Rapid and Reliable Structural Modeling of Adaptive Immune Receptors Using an Optimized AlphaFold3 workflow

Alexandre Jann[1,2], Marta A. S. Perez[1,2,*] and Vincent Zoete[1,2,*]

[1] Ludwig Institute for Cancer Research, Lausanne Branch, Department of Oncology, Lausanne University Hospital (CHUV) and University of Lausanne (UNIL), Agora Cancer Research Center, Lausanne, Switzerland.

[2]Molecular Modeling Group, SIB Swiss Institute of Bioinformatics, University of Lausanne, Quartier UNIL-Sorge, Bâtiment Amphipole, CH-1015 Lausanne, Switzerland.

*Correspondence to: marta.perez@sib.swiss, vincent.zoete@unil.ch

Abstract

AlphaFold (AF), a deep-learning based protein modelling approach, has revolutionized structural biology by generally achieving near-experimental accuracy in protein structure prediction. An impactful application of AF is the modelling of antibodies (Abs) and T-cell receptors (TCRs), key mediators of cellular immunity, whose structural specificity underlies responses in cancer, infection, and autoimmune diseases. In this work, we analyse AF3 performance by systematically examining how MSA composition and the number of inference phases affect prediction accuracy and computational efficiency. Using reduced UniRef90 subsets, we present a ~45-fold accelerated and highly accurate variant of the AF3 workflow specifically optimized for the modelling of the Abs and TCRs receptor domains, enabling rapid, reliable structural predictions at a scale suitable for high-throughput immunological studies. Our findings provide a foundation for faster therapeutic discovery and deeper molecular mechanism understanding of immune recognition.

Introduction

The ability to accurately predict protein structures has been revolutionized by AlphaFold (AF)[1], a deep learning (DL) framework that generally demonstrates near-experimental accuracy across most of the human proteome. AlphaFold3 (AF3)[2], released in 2024, is the most recent and advanced version of AF for biomolecular structure prediction and interaction analysis. Unlike AlphaFold2 (AF2)[3], AF3 significantly reduces the reliance on Multiple Sequence Alignment (MSA) representations. While MSA information is still used, its role is greatly diminished in favor of single and pair representations, which helps to simplify the model's architecture. The single representation refers to per-residue or per-atom embeddings that capture local biochemical properties. The pair representation encodes features between every pair of residues or atoms, capturing spatial proximity, geometric constraints, and co-evolutionary signals. AF3 adopts new components, such as a Pairformer and a Diffusion-based Generative Model, to replace the Evoformer stack used in AF2. These new modules are responsible for updating both pairwise and single representations in a more flexible and generalizable manner, allowing AF3 to support heterogeneous biomolecular assemblies.

Despite the reduced influence of the MSA on model construction in AF3, the MSA component remains the slowest step in the workflow, representing more than 90% of the end-to-end execution time. Consequently, MSA generation imposes a disproportionate runtime and emerges as a bottleneck in AF3, notably for studies requiring the generation of large numbers of 3D structures.

T-cell Receptors (TCRs) and Antibodies (Abs) are key mediators of cellular immunity, whose structural specificity underlies responses in cancer, infection, and autoimmune diseases. TCRs and Abs exhibit exceptional sequence diversity and flexible regions, posing challenges in the creation of 3D models. Abs and TCRs are distinct from other proteins due to their conserved structure in the framework region but large structure variability in the Complementarity-Determining Regions (CDR) regions. Our group develops 3D-

based predictors of TCR and Ab repertoires analysis and specificity prediction[4, 5], which rely on the rapid modeling of large numbers of receptor domains.

In this paper, we conducted a comprehensive performance analysis of AF3 for modelling TCRs and Abs. We focus on two phases of the AF3 workflow: (i) the MSA phase, which prepares evolutionary features from homologous sequences, and (ii) the inference phase, which performs structure prediction using deep neural network modules, primarily the Pairformer and Diffusion module, to iteratively update pair representations that capture atomic interaction relationships. By benchmarking end-to-end execution time across multiple AF3 workflow variants, with different MSA databases and number of seeds, we demonstrate that the AF3 efficiency can be improved by using subsets of the UniRef90 database [6] in the MSA phase. On our desktop system, execution time is ~45 fold accelerated while keeping the TCRs and Abs models at near-experimental accuracy. Developing specialized AF3 variants that retain AF's predictive accuracy while making it extremely fast is critical for advancing structural immunology and enabling large-scale modeling of immune repertoires. Our approach could be readily generalized to other protein families with large numbers of experimentally determined 3D structures in the Protein Data Bank (PDB), such as kinases or proteases.

Results

In this results section, we present AF3 workflow variants specifically developed for the modelling of the receptor domain of the Abs and TCRs. We show that the AF3 end-to-end runtime can be substantially reduced by generating a curated subset of UniRef90 sequences, rather than relying on the full databases used by AF3. Using the AF3 ranking score, we further demonstrate that a single seed is sufficient to obtain highly accurate models, eliminating the need for multiple runs. Notably, we show that our AF3 workflow

variant achieves near-experimental accuracy, even for challenging CDRs such as antibody CDRH3 and T-cell receptor CDR3β, while relying solely on the reduced database and a single seed.

Less than 3% of the UniRef90 sequences are used in AF3 MSA for TCR and Ab modelling

AF uses multiple sequence alignments (MSAs) to extract information from evolutionarily related sequences corresponding to the input proteins. Homologous sequences are retrieved for each protein chain in a very time-consuming process, reading sequences from huge databases such as UniRef90, which contains 153,742,194 sequences, to ensure general applicability. AF3 generates deep, high-quality MSAs using JackHMMER[7].

To evaluate the Uniref90 sequences selected by AF3 for the modelling of TCRs we used a non-redundant set of 3'213 TCRs from the VDJ database [8, 9] (**Methods**). By sequentially generating MSAs for individual TCRs and aggregating the resulting sequences, we found that the size of the cumulative sequence database converges after roughly 1000 TCRs, becoming sufficient to reproduce the MSAs of all subsequent TCRs (**Figure 1 (A-C)**). Across the entire VDJ set, 261,055 sequences are used for TCRα (**Figure 1A**) and **Table 1**), 208,047 for TCRβ (**Figure 1B)** and **Table 1**), and 261,780 for TCRαβ (**Figure 1 C)**) and **Table 1**), indicating that the α chain incorporates more UniRef90 sequences than the β chain, with substantial overlap between the sequences used for both chains. Less than 3% of the UniRef90 sequences are used in the MSA for the entire set suggesting that the size of the reference database used for MSA could be substantially reduced if AF3 is applied to this specific protein family. These sequences were compiled into a new database called UniRef-TCR.

We also performed the same exercise for a set of 152 reference Abs and aggregated a total of 17,887 unique sequences from the UniRef90 database after sequentially treating the heavy chain, 18,312 for the light chain and 35,607 for both heavy and light chains. In contrast with TCRs MSA, only a very small overlap of 592 sequences was found between the sequence datasets related to heavy and light chains, resulting in a total of

35,607 sequences necessary to treat both chains (**Table 1)**. These sequences were aggregated into a new database called UniRef-Ab. The convergence of sequence accumulation during the sequential processing of antibodies, which supports the sufficiency of the final sequence datasets, could in principle depend on the order in which antibodies are analyzed. To assess if convergence is robust to Abs processing order, we performed 100 accumulation runs, each initiated from a different random shuffling of the reference antibody set. The resulting standard deviation of the cumulative number of sequences as a function of the number of reference Abs processed is shown in **Figure 1 (D–F)**. These results demonstrate that convergence is independent of antibody processing order, supporting the comprehensiveness of the final sequence datasets. We expect this behavior to extend to the construction of the TCR datasets as well. Notably, the accumulation analysis (**Figure 1 (D–F)**) reveals that, on average, approximately 80% of all unique sequences are already captured after processing the first five reference Abs, indicating substantial sequence sharing across multiple Abs.

