https://doi.org/10.26565/2220-637X-2024-43-01 ISSN 2220-637X
УДК: 544.165+577.322

BENCHMARKING GOOGLE DEEPMIND`S AlphaFold 3 PERFORMANCE FOR PROTEIN 3D-STRUCTURE PREDICTION

Y. Duma[a] , A. Kyrychenko[b]

V. N. Karazin Kharkiv National University, School of Chemistry, 4 Svobody sqr., Kharkiv, 61022 Ukraine

a) yelyzaveta.duma@student.karazin.ua https://orcid.org/0009-0004-0873-6490
b) a.v.kyrychenko@karazin.ua https://orcid.org/0000-0002-6223-0990

The 3D structure of proteins is directly linked to their function, making its determination crucial for understanding biological processes and addressing issues related to human health and life sciences. Despite the continuous experimental acquisition of new protein structures, there remains a significant gap between the number of protein sequences available and those that have an established experimental high-resolution tertiary structure. Several computational approaches have focused on predicting protein structures using either templates or empirical force field modeling. In recent years, various methods have been combined to address the individual limitations of these approaches, leading to the development of AlphaFold 3 (AF3) by Google DeepMind. AF3 enables prediction of 3D protein structures with high accuracy based on its amino acid sequence. In this study, we benchmarked applicability, performance, and limitations of AF3 for predicting 3D structure of a broad series of proteins, including SARS-CoV-2 coronavirus proteins, other bacterial and viral proteins, as well as some plant enzymes. We found that AlphaFold 3 could capture the overall backbone features of the most examined proteins in terms of small deviation from available X-ray structures. Some minor miss-folding of N- and C-terminal segments were common, which, often, did not affect biological roles of the studied proteins. In cases involving protein dimers or higher-order oligomers, there are notable differences between the predicted AF3 models of a single-chain monomer and their corresponding experimental structures. These discrepancies are particularly evident in regions related to protein dimerization, assembly, and binding interfaces. Ultimately, while capturing the overall fold, predicting the complex structure of the Spike glycoprotein is still beyond the current capabilities of AF3.

Keywords: protein folding, enzyme, 3D structure, drug design, artificial intelligence, AlphaFold 3.

Introduction

Tertiary protein structures are typically determined using techniques, such as X-ray crystallography, electron cryomicroscopy (cryo-EM), and nuclear magnetic resonance (NMR). However, these methods are complex, time-consuming, and expensive, and they often do not capture the protein in its native form. Given these challenges, it is not surprising that the number of proteins with resolved tertiary structures is relatively small — only ~202000 entries in the Protein Data Bank (PDB) — compared to the vast number of known protein sequences, which stands at 240 million entries in UniProt DataBase. This disparity presents an open challenge, highlighting the need for innovative approaches in protein structure prediction, particularly methods that utilize computational models to predict 3D protein structures from polypeptide sequences .

Several algorithms and web servers have been developed to enhance protein structure prediction . The significance of these efforts is highlighted by their impact on various applications, including rational drug design, mutational studies, structural comparisons, evolutionary analysis, and folding studies . However, despite significant advancements, leading to gradual improvements in accuracy, the fundamental problem of predicting protein structures remained largely unsolved for decades . Since the introduction of AlphaFold by Google DeepMind in 2018 and AlphaFold 1 in 2019, these systems have rapidly outperformed other computational methods in predicting protein structures. Following this, AlphaFold 2 was released in 2020 and revolutionized the field of protein structure prediction . The latest version, AlphaFold 3, was presented in May 2024 and represents a significant advancement . The major improvements in AlphaFold 3 stem from a substantial evolution in its network architecture and training procedures, allowing it to represent a wider range of chemical structures and enhancing the efficiency of learning from available data. Additionally, AlphaFold 3 incorporates a diffusion model to predict the coordinates of each atom, replacing the structure module used in Al-

phaFold 2, which operated at the amino acid level . So, recent studies have demonstrated that AF3 can achieve experimental accuracy in an increasing number of cases . Finally, considering all these efforts, AlphaFold`s developers, David Baker, Demis Hassabis and John Jumper, have been awarded the Nobel Prize for Chemistry 2024 for their work on computational protein design and protein structure prediction.

In this study, we benchmarked AlphaFold 3's applicability and limitations for predicting 3D tertiary structure of a broad series of proteins, including SARS-CoV-2 virus enzymes, bacterial and viral proteins, as well as some plant enzymes.

Results and Discussion

To benchmark performance of AF3 for protein structure prediction we selected diverse families of functional proteins and enzymes, such as SARS-CoV-2 proteins, some proteins from *Mycobacterium tuberculosis*, as well as other crucial enzymes. Selection criteria were based on biological importance and availability of experimental high-resolution 3D structures. The selected proteins are composed of various conservative secondary structures, such as α-helixes and β-pleated sheets, offering some computational challenge for their structure evaluations.

Key proteins of the SARS-CoV-2 coronavirus

Coronavirus disease 2019 (COVID-19) is caused by a novel member of the *Coronaviridae* family known as Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). The genome of this virus contains both structural and non-structural proteins (NSps). The structural proteins include Spike (S), membrane (M), hemagglutinin-esterase (H), and envelope (E) proteins, while the NSPs consist of accessory proteins and replicase proteins (Fig. 1) .

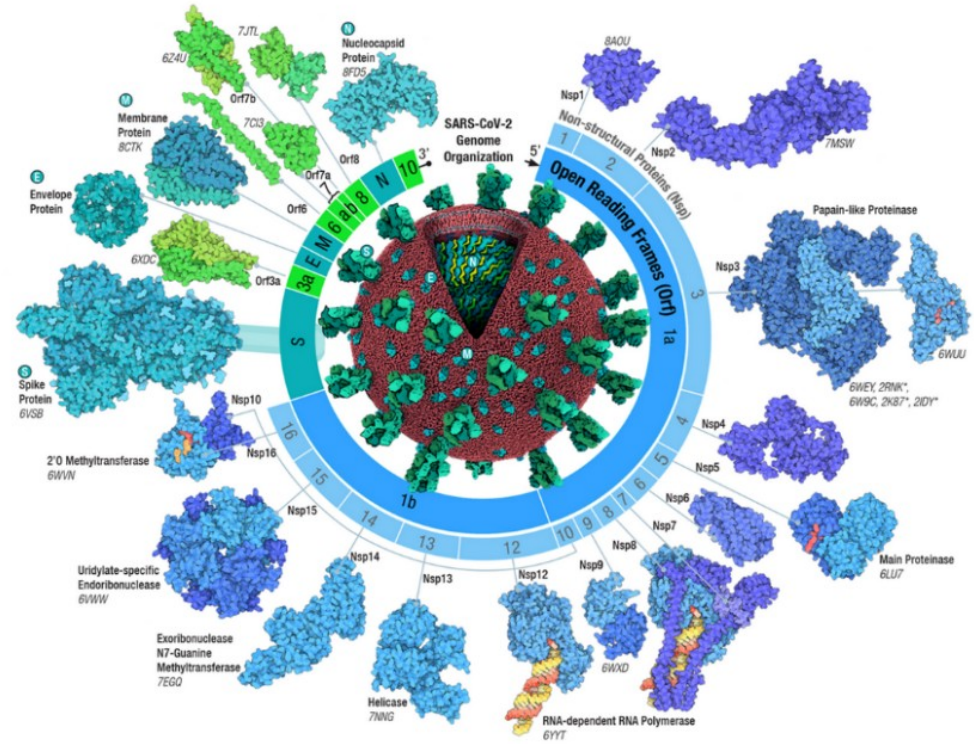

Figure 1. Architecture of the SARS-CoV-2 genome and proteome, demonstrating key enzymes of the coronavirus. Adopted with permission from . Copyright © 2021 Wiley Periodicals LLC.

The structural and Nsp components of SARS-CoV-2 are crucial for its infectivity, and some may play significant roles in the development of chronic diseases such as cancer, coagulation disorders, neurodegenerative disorders, and cardiovascular diseases . Additionally, the proteins of SARS-CoV-2 interact with targets like the angiotensin-converting enzyme 2 (ACE2) receptor.

A structural comparison of several essential SARS-CoV-2 proteins, including ten variants of the Spike protein and cysteine proteases, was recently conducted using an earlier version of AlphaFold 2 . In this study, some of these structures were reanalyzed using the newer version, AlphaFold 3.

The first set of the studied proteins were selected from SARS-CoV-2 coronavirus' lifecycle. SARS-CoV-2 has a genome that consists of a long RNA strand, making it one of the largest among all RNA viruses. When the virus infects a cell, this genome acts like messenger RNA (mRNA) and directs the synthesis of long polyproteins (Fig. 1). These polyproteins contain the components necessary for the virus to replicate and produce new viruses. The proteins include a replication/transcription complex that generates more RNA, several structural proteins that assemble new virions, and two cysteine proteases. These proteases are crucial as they cut the polyproteins into the functional pieces needed for the virus's lifecycle. In this set, the main protease M^{pro} (PDB: 6LU7), papain-like protease PL^{pro} (PDB: 7SGW), Nsp12-Nsp7-Nsp8 complex of non-structural proteins (Nsp) (PDB: 7BV1) were considered (Fig. 2). Additionally, Spike protein of recent Omicron variant of the virus is also considered.

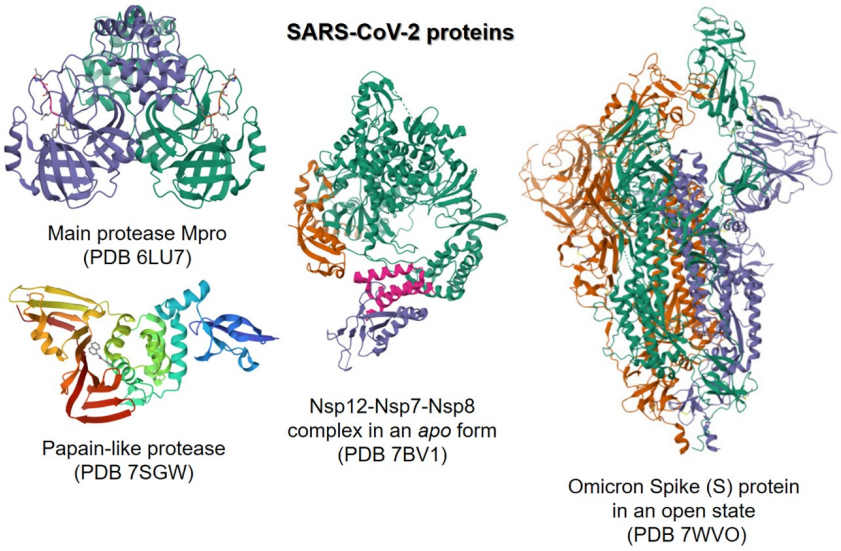

Figure 2. The X-ray or cryo-EM structures of some essential proteins of the SARS-CoV-2 virus. The corresponding PDB codes are given in brackets.

