProt sequence identifiers
- number of residues that are missing from the structure and are present in the UniProt sequence

The match assessment message is displayed below the summary table and contains two possible assessments:

- *Your structure matches the UniProt sequence identifiers, safe to map features!* is displayed in case all the residue identifiers of the structure match all the identifiers of the UniProt sequence. This corresponds to a perfect identifier match, and the recommendation is to freely use the UniProt annotations for those residues that are not missing in the structure.
- *Differences between the UniProt residue identifiers and structure residue identifiers – map UniProt annotations carefully.* is displayed in case the match between identifiers is not perfect. The recommendation is to use the UniProt annotations carefully for the supplied UniProt ID on this structure.

ID match/mismatch labels are also visually displayed in separate tracks on top of the alignment in the interactive dashboard (Figure 2B).

The detailed residue table (Figure 2B) shows the following information for each residue in the alignment result:

- Alignment ID – residue identifier in the alignment (starting from 1)
- UniProt residue AA – amino acid at that position of the UniProt sequence
- UniProt residue ID – identifier in the UniProt sequence (starting from 1 of the first residue in the UniProt sequence)
- Structure residue AA – amino acid on that position in the structure file
- Structure residue ID – identifier in the structure file (extracted from the ATOM lines of the structure file)
- Label – *ID match*/*ID mismatch*/*missing*

Insertions are defined as residues present in the experimental structure but absent from the UniProt reference sequence, while deletions correspond to UniProt residues missing from the structure. Inserted residues are excluded from the renumbered structure to preserve a one-to-one mapping with UniProt, and users are warned when this results in the removal of residues from the original structure. The table with residue

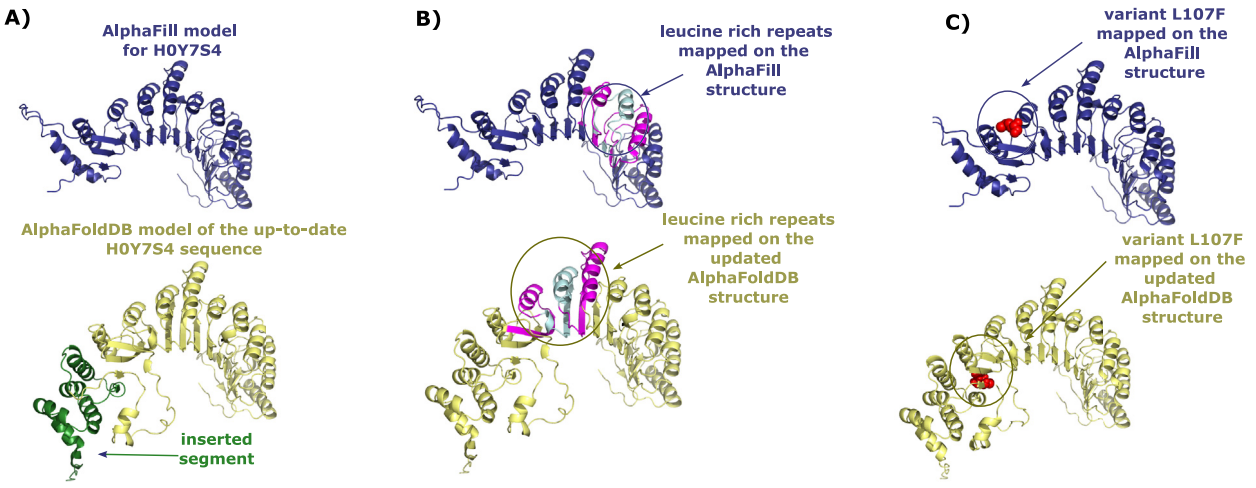


Figure 3. PRAMEF26 (H0Y7S4) example. (A) AlphaFill model (blue) and up-to-date AlphaFoldDB model (yellow). The inserted segment is highlighted in green. (B) AlphaFill and up-to-date AlphaFoldDB models with LRR domains highlighted. (C) AlphaFill and up-to-date AlphaFoldDB models with the L107F variant highlighted.

mappings is available for download in TSV and JSON formats. The renumbered structure file for the selected chain is available for download in PDB format.

