AlphaFold2 and AlphaFold3 leads to significantly different results in human-parasite interaction prediction

Burcu Ozden[1,2], Yesid Cuesta-Astroz[3,4], Ezgi Karaca[1,2]

[1]Izmir Biomedicine and Genome Center; [2]Dokuz Eylul University; Izmir, Türkiye.
[3]Escuela de Microbiología, Universidad de Antioquia, Medellín; [4]Instituto Colombiano de Medicina Tropical, Universidad CES, Sabaneta; Colombia.

Correspondence: ezgi.karaca@ibg.edu.tr

Abstract

Parasitic diseases pose a significant global health challenge with substantial socioeconomic impact, especially in developing countries. Combatting these diseases is difficult due to two factors: the resistance developed by parasites to existing drugs and the underfunded research efforts to resolve host-parasite interactions. To address the latter, we focused on 276 human-parasite interaction predictions involving 15 parasitic species known to cause infections in humans. We aimed to validate these interactions by using AlphaFold2-Multimer v2.2 (AF2) and AlphaFold3 (AF3) modeling. Overall, our findings show that, in the context of human-parasite interactions, AF2 and AF3 produce distinct models and score distributions, with no significant correlation between their top-ranked predictions. We recognize that predicting host-parasite interactions is particularly challenging due to the absence of typical coevolutionary signals. Nevertheless, the differences in scores and models between AF2 and AF3 are striking and warrant further investigation, which could only be performed upon the release of AF3.

Introduction

Parasitic diseases pose a significant global health challenge with substantial socioeconomic impact, especially in developing countries (Dampier et al., 2009; Mehmood et al., 2017; Merrifield et al., 2016). Combatting these diseases is difficult due to two factors: the resistance developed by parasites to existing drugs and the underfunded research efforts to resolve host-parasite interactions (Cuesta-Astroz & Oliveira, 2018). To address the latter, we focused on 276 human-parasite interaction predictions involving 15 parasitic species known to cause infections in humans (**Table S1**) (Cuesta-Astroz et al., 2019). We aimed to validate these interactions by using AlphaFold2-Multimer v2.2 (AF2) and AlphaFold3 (AF3) modeling (Cuesta-Astroz et al., 2019; Jumper et al., 2021; Abramson et al., 2024). Our initial hypothesis was that the interactions with an AF confidence score ≥ 0.8 could serve as high confident templates for further research. To test this, we modeled all interaction predictions with both methods and isolated their highest-scoring models (*ranked_0.pdb*) for further analysis (see **Table S2** for raw data).

Results

Figure 1 presents the AF2/3 confidence, ipTM, and pTM score distributions for all the predicted human-parasite interactions belonging to each species. Here, AF2 consistently produced higher mean scores than AF3, though there were notable variations in confidence and ipTM scores across different parasitic species.