Some recent studies utilizing X-ray crystallography cryo-electron microscopy have reported the high-resolution structure of some essential SARS-CoV-2 proteases, such as main and papain-like proteases, Nsp complex was well as Spike-proteins, as summarized in Fig. 2. Therefore, these selected proteins of the SARS-CoV-2 family were subjected to the AF3 evaluation using their FASTA sequences (Table 1).

Table 1. Amino acid sequences of the M^{pro} and PL^{pro} proteases used for AF3-prediction.

Protein, PDB Code	FASTA Sequence
Main protease M^{pro}, 6LU7	SGFRKMAFPSGKVEGCMVQVTCGTTTLNGLWLDDVVYCPRHVICTSEDML-NPNY EDLLIRKSNHNFLVQAGNVQLRVIGHSMQNCVLKLKVDTANPKTP-KYKFVRIQPGQ TFSVLACYNGSPSGVYQCAMRPNFTIKGS-FLNGSCGSVGFNIDYDCVSFCYMHHM ELPTGVHAGTDLEGNFYGPFVDRQ-TAQAAGTDTTITVNVLAWLYAAVINGDRWFLN RFTTTLNDFNL-VAMKYNYEPLTQDHVDILGPLSAQTGIAVLDMCASLKELLQNGMN GRTIL-GSALLEDEFTPFDVVRQCSGVTFQ
Papain-like protease PL^{pro},	SNAEVRTIKVFTTVDNINLHTQVVDMSMTYGQQFGPTYLDGADVTKIKPHN-SHEGK TFYVLPNDDTLRVEAFEYYHTTDPSFLGRYM-SALNHTKKWKYPQVNGLTSIKWADN NCYLATALLTLQQIELKFNPPALQ-

7SGW	DAYYRARAGEAANFCALILAYCNKTVGELGDVR ETMSYLFQHANLDSCK-RVLNVVCKTCGQQQTTLKGVEAVMYMGTLSYEQFKKGVQI PCTCGKQATKYLVQQESPFVMMSAPPAQYELKHGTFTCASEYT-GNYQCGHYKHITS KETLYCIDGALLTKSSEYKGPITDVFYKENSYTTTIK

Main protease M^{pro}. The main protease, known as M^{pro} (a 3C-like protease), is responsible for cleaving eleven specific sites on the two SARS-CoV-2 polyproteins. M^{pro} is comprised of three domains: a chymotrypsin-like domain (Domain I), a 3C-protease-like domain (Domain II), and a third domain (Domain III) that contains five α-helices . The binding pocket of M^{pro} is situated between Domains I and II and features a catalytic Cys-His dyad, consisting of Cys145 and His41 . Therefore, M^{pro} has attracted essential attention and is considered an ideal target for developing antiviral agents .

Fig. 3 shows the overlap of the AF3-predictuted structure (*green*) of the M^{pro} protease with its native X-ray structure (PDB 6LU7, *blue*) in two different views. We estimate quantitative measure of AF3 performance as the C_α root mean square deviation (C_α RMSD) between the predicted AF3 model and experimental X-ray structure for entire protein, calculated by PyMOL tool *super*. The AF3 predicted models show an average global C_α RMSD of 0.411 Å from the experimental structures, suggesting that the global structure properties of M^{pro} have been captured by AlphaFold3. In addition, the RMSD of main chain atoms between the predicted AF3 model and experimental X-ray structure for M^{pro} were also calculated by Discovery Studio software and found to be 2.796 Å, as summarized in Table 2. Some key metrics of AF3 confidence measures for its structural predictions are also given in Table 2.

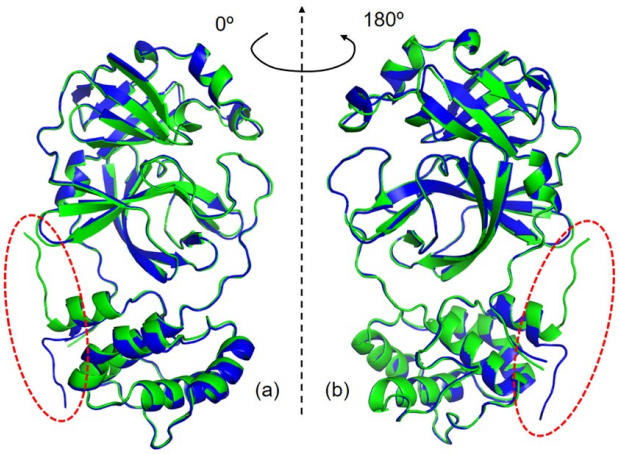

Figure 3. Overlay of the native X-ray structure of the M^{pro} protease (PDB 6LU7, *blue*) and the AF3-predictuted structure (*green*) in a side (a) and 180º-rotated (b) views, respectively. Some disagreement between the AF3-predicted model and the experimental structure are highlighted in *red* circles.

Papain-like protease PL^{pro}. In SARS-CoV-2, the Nsp3 protein consists of 1,945 residues (approximately 212 kDa), and PL^{pro} is a domain of this large, multidomain protein that plays a crucial role in the replication-transcription complex (RTC) . PL^{pro} is located in Nsp3, positioned between the SARS unique domain (SUD/HVR) and a nucleic acid-binding domain (NAB). This enzyme is highly conserved and is found in all coronaviruses .

PL^{pro} exhibits multiple functions, including proteolytic activity, processing of Pp1a/Pp1ab, and deubiquitination, which involves efficiently disassembling mono-ubiquitin, polyubiquitin chains, and branched polyubiquitin chains . It features a canonical cysteine protease catalytic triad, making it a target for anti-COVID-19 drug development .

Fig. 4 demonstrates comparison of the AF3-predictuted structure (*green*) of the PL^{pro} protease with its native X-ray structure (PDB 7SGW, *blue*) in two different views. The RMSD between the predicted AF3 model and experimental X-ray structure for PL^{pro} was found to be 0.456 Å (PyMOL) and

0.522 Å (Discover Studio), suggesting that AlphaFold3 maintains its excellent performance in predicting overall 3D structures of monomeric single-chain proteins.

Table 2. Comparison of AF3-predicted models for the studied proteins, focusing on AF3 confidence metrics and their deviation from experimental results.

Protein, PBD	AF3 confidence metrics			RMSD (Å)	
	Chain PTM[1]	Chain pair PAE min[2]	Ranking Score[3]	C_α (PyMOL)	Main chain (Discovery Studio)
Main protease M^{pro}, 6LU7	0.93	0.76	0.93	0.411	2.796
Papain-like protease PL^{pro}, 7SGW	0.92	0.76	0.92	0.456	0.522
Nsp12-Nsp7-Nsp8 complex, 7BV1	0.94	0.76	0.95	0.614	1.303
Omicron Spike (S) protein, 7WVO	0.76	0.76	0.81	2.046	9.488
Oxidoreductase DprE1, 4FF6	0.90	0.76	0.94	0.236	1.355
Dihydrofolate reductase DHFR, 1DF7	0.94	0.76	0.94	0.408	0.486
Pantothenate kinase PanK, 4BFT	0.86	0.76	0.89	0.629	2.560
Enoyl-ACP reductase InhA, 4TRO	0.95	0.76	0.95	0.385	1.421
Sirtuin SIRT1, 4I5I	0.83	0.76	0.87	1.279	3.301
Sirtuin SIRT2, 4RMG	0.92	0.76	0.92	0.586	1.938
Diphtheria toxin, 7K7E	0.89	0.76	0.89	0.792	16.034
Paenibacillus polymyxa β-Glucosidase B, 2O9R	0.96	0.76	0.97	0.371	0.736

[1] – A per-chain pTM score, which suggests confidence for the individual chain structure.

[2] – Minimum predicted aligned error (PAE) for chain pairs, indicating reliability of structural alignment.

[3] - A general confidence score, which summarizes different predictions.

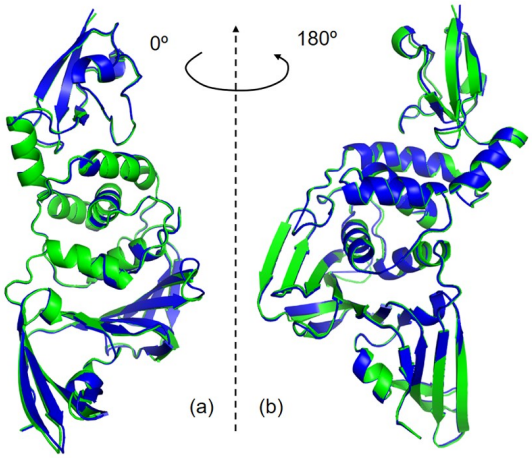

Figure 4. Overlay of the native Cryo-EM structure (PDB 7SGW, *blue*) of the PL^{pro} protease and the AF3-predictuted structure (*green*) in a side (a) and 180°-rotated (b) views, respectively.

Nsp12-Nsp7-Nsp8 complex. The genome organization of SARS-CoV-2 comprises sixteen non-structural proteins, Nsp1–16, which are highly conserved among coronaviruses (Fig. 1) . Non-structural proteins from this family present essential challenge for 3D structure prediction.

Fig. 5 illustrates the AF3-predicted model of the Nsp12-Nsp7-Nsp8 complex, which is part of the RNA-dependent RNA polymerase (RdRp) enzyme of SARS-CoV-2. When compared to the experimental Cryo-EM structure, which was measured at a resolution of 2.8 Å, the AF3 model displayed a more compact fold, with two terminal protein segments predicted to differ in several aspects. While the C_α RMSD measured using PyMOL was relatively small at 0.614 Å, the main chain RMSD showed a deviation of 1.303 Å (Table 2). This suggests that predicting the structures of viral nonstructural proteins remains a significant computational challenge.