Another new database, UniRef-TextMining, containing 41,055 sequences, was generated by text-mining the UniRef90 database for terms associated with TCRs or Abs.

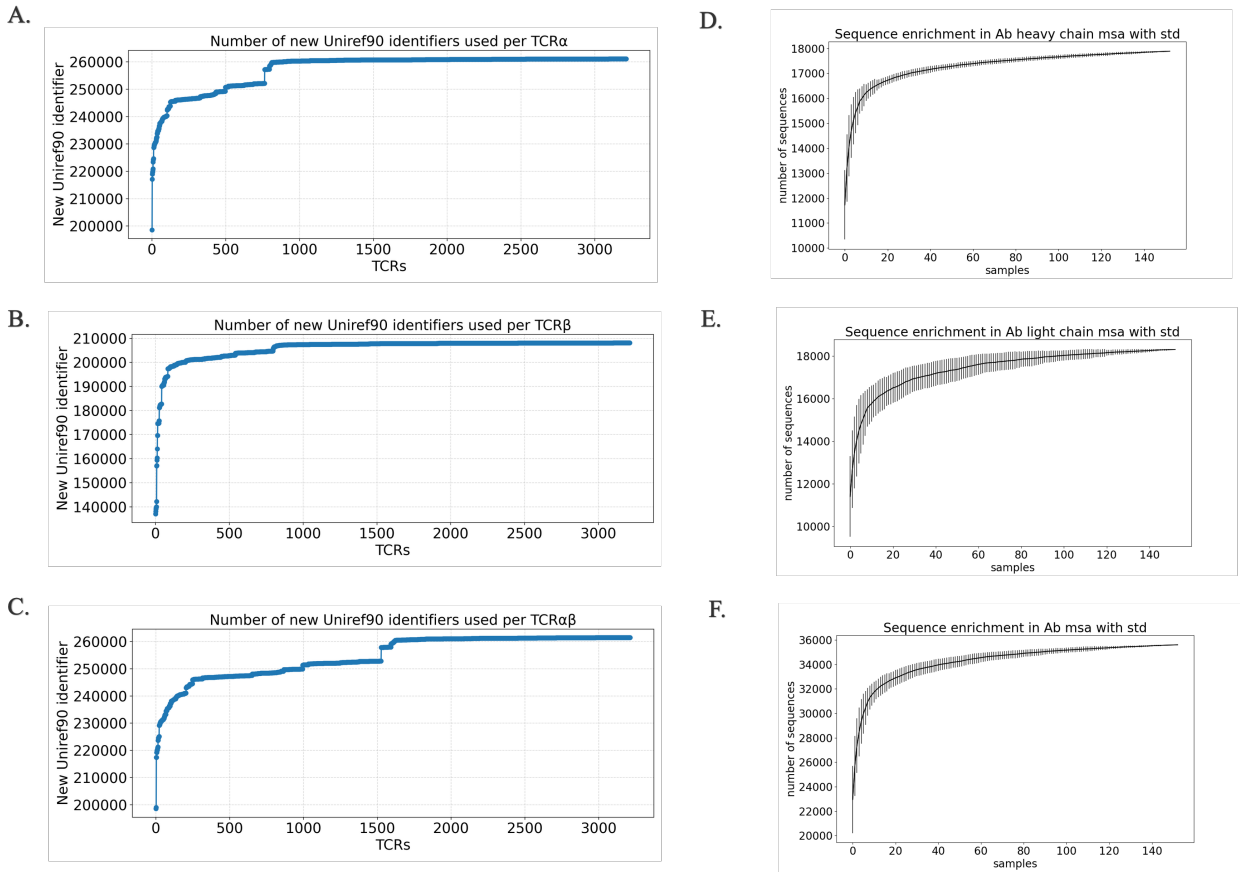


Figure 1. Convergence of the aggregated MSA-derived sequence databases. (A-C) Cumulative number of unique sequences obtained from UniRef90 by sequentially generating multiple sequence alignments (MSAs) for individual T cell receptors (TCRs) and aggregating the resulting sequences, plotted as a function of the number of TCRs processed, for a set 3,213 unique TCRs from VDJ database. A) and B) show analyses for TCR α chains and β chains, respectively, and C) present analyses of paired αβ TCRs. (D-F) Sequence accumulation obtained from the sequential MSA analysis of a set of 152 reference Abs, and respective standard deviation computed over 100 random shuffling of the order in which the antibodies were processed. (D) corresponds to the Ab's heavy chain MSA analysis, (E) to Ab's light chain and F) Ab's heavy and light chain together.

Table 1. Number of UniRef90 sequences in the AF3 Uniref90 database, number of sequences from UniRef90 used either in the MSA of TCRα, TCRαβ and TCRαβ paired for a set of 3,213 unique TCRs

126 from VDJ database or used in the MSA of AbH, AbL and AbHL paired for a set of 152 Abs. Number of
127 sequences in a non-exhaustive database containing just chains belonging to TCRs and Abs.

Description	**Number of sequences**
UniRef90-AF3	153,742,194 (>100 million)
TCRαβ MSA , UniRef90-TCR	261,780 (249,782) *
TCRα MSA	261,055
TCRβ MSA	208,047
AbH MSA	17,887
AbL MSA	18,312
AbHL MSA, UniRef90-Ab	35,607
UniRef-textmining	41,055

128

129 These specialized databases, UniRef-TCR, UniRef-Ab and UniRef-textminining were used as the only
130 databases to generate MSA instead of the databases usually used during a normal AF3 run . The four
131 databases used by AF3 to gather homologous sequences for building the MSA are (i) UniRef90, a clustered
132 version of UniProt with sequences merged at 90 % identity for broad evolutionary coverage (described as
133 UniRef90-AF3 in **Table1**), (ii) MGnigy, large metagenomic sequence database that greatly increases depth
134 of evolutionary information, (iii) BFD (Big Fantastic Database), one of the largest collection of protein
135 families and (iv) the full UniProt protein sequence database[3]. AF3 can also generate MSAs for RNA
136 chains using nucleotide resources (not described because were not used).

Focused sequence databases can be used for Ab and TCR modelling while keeping accuracy near experimental resolution

To evaluate the impact of reducing the reference database used for MSA, we assessed model performance and accuracy across different AF3 workflow variants that incorporate sequence databases of varying sizes. Model accuracy was quantified by calculating the root mean square deviation (RMSD) between the predicted 3D structure generated with a given AF3 variant (both the original and the reduced-database versions) and the corresponding experimental PDB structure. **Figure 2.**

Our systems of interest, TCRs and Abs, were selected because of their structural similarities, and the well-known challenges associated with accurately modeling their loops[10-12]. TCRs and Abs share a broadly similar 3D architecture (**Figure 2**): both are heterodimers containing six complementarity-determining regions (CDRs) that form their binding interfaces. As illustrated in **Figure 2**, CDRH1, CDRH2, and CDRH3 from the Ab heavy chain (H) correspond to CDRβ1, CDRβ2, and CDRβ3 of the TCR β chain; likewise, the three CDRs of the Ab light chain (L) align with those of the TCR α chain. Despite these structural parallels, Abs can bind to an extraordinarily wide range of molecular targets, including peptides, proteins, and haptens, whereas αβ TCRs recognize peptide antigens presented by MHC molecules.