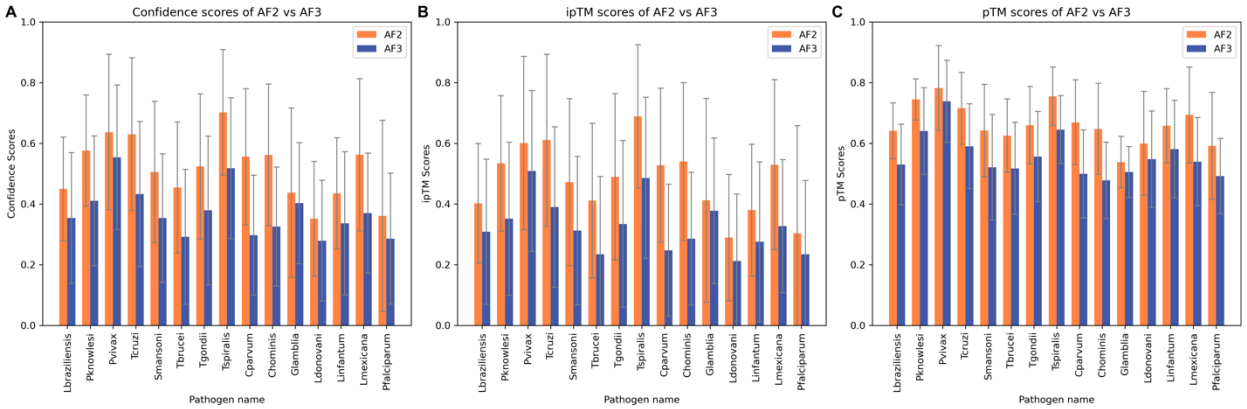

Figure 1. Score distributions of AF2 and AF3 for each parasitic species. Mean confidence **(A)**, ipTM **(B)**, and pTM scores along with their standard deviations for each parasitic species (for *L. braziliensis* n=4; *P. knowlesi* n=18; *P. vivax* n=20; *T. cruzi* n=54; *S. mansoni* n=17; *T. brucei* n=15; *T. gondii* n=25; *T. spiralis n*=7; *C. parvum* n=10; *C. hominis* n=6; *G. lamblia* n=11; *L. donovani* n=12; *L. infantum* n=21; *L. mexicana* n=35; *P. falciparum* n=21).

A correlative analysis between AF2 and AF3 scoring terms revealed considerable differences (**Figure 2**). AF2 vs. AF3 confidence scores showed a moderate correlation according to all probed metrics (^{2}=0.42, =4.64×10^{-34} for confidence; ^{2}=0.42, =6.21×10^{-34} for ipTM; ^{2}=0.43, =3.37×10^{-35} for pTM). We identified 42 out of 276 predicted interactions as structurally viable (models with confidence scores ≥

0.8 obtained either from AF2 or AF3, **Figure 2A**). Interestingly, **only about 4% of the models achieved a confidence score ≥ 0.8 in both AF2 and AF3 models**.

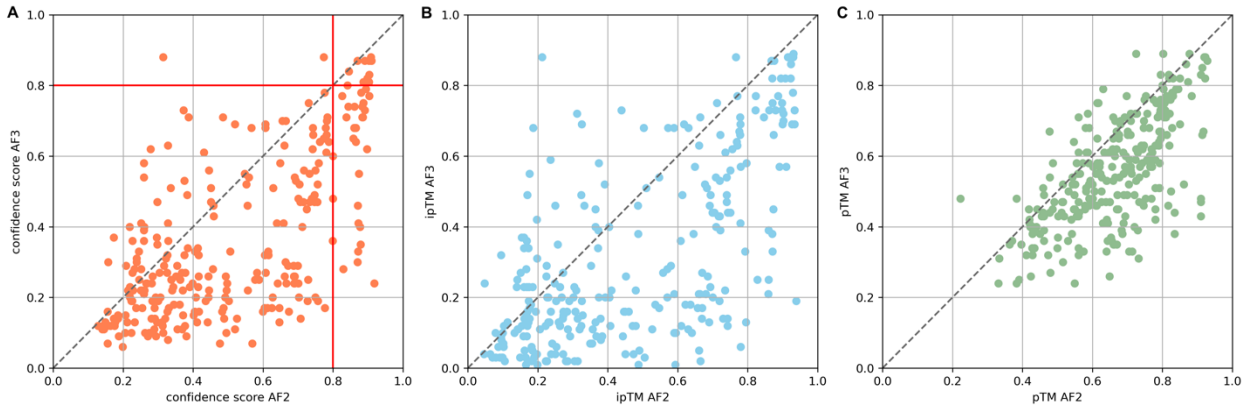

Figure 2. Comparison of AF2 vs. AF3 Scores. Scatter plots of AF2 vs. AF3 scores for **(A)** Confidence, **(B)** ipTM, and **(C)** pTM metrics. Higher scores indicate better model quality. In panel A, the red lines on the confidence score plot designate the threshold for high-quality models (≥ 0.8) for both AF2 and AF3.

We finally focused on examining the structural similarities between the AF2 and AF3 predictions for the 42 *'trustworthy'* human-pathogen interactions. For this, we calculated the DockQ scores between each interaction modeled by AF2 and AF3 (**Table 1**). **Of the 42 interactions, only half were highly similarly modeled by AF2 and AF3** (**Figure 3A**).