Spike (S) protein. The SARS-CoV-2 virus is covered by a large number of glycosylated Spike (S) proteins, which bind to the host cell receptor known as angiotensin-converting enzyme 2 (ACE2). This binding mediates the entry of the virus into host cells (see Fig. 1) . The S protein is crucial for receptor recognition and the process of cell membrane fusion, consisting of two subunits: S1 and S2. The S1 subunit contains a receptor-binding domain that recognizes and attaches to ACE2, while the S2 subunit facilitates the fusion of the viral and host cell membranes by forming a six-helical bundle through its two-heptad repeat domain . When the S protein binds to the receptor, it is activated by the host cell's TM protease serine 2 (TMPRSS2), a type 2 transmembrane serine protease. This activation promotes the entry of the virus into the cell. Once inside, the viral RNA is released, and the cell translates polyproteins from the RNA genome. The replication and transcription of the viral RNA genome occur via protein cleavage and the assembly of the replicase–transcriptase complex. Consequently, the design of antibodies or small-molecule entry blockers targeting S proteins is highly significant for preventing the virus and treating COVID-19 .

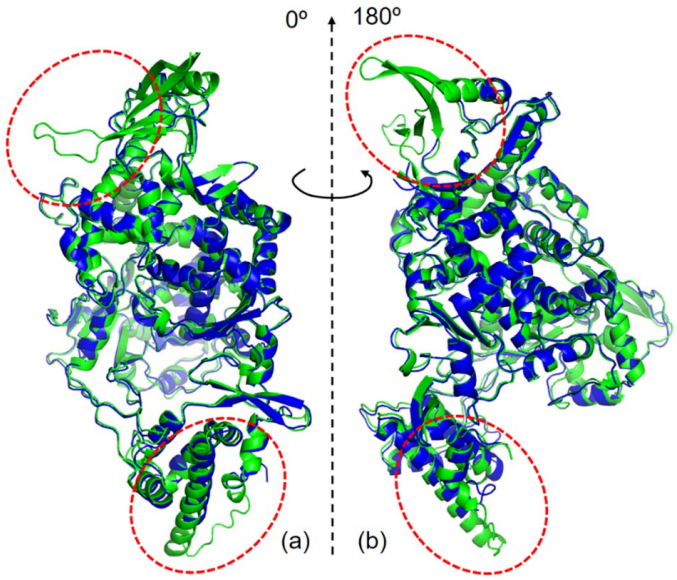

Figure 5. Overlay of the native Cryo-EM structure of the Nsp12-Nsp7-Nsp8 RdRp complex in an *apo*-form (PDB 7BV1, *blue*) and the AF3-predictuted structure (*green*) in a side (a) and 180°-rotated (b) views, respectively. Essential disagreements between the AF3-predicted model and the experimental structure are highlighted in *red* dotted circles.

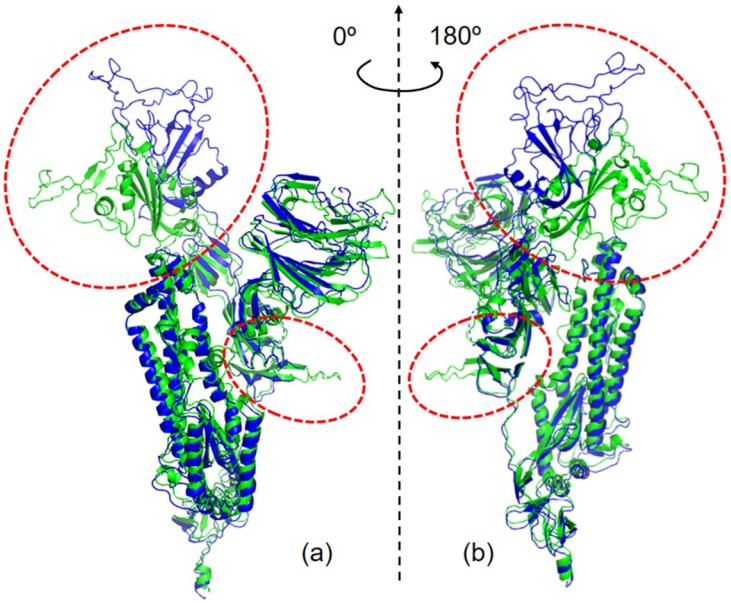

Figure 6. Overlay of the native X-ray structure of the Omicron Spike (S) protein in an open state (PDB 7WVO, *blue*) and the AF3-predictuted structure (*green*) in a side (a) and 180°-rotated (b) views, respectively. Essential disagreements between the AF3-predicted model and the experimental structure are highlighted in *red* dotted areas.

Considering the complex structure of the S-protein, which consists of 1,258 amino acids (Fig. 2), this case represents one of the most challenging tests for AF3 to date. Figure 6 illustrates the model of the S-protein predicted by AF3 and compares it with the corresponding cryo-EM structure. It is evident that AF3 successfully captured most secondary structure features, such as α-helices and β-sheets. However, some crucial segments of the S-protein, particularly the receptor-binding domain (RBD) that plays a central role in ACE2 recognition, were predicted with unexpectedly low accuracy (Table 2). Therefore, we can conclude that predicting the structure of the Spike glycoprotein remains beyond the capabilities of AF3 at this time.

Essential proteins from *Mycobacterium tuberculosis*

Mycobacterium tuberculosis (Mtb), the bacterium responsible for tuberculosis (TB), is a resilient pathogen that has latently infected about one-third of the world's population. However, traditional TB treatment regimens are increasingly inadequate in addressing the rising threat of drug resistance. This situation has prompted the development of innovative anti-tuberculosis agents, particularly those targeting new proteins. The Mtb genome contains approximately 4,000 predicted proteins, many of which are enzymes involved in various cellular metabolic processes. Notably, over 200 of these proteins play a role in fatty acid biosynthesis. This process is crucial for building the cell envelope and is closely linked to the pathogenesis and drug resistance of mycobacteria .

Therefore, the second set of the proteins were selected from *Mycobacterium tuberculosis* (Mtb) family , as summarized in Fig. 7.

Oxidoreductase DprE1. Decaprenylphosphoryl-β-D-ribose 2'-oxidoreductase (DprE1) is a highly conserved enzyme that contains FAD. Together with its downstream NADH-dependent reductase, DprE2, it facilitates the conversion of decaprenylphosphoryl-β-D-ribofuranose (DPR) into decaprenylphosphoryl-β-D-arabinofuranose (DPA). DprE1 is a promising target for the development of new anti-tubercular drugs, and researchers are actively investigating various inhibitors of DprE1 .

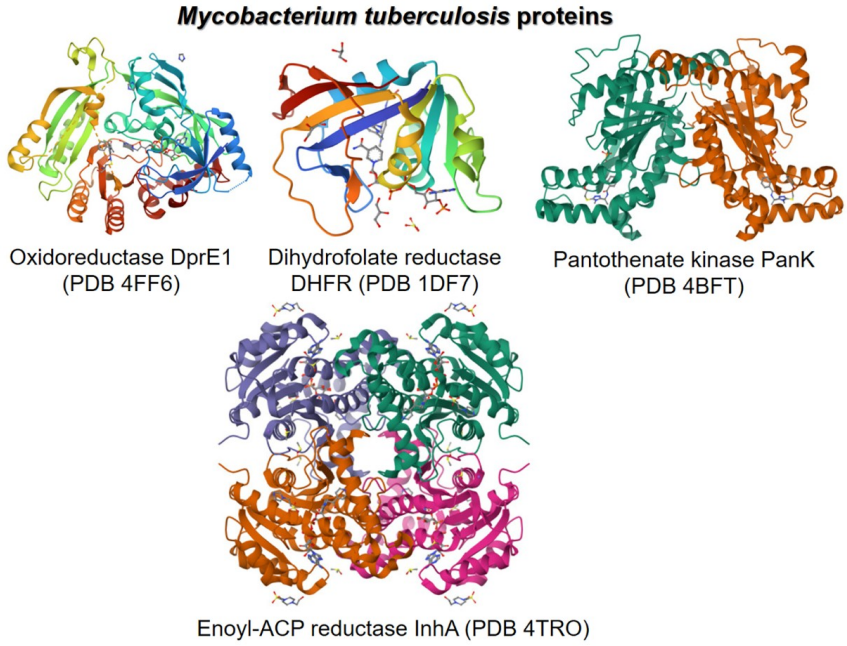

Figure 7. The X-ray or cryo-EM structures of some essential proteins of the *Mycobacterium tuberculosis*. The corresponding PDB codes are given in brackets.

Fig. 8 shows the AF3-predicted model of DprE1 enzyme that revealed the very good agreement with the corresponding X-ray structure. The PyMOL C_α RMS deviation was found to be 0.236 Å only. Some larger RMSD of 1.355 Å for the main chain DprE1 is because of contribution of non-structured loops (Fig. 8a-b). It should also be noted that the AF3-predicted model was able to reconstruct a large protein segment of 269-283, missed in the original X-ray structure.

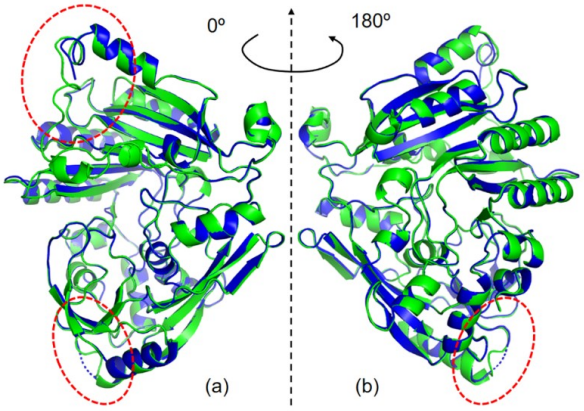

Figure 8. Overlay of the native X-ray structure of the oxidoreductase DprE1 measured with resolution of 2.6 Å (PDB 4FF6, *blue*) and the AF3-predictuted structure (*green*) in a side (a) and 180º-rotated (b) views, respectively. Some disagreements between the AF3-predicted model and the experimental structure are highlighted in *red* dotted areas.

Dihydrofolate reductase (DHFR). The dihydrofolate reductase (DHFR) enzyme catalyzes the NADPH-dependent reduction of dihydrofolate to tetrahydrofolate. This reaction produces a precursor for cofactors necessary for synthesizing thymidylate, purine nucleotides, methionine, serine, and glycine, which are required for DNA, RNA, and protein synthesis . Therefore, DHFR is a promising target for developing novel antitubercular agents .

Fig. 9 demonstrates the strong correlation between the AF3-predicted model of DHFR and the available X-ray structure. The calculated RMSDs of 0.408 and 0.486 Å rank DHFR among the best predicted models analyzed in Table 2.