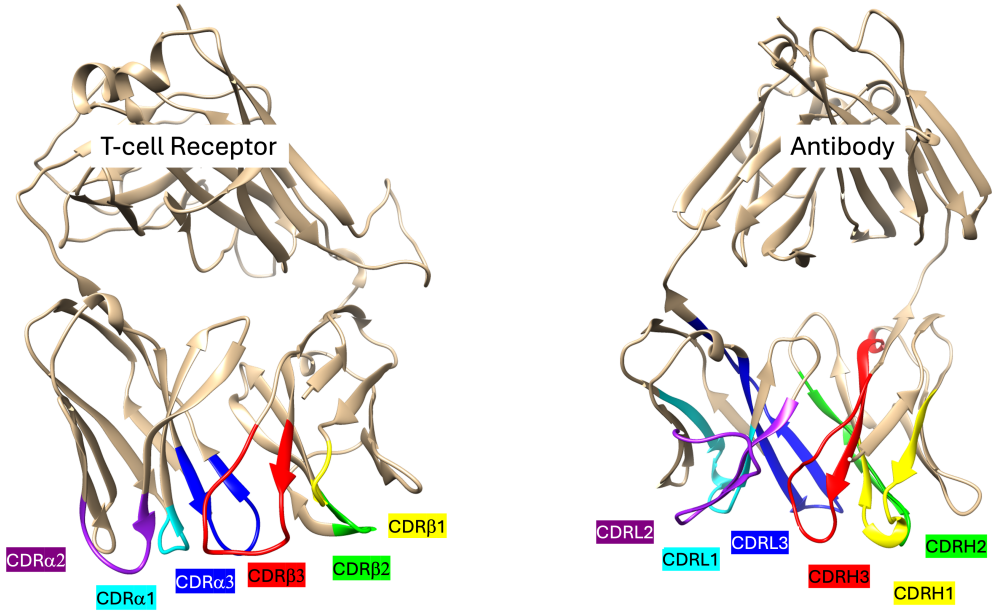


Figure 2. Structural parallels between TCRs (on the left) and antibodies (on the right). Analogous chains and CDRs are color-coded consistently.

RMSDs were calculated for the entire system as well as for the CDR loops, which are the most challenging regions and contain the key structural information.

As shown in **Figure 3**, AF3 workflow variants with MSA based on reduced sequence datasets performed comparably to the original AF3 for TCR modelling, including for the challenging CDR3β loop. On the contrary, MSA-free predictions performed markedly worse **(Figure 3 and Figure 4)**, with an average RMSD considering all heavy atoms in the entire TCR of 11.0 ±6.5 Å, underscoring that some degree of evolutionary information remains essential for accurate TCR structure prediction. Notably, the UniRef-textmining based variant yields the lowest average RMSD for CDR3β among all tested configurations **(Figure 3)**, a particularly relevant result given that CDR3β is widely recognized as the most difficult TCR loop to model accurately[11, 12]. The RMSD differences across variants with reduced sequence datasets

are not statistically significant (p <= 1, annotated as "ns" in **Figure 3**), indicating that reduced sequence datasets for MSAs can effectively compete with the full AF3 reference database. Crucially, these reduced-MSA variants offer substantial computational advantages, enabling up to a 45-fold reduction in runtime. The AF3 variant with a reduced database of sequences, decreases the computation time from 15 minutes to less than 40 seconds (on a recent Dell Precision 3680 with 1 GPU NVIDIA RTX A4500 Ada Generation; 1 cpu used). RMSD values for each TCR in the validation set across all AF3 variants are provided in the **Supporting Information.**

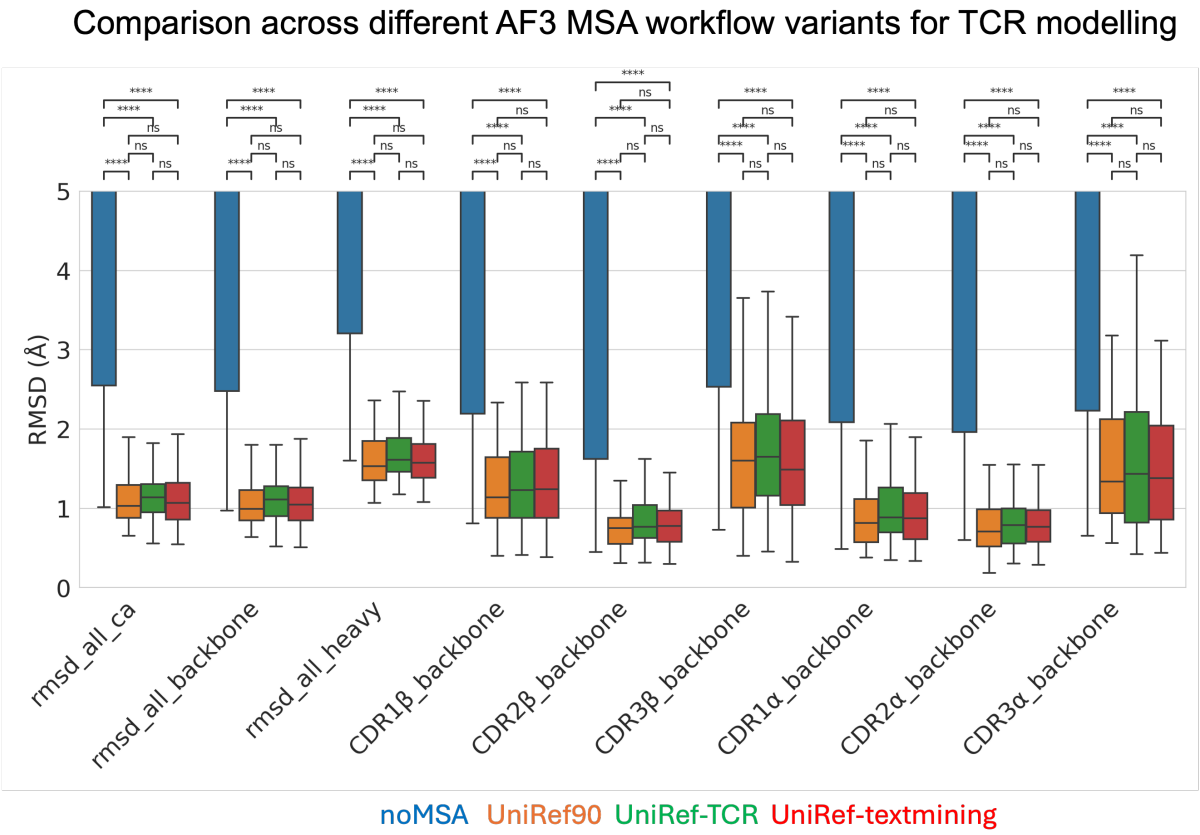


Figure 3. RMSD analysis using different AF3 MSA variants for TCR modelling. We compared results of AF3 runs using four MSA configurations: no MSA (in blue), the default AF3 that uses among others UniRef90 (in orange), a UniRef90 subset selected according to AF3's choices during TCR modelling from the VDJ database, i.e. UniRef-TCR (in green), and a UniRef90 subset curated using TCR and antibody

sequences, UniRef-textmining (in red). The evaluation used a set of 77 TCRs for which just one TCR was accidentally used to also generate the sequence datasets for the MSA. RMSDs were calculated for the entire TCR receptor domain based on carbon alpha (rmsd_all_ca), backbone (rmsd_all_backbone) and all heavy atoms (rmsd_all_heavy). Backbone RMSD for the CDR1β (CDR1β_backbone), CDR2β (CDR2β_backbone), and CDR3β (CDR3β _backbone), CDR1α (CDR1α_backbone), CDR2α (CDR2α_backbone) and CDR3α (CDR3α _backbone). P-values calculated with Mann-Whitney-Wilcoxon test two-sided with Bonferroni correction. The annotations used in the graph are ns: p <= 1.00e+00, *: 1.00e-02 < p <= 5.00e-02, **: 1.00e-03 < p <= 1.00e-02, ***: 1.00e-04 < p <= 1.00e-03 and ****: p <= 1.00e-04.