Example

PRAME family member 26 (UniProt ID: H0Y7S4)

PRAME (PReferentially expressed Antigen of MElanoma) is a protein family of leucine-rich repeat (LRR) proteins expressed in several cancer cells and recognized by cytolytic T lymphocytes [26]. PRAMEF26, UniProt entry H0Y7S4, is a well characterized member of the PRAME family, and represented in AlphaFill by an outdated sequence model. Running entry H0Y7S4 through the 3DSeqCheck, we observe that the current UniProt entry contains 478 amino acids, while the AlphaFill structure contains 382 with a region of 96 consecutive amino acids deleted in the C-terminus of the structure, with respect to the current UniProt entry (Figure 1A). In addition, there are two amino acids at the beginning of the AlphaFill model that differ from the UniProt sequence (Figure 1B). Due to the deletion at the start of the sequence, the residue identifiers in the AlphaFill structure (Figure 2A) are in complete disagreement with the UniProt identifiers.

In this example, we showcase the problems users may encounter when using an outdated AlphaFill structure in a widely used molecular graphics program such as PyMol [27]. We compare the AlphaFill model with the up-to-date AlphaFoldDB model of the sequence corresponding to the latest UniProt entry (Figure 3A). The recent UniProt sequence update introduces an additional alpha-helical domain at the N terminus of PRAMEF26 structure (Figure 3A, highlighted green). Furthermore, the deletion causes the sequence annotations to be shifted by 96 amino acids (Figure 2B). This shift affects the annotation of LLR (Figure 3C) domains as well as the numbering of the leucine at position 107, which harbors the amino acid substitution p.L107F. In the out-of-date AlphaFill model, leucine 107 corresponds to position 11 and position 107 is occupied by a methionine (Figure 3D). Discrepancy in the AlphaFill model and the UniProt sequence would go unnoticed by molecular graphics programs such as PyMol [27] or Chimera [28] in which the structural features (i.e., variants, domains, binding sites) are easily highlighted. Furthermore, it can lead to inaccuracies when using downstream structure-based algorithms such as Missense3D [8], potentially misrepresenting the functional or structural impacts of variants. Addressing these issues is critical for maximizing the utility of structure models and ensuring their integration into downstream tools and resources.

Conclusion

3DSeqCheck is a web application designed for rapid and easy validation of the consistency of a protein structure against the latest UniProt entries. The interactive sequence alignment overview and residue-by-residue identifier comparison present 3DSeqCheck as a resource for the community using 3D structures in downstream analyses.

With the example of PRAMEF26 protein (UniProt entry H0Y7S4) we show how discrepancies between UniProt sequences and corresponding protein structures pose risks on two levels. First, a structure file may contain a different sequence, or only a fragment of the query UniProt sequence. Users may draw wrong conclusions about the protein, as structural domains could be inaccurately represented or incompletely modeled. Second, even when the sequences are in

agreement, there is a risk that residue identifiers used in the structure file do not agree with the identifiers within the UniProt sequence. Users risk mapping annotations from UniProt to the wrong portions of the structure. 3DSeqCheck allows users to interactively inspect such residue identifier disagreement in the bottom panel of the results page (Figure 2).

The potential discrepancies between a structural file and UniProt residue identifiers could result in errors when using molecular graphics programs such as PyMOL[27] and Chimera[28] where residue identifiers are used to locate and highlight positions of particular features (e.g., variants, binding sites, post-translational modifications). 3DSeqCheck is useful in these cases as it allows users to download a renumbered structure file that corresponds to the UniProt sequence.

3DSeqCheck highlights and explains the underlying differences between the modeled structure and the UniProt entry. Nevertheless, correct mapping remains critical—for example, when investigating the structural context of specific variants, since even small inconsistencies can compromise biological interpretation. While the latest release of AlphaFoldDB provides up-to-date models, it is not clear how frequent these updates will be in the future; other sources of models accessible through 3DBeacons, such as AlphaFill, ModelArchive and Isoform.io, suffer from the same synchronization issue. 3DSeqCheck can be applied to a structure file from any source, thereby offering broader utility across modeling resources and ensuring greater reliability in the integration of structural data with sequence-based annotations, and thus it facilitates a more accurate use of protein structure data.

As protein structure modeling becomes more accurate, the models are increasingly being integrated into a variety of different bioinformatics pipelines. To ensure the reliability of downstream analyses, it is essential that modeled structures remain consistent with major reference databases and tools [9]. 3DSeqCheck is a contribution to the software ecosystem allowing for robust integration of protein structure data with functional annotations.