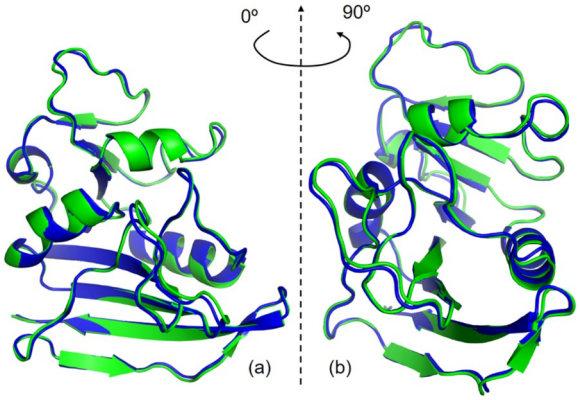

Figure 9. Overlay of the native X-ray structure of the dihydrofolate reductase DHFR (PDB 1DF7, *blue*) and the AF3-predictuted structure (*green*) in a side (a) and 90º-rotated (b) views, respectively.

Pantothenate kinase PanK. Pantothenate kinase (PanK) is responsible for the phosphorylation of pantothenate, which is the first committed and rate-limiting step in the biosynthesis of coenzyme A (CoA). Research has shown that the type I isoform, encoded by the coaA gene, is an essential pantothenate kinase in Mycobacterium tuberculosis. This crucial information has been utilized to screen large libraries for the identification of different classes of PanK inhibitors that operate through various mechanisms .

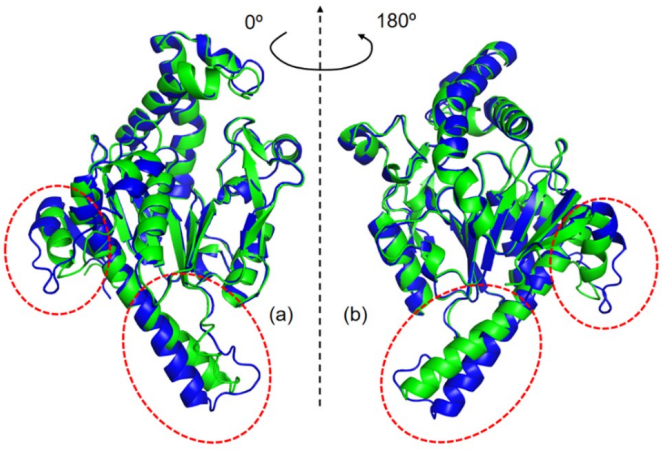

Figure 10. Overlay of the native X-ray structure of the Pantothenate kinase PanK measured with resolution of 2.29 Å (PDB – 4BFT), *blue*) and the AF3-predictuted structure (*green*) in a side (a) and 180º-rotated (b) views, respectively. Essential disagreements between the AF3-predicted model and the experimental structure are highlighted in *red* dotted areas.

Fig. 10 illustrates the strong correlation between the AF3-predicted model of PanK and the available X-ray structure in most segments, with the exception of α-helix 4. In the X-ray structure, α-helix 4 is crucial for binding between PanK chains A and B, as it forms a helix-helix dimer (see Figure 7). Consequently, the proper conformation of α-helix 4 is maintained in the PanK dimer, which is why this protein segment is not accurately predicted single-chain PanK by AF3. The calculated Cα RMSD is 0.629 Å, while the main chain RMSD of 2.560 Å indicates this deviation (refer to Table 2).

Enoyl-acyl carrier protein reductase InhA. Enoyl-ACP reductase (InhA) is a vital biosynthetic enzyme involved in the production of mycolic acids (MAs). It catalyzes the NADH-dependent reduction of long-chain trans-2-enoyl-ACP in the type II fatty acid synthesis of *Mycobacterium tuberculosis* (*Mtb*). Importantly, *Mtb* InhA does not have a human counterpart, which may reduce the risk of toxicity from inhibitors targeting this enzyme. As a result, InhA has been established as a reliable

target for the design and discovery of novel agents against *Mtb*, particularly for frontline and second-line antitubercular drugs such as isoniazid and ethionamide .

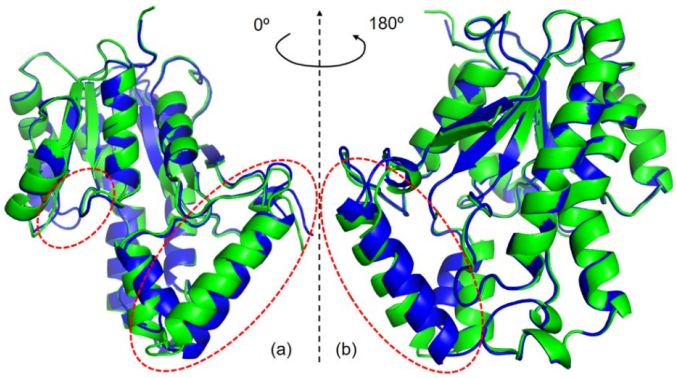

Figure 11. Overlay of the native X-ray structure of the Enoyl-ACP reductase InhA measured with resolution of 1.40 Å (PDB – 4TRO), *blue*) and the AF3-predictuted structure (*green*) in a side (a) and 180°-rotated (b) views, respectively. Essential disagreements between the AF3-predicted model and the experimental structure are highlighted in *red* dotted areas.

Fig. 11 illustrates the AF3-predicted model of InhA, which resemble some prediction features of the above mentioned PanK. In particular, AF3 performs well in most protein segments, with the exception of α-helixes 6 and 7. These α-helixes are crucial for assembling the quaternary structure of InhA composed of four chains A-C, as shown in Fig. 7. Consequently, this case further demonstrate some limitation of AF3 in reproducing 3D-structure of a single-chain protein when it is a part of complex quaternary structure, such as dimer or tetramer. The AF3 confidence score of 0.95 for InhA is also among the best, as seen in Table 2.

Results of calculations

In our recent studies, we also considered several important enzymes, which are now selected for benchmarking using AlphaFold3. These enzymes are promising targets for drug design and development, making the prediction of their structure and function critically important. Among them are silent information regulators, a plant hydrolase, and the full-length diphtheria toxin (Fig. 12).

Sirtuins. Silent Information Regulator Two Homologue One (SIRT1) is the most extensively studied member of the Sirtuin family, which is related to the class III histone deacetylases and utilizes NAD^+ as its cofactor. SIRT1 is one of seven human Sirtuins, which are conserved from archaea to mammals . These proteins play key roles in regulating transcription, genome stability, longevity, and metabolism . SIRT1 influences transcription by deacetylating various transcription factors, including PPARγ, NFκB, and the tumor suppressor protein p53. Research indicates that SIRT1, along with its histone and non-histone targets, is involved in a wide range of biological processes. These include metabolic diseases, age-related conditions, viral infections, inflammation, and the growth and metastasis of tumor cells. Consequently, modulating SIRT1 expression could offer new insights into the development of therapeutics, whether they are derived from natural sources or synthetic origins .

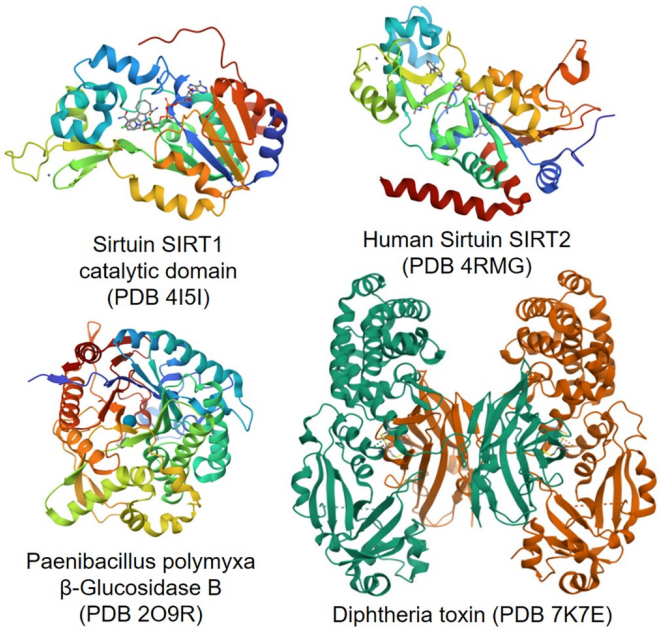

Figure 12. The X-ray or cryo-electron microscopy structures of SIRT1, SIRT2, diphtheria toxin and β-glucosidase, respectively. The corresponding PDB codes are given in brackets.

Despite the relatively simple folding of SIRT1's catalytic domain, predicting its structure with AF3 faced several systematic challenges. Fig. 13 compares the structures predicted by AF3 with those obtained through X-ray crystallography. However, there are significant deviations and inconsistencies between the two.

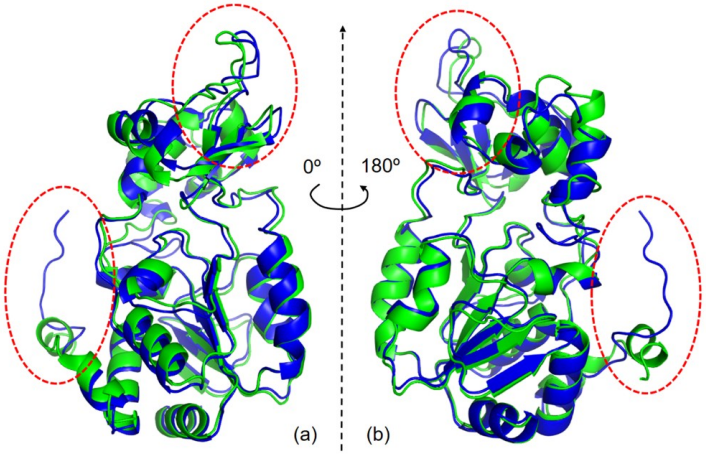

Figure 13. Overlay of the native X-ray structure of the catalytic domain of Sirtuin SIRT1 (PDB 4I5I, *blue*) and the AF3-predictuted structure (*green*) in a side (a) and 180º-rotated (b) views, respectively. Essential disagreements between the AF3-predicted model and the experimental structure are highlighted in *red* dotted areas.

Most of these discrepancies occur in the flexible, unstructured loops that connect various secondary structural elements, such as α-helices. So, SIRT1 is characterized by the largest RMSD of 1.279 Å and 3.301 Å observed among all the studied proteins, as detailed in Table 2.

AF3 faced similar systematic challenges when predicting the 3D structure of human Sirtuin SIRT2, as summarized in Fig. 14. Several unstructured loops were not accurately reproduced. Additionally, the experimental X-ray structure omitted some sequence segments, specifically 47-51 and 379-381. As a result, the reconstruction of these unstructured segments by AF3 produced an artificial prediction of α-helices.