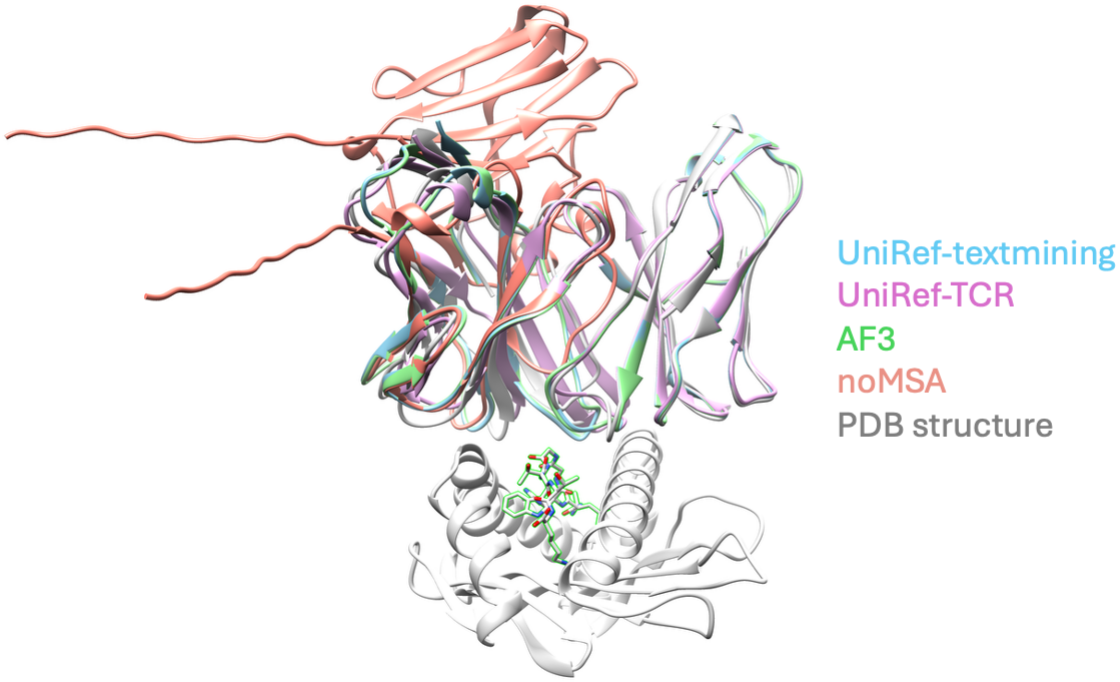


Figure 4. Superimposed structural models, using different MSA variants, with the experimental structure PDB code 7RRG.

The correlation between TCR RMSDs obtained from the native and AF3 workflow variants can be seen in **Figure 5** and **Supporting Information.** We observe that correlation between the RMSD obtained from AF3 (full databases) and AF3 workflow variant with UniRef-TCR database models is higher than 0.75 for

all the different RMSD analysis including loop-specific RMSDs, indicating that expected prediction accuracy for each TCR is preserved regardless of whether the native or accelerated workflow is used, with both workflows reproducing the same overall error patterns and relative loop-specific modeling difficulties.

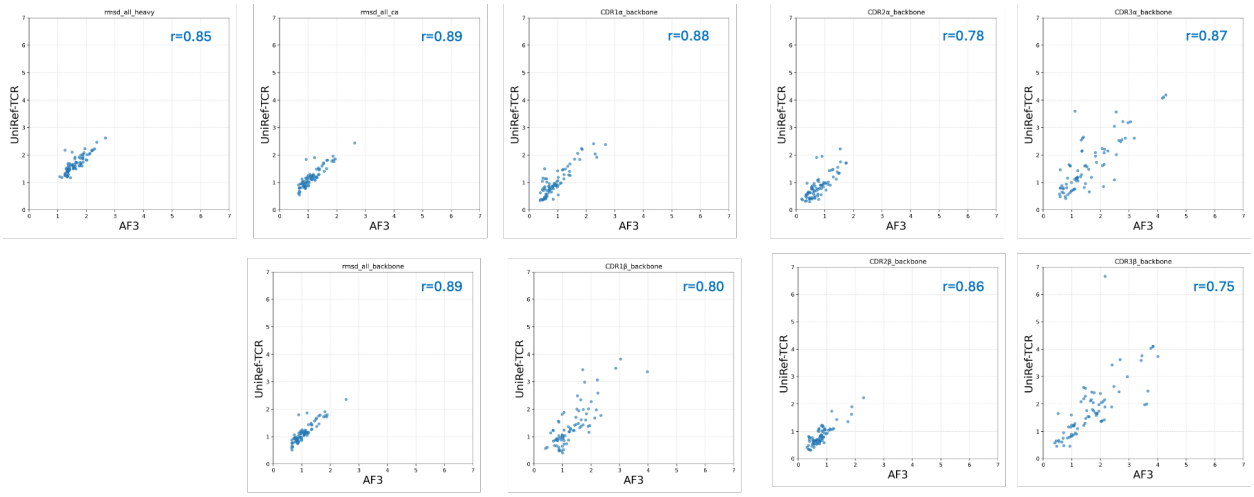
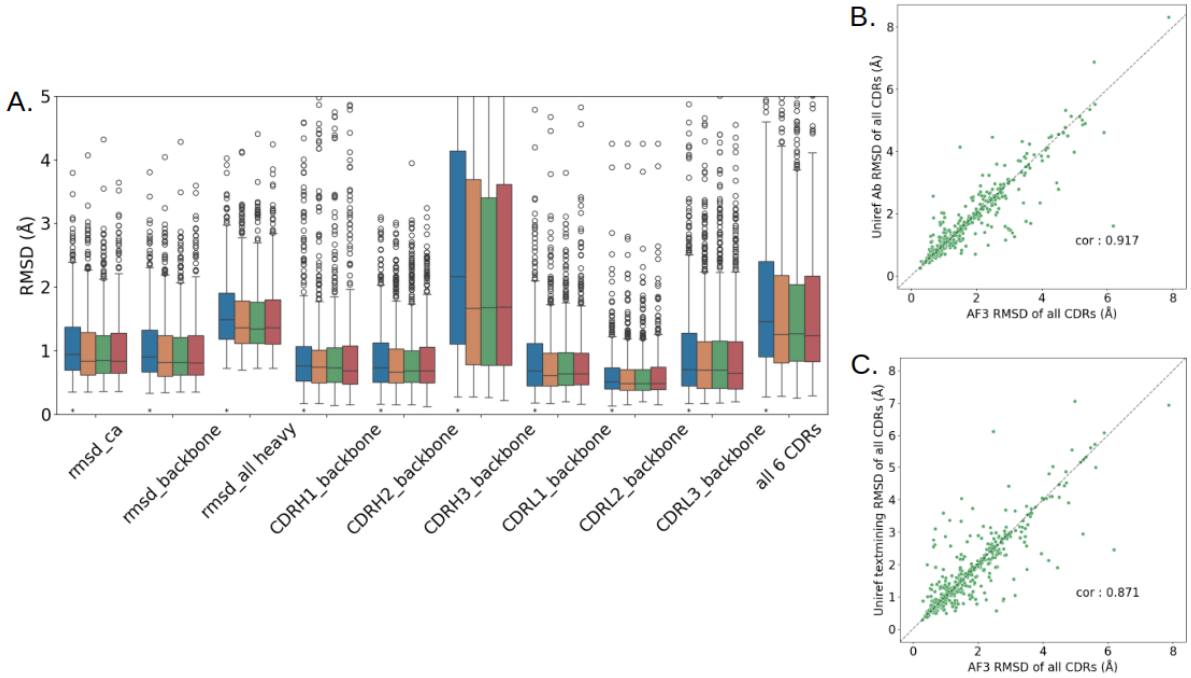


