Applying generative language models to design new antibody sequences to target Influenza´s hemagglutinin

Carlos Gomes[1[0000-0003-4524-8109]], Pedro Moreira[2[0000-0001-7734-1792]], Diana Lousa [2[0000-0002-2309-0980]] and Miguel Rocha [3[0000-0001-8439-8172]]

[1] School of Engineering, University of Minho, Braga, Portugal
[2] António Xavier Institute of Chemical and Biological Technology, University Nova de Lisboa, Oeiras, Portugal
[3] Center of Biological Engineering, University of Minho, Braga, Portugal

Abstract. Influenza is a highly contagious viral disease that has a significant impact on the health of the world's population and has a very fast evolutionary capacity. So, in this project, using various bioinformatics tools such as SAbDab, AlphaFold, ProPythia, and many others, we intend to design new antibody sequences to target Influenza hemagglutinin.

Keywords: Influenza, Hemagglutinin, Generative Language Models, Bioinformatics.

1 Motivation and Objectives

Influenza is a highly contagious viral disease that infects millions of human beings every year. Its symptoms range from mild to severe and can include fever, muscle aches extreme fatigue and serious respiratory complications such as pneumonia. Although most people recover from the flu without serious complications, it still represents a significant public health burden, especially in high-risk groups [1]. Despite advances in the prevention and treatment of Influenza, including the availability of seasonal vaccines and antivirals, the Influenza virus continues to pose a significant challenge to public health due to its ability to mutate rapidly and evade existing immunity[2]. Therefore, it is essential to continue investing in research and development into new prevention and treatment strategies.

So, the main objective of this project is to obtain new antibody sequences to target the hemagglutinin of the Influenza virus.

2 State of the Art

2.1 Influenza

Epidemiology and Transmission. Influenza is a highly contagious viral infection that manifests itself in seasonal epidemics, mainly in winter [1]. The Influenza virus can affect all organs of the body and manifests itself as an acute febrile illness with varying degrees of systemic and respiratory symptoms. The main symptoms include fever, chills, headaches, weakness, redness of the eyes, sore throat, runny nose, and dry cough, and when complications from the infection are severe, they can be life-threatening for high-risk individuals or groups [3].

The main route of transmission is through the inhalation of infectious respiratory particles (large droplet transmission) when an infected person coughs or sneezes. There is also evidence of airborne transmission (small particles transmitted by speech or exhalation) and by fomites [4]. The typical incubation period is 24 to 48 hours. Patients are infectious one to two days before the onset of symptoms and for five to seven days afterwards. Children and immunosuppressed people may experience prolonged viral transmission [5].

For most outpatients, the diagnosis is made clinically, and laboratory confirmation is not necessary. Laboratory tests can be useful in hospitalized patients with suspected flu and in patients for whom a confirmed diagnosis will change treatment decisions. Rapid molecular assays are the preferred diagnostic tests because they can be carried out at the point of care, are highly accurate and have rapid results. Treatment with one of the four approved anti-Influenza drugs can be considered if the patient presents within 48 hours of the onset of symptoms. The benefit of treatment is greatest when antiviral therapy is started within 24 hours of the onset of symptoms. These drugs can reduce the risk of serious complications. There is also the possibility of annual vaccination as a form of flu prevention, and it is recommended for all people over six months of age who have no contraindications [6].

Etiology. Influenza viruses evolve quickly by frequent antigenic variation. Antigenic drift and shift are terms used to describe how the virus mutates and results in new strains. There is a significant change in the virus's genome in antigenic shift resulting in new hemagglutinin (HA) and neuraminidase (NA) protein expression [2]. This means that, despite improvements in prevention, control and case management, the antigenic shift continues to make Influenza a disease transmitted worldwide [7].

Influenza has two surface glycoproteins, Hemagglutinin and Neuraminase [8] .The virus infects the host by binding to the host cell and penetrating the membrane. As will be explored in the next subchapter, hemagglutinin binds to cell surface receptors and initiates the entry of the virus into these cells. Neuraminase is an enzyme that aids viral replication and allows the virus to be released from the host cell. So, viral glycoproteins play an essential role in the virulence and pathogenesis of the Influenza virus [7].

The role of hemagglutinin. The surface of Influenza virions is dominated by HA, which outnumbers NA by five to ten-fold on average (1) [9]. Hemagglutinin can agglutinate red blood cells, and this ability can be attributed to its receptor-binding function. HA binds to sialylated glycan receptors on host cells to initiate viral entry and carries the machinery for membrane fusion [10].