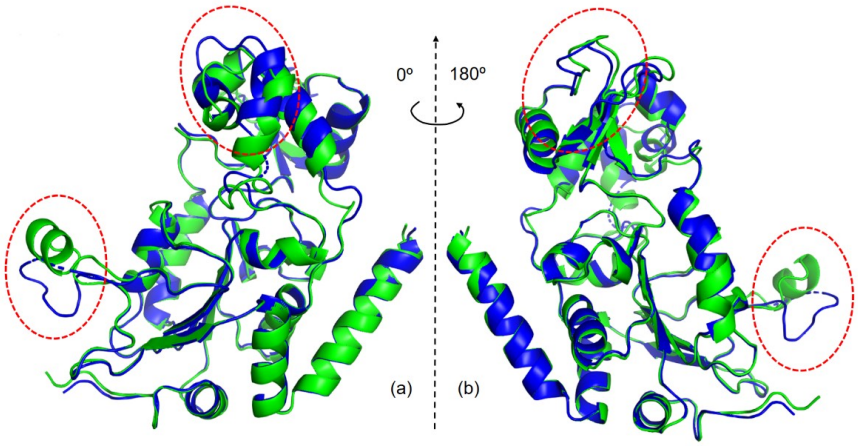

Figure 14. Overlay of the native X-ray structure of the human Sirtuin SIRT2 (PDB 4RMG, *blue*) and the AF3-predictuted structure (*green*) in a side (a) and 180º-rotated (b) views, respectively. Essential disagreements between the AF3-predicted model and the experimental structure are highlighted in *red* dotted areas.

Diphtheria toxin. Diphtheria toxin is an exotoxin produced by *Corynebacterium* that causes disease in humans by inhibiting protein synthesis. It enters the cell through a process called receptor-mediated endocytosis. Once inside, the acidification of the endosome triggers a series of conformational changes . This results in the refolding and insertion of the translocation (T) domain into the membrane, ultimately allowing the catalytic domain to be transported into the cytoplasm . Additionally, the T-domain has become a convenient model for studying pH-triggered protein-membrane interactions .

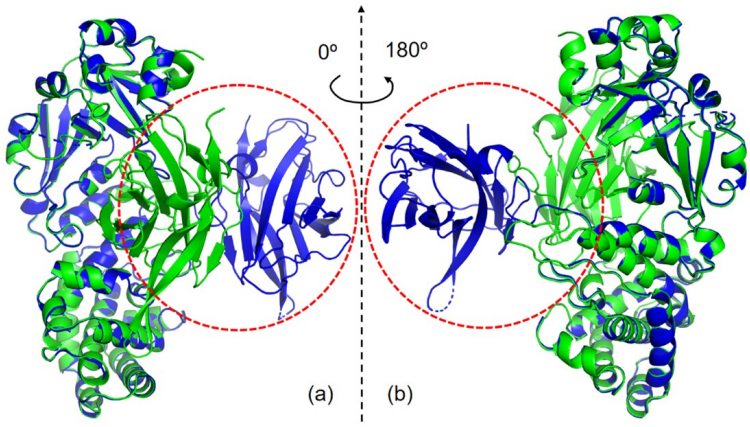

Figure 15. Overlay of the native X-ray structure of the diphtheria toxin (PDB 7K7E, *blue*) and the AF3-predictuted structure (*green*) in a side (a) and 180º-rotated (b) views, respectively. Essential disagreements between the AF3-predicted model and the experimental structure are highlighted in *red* dotted areas.

In its crystalline form, the diphtheria toxin exists as a homodimer, with its quaternary structure maintained by strongly bound interfacial segments (see Fig. 15). Predicting the 3D structure of a single chain monomer of the toxin presented expected challenges with AF3, as illustrated in Fig. 15. AF3 predicts a compact fold for the toxin; however, the C-terminal domain in its native conformation unfolds towards another part of the homodimer (see Fig. 15). All these deviations are highlighted in Table 2.

***Paenibacillus polymyxa* β-Glucosidase B**. Bacterial species that degrade cellulosic substrates produce various enzymes. *Paenibacillus polymyxa* encodes β-glucosidase B (BglB), a monomeric enzyme that functions as an exo-beta-glucosidase, hydrolyzing cellobiose and higher-degree cellodextrins . It is a 1,4-β-glucosidase. β-Glucosidase is an important target in developing substrate-specific

inhibitors, passive binders and studying protein-ligand interactions . The X-ray structure of BglB (2O9R) is available making it possible to do direct comparison to the AF3 prediction.

Figure 16. Overlay of the native X-ray structure of *Paenibacillus polymyxa* β-Glucosidase B (PDB 2O9R, *blue*) and the AF3-predictuted structure (*green*) in a side (a) and top (b) views, respectively.

Fig. 16 demonstrates the impressive performance of AF3 in predicting the 3D structure of the water-soluble single-chain enzyme, such as β-glucosidase (BglB). This enzyme consists of a bundle of nine α-helices that encircle a deeply buried catalytic site. AF3 successfully predicted the 3D structure of BglB with exceptional accuracy, as indicated by a small C_α RMSD (Root Mean Square Deviation) of 0.371 Å. Even with the presence of numerous connecting loops, AF3 surprisingly reproduced their structure well, achieving a main chain RMSD of only 0.736 Å (Table 2). The AF3 confidence score of 0.97 for this protein is also among the best, as seen in Table 2.

Conclusion

Computational prediction of protein structure from its amino acid sequences has been a central problem in molecular biology since the 1960s, when it was demonstrated that a protein's amino acid sequence determines its 3D-structure . Beyond its theoretical significance in understanding the fundamental principles governing biological macromolecules, solving this problem offers some practical implications in protein design for various medical and industrial applications .

In this study, we benchmarked applicability of AlphaFold3 for predicting 3D structure of some proteins of varying sizes and functions, including SARS-CoV-2 coronavirus proteins, bacterial and viral proteins, as well as plant enzymes. We found that AF3 could capture the overall backbone features of the most examined proteins with small RMS deviation <0.5 Å from available X-ray structures. Some common deviation issues were observed for N- and C-terminal segments, which, however, did not affect biological roles of the studied proteins. In cases involving protein dimers or higher-order oligomers, we observed notable differences between the predicted AF3 models of a single-chain monomer and their corresponding experimental structures. These structural discrepancies are particularly apparent in regions associated with protein dimerization, assembly, and binding interfaces. Ultimately, while capturing the overall fold, predicting the complex structure of the Spike glycoprotein remains beyond the current capabilities of AF3.

From a practical standpoint, AF3 is becoming a highly promising computational tool for predicting the three-dimensional structure of proteins, especially when available X-ray or Cryo-EM structures are poorly resolved or have missing sequence segments (as demonstrated in Figs. 8 and 14) . Additionally, there is a new, encouraging direction in the virtual screening of protein mutations to evaluate their impact on protein folding, protein-protein interactions, and overall activity .

Finally, the appearance of AlphaFold 3 has spurred rapid development throughout the field of Computational Biology, inspiring the creation of competing and complementary tools, such as RoseTTAFold by the Baker Lab . OpenFold is an important open-source implementation of AlphaFold, facilitating further customization and innovation, which is expected to accelerate the field in the coming years .

Acknowledgement

A. K. acknowledges the grant № 87/0062 (2021.01/0062) "Molecular design, synthesis and screening of new potential antiviral pharmaceutical ingredients for the treatment of infectious diseases COVID-19" from the National Research Foundation of Ukraine.

References

1. Dill K. A., Ozkan S. B., Weikl T. R., Chodera J. D., Voelz V. A. The protein folding problem: When will it be solved? *Curr Opin Struct Biol* **2007**, *17* (3), 342-346. https://doi.org/10.1016/j.sbi.2007.06.001
2. Nassar R., Dignon G. L., Razban R. M., Dill K. A. The protein folding problem: The role of theory. *J. Mol. Biol.* **2021**, *433* (20), 167126. https://doi.org/10.1016/j.jmb.2021.167126
3. Weller J. A., Rohs R. Structure-based drug design with a deep hierarchical generative model. *J. Chem. Inf. Model.* **2024**, *64* (16), 6450-6463. https://doi.org/10.1021/acs.jcim.4c01193
4. Chen L., Li Q., Nasif K. F., Xie Y., Deng B., Niu S., Pouriyeh S., Dai Z., Chen J., Xie C. Y. AI-driven deep learning techniques in protein structure prediction. *Int. J. Mol. Sci.* **2024**, *25* (15), 8426. https://doi.org/10.3390/ijms25158426
5. Bryant P., Pozzati G., Elofsson A. Improved prediction of protein-protein interactions using AlphaFold2. *Nat. Commun.* **2022**, *13* (1), 1265. https://doi.org/10.1038/s41467-022-28865-w
6. Huang P.-S., Boyken S. E., Baker D. The coming of age of *de novo* protein design. *Nature* **2016**, *537* (7620), 320-327. https://doi.org/10.1038/nature19946
7. Zonta F., Pantano S. From sequence to mechanobiology? Promises and challenges for AlphaFold 3. *Mechanobiol. Med.* **2024**, *2* (3), 100083. doi: https://doi.org/10.1016/j.mbm.2024.100083
8. 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., Back T., Petersen S., Reiman D., Clancy E., Zielinski M., Steinegger M., Pacholska M., Berghammer T., Bodenstein S., Silver D., Vinyals O., Senior A. W., Kavukcuoglu K., Kohli P., Hassabis D. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583-589. https://doi.org/10.1038/s41586-021-03819-2
9. Nunes-Alves A., Merz K. AlphaFold2 in molecular discovery. *J. Chem. Inf. Model.* **2023**, *63* (19), 5947-5949. https://doi.org/10.1021/acs.jcim.3c01459
10. 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., Beattie C., Bertolli O., Bridgland A., Cherepanov A., Congreve M., Cowen-Rivers A. I., Cowie A., Figurnov M., Fuchs F. B., Gladman H., Jain R., Khan Y. A., Low C. M. R., Perlin K., Potapenko A., Savy P., Singh S., Stecula A., Thillaisundaram A., Tong C., Yakneen S., Zhong E. D., Zielinski M., Žídek A., Bapst V., Kohli P., Jaderberg M., Hassabis D., Jumper J. M. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **2024**, *630* (8016), 493-500. https://doi.org/10.1038/s41586-024-07487-w
11. McDonnell R. T., Henderson A. N., Elcock A. H. Structure prediction of large RNAs with AlphaFold 3 highlights its capabilities and limitations. *J. Mol. Biol.* **2024**, *436* (22), 168816. https://doi.org/10.1016/j.jmb.2024.168816
12. Kondo H. X., Takano Y. Structure comparison of heme-binding sites in heme protein predicted by AlphaFold3 and AlphaFold2. *Chem. Lett.* **2024**, *53* (8), upae148. https://doi.org/10.1093/chemle/upae148
13. He X.-h., Li J.-r., Shen S.-y., Xu H. E. AlphaFold3 versus experimental structures: Assessment of the accuracy in ligand-bound G protein-coupled receptors. *Acta Pharm. Sinica* **2024**, doi: 10.1038/s41401-024-01429-y. https://doi.org/10.1038/s41401-024-01429-y
14. Lynch M. L., Snell E. H., Bowman S. E. J. Structural biology in the time of COVID-19: Perspectives on methods and milestones. *IUCrJ* **2021**, *8* 335-341. https://doi.org/10.1107/S2052252521003948
15. Lubin J. H., Zardecki C., Dolan E. M., Lu C., Shen Z., Dutta S., Westbrook J. D., Hudson B. P., Goodsell D. S., Williams J. K., Voigt M., Sarma V., Xie L., Venkatachalam T., Arnold S., Alfaro