CRediT authorship contribution statement

Anja Conev: Writing – review & editing, Writing – original draft, Software, Conceptualization. **Suhail A. Islam:** Writing – review & editing, Software, Investigation. **Ifigenia Tsitsa:** Writing – review & editing, Investigation. **Alessia David:** Writing – review & editing, Supervision, Funding acquisition. **Michael J.E. Sternberg:** Writing – review & editing, Supervision, Funding acquisition.

Funding

AC and IT were supported by a Wellcome Trust grant 218242/Z/19/z. AD was supported by an MRC grant MR/Y031091/1. SAI and MJES were supported by Biotechnology and Biological Sciences Research Council (BBSRC) grants BB/T010487/1 and BB/V018558/1. This research was funded in part by the Wellcome Trust [Grant number 218242/Z/19/z]. For the purpose of open access, the authors have applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

DATA AVAILABILITY

Code on GitHub. web server freely available.

DECLARATION OF COMPETING INTEREST

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Anja Conev, Ifigenia Tsitsa reports financial support was provided by Wellcome Trust. Alessia David reports financial support was provided by Medical Research Council and holds a Wellcome Trust grant. Michael J.E. Sternberg reports financial support was provided by Biotechnology and Biological Sciences Research Council and holds a Wellcome Trust grant. Suhail Islam reports financial support was provided by Biotechnology and Biological Sciences Research Council. Michael J.E. Sternberg reports a relationship with DeepMind that includes: funding grants. MJES is a member of the editorial board of JMB and receives an honorarium from Elsevier. DeepMind, who generated the structure in the AlphaFold database, have provided funding for studentships at Imperial College London, which could have included a Masters' course of which MJES is the Director. The Authors declare no other competing interests. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank Ryan Pye for helpful discussion.

Appendix A. Supplementary material

Supplementary material to this article can be found online at https://doi.org/10.1016/j.jmb.2025.169620.

Received 30 October 2025;
Accepted 26 December 2025;
Available online 30 December 2025

Keywords:
UniProt;
AlphaFoldDB;
protein structure;
molecular graphics;
structure check

† Joint last authors.

References

[1]. Jumper, J. et al, (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* **596** (7873), 7873. https://doi.org/10.1038/s41586-021-03819-2.

[2]. Varadi, M. et al, (2022). AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50** (D1), D439–D444. https://doi.org/10.1093/nar/gkab1061.

[3]. David, A., Islam, S., Tankhilevich, E., Sternberg, M.J.E., (2022). The AlphaFold database of protein structures: a biologist's guide. *J. Mol. Biol.* **434**, (2)167336 https://doi.org/10.1016/j.jmb.2021.167336.

[4]. Lee, D., Redfern, O., Orengo, C., (2007). Predicting protein function from sequence and structure. *Nature Rev. Mol. Cell Biol.* **8** (12), 995–1005. https://doi.org/10.1038/nrm2281.

[5]. Eberhardt, J., Santos-Martins, D., Tillack, A.F., Forli, S., (Aug. 2021). AutoDock Vina 1.2.0: new docking methods, expanded force field, and Python bindings. *J. Chem. Inf. Model.* **61** (8), 3891–3898. https://doi.org/10.1021/acs.jcim.1c00203.

[6]. Conev, A., Chen, J., Kavraki, L.E., (2025). DINC-ensemble: a web server for docking large ligands incrementally to an ensemble of receptor conformations. *J. Mol. Biol.* **437**, (15) 169163 https://doi.org/10.1016/j.jmb.2025.169163.

[7]. Malladi, S. et al, (2022). GWYRE: a resource for mapping variants onto experimental and modeled structures of human protein complexes. *J. Mol. Biol.* **434**, (11)167608 https://doi.org/10.1016/j.jmb.2022.167608.

[8]. Pennica, C., Hanna, G., Islam, S.A., Sternberg, M.J.E., David, A., (2023). Missense3D-PPI: a web resource to predict the impact of missense variants at protein interfaces using 3D structural data. *J. Mol. Biol.* **435**, (14)168060 https://doi.org/10.1016/j.jmb.2023.168060.

[9]. Tsitsa, I., Conev, A., David, A., Islam, S.A., Sternberg, M. J.E., (2025). The aging of the AlphaFold database. *Nature Struct. Mol. Biol.* **32**, 2374–2376. https://doi.org/10.1038/s41594-025-01725-z.