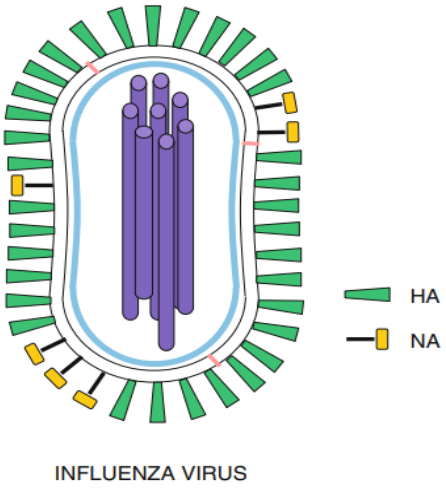

Fig. 1. Cartoon showing the architecture of Influenza virus. HA stands for Hemagglutinin while NA stands for Neuraminidase [11].

Hemagglutinins are found from the virus surface membrane as glycoprotein spikes, and each spike contains three identical subunits formed by two glycopolypeptides. Each of these subunits is divided into four subdomains (2) [12]. It has two main roles in Influenza infections: attaching the virus to the host receptors and promoting membrane fusion between the virus and host membranes. They recognize cell surface glycoconjugates containing sialic acid as receptors, but have limited affinity for them and, consequently, the binding of the virus to cells requires its interaction with several HAs of the virus. The receptor-bound virus is transferred to the endosomes where membrane fusion by HAs is activated at a pH between 5 and 6.5, depending on the virus strain. The fusion activity requires extensive rearrangements in the HA conformation, which include the extrusion of a buried "fusion peptide" to bind to the endosomal membrane, form a bridge to the virus membrane and finally bring the two membranes closer together [12]. Influenza viruses are classified based on their antigenicity, which is determined by their surface glycoproteins. There are four characterized types of flu virus, A, B, C and D. Influenza A and B viruses have two surface glycoproteins, while influenza C and D viruses have only one surface glycoprotein, hemagglutinin-esterase fusion. Both influenza A and B viruses infect humans and can cause severe illness or death. In contrast, Influenza C virus only causes mild symptoms in most cases. Human infection with Influenza D virus has not been observed. Therefore, most Influenza research has

been focused on influenza A and B viruses. Between those two viruses, the main difference is that influenza B virus is only found in humans, whereas the primary natural reservoir for influenza A virus is aquatic birds. As a result, influenza A virus usually receives more attention and has been studied more extensively [13].

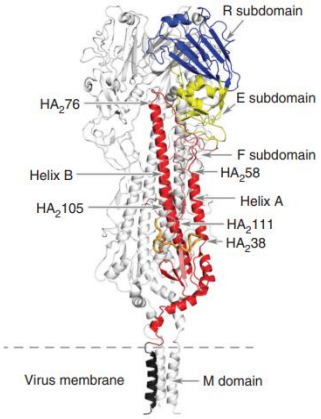

Fig. 2. The polypeptide and subdomain structure of hemagglutinin [12].

2.2 Antibodies and their role in defending against the virus

Antibodies are the basis of the immune response in vertebrates. These proteins form complexes with potentially pathogenic molecules called antigens and inhibit their function or recruit other components of the immune machinery to destroy them [14]. In addition to the biological importance of antibodies, their ability to be raised against an almost unlimited number of molecules has made them useful laboratory tools and increasingly useful as therapeutic agents in humans [15].

As mentioned above, in the case of Influenza, the HA and NA proteins are highly immunogenic, and antibodies directed at both glycoproteins can be isolated after natural infection or vaccination. By binding to the viral surface proteins HA and NA, antibodies can block essential steps in the virus replication cycle, thus limiting the spread of infection. Due to the host's immune pressure and the error-prone RNA polymerase, HA and NA are very plastic and show differences in antigenic properties. For instance, in the case of Influenza A, the virus can split into 18 HA subtypes and 11 NA subtypes [16]. Previously, the protective efficiency of antibodies was measured by their ability to prevent HA binding through the neutralization assay and the hemagglutination inhibition assay [17], and antibodies without these functions did not receive as much attention. However, increasing evidence suggests that non-neutralizing antibodies (nnAbs) can also confer protection through multiple mechanisms without interrupting virus entry or membrane fusion, such as complement activation, increased phagocytosis, targeting of internal viral proteins and activation of crystallizable fragment functions [18].