Figure 5. RMSD correlations for the native AF3 workflow and one variant using UniRef-TCR. RMSD were calculated between the experimental structure and the model generated by a given AF3 workflow, on the entire TCR based on heavy atoms (rmsd_all_heavy), on Cα atoms (rmsd_all_Ca), and on backbone atoms (rmsd_backbone), as well as for the backbone of the CDR1β (CDR1β _backbone), CDR2β (CDR1β _backbone), CDR3β (CDR1β _backbone), CDR1α (CDR1α_backbone), CDR2α (CDR2α_backbone), and CDR3α (CDR3α_backbone) loops.

The AF3 MSA pipeline was run either with its default databases or exclusively with the specialized databases on a set of 525 Abs from SAbDab (thus with an experimental 3D structure) that were not part of the AF3 training set. The inference was performed with 10 seeds. We then compared the RMSD of the backbone atoms of the models to their crystal structures. Whatever the native or accelerated workflows used, we observed a much higher RMSD for CDRH3 than any other CDR – the latter having a median RMSD below 1Å (**Figure 6A**). When comparing the RMSD for all CDRs between the models created with the accelerated (using the Uniref-Ab sequence dataset) and native AF3 workflows, the first showed marginally higher RMSD despite having a median RMSD with the same order of magnitude, indicating

212 that the predictive performance of the two workflows is effectively indistinguishable. A high correlation
213 could be observed between the 6 CDRs RMSD of the models generated with the native and this accelerated
214 AF3 workflows (pearson correlation: 0.917) as well as between the native AF3 workflow and that using
215 the AF3-textmining sequence dataset (pearson correlation: 0.871), **Figures 6B and C,** respectively. The
216 correlation between CDRs' RMSD obtained from the native and AF3 workflow variant UniRef-Ab can be
217 seen in **Supporting Information**. Thus, similarly to TCRs, the use of reduced-MSA protocols to model Ab
218 structures reproduce the same overall error patterns and relative loop-specific modelling difficulties as the
219 original AF3 workflow.



220

221 **Figure 6**. RMSD analysis using different AF3 workflow variants for Ab modelling. A) We compared
222 results of AF3 runs using four MSA configurations: no MSA (in blue), the default AF3 that uses among
223 others UniRef90 (in orange), a UniRef90 subset selected according to AF3's MSA on Ab modelling of 152
224 structures, i.e. UniRef-Ab (in green), and a UniRef90 subset curated using TCR and antibody sequences,
225 UniRef-textmining (in red). For evaluation, we used 525 Abs from SAbDab that were not included in AF3's
226 training set. RMSDs were calculated for the entire Ab's Fv domain based on carbon alpha (rmsd_ca),

backbone (rmsd_backbone), and all heavy atoms (rmsd_all heavy). Backbone RMSD for the CDR H1 (CDRH1_backbone), CDR H2 (CDRH2_backbone), and CDR H3 (CDRH3_backbone), CDR L1 (CDRL1_backbone), CDR L2 (CDRL2_backbone), and CDR L3 (CDRL3_backbone). The annotation: * represents a p-value < 0.025 in a one-sided wilcoxon test compared to the original AF3 workflow. For a non-cropped Y axis see **Supporting Information**. In B) and C) we observed a high correlation between the RMSD of all 6 CDRs between AF3 and Uniref-Ab or Uniref-textmining workflow variants.

One seed is enough for high-accuracy modelling of TCR and Ab receptor domains

We analysed the relationship between the AF3 ranking scores - generated across all models, using 20 seeds and five inferences per seed applied to the TCR validation set – and the model accuracy (**Figure 7**). We have observed some outliers, highlighted with an orange cross for which the RMSD using the 6CDRs was higher than 15 Angstrom and the ranking score was lower than 0.4. Surprisingly, for example, for TCR with PDB code 7N2Q and for seed 56970, the ranking scores range between 0.35 and 0.44 and the RMSD ranges between 0.7 and 16.1 Angstrom. See **Table 2**. Our results showed that a high ranking score does not reliably correspond to an experimentally accurate prediction, nor does the top-ranked model necessarily represent the best-quality structure among those generated. Furthermore, increasing the number of seeds does not improve model accuracy, indicating that additional sampling offers no benefit in this context **(Figure 8).**

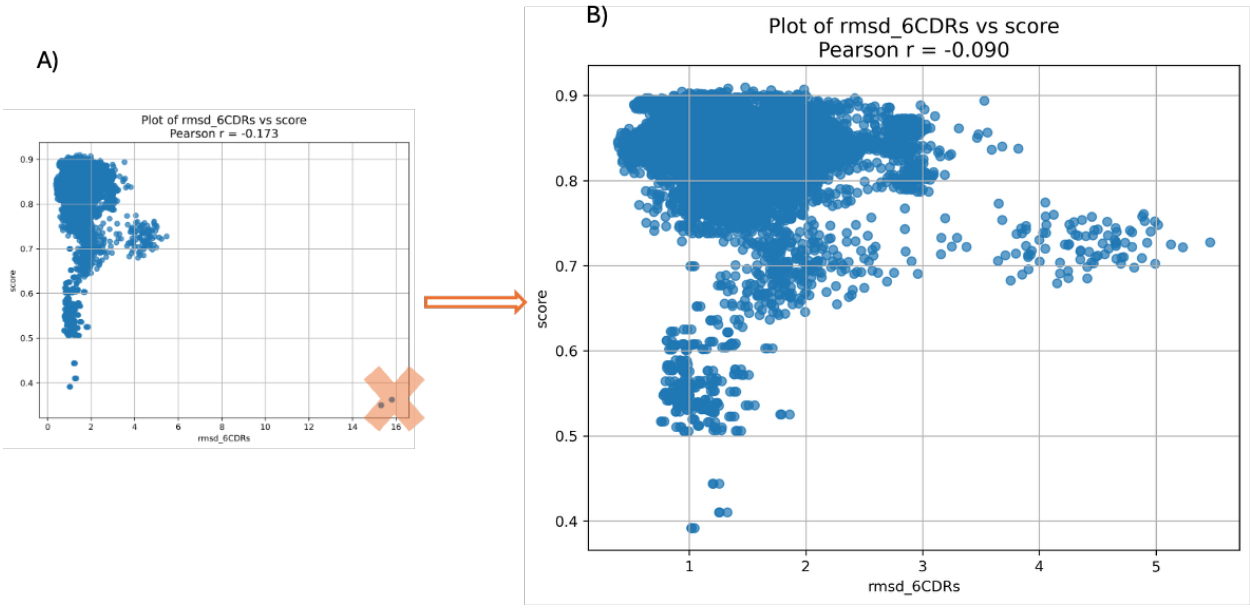


Figure 7. (A-B) Ranking score vs backbone RMSD calculated for the 6 CDR loops using UniRef-TCR sequence dataset. 20 seeds and 5 inferences per seed were used on a set of 77 TCR structures with known 3D that were not included in AF3 training. A) with outliers and B) without outliers. Ranking score = 0.8 × ipTM + 0.2 × pTM + 0.5 × disorder factor – 100 × has_clash as described in the code *https://github.com/google-deepmind/alphafold3*.