[10]. The UniProt Consortium, (2023). UniProt: the universal protein knowledgebase in 2023. *Nucleic Acids Res.* **51** (D1), D523–D531. https://doi.org/10.1093/nar/gkac1052.

[11]. Waterhouse, A. et al, (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46** (W1), W296–W303. https://doi.org/10.1093/nar/gky427.

[12]. Varadi, M. et al, (2022). 3D-beacons: decreasing the gap between protein sequences and structures through a federated network of protein structure data resources. *Gigascience* **11**, giac118. https://doi.org/10.1093/gigascience/giac118.

[13]. Tordai, H., Suhajda, E., Sillitoe, I., Nair, S., Varadi, M., Hegedus, T., (2022). Comprehensive collection and prediction of ABC transmembrane protein structures in the AI era of structural biology. *Int. J. Mol. Sci.* **23** (16), 8877. https://doi.org/10.3390/ijms23168877.

[14]. Sommer, M.J. et al, (2022). Structure-guided isoform identification for the human transcriptome. *eLife* **11**, e82556. https://doi.org/10.7554/eLife.82556.

[15]. Tauriello, G. et al, (2025). ModelArchive: a deposition database for computational macromolecular structural models. *J. Mol. Biol.* **437**, (15)168996 https://doi.org/10.1016/j.jmb.2025.168996.

[16]. Dey, S., Ritchie, D.W., Levy, E.D., (2018). PDB-wide identification of biological assemblies from conserved quaternary structure geometry. *Nature Methods* **15** (1), 67–72. https://doi.org/10.1038/nmeth.4510.

[17]. Hekkelman, M.L., de Vries, I., Joosten, R.P., Perrakis, A., (2023). AlphaFill: enriching AlphaFold models with ligands and cofactors. *Nature Methods* **20** (2), 205–213. https://doi.org/10.1038/s41592-022-01685-y.

[18]. Dana, J.M. et al, (2019). SIFTS: updated Structure Integration with Function, Taxonomy and Sequences resource allows 40-fold increase in coverage of structure-based annotations for proteins. *Nucleic Acids Res.* **47** (D1), D482–D489. https://doi.org/10.1093/nar/gky1114.

[19]. Velankar, S. et al, (2013). SIFTS: structure integration with function, taxonomy and sequences resource. *Nucleic Acids Res.* **41** (D1), D483–D489. https://doi.org/10.1093/nar/gks1258.

[20]. Velankar, S. et al, (2005). E-MSD: an integrated data resource for bioinformatics. *Nucleic Acids Res.* **33** (suppl_1), D262–D265. https://doi.org/10.1093/nar/gki058.

[21]. Faezov, B., Jr, R.L.D., (2021). PDBrenum: a webserver and program providing protein data bank files renumbered according to their UniProt sequences. *PLoS One* **16**, (7)e0253411 https://doi.org/10.1371/journal.pone.0253411.

[22]. Burley, S.K. et al, (2021). RCSB protein data bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Res.* **49** (D1), D437–D451. https://doi.org/10.1093/nar/gkaa1038.

[23]. Ahmad, S. et al, (2025). The UniProt website API: facilitating programmatic access to protein knowledge. *Nucleic Acids Res.* **53** (W1), W547–W553. https://doi.org/10.1093/nar/gkaf394.

[24]. Cock, P.J.A. et al, (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25** (11), 1422–1423. https://doi.org/10.1093/bioinformatics/btp163.

[25]. Salazar, G.A. et al, (2023). Nightingale: web components for protein feature visualization. *Bioinform. Adv.* **3**, (1) vbad064 https://doi.org/10.1093/bioadv/vbad064.

[26]. Hermes, N., Kewitz, S., Staege, M.S., (2016). Preferentially expressed antigen in melanoma (PRAME) and the PRAME family of leucine-rich repeat proteins. *Curr. Cancer Drug Targets* **16** (5), 400–414. https://doi.org/10.2174/1568009616666151222151818.

[27]. Pymol: an open-source molecular graphics tool – ScienceOpen. Accessed: Sep. 22, 2025. [Online]. Available: https://www.scienceopen.com/book?vid=4362f9a2-0b29-433f-aa65-51db01f4962f.

[28]. Pettersen, E.F. et al, (2004). UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25** (13), 1605–1612. https://doi.org/10.1002/jcc.20084.