Alvarado L. H., Catalfano K., Khan A., McCarthy E., Staggers S., Tinsley B., Trudeau A., Singh J., Whitmore L., Zheng H., Benedek M., Currier J., Dresel M., Duvvuru A., Dyszel B., Fingar E., Hennen E. M., Kirsch M., Khan A. A., Labrie-Cleary C., Laporte S., Lenkeit E., Martin K., Orellana M., Ortiz-Alvarez de la Campa M., Paredes I., Wheeler B., Rupert A., Sam A., See K., Soto Zapata S., Craig P. A., Hall B. L., Jiang J., Koeppe J. R., Mills S. A., Pikaart M. J., Roberts R., Bromberg Y., Hoyer J. S., Duffy S., Tischfield J., Ruiz F. X., Arnold E., Baum J., Sandberg J., Brannigan G., Khare S. D., Burley S. K. Evolution of the SARS-CoV-2 proteome in three dimensions (3D) during the first 6 months of the COVID-19 pandemic. *Proteins: Struct., Funct., Bioinf.* **2022**, *90* (5), 1054-1080. https://doi.org/10.1002/prot.26250

16. Yang Q., Jian X., Syed A. A. S., Fahira A., Zheng C., Zhu Z., Wang K., Zhang J., Wen Y., Li Z., Pan D., Lu T., Wang Z., Shi Y. Structural comparison and drug screening of spike proteins of ten SARS-CoV-2 variants. *Research* **2022**, *2022,* 9781758. https://doi.org/10.34133/2022/9781758
17. Parigger L., Krassnigg A., Grabuschnig S., Gruber K., Steinkellner G., Gruber Christian C. AI-assisted structural consensus-proteome prediction of human monkeypox viruses isolated within a year after the 2022 multi-country outbreak. *Microbiology Spectrum* **2023**, *11* (6), e02315-23. https://doi.org/ 10.1128/spectrum.02315-23
18. Lubin J. H., Martinusen S. G., Zardecki C., Olivas C., Bacorn M., Balogun M., Slaton E. W., Wu A. W., Sakeer S., Hudson B. P., Denard C. A., Burley S. K., Khare S. D. A comprehensive survey of coronaviral main protease active site diversity in 3D: Identifying and analyzing drug discovery targets in search of broad specificity inhibitors for the next coronavirus pandemic. *bioRxiv* **2023**, 2023.01.30.526101. https://doi.org/10.1101/2023.01.30.526101
19. Jin Z., Du X., Xu Y., Deng Y., Liu M., Zhao Y., Zhang B., Li X., Zhang L., Peng C., Duan Y., Yu J., Wang L., Yang K., Liu F., Jiang R., Yang X., You T., Liu X., Yang X., Bai F., Liu H., Liu X., Guddat L. W., Xu W., Xiao G., Qin C., Shi Z., Jiang H., Rao Z., Yang H. Structure of M^{pro} from SARS-CoV-2 and discovery of its inhibitors. *Nature.* **2020**, *582* (7811), 289-293. https://doi.org/10.1038/s41586-020-2223-y
20. Rut W., Lv Z., Zmudzinski M., Patchett S., Nayak D., Snipas S. J., El Oualid F., Huang T. T., Bekes M., Drag M., Olsen S. K. Activity profiling and crystal structures of inhibitor-bound SARS-CoV-2 papain-like protease: A framework for anti–COVID-19 drug design. *Sci. Adv.* **2020**, *6* (42), eabd4596. https://doi.org/10.1126/sciadv.abd4596
21. Yin W., Mao C., Luan X., Shen D.-D., Shen Q., Su H., Wang X., Zhou F., Zhao W., Gao M., Chang S., Xie Y.-C., Tian G., Jiang H.-W., Tao S.-C., Shen J., Jiang Y., Jiang H., Xu Y., Zhang S., Zhang Y., Xu H. E. Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. *Science* **2020**, *368* (6498), 1499. https://doi.org/10.1126/science.abc1560
22. Tan B., Joyce R., Tan H., Hu Y., Wang J. SARS-CoV-2 main protease drug design, assay development, and drug resistance studies. *Acc. Chem. Res.* **2023**, *56* (2), 157-168. https://doi.org/10.1021/acs.accounts.2c00735
23. Yevsieieva L. V., Lohachova K. O., Kyrychenko A., Kovalenko S. M., Ivanov V. V., Kalugin O. N. Main and papain-like proteases as prospective targets for pharmacological treatment of coronavirus SARS-CoV-2. *RSC Adv.* **2023**, *13* (50), 35500–35524. https://doi.org/10.1039/d3ra06479d
24. Hu Q., Xiong Y., Zhu G.-H., Zhang Y.-N., Zhang Y.-W., Huang P., Ge G.-B. The SARS-CoV-2 main protease (Mpro): Structure, function, and emerging therapies for COVID-19. *MedComm* **2022**, *3* (3), e151. doi: https://doi.org/10.1002/mco2.151
25. Lohachova K. O., Sviatenko A. S., Kyrychenko A., Ivanov V. V., Langer T., Kovalenko S. M., Kalugin O. N. Computer-aided drug design of novel nirmatrelvir analogs inhibiting main protease of coronavirus SARS-CoV-2. *J. Appl. Pharm. Sci.* **2024**, *14* (5), 232-239. https://doi.org/10.7324/JAPS.2024.158114
26. Zagórska A., Czopek A., Fryc M., Jończyk J. Inhibitors of SARS-CoV-2 main protease (Mpro) as anti-coronavirus agents. *Biomolecules.* **2024**, *14* (7), 797. https://doi.org/10.3390/biom14070797
27. Yang Y., Luo Y.-D., Zhang C.-B., Xiang Y., Bai X.-Y., Zhang D., Fu Z.-Y., Hao R.-B., Liu X.-L. Progress in research on inhibitors targeting SARS-CoV-2 main protease (Mpro). *ACS Omega* **2024**, *9* (32), 34196-34219. https://doi.org/10.1021/acsomega.4c03023

28. Xiao Y.-Q., Long J., Zhang S.-S., Zhu Y.-Y., Gu S.-X. Non-peptidic inhibitors targeting SARS-CoV-2 main protease: A review. *Bioorg. Chem.* **2024**, *147* 107380. doi: https://doi.org/10.1016/j.bioorg.2024.107380
29. Yevsieieva L., Trostianko P., Kyrychenko A., Ivanov V., Kovalenko S., Kalugin O. Design of non-covalent dual-acting inhibitors for proteases M^{pro} and PL^{pro} of coronavirus SARS-CoV-2 through evolutionary library generation, pharmacophore profile matching, and molecular docking calculations. *Sci. Rise. Pharm. Sci.* **2024**, (6(52)), 15-26. https://doi.org/10.15587/2519-4852.2024.313808
30. Ullrich S., Nitsche C. SARS-CoV-2 papain-like protease: Structure, function and inhibition. *ChemBioChem* **2022**, *23* (19), e202200327. https://doi.org/10.1002/cbic.202200327
31. Gao X., Qin B., Chen P., Zhu K., Hou P., Wojdyla J. A., Wang M., Cui S. Crystal structure of SARS-CoV-2 papain-like protease. *Acta Pharmac. Sinica B* **2021**, *11* (1), 237-245. https://doi.org/10.1016/j.apsb.2020.08.014
32. Osipiuk J., Azizi S.-A., Dvorkin S., Endres M., Jedrzejczak R., Jones K. A., Kang S., Kathayat R. S., Kim Y., Lisnyak V. G., Maki S. L., Nicolaescu V., Taylor C. A., Tesar C., Zhang Y.-A., Zhou Z., Randall G., Michalska K., Snyder S. A., Dickinson B. C., Joachimiak A. Structure of papain-like protease from SARS-CoV-2 and its complexes with non-covalent inhibitors. *Nat. Commun.* **2021**, *12* (1), 743. https://doi.org/10.1038/s41467-021-21060-3
33. Taylor A. J., Amporndanai K., Rietz T. A., Zhao B., Thiruvaipati A., Wei Q., South T. M., Crow M. M., Apakama C., Sensintaffar J. L., Phan J., Lee T., Fesik S. W. Fragment-based screen of SARS-CoV-2 papain-like protease (PLpro). *ACS Med. Chem. Lett.* **2024**, *15* (8), 1351-1357. https://doi.org/10.1021/acsmedchemlett.4c00238
34. Magwaza N. N., Mushebenge A. G., Ugbaja S. C., Mbatha N. A., Khan R. B., Kumalo H. M. Mechanistic insights into targeting SARS-CoV-2 papain-like protease in the evolution and management of COVID-19. *BioChem* **2024**, *4* (3), 268-299. https://doi.org/10.3390/biochem4030014
35. Han H., Gracia A. V., Røise J. J., Boike L., Leon K., Schulze-Gahmen U., Stentzel M. R., Bajaj T., Chen D., Li I. C., He M., Behrouzi K., Khodabakhshi Z., Nomura D. K., Mofrad M. R. K., Kumar G. R., Ott M., Murthy N. A covalent inhibitor targeting the papain-like protease from SARS-CoV-2 inhibits viral replication. *RSC Adv.* **2023**, *13* (16), 10636-10641. https://doi.org/10.1039/D3RA00426K
36. Stasiulewicz A., Maksymiuk A. W., Nguyen M. L., Bełza B., Sulkowska J. I. SARS-CoV-2 papain-like protease potential inhibitors — *in silico* quantitative assessment. *Int. J. Mol. Sci.* **2021**, *22* (8), 3957. https://doi.org/10.3390/ijms22083957
37. Shan H., Liu J., Shen J., Dai J., Xu G., Lu K., Han C., Wang Y., Xu X., Tong Y., Xiang H., Ai Z., Zhuang G., Hu J., Zhang Z., Li Y., Pan L., Tan L. Development of potent and selective inhibitors targeting the papain-like protease of SARS-CoV-2. *Cell Chem Biol* **2021**, *28* (6), 855-865.e9. https://doi.org/10.1016/j.chembiol.2021.04.020
38. Ma C., Sacco M. D., Xia Z., Lambrinidis G., Townsend J. A., Hu Y., Meng X., Szeto T., Ba M., Zhang X., Gongora M., Zhang F., Marty M. T., Xiang Y., Kolocouris A., Chen Y., Wang J. Discovery of SARS-CoV-2 papain-like protease inhibitors through a combination of high-throughput screening and a FlipGFP-based reporter assay. *ACS Centrl. Sci.* **2021**, *7* (7), 1245-1260. https://doi.org/10.1021/acscentsci.1c00519
39. Huynh T., Cornell W., Luan B. *In silico* exploration of inhibitors for SARS-CoV-2's papain-like protease. *Front. Chem.* **2021**, *8* 624163. https://doi.org/10.3389/fchem.2020.624163
40. Kakavandi S., Zare I., VaezJalali M., Dadashi M., Azarian M., Akbari A., Ramezani Farani M., Zalpoor H., Hajikhani B. Structural and non-structural proteins in SARS-CoV-2: Potential aspects to COVID-19 treatment or prevention of progression of related diseases. *Cell Comm. Signal.* **2023**, *21* (1), 110. https://doi.org/10.1186/s12964-023-01104-5
41. Huang Y., Yang C., Xu X.-f., Xu W., Liu S.-W. Structural and functional properties of SARS-CoV-2 Spike protein: Potential antivirus drug development for COVID-19. *Acta Pharmacol. Sin.* **2020**, *41* (9), 1141-1149. https://doi.org/10.1038/s41401-020-0485-4
42. Wang L., Wu Y., Yao S., Ge H., Zhu Y., Chen K., Chen W.-Z., Zhang Y., Zhu W., Wang H.-Y., Guo Y., Ma P.-X., Ren P.-X., Zhang X.-l., Li H.-Q., Ali M. A., Xu W.-Q., Jiang H.-l., Zhang L.-K., Zhu L.-l., Ye Y., Shang W.-j., Bai F. Discovery of potential small molecular SARS-CoV-2