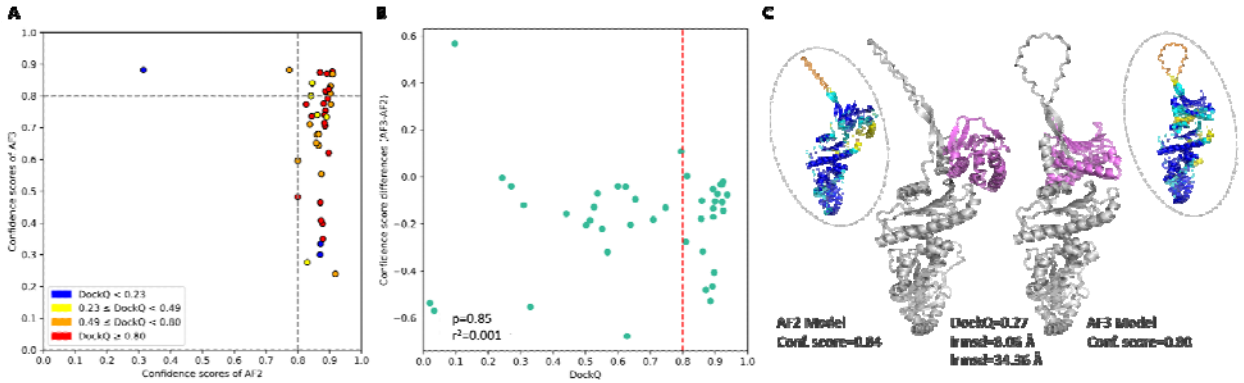

Figure 3. Correlation between AF confidence scores and structural similarities. (A) Structural similarities assessed using the DockQ score for models with a confidence score higher than 0.8 in AF2/AF3. Significantly different models (DockQ < 0.23) are shown in blue, acceptable models (0.23 ≤ DockQ < 0.49) in yellow, moderately similar models (0.49 ≤ DockQ < 0.80) in orange, and highly agreeing models (DockQ ≥ 0.80) in red. **(B)** Correlation between DockQ score and confidence score differences between AF2 and AF3. The red dashed line represents the high-quality threshold for the DockQ score. **(C)** Predicted human - *L. mexicana* interaction (human UniProt ID: ENSP00000335153, *L. mexicana* UniProt protein ID: PF02518_XP_003872486.1-PF00183) is shown with chains colored in pink (human) and gray (parasite) together with encircled pLDDT mapped complexes. The confidence scores for the AF2 and AF3 models are 0.84 and 0.80, respectively, with DockQ at 0.27, irmsd at 8.1 Å, and lrmsd at 34.4 Å.

In several cases, AF2 and AF3 predicted completely different binding modes, yet both models achieved high confidence scores (yellow and blue points in **Figure 3A**). To quantitatively assess this observation, we calculated the differences in confidence scores between AF2 and AF3 models and plotted them against their DockQ scores. This yielded an r^2 of 0.001 (**Figure 3B**), highlighting the lack of a significant relationship between AF scores and model similarity. As an example, the predicted interfaces of a human and *L. mexicana* complex with 0.84 AF2 and 0.80 AF3 confidence scores could have an interface offset as large as 8 Å (irmsd).

Table 1. Structural Comparison of the Best AF2 and AF3 Models with confidence scores above 0.8, including confidence score differences, and DockQ, interface root-mean-square deviation (irmsd), and ligand root-mean-square deviation (Lrmsd) calculated using the DockQ tool (Basu & Wallner, 2016).