Table 2. Ranking score and RMSD for a PDB example: TCR 7n2q – 3 inferences for seed 56970. The ranking scores range between 0.35 and 0.44 and the RMSD for the 6CDRs between 0.7 and 16.1 Angstrom.

Seed Number (PDBid 7N2Q)	Inference	RMSD 6CDRs	Ranking score
56970	0	15.8	0.35
56970	1	16.1	0.36
56970	2	0.7	0.44

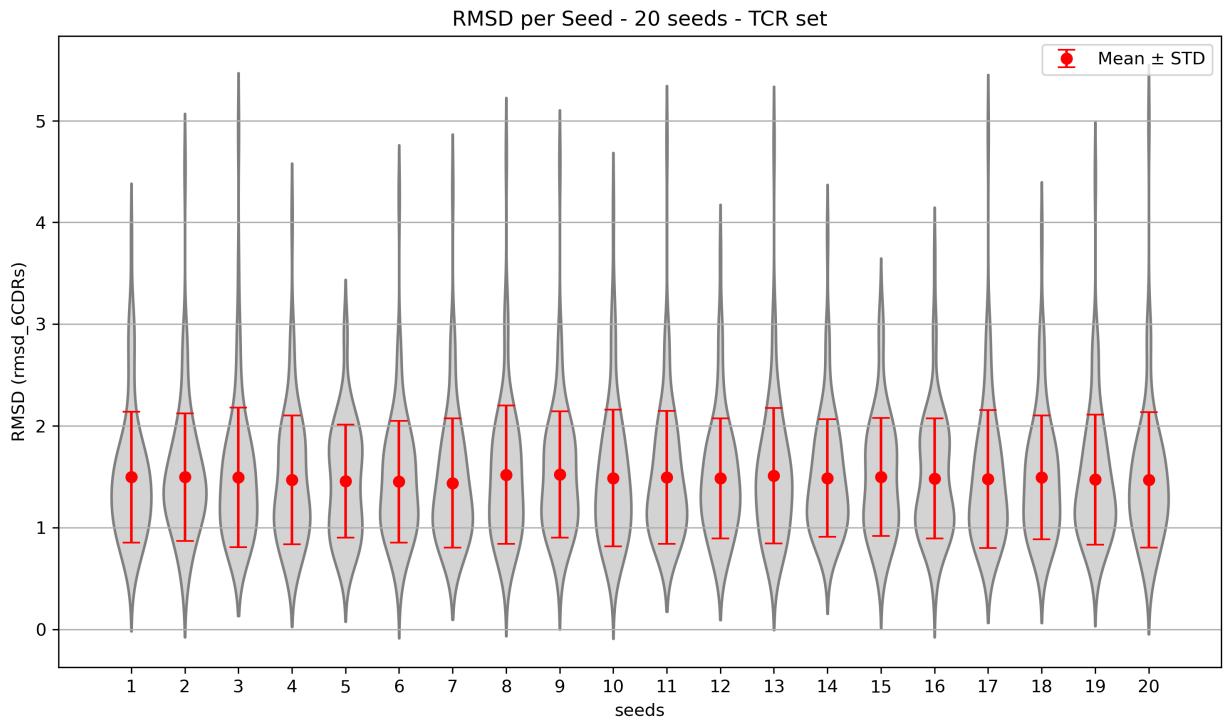


Figure 8. RMSD of the 6 CDR loops as a function of the number of inference seeds used. Each run with n+1 seeds includes all seeds from the run with n seeds, plus one additional randomly selected seed. At each iteration, the RMSD is analysed relative to the best ranking score obtained across all preceding seeds and the newly added seed. RMSD analysis of TCR modelling using an AF3 variant with the UniRef90-TCR sequence dataset. A maximum of 20 random seeds were used.

We also analysed the AF3 ranking scores generated across all models, using 10 seeds and five inferences per seed for the Ab validation set (**Figure 9**). We found that for the same Ab, several models achieved low RMSD for CDRH3 or all CDRs but were not top ranked. So, suboptimal model selection (i.e. scoring failure) is observed. The AF3 ranking score showed poor correlation with CDR accuracy as measured by the RMSD, both for all CDRs (Pearson $r = -0.297$) and for CDRH3 ($r = -0.247$). Increasing the number of seeds to 20 did not improve CDR modelling or model selection, and the same trend was observed with the AF3 variants with reduced UniRef-Ab and UniRef-textmining sequence datasets (**Supplementary**

Figures). These results indicate that the default AF3 ranking score, dominated by the ipTM term, inadequately reflects Ab model quality.