entry blockers targeting the Spike protein. *Acta Pharmacol. Sin.* **2022**, *43* (4), 788-796. https://doi.org/10.1038/s41401-021-00735-z

43. Hong Q., Han W., Li J., Xu S., Wang Y., Xu C., Li Z., Wang Y., Zhang C., Huang Z., Cong Y. Molecular basis of receptor binding and antibody neutralization of Omicron. *Nature* **2022**, *604* (7906), 546-552. https://doi.org/10.1038/s41586-022-04581-9
44. Yan W., Zheng Y., Dou C., Zhang G., Arnaout T., Cheng W. The pathogenic mechanism of *Mycobacterium tuberculosis*: Implication for new drug development. *Mol. Biomed.* **2022**, *3* (1), 48. https://doi.org/10.1186/s43556-022-00106-y
45. Bose P., Harit A. K., Das R., Sau S., Iyer A. K., Kashaw S. K. Tuberculosis: Current scenario, drug targets, and future prospects. *Med. Chem. Res.* **2021**, *30* (4), 807-833. https://doi.org/10.1007/s00044-020-02691-5
46. Liu L., Kong C., Fumagalli M., Savková K., Xu Y., Huszár S., Sammartino J. C., Fan D., Chiarelli L. R., Mikušová K., Sun Z., Qiao C. Design, synthesis and evaluation of covalent inhibitors of DpRe1 as antitubercular agents. *Eur. J. Med. Chem.* **2020**, *208* 112773. doi: https://doi.org/10.1016/j.ejmech.2020.112773
47. Snizhko A. D., Kyrychenko A. V., Gladkov E. S. Synthesis of novel derivatives of 5,6,7,8-tetrahydro-quinazolines using of α-aminoamidines and *in silico* screening of their biological activity. *Int. J. Mol. Sci.* **2022**, *23* (7), 3781. https://doi.org/10.3390/ijms23073781
48. Hong W., Wang Y., Chang Z., Yang Y., Pu J., Sun T., Kaur S., Sacchettini J. C., Jung H., Lin Wong W., Fah Yap L., Fong Ngeow Y., Paterson I. C., Wang H. The identification of novel *Mycobacterium tuberculosis* DHFR inhibitors and the investigation of their binding preferences by using molecular modelling. *Sci. Rep.* **2015**, *5* (1), 15328. https://doi.org/10.1038/srep15328
49. Dal Molin M., Selchow P., Schäfle D., Tschumi A., Ryckmans T., Laage-Witt S., Sander P. Identification of novel scaffolds targeting *Mycobacterium tuberculosis*. *J. Mol. Med.* **2019**, *97* (11), 1601-1613. https://doi.org/10.1007/s00109-019-01840-7
50. Reddy B. K. K., Landge S., Ravishankar S., Patil V., Shinde V., Tantry S., Kale M., Raichurkar A., Menasinakai S., Mudugal Naina V., Ambady A., Ghosh A., Tunduguru R., Kaur P., Singh R., Kumar N., Bharath S., Sundaram A., Bhat J., Sambandamurthy Vasan K., Björkelid C., Jones T. A., Das K., Bandodkar B., Malolanarasimhan K., Mukherjee K., Ramachandran V. Assessment of *Mycobacterium tuberculosis* pantothenate kinase vulnerability through target knockdown and mechanistically diverse inhibitors. *Antimicrob. Agents Chemother.* **2014**, *58* (6), 3312-3326. https://doi.org/10.1128/aac.00140-14
51. Björkelid C., Bergfors T., Raichurkar A. K. V., Mukherjee K., Malolanarasimhan K., Bandodkar B., Jones T. A. Structural and biochemical characterization of compounds inhibiting *Mycobacterium tuberculosis* pantothenate kinase. *J. Biol. Chem.* **2013**, *288* (25), 18260-18270. https://doi.org/10.1074/jbc.M113.476473
52. Chollet A., Mourey L., Lherbet C., Delbot A., Julien S., Baltas M., Bernadou J., Pratviel G., Maveyraud L., Bernardes-Génisson V. Crystal structure of the enoyl-ACP reductase of mycobacterium tuberculosis (InhA) in the apo-form and in complex with the active metabolite of isoniazid pre-formed by a biomimetic approach. *J. Struct. Biol.* **2015**, *190* (3), 328-337. https://doi.org/10.1016/j.jsb.2015.04.008
53. Marchenko K. I., Kyrychenko A. V., Kolos N. M. Synthesis and modification of 7-aroyl derivatives of 4,7-dihydro-[1,2,4]triazolo-[1,5-a]-pyrimidine as potent inhibitors of Sirtuin-2. *Funct. Mater.* **2024**, *31* (2), 260-268. https://doi.org/10.15407/fm31.02.260
54. Zhao X., Allison D., Condon B., Zhang F., Gheyi T., Zhang A., Ashok S., Russell M., MacEwan I., Qian Y., Jamison J. A., Luz J. G. The 2.5 Å crystal structure of the Sirt1 catalytic domain bound to nicotinamide adenine dinucleotide (NAD^+) and an indole (ex527 analogue) reveals a novel mechanism of histone deacetylase inhibition. *J. Med. Chem.* **2013**, *56* (3), 963-969. https://doi.org/10.1021/jm301431y
55. Kumar A., Chauhan S. How much successful are the medicinal chemists in modulation of Sirt1: A critical review. *Eur. J. Med. Chem.* **2016**, *119* 45-69. https://doi.org/10.1016/j.ejmech.2016.04.063
56. Lipson V. V., Yaremenko F. G., Vakula V. M., Kovalenko S. V., Kyrychenko A. V., Desenko S. M., Musatov V. I., Borysko P. O., Zozulya S. O. Imidazole derivatives as potent inhibitors of Sirtuin-1. *Funct. Mater.* **2023**, *30* (4), 486-493. https://doi.org/10.15407/fm30.04.486