No	Parasite name	Host-parasite ids	AF2 confrank	AF3 confrank	Confrank diff. (AF3-AF2)	DockQ	irmsd	Lrmsd
1	*G.lamblia*	ENSP00000357981-PF08246_XP_001706373.1	0.90	0.77	-0.13	0.75	1.69	2.75
2	*G.lamblia*	ENSP00000357981-PF08246_XP_001708083.1	0.90	0.62	-0.28	0.81	1.01	3.73
3	*G.lamblia*	ENSP00000357981-PF08246_XP_001704877.1	0.89	0.79	-0.10	0.86	0.89	2.38
4	*G.lamblia*	ENSP00000357981-PF028246_XP_001709033.1	0.89	0.75	-0.13	0.89	0.79	2.48
5	*G.lamblia*	ENSP00000357981-PF08246_XP_001704582.1	0.88	0.78	-0.11	0.90	0.58	1.44
6	*L.braziliensis*	ENSP00000357998_XP_001566161.1	0.89	0.71	-0.18	0.86	1.05	1.72
7	*L.donovani*	ENSP00000366179_XP_003862230.1	0.86	0.68	-0.18	0.71	1.56	3.16
8	*L.infantum*	ENSP00000249356_XP_001470161.1	0.88	0.71	-0.17	0.89	0.65	1.85
9	*L.infantum*	ENSP00000309334_XP_001463268.1	0.84	0.74	-0.11	0.92	0.52	1.82
10	*L.mexicana*	ENSP00000365991_XP_003876961.1	0.89	0.81	-0.07	0.91	0.61	1.45
11	*L.mexicana*	ENSP00000357998_XP_003876961.1	0.88	0.74	-0.14	0.93	0.63	1.01
12	*L.mexicana*	ENSP00000325875-PF02518_XP_003872486.1-PF00183	0.85	0.84	-0.01	0.24	9.84	31.98
13	*L.mexicana*	ENSP00000335153-PF02518_XP_003872486.1-PF00183	0.84	0.80	-0.04	0.27	8.06	34.36
14	*P.falciparum*	ENSP00000249356_PFI0875w	0.91	0.88	-0.03	0.92	0.58	0.98
15	*P.falciparum*	ENSP00000365991_PFI0875w	0.90	0.83	-0.07	0.54	1.14	23.44
16	*P.falciparum*	ENSP00000366179_PFI0875w	0.90	0.81	-0.10	0.65	0.53	22.45
17	*P.falciparum*	ENSP00000357998_PFI0875w	0.90	0.87	-0.03	0.92	0.63	1.31
18	*P.falciparum*	ENSP00000265028_PFI0875w	0.89	0.73	-0.16	0.44	1.47	27.55
19	*P.falciparum*	ENSP00000277865-PF02812_PF14_0286	0.87	0.33	-0.54	0.02	16.71	36.46
20	*P.falciparum*	ENSP00000327589-PF02812_PF14_0286	0.87	0.30	-0.57	0.03	15.90	33.99
21	*P.knowlesi*	ENSP00000249356_PKH_071520	0.91	0.87	-0.04	0.62	0.60	25.28
22	*P.knowlesi*	ENSP00000354554_PKH_124300	0.88	0.40	-0.48	0.87	0.76	3.10
23	*P.vivax*	ENSP00000249356_PVX_099315	0.91	0.87	-0.04	0.60	0.77	25.30
24	*P.vivax*	ENSP00000262061_PVX_113995	0.87	0.55	-0.32	0.57	1.78	7.06
25	*P.vivax*	ENSP00000354554_PVX_117325	0.87	0.41	-0.47	0.89	0.68	2.52
26	*S.mansoni*	ENSP00000348924_Smp_035980__mRNA	0.87	0.64	-0.22	0.55	9.01	4.47
27	*S.mansoni*	ENSP00000353074_Smp_035980__mRNA	0.87	0.68	-0.19	0.51	7.60	5.72
28	*S.mansoni*	ENSP00000244601_Smp_035980__mRNA	0.86	0.74	-0.12	0.31	13.85	13.54
29	*S.mansoni*	ENSP00000366962_Smp_035980__mRNA	0.86	0.65	-0.21	0.50	8.30	6.12
30	*T.brucei*	ENSP00000252725_EAN78692	0.83	0.77	-0.05	0.89	0.87	0.74
31	*T.cruzi*	ENSP00000296581_XP_808372.1	0.87	0.87	0.00	0.81	1.07	2.04
32	*T.cruzi*	ENSP00000380349_XP_816718.1	0.80	0.48	-0.32	0.86	0.77	1.89
33	*T.gondii*	ENSP00000358931_TGME49_008820	0.92	0.24	-0.68	0.63	2.11	4.85
34	*T.gondii*	ENSP00000369456_TGME49_002300	0.90	0.82	-0.08	0.94	0.52	2.30
35	*T.gondii*	ENSP00000249356_TGME49_111720	0.89	0.87	-0.02	0.92	0.64	0.85
36	*T.gondii*	ENSP00000354554_TGME49_046540	0.88	0.35	-0.53	0.89	0.71	2.80