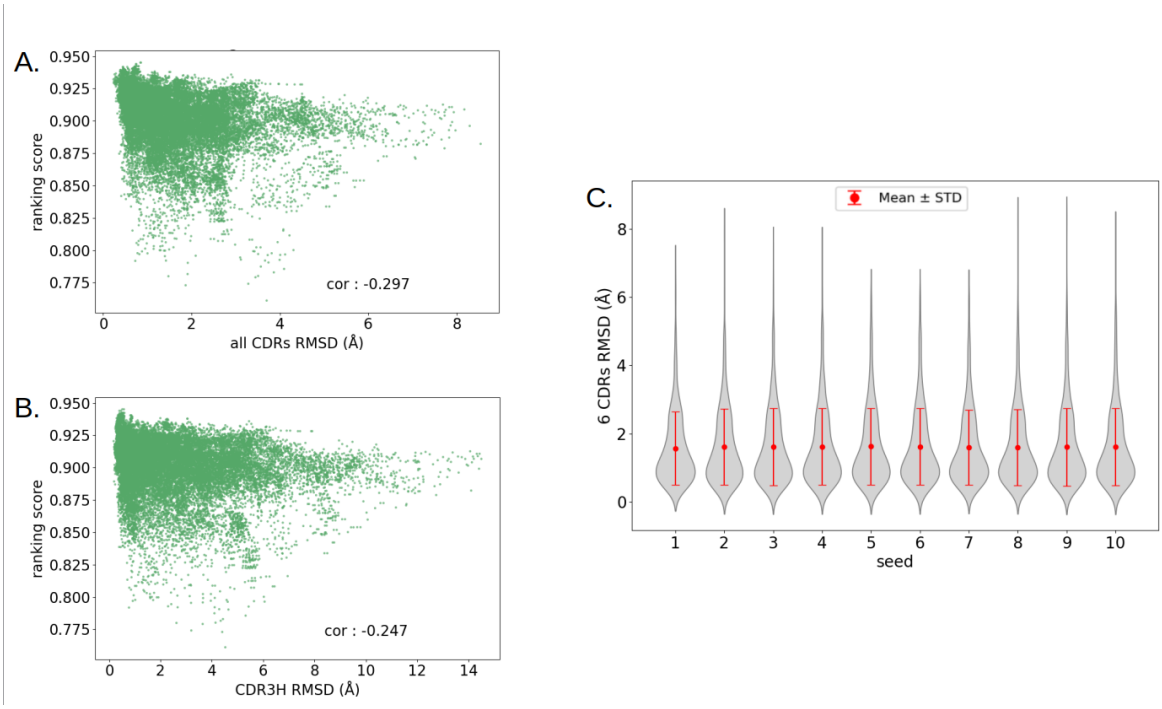


Figure 9. Analysis of the relationship between AF3's ranking score and CDRs modelling accuracy. We assessed the correlation between the ranking score and A) the RMSD of all 6 CDRs and B) the RMSD of CDR H3. C) We plotted the evolution of the RMSD of all CDRs over the generation of models for 10 seeds, taking for each seed the best ranking score among the generated models up to that seed.

Discussion

Despite the reduced influence of the MSA on model construction in AF3, the MSA component remains the slowest step in the workflow, representing more than 90% of the end-to-end execution time. Consequently, MSA generation imposes a disproportionate runtime and emerges as a bottleneck in AF3.

To largely decrease the end to end execution time of AF3 to generate structural models of TCR or Ab receptor domains, four different AF3 MSA variants were developed: no MSA, the original AF3 setup with

287 the full UniRef90 database, a UniRef90 subset selected according to AF3's choices during TCR and Ab
288 modelling (UniRef90-TCR and UniRef90-Ab, respectively) and a non-exhaustive UniRef90 subset curated
289 using TCR and Ab sequences (UniRef90-textmining).

290 All the models obtained with the AF3 MSA-free variant are significantly worse than the ones obtained with
291 the reduced databases and original AF3 (p-value < 0.0001) for the modelling of TCRs, indicating that MSA
292 is needed for the modelling of the TCRs. Conversely, we observed that for the Ab modelling the MSA-free
293 variant can be used while keeping accuracy near experimental resolution (except for one model of pdb id :
294 8DAO, CDRs RMSD > 21Å, see **Supporting Information**) – at the cost of a very marginal yet significant
295 loss in accuracy compared to the native AF3 workflow. We hypothesize that these differences arise partly
296 because far fewer TCR experimental 3D structures were available for AF3 training compared to antibodies.
297 Although some TCRs are better modelled than others in the MSA-free variant, we do not observe that the
298 ones with low RMSD have better representation of their genes in the 3D structures that were used for AF3
299 training (data not shown).

300 Despite using highly size-reduced subsets of UniRef90 for MSA, all our accelerated Af3 workflows lead
301 to results comparable to the native AF3 workflow. The RMSD differences are comparable indicating that
302 reduced-database MSAs can effectively compete with the full AF3 reference database for some conserved
303 families/domains of proteins, including TCRs and Abs.

304 We observed that the Pearson correlation between the original AF3 model and each reduced-MSA AF3
305 variant for TCR and Ab modelling exceeds 0.75 across all RMSD evaluations, including the per-loop
306 analyses. In addition to reflecting methodological consistency, a high Pearson correlation suggests three
307 important points. First, the structural variability captured by the models is largely driven by intrinsic
308 features of the TCRs and Abs—such as loop length, conformational flexibility, or limited evolutionary
309 signal—rather than by the specific size of the MSA database used. Second, the strong correlation implies

310 that reducing the MSA complexity does not introduce new biases or failure modes; the variants remain
311 aligned with AF3's ranking of "easy" versus "difficult" cases to model. Third, this stability supports the
312 interpretability and robustness of the reduced-database approach, reinforcing that computational efficiency
313 can be gained without altering the fundamental behaviour of the prediction algorithm.

314 Crucially, these reduced-MSA variants offer substantial computational advantages, enabling up to a 45-
315 fold reduction in runtime. AF3 variant with a reduced database of sequences, decreases the computation
316 time from 17 minutes (11 minutes for Abs) to less than 40 seconds (Machine Dell Precision 3680 with 1
317 GPU NVIDIA RTX 4500 Ada Generation, 1 cpu used).

318 Our results showed that the AF3 ranking score does not correlate well with loop RMSD for either TCRs or
319 Abs. We therefore focused our analysis on loop regions, which represent the most challenging to model in
320 these structures, whereas framework regions are highly conserved and typically achieve RMSDs of ~1 Å.
321 Across different numbers of seeds and inference runs, we consistently observed that the top-ranked models
322 were not those with the lowest loop RMSD, leading to suboptimal model selection (i.e. scoring failure).
323 Importantly, increasing the number of seeds did not improve the average RMSD, indicating that generating
324 models for more than one seed provides little to no benefit. In this context, AF3 with a reduced MSA and
325 a single seed represents an excellent trade-off between computational efficiency and accuracy for TCR and
326 Ab modeling.

327 We decided to focus on AF3 workflow variants rather than benchmarking against alternative methods, as
328 our primary objective was to accelerate AF3 while preserving its predictive accuracy. Notably, while MSA
329 information is still used in AF3, its contribution appears substantially reduced compared with AF2,
330 suggesting additional opportunities for optimizing the MSA stage without sacrificing accuracy. Although
331 several faster competing approaches exist and have demonstrated strong performance in their original
332 publications, they have already been extensively compared to AF3 by their developers. AF3 relies on

333 JackHMMER to generate MSAs, which is computationally expensive but produces deep, high-quality
334 alignments. Faster alternatives, such as MMseqs2—used in the ColabFold API based on AF2—could
335 further reduce runtime. However, the use of such a tool puts an extra layer of complexity on the user's
336 workflow, while our method is simple and merely requires the user to use the dedicated AF3 options for
337 specific databases and to create a specialized database of sequences.

338 In conclusion, we analyzed AF3 performance for TCR and Ab modeling, varying the MSA generation and
339 inference phase. Using reduced subset of sequences from UniRef90 greatly speeds up AF3, achieving ~45×
340 faster runtimes on a desktop while maintaining near-experimental accuracy, enabling scalable structural
341 immunology.
342

343 # Methods

344 ## TCR and Antibody set used for MSA analysis

345 A non-redundant set of 3,213 TCR sequences, derived from the VDJ database[8, 9], was used to identify
346 AF3-selected sequences for multiple sequence alignment (MSA), covering specificities for 200 different
347 antigens. Detailed information on all TCRs, including gene composition and peptide specificity, is provided
348 in Suplementary Table 9 of Mayol-Rullan *et al[5]*.

349 A set of Ab sequences was derived from apo-form structures found in SAbDab [13, 14] and deposited
350 between the 1st of January 2022 and the 1st of September 2025. Structures with a resolution above 3Å and
351 or an R-free value superior to 0.3 were excluded. Only structures with two different chains as their heavy
352 and light chains were kept and if several copies of the same Ab were found in the same PDB entry, only
353 the first one would be kept. Two sequences could not be processed by ANARCI [15] for conversion to a
354 specific scheme numbering resulting in the Ab sequences a total of 152 sequences.