57. Lipson V. V., Yaremenko F. G., Vakula V. M., Kovalenko S. V., Kyrychenko A. V., Desenko S. M., Borysko P. O., Zozulya S. O. Discovery of novel N-acylhydrazone derivatives as potent inhibitors of Sirtuin-1. *SynOpen* **2024**, *08* (02), 100-108. https://doi.org/10.1055/s-0043-1763747
58. Kyrychenko A. V., Ladokhin A. S. Fluorescence tools for studies of membrane protein insertion. *Biopolym. Cell.* **2018**, *34* (4), 251-271. https://doi.org/10.7124/bc.00097F
59. Ladokhin A. S., Kyrychenko A., Rodnin M. V., Vasquez-Montes V. Conformational switching, refolding and membrane insertion of the diphtheria toxin translocation domain. In *Methods Enzymol.*, Academic Press: 2021; Vol. 649, pp 341-370. https://doi.org/10.1016/bs.mie.2020.12.016
60. Kyrychenko A., Posokhov Y. O., Rodnin M. V., Ladokhin A. S. Kinetic intermediate reveals staggered pH-dependent transitions along the membrane insertion pathway of the diphtheria toxin T-domain. *Biochemistry.* **2009**, *48* (32), 7584-7594. https://doi.org/10.1021/bi9009264
61. Rodnin M. V., Kyrychenko A., Kienker P., Sharma O., Posokhov Y. O., Collier R. J., Finkelstein A., Ladokhin A. S. Conformational switching of the diphtheria toxin T domain. *J. Mol. Biol.* **2010**, *402* (1), 1-7. https://doi.org/10.1016/j.jmb.2010.07.024
62. Rodnin Mykola V., Kyrychenko A., Kienker P., Sharma O., Vargas-Uribe M., Collier R. J., Finkelstein A., Ladokhin Alexey S. Replacement of C-terminal histidines uncouples membrane insertion and translocation in diphtheria toxin T-domain. *Biophys. J.* **2011**, *101* (10), L41-L43. https://doi.org/10.1016/j.bpj.2011.10.018
63. Kurnikov I. V., Kyrychenko A., Flores-Canales J. C., Rodnin M. V., Simakov N., Vargas-Uribe M., Posokhov Y. O., Kurnikova M., Ladokhin A. S. pH-triggered conformational switching of the diphtheria toxin T-domain: The roles of N-terminal histidines. *J. Mol. Biol.* **2013**, *425* (15), 2752-2764. https://doi.org/10.1016/j.jmb.2013.04.030
64. Vargas-Uribe M., Rodnin M. V., Öjemalm K., Holgado A., Kyrychenko A., Nilsson I., Posokhov Y. O., Makhatadze G., von Heijne G., Ladokhin A. S. Thermodynamics of membrane insertion and refolding of the diphtheria toxin T-domain. *J Membr Biol* **2015**, *248* (3), 383-394. https://doi.org/10.1007/s00232-014-9734-0
65. Kyrychenko A., Posokhov Y. O., Vargas-Uribe M., Ghatak C., Rodnin M. V., Ladokhin A. S. Fluorescence applications for structural and thermodynamic studies of membrane protein insertion. In *Reviews in fluorescence 2016*, Geddes, C. D., Ed. Springer International Publishing: Cham, Switzerland, 2017; pp 243-274. https://doi.org/10.1007/978-3-319-48260-6_10
66. Rodnin M. V., Vasques-Montes V., Kyrychenko A., Oliveira N. F. B., Kashipathy M. M., Battaile K. P., Douglas J., Lovell S., Machuqueiro M., Ladokhin A. S. Histidine protonation and conformational switching in diphtheria toxin translocation domain. *Toxins* **2023**, *15* (7), Art. num. 410. https://doi.org/10.3390/toxins15070410
67. Rodnin M. V., Kashipathy M. M., Kyrychenko A., Battaile K. P., Lovell S., Ladokhin A. S. Structure of the diphtheria toxin at acidic pH: Implications for the conformational switching of the translocation domain. *Toxins* **2020**, *12* (11), Art. num. 704. https://doi.org/10.3390/toxins12110704
68. Isorna P., Polaina J., Latorre-García L., Cañada F. J., González B., Sanz-Aparicio J. Crystal structures of paenibacillus polymyxa β-glucosidase B complexes reveal the molecular basis of substrate specificity and give new insights into the catalytic machinery of family I glycosidases. *J. Mol. Biol.* **2007**, *371* (5), 1204-1218. https://doi.org/10.1016/j.jmb.2007.05.082
69. Chepeleva L. V., Tarasenko D. O., Chumak A. Y., Demidov O. O., Snizhko A. D., Kolomoitsev O. O., Kotliar V. M., Gladkov E. S., Tatarets A. L., Kyrychenko A. V., Roshal A. D. 4'-benzyloxyflavonol glucoside as fluorescent indicator for β-glucosidase activity. *Funct. Mater.* **2023**, *30* (4), 494-505. https://doi.org/10.15407/fm30.04.494
70. Chepeleva L. V., Demidov O. O., Snizhko A. D., Tarasenko D. O., Chumak A. Y., Kolomoitsev O. O., Kotliar V. M., Gladkov E. S., Kyrychenko A., Roshal A. D. Binding interactions of hydrophobically-modified flavonols with β-glucosidase: Fluorescence spectroscopy and molecular modelling study. *RSC Adv.* **2023**, *13* (48), 34107-34121. https://doi.org/10.1039/D3RA06276G
71. Demidov O. O., Gladkov E. S., Kyrychenko A. V., Roshal A. D. Synthetic and natural flavonols as promising fluorescence probes for β-glucosidase activity screening *Funct. Mater.* **2022**, *29* (2), 252-262. https://doi.org/10.15407/fm29.02.252
72. Zhang Y., Vass M., Shi D., Abualrous E., Chambers J. M., Chopra N., Higgs C., Kasavajhala K., Li H., Nandekar P., Sato H., Miller E. B., Repasky M. P., Jerome S. V. Benchmarking refined and

unrefined AlphaFold2 structures for hit discovery. *J. Chem. Inf. Model.* **2023**, *63* (6), 1656-1667. https://doi.org/10.1021/acs.jcim.2c01219

73. Díaz-Rovira A. M., Martín H., Beuming T., Díaz L., Guallar V., Ray S. S. Are deep learning structural models sufficiently accurate for virtual screening? Application of docking algorithms to AlphaFold2 predicted structures. *J. Chem. Inf. Model.* **2023**, *63* (6), 1668-1674. https://doi.org/10.1021/acs.jcim.2c01270
74. Beuming T., Martín H., Díaz-Rovira A. M., Díaz L., Guallar V., Ray S. S. Are deep learning structural models sufficiently accurate for free-energy calculations? Application of FEP$^+$ to AlphaFold2-predicted structures. *J. Chem. Inf. Model.* **2022**, *62* (18), 4351-4360. https://doi.org/10.1021/acs.jcim.2c00796
75. Bernatavicius A., Šícho M., Janssen A. P. A., Hassen A. K., Preuss M., van Westen G. J. P. AlphaFold meets *de novo* drug design: Leveraging structural protein information in multitarget molecular generative models. *J. Chem. Inf. Model.* **2024**, *64* (21), 8113-8122. https://doi.org/10.1021/acs.jcim.4c00309
76. Gu S., Yang Y., Zhao Y., Qiu J., Wang X., Tong H. H. Y., Liu L., Wan X., Liu H., Hou T., Kang Y. Evaluation of AlphaFold2 structures for hit identification across multiple scenarios. *J. Chem. Inf. Model.* **2024**, *64* (9), 3630-3639. https://doi.org/10.1021/acs.jcim.3c01976
77. Pei J., Andreeva A., Chuguransky S., Lázaro Pinto B., Paysan-Lafosse T., Dustin Schaeffer R., Bateman A., Cong Q., Grishin N. V. Bridging the gap between sequence and structure classifications of proteins with AlphaFold models. *J. Mol. Biol.* **2024**, *436* (22), 168764. https://doi.org/10.1016/j.jmb.2024.168764
78. Wee J., Wei G.-W. Evaluation of AlphaFold 3's protein–protein complexes for predicting binding free energy changes upon mutation. *J. Chem. Inf. Model.* **2024**, *64* (16), 6676-6683. https://doi.org/10.1021/acs.jcim.4c00976
79. Baek M., DiMaio F., Anishchenko I., Dauparas J., Ovchinnikov S., Lee G. R., Wang J., Cong Q., Kinch L. N., Schaeffer R. D., Millán C., Park H., Adams C., Glassman C. R., DeGiovanni A., Pereira J. H., Rodrigues A. V., van Dijk A. A., Ebrecht A. C., Opperman D. J., Sagmeister T., Buhlheller C., Pavkov-Keller T., Rathinaswamy M. K., Dalwadi U., Yip C. K., Burke J. E., Garcia K. C., Grishin N. V., Adams P. D., Read R. J., Baker D. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **2021**, *373* (6557), 871-876. https://doi.org/10.1126/science.abj8754
80. Ahdritz G., Bouatta N., Floristean C., Kadyan S., Xia Q., Gerecke W., O'Donnell T. J., Berenberg D., Fisk I., Zanichelli N., Zhang B., Nowaczynski A., Wang B., Stepniewska-Dziubinska M. M., Zhang S., Ojewole A., Guney M. E., Biderman S., Watkins A. M., Ra S., Lorenzo P. R., Nivon L., Weitzner B., Ban Y.-E. A., Chen S., Zhang M., Li C., Song S. L., He Y., Sorger P. K., Mostaque E., Zhang Z., Bonneau R., AlQuraishi M. Openfold: Retraining AlphaFold2 yields new insights into its learning mechanisms and capacity for generalization. *Nat. Methods* **2024**, *21* (8), 1514-1524. https://doi.org/10.1038/s41592-024-02272-z

Received 11.09.2024 *Accepted 19.12.2024*

Є. В. Дума, О. В. Кириченко. Порівняння ефективності прогнозування 3D-структури білка з використанням AlphaFold 3 від Google DeepMind.

Харківський національний університет імені В.Н. Каразіна, хімічний факультет, майдан Свободи, 4, Харків, 61022, Україна

Тривимірна структура білків безпосередньо пов'язана з їхньою функцією, що робить її визначення вирішальним для розуміння біологічних процесів і вирішення проблем, пов'язаних зі здоров'ям людини та науками про життя. Незважаючи на постійне експериментальне отримання нових білкових структур, залишається значний розрив між кількістю доступних білкових послідовностей та експериментальних третинних структур.

. Останні дослідження зосереджені на прогнозуванні білкових структур за допомогою обчислювальних методів, заснованих або на шаблонах, або на моделюванні емпіричного силового поля. В останні роки різні методи були об'єднані для вирішення окремих обмежень цих підходів, що призвело до розробки AlphaFold3 Google DeepMind. Цей інноваційний інструмент дозволяє прогнозувати тривимірні структури білка з високою точністю на основі його амінокислотної послідовності. У цьому дослідженні ми перевірили застосовність, продуктивність і обмеження AlphaFold3 (AF3) для прогнозування тривимірної структури ши-

рокого ряду білків, включаючи білки коронавірусу SARS-CoV-2, інші бактеріальні та вірусні білки, а також рослинні ферменти. Ми виявили, що AlphaFold3 добре відтворює 3D-будову більшості з досліджених білків з точки зору відхилення від доступних рентгенівських структур. Деякі незначні помилки згортання N- і C-кінцевих сегментів були поширеними, що часто не впливало на біологічну роль досліджуваних білків. У випадках, пов’язаних із білковими димерами або олігомерами вищого порядку, спостерігалися помітні відмінності між прогнозованими моделями AF3 для одноланцюгових мономерів та їх відповідними експериментальними структурами. Ці розбіжності були особливо очевидні в областях, пов'язаних з димеризацією білка, білковою асоціацією і на границі стикування білків. Зрештою, незважаючи на відтворення глобального фолдінгу, прогнозування складної структури Spike глікопротеїнів все ще виходить за межі поточних можливостей AF3.

Ключові слова: *згортання білка, фермент, тривимірна структура, дизайн ліків, штучний інтелект, AlphaFold3*

Надіслано до редакції 11.09.2024 *Прийнято до друку 19.12.2024*

Kharkiv University Bulletin. Chemical Series. Issue 43 (66), 2024