37	*T. gondii*	ENSP00000309565_TGME49_046540	0.87	0.46	-0.41	0.90	0.74	0.77
38	*T. gondii*	ENSP00000307786_TGME49_046540	0.83	0.28	-0.55	0.33	3.58	6.03
39	*T. gondii*	ENSP00000227378_TGME49_111720	0.80	0.60	-0.20	0.64	1.51	4.38
40	*T. spiralis*	ENSP00000396937_EFV54008	0.84	0.71	-0.13	0.53	1.25	34.61
41	*T. cruzi*	ENSP00000337459_XP_810491.1	0.31	0.88	0.57	0.10	6.09	15.41
42	*T. cruzi*	ENSP00000300413_XP_808372.1	0.77	0.88	0.11	0.80	1.21	1.76

Conclusion

Our findings show that, in the context of human-parasite interactions, AF2 and AF3 produce distinct models and score distributions, with no significant correlation between their top-ranked predictions. We recognize that predicting host-parasite interactions is particularly challenging due to the absence of typical coevolutionary signals. Nevertheless, the differences in scores and models between AF2 and AF3 are striking and warrant further investigation, which could only be performed upon the release of AF3. Until then, we plan to extend our benchmarking with the results of AF2 v2.1 and AF2 v2.3 modeling.

Important Note

This study builds on our previous TROPIC project, funded by Turkish and Colombian Scientific Research Councils (TÜBİTAK and MINCIENCIAS). During the TROPIC project, we performed host-parasite interaction modeling with AF2-Multimer v2.2 (the best option at the time) and made our results publicly available at https://github.com/CSB-KaracaLab/TROPIC . As AF2v2.2 were readily available, we used its scores/models in our benchmarking.

Acknowledgements

AF2 and AF3 modeling were conducted using the TUBITAK ULAKBIM, High Performance and Grid Computing Center (TRUBA resources) and the AF3 webserver (https://alphafoldserver.com), respectively. AF3 models were generated on the AF3 webserver between 24 June 2024 and 03 July 2024. The raw data for our interaction predictions are provided in the Supporting Information to assist further analysis. Finally, we finally would like to thank Beyza Kaynarca, Buse Şahin, and Yağmur Cavaş for their assistance in running the AF2 and AF3 modeling.

REFERENCES

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., Bodenstein, S. W., Evans, D. A., Hung, C.-C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., Arvaniti, E., … Jumper, J. M. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, *630*(8016), 493–500. https://doi.org/10.1038/s41586-024-07487-w

Basu, S., & Wallner, B. (2016). DockQ: A quality measure for protein-protein docking models. *PLoS ONE*, *11*(8). https://doi.org/10.1371/journal.pone.0161879

Cuesta-Astroz, Y., & Oliveira, G. (2018). *Computational and Experimental Approaches to Predict Host–Parasite Protein–Protein Interactions* (pp. 153–173). https://doi.org/10.1007/978-1-4939-8618-7_7

Cuesta-Astroz, Y., Santos, A., Oliveira, G., & Jensen, L. J. (2019). Analysis of Predicted Host–Parasite Interactomes Reveals Commonalities and Specificities Related to Parasitic Lifestyle and Tissues Tropism. *Frontiers in Immunology*, *10*. https://doi.org/10.3389/fimmu.2019.00212

Dampier, W., Evans, P., Ungar, L., & Tozeren, A. (2009). Host sequence motifs shared by HIV predict response to antiretroviral therapy. *BMC Medical Genomics*, *2*. https://doi.org/10.1186/1755-8794-2-47

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

Mehmood, K., Zhang, H., Sabir, A. J., Abbas, R. Z., Ijaz, M., Durrani, A. Z., Saleem, M. H., Ur Rehman, M., Iqbal, M. K., Wang, Y., Ahmad, H. I., Abbas, T., Hussain, R., Ghori, M. T., Ali, S., Khan, A. U., & Li, J. (2017). A review on epidemiology, global prevalence and economical losses of fasciolosis in ruminants. *Microbial Pathogenesis*, *109*, 253–262. https://doi.org/10.1016/j.micpath.2017.06.006

Merrifield, M., Hotez, P. J., Beaumier, C. M., Gillespie, P., Strych, U., Hayward, T., & Bottazzi, M. E. (2016). Advancing a vaccine to prevent human schistosomiasis. *Vaccine*, *34*(26), 2988–2991. https://doi.org/10.1016/j.vaccine.2016.03.079