355

Custom Uniref90-derived databases creation:

Multiple sequence alignment (MSA) of the TCR sequences was performed using JackHMMER [7] (EVALUE=1e-3, NUM_ITER=3) against the UniRef90 database[6, 16]. All unique aligned sequences were aggregated into a new database called **UniRef-TCR** composed of 261’780 sequences, among which 261’055 were aligned with the alpha chain and 208’047 with the beta chain.

Ab sequences were used as input for AF3 MSA pipeline (--norun_inference parameter). Unpaired MSA sequences from both the heavy and light chains from the generated json files were extracted and unique sequences were concatenated into a new database called **UniRef-Ab**. Leading to a reduced database of 35’607 sequences, with 17’887 of them aligned to the heavy chain and 18’312 to the light chain.

Another subset of UniRef90 was created by searching in the database for the words "t-cell receptor", "t cell receptor", "Tcell receptor", "immunoglobulin", "antibody", "nanobody" in the sequences header. This resulted in a non-exhaustive set of sequences with a total of 41’055 sequences gathered this way. We called this new database **UniRef-textmining.**

Validation dataset:

TCR structures bound to a pMHC class I uploaded to the Protein Data Bank (PDB) between the 1st of January 2022 and 12 November 2025 (outside of AlphaFold3 training set) were collected. As our objective is to model TCR receptors alone to predict their specificity just TCR-pMHC class I structures were considered and 77 structures were retained. The PDB codes of these structures are available in **Supporting Information.**

Ab structures bound to a peptide or protein antigen according to SAbDab and added between the 1st of January 2022 to 15th of October 2025 were collected. The same constraints as for the training set, regarding

the resolution, R-free value and chains were applied. Moreover, only the structures without missing residues were analysed resulting in a total of 525 structures. The PDB code for the Ab structures used can be seen in **Supporting Information**.

AlphaFold3 model generation with custom databases:

Structures modelled with AlphaFold3, used 5 inferences per seed and the 20 following randomly chosen seeds: 9986, 67719, 29960, 58505, 56970, 55176, 90764, 44813, 71819, 36501, 58009, 18970, 90779, 1948, 91807, 65829, 92455, 87079, 11815, 39719. When custom sequence databases were used, if a AF3 original databases needed to be removed, only the first two lines (the sequence header and the first line of the sequence) were kept as AlphaFold3 would throw an error when trying to use empty files as databases. The removed databases were bfd-first_non_consensus_sequences.fasta, mgy_clusters_2022_05.fa and uniprot_all_2021_04.fa.

The UniRef-TCR, UniRef-Ab selection and UniRef-textmining custom databases were temporarily renamed as the larger database they were issued from: uniref90_2022_05.fa when producing MSA through AF3. AF3 MSA pipeline and inference were run independently for optimization purposes. To accelerate model generation, the slurm queueing system was used to run batches of 4 MSA pipeline in parallel using 6 cpus each. Inference was run on a single Nvidia RTX 4500 Ada Generation GPU. Generating both chains of a Variable Fragment (Fv) of an antibody took on average 30 seconds for a single seed and 170 seconds for 10 seeds. Generating both chains of TCR took on average 40 seconds for a single seed.

Root Mean Square Deviation (RMSD) calculation:

To superimpose and compute the RMSD of TCRs and Ab models, the Biopython [17] package was used. Any crystal structures with missing residues were removed from the RMSD analysis. For a more accurate and consistent comparison, models were superimposed on their framework residues (thus excluding the structurally variable CDR loops) rather than on the whole structure and RMSD was computed either on single Complementary Determining Region's (CDR) residues, on a set of all CDRs residues or on all residues of the model. To determine CDRs and framework residues the ANARCI software was used[15]. For TCRs, the IMGT numbering scheme was used while the Chothia numbering scheme was preferred for antibodies.

Considering the non-normality of the RMSD distribution, a non-parametrical Wilcoxon statistical test (stats.wilcoxon from python's SciPy package[18]) was used to compare between pairs of RMSD.

Analysis of CDRs RMSD across multiple seeds:

To compare the evolution of the RMSD of CDRs across multiple seeds we imitated AlphaFold3 behaviour by taking the model with the highest ranking score for each seeds taken into consideration (e.g., after 4 seeds, the highest score among the 20 inferences). RMSD of all CDRs residues were then computed.

Funding

This study was supported by a SNSF grant to VZ (#320030-236248).

References

1. Senior, A.W., et al., *Improved protein structure prediction using potentials from deep learning.* Nature, 2020. **577**(7792): p. 706-710.
2. Abramson, J., et al., *Accurate structure prediction of biomolecular interactions with AlphaFold 3.* Nature, 2024. **630**(8016): p. 493-500.

3. Jumper, J., et al., *Highly accurate protein structure prediction with AlphaFold.* Nature, 2021. **596**(7873): p. 583-589.
4. Perez, M.A.S., et al., *Predicting Antigen-Specificities of Orphan T Cell Receptors from Cancer Patients with TCRpcDist.* Adv Sci (Weinh), 2024. **11**(40): p. e2405949.
5. Mayol-Rullan, F., et al., *A fingerprint approach to pioneer structure-based T cell receptor repertoire analysis and specificity prediction.* Front Immunol, 2025. **16**: p. 1688805.
6. Suzek, B.E., et al., *UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches.* Bioinformatics, 2015. **31**(6): p. 926-32.
7. Potter, S.C., et al., *HMMER web server: 2018 update.* Nucleic Acids Res, 2018. **46**(W1): p. W200-W204.
8. Shugay, M., et al., *VDJdb: a curated database of T-cell receptor sequences with known antigen specificity.* Nucleic Acids Res, 2018. **46**(D1): p. D419-D427.
9. Goncharov, M., et al., *VDJdb in the pandemic era: a compendium of T cell receptors specific for SARS-CoV-2.* Nat Methods, 2022. **19**(9): p. 1017-1019.
10. Regep, C., et al., *The H3 loop of antibodies shows unique structural characteristics.* Proteins, 2017. **85**(7): p. 1311-1318.
11. Marin, F.I. and P. Marcatili, *Computational Modeling of Antibody and T-Cell Receptor (CDR3 Loops).* Methods Mol Biol, 2023. **2552**: p. 83-100.
12. Abanades, B., et al., *ImmuneBuilder: Deep-Learning models for predicting the structures of immune proteins.* Commun Biol, 2023. **6**(1): p. 575.
13. Dunbar, J., et al., *SAbDab: the structural antibody database.* Nucleic Acids Res, 2014. **42**(Database issue): p. D1140-6.
14. Schneider, C., M.I.J. Raybould, and C.M. Deane, *SAbDab in the age of biotherapeutics: updates including SAbDab-nano, the nanobody structure tracker.* Nucleic Acids Res, 2022. **50**(D1): p. D1368-D1372.
15. Dunbar, J. and C.M. Deane, *ANARCI: antigen receptor numbering and receptor classification.* Bioinformatics, 2016. **32**(2): p. 298-300.
16. Pundir, S., et al., *Searching and Navigating UniProt Databases.* Curr Protoc Bioinformatics, 2015. **50**: p. 1.27.1-1.27.10.
17. Cock, P.J., et al., *Biopython: freely available Python tools for computational molecular biology and bioinformatics.* Bioinformatics, 2009. **25**(11): p. 1422-3.
18. Virtanen, P., et al., *SciPy 1.0: fundamental algorithms for scientific computing in Python.* Nat Methods, 2020. **17**(3): p. 261-272.