In addition to the biological importance of antibodies, their ability to be created against an almost unlimited number of molecules has made them useful laboratory tools and increasingly useful as therapeutic agents in humans. This biopharmaceutical application has increased the desire to understand how the antibody's binding, stability and immunogenic properties are determined and how they can be modified [14].
Due to their efficiency, analysis and computational tools are increasingly being used to aid the antibody engineering process. Currently, many of these tools use only antibody data, as opposed to general protein data, as this has been shown to increase performance [19].

2.3 Relevant software, packages, databases and models

In this project, different databases and different software will be used to fulfill the proposed objectives. Starting with the state of the art, we mainly used databases such as PubMed. This is an open-access site that has a wide range of articles on topics ranging from life sciences to bioengineering and is therefore important for searching for scientific evidence [20]. As for obtaining information on the antibodies that will be most useful for this work SAbDab was used. SAbDab is an online antibody structural database that brings together all publicly available antibody structures. The data is enriched with various properties, including experimental information, genetic details, antigen details, and where available the antibody-antigen binding affinity [14]. In addition, the classification of the structures of the complementarity determining regions (CDRs) of antibodies is very important for predicting the structure of antibodies and for their computational design. PyIgClassify is a database and webserver that gives access to CDR structures present in the Protein Data Bank (PDB) [21] and will therefore be important in this work. The Protein Data Bank is a global repository of experimentally determined 3D structures of biological macromolecules, experimental data, and the associated metadata [22].

We will also use protein language models to generate new antibody sequences. These models are deep learning models based on natural language processing methods. They are trained using large sets of protein sequences and find long-range dependencies in a protein sequence [23]. These pre-trained models can predict structure in an unsupervised way either taking as input a single sequence [24] or a multiple sequence alignment [25]. In this work, we will use a protein language modelling approach called Evolutionary Scale Modelling (ESM). This model assumes that natural evolution must explore a vast landscape of possible sequences for desirable but rare mutations, which suggests that learning natural evolutionary strategies can guide artificial evolution. Therefore, it can be used to efficiently evolve human antibodies by suggesting mutations that are evolutionarily plausible, although it is not necessary to provide the model with any information about the target antigen, binding specificity or protein structure [26]

Other tools will be used to extract properties from the results obtained. ProPythia is a Python package that will be used to extract some of the properties. It offers a range of functions for calculating several physicochemical properties and other representations of proteins. It allows the user to pre-process the dataset, manage and modify

sequences, clustering, manifold learning, feature selection and dimensionality reduction with a variety of diagrams to facilitate user interpretation. It also allows for the training and optimization of traditional Machine Learning models to make predictions on unseen data and respective feature analysis [27].

3 Methodology

The aim of this project is to obtain new antibody sequences to target the hemagglutinin of the Influenza virus. To achieve this, the following work plan has been drawn up.

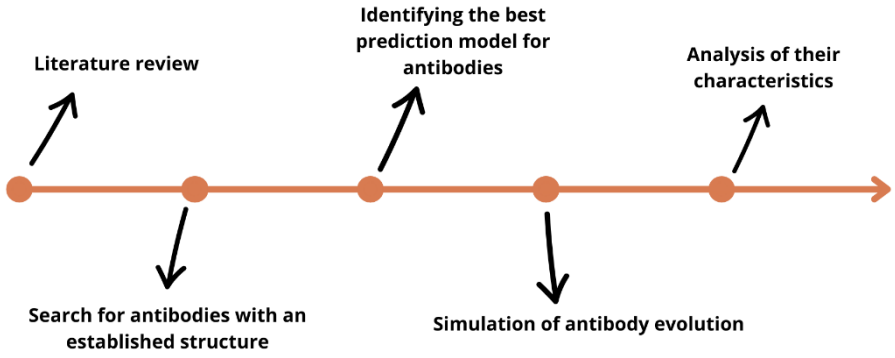

Fig. 3. Project methodology diagram.

First, we will start with a literature review. At the start of the project, it is essential to do a comprehensive literature review. This involves an analysis of studies already done on the interaction of antibodies and Influenza hemagglutinin, as well as bioinformatics and molecular modeling techniques.

Then, we will search for Antibodies to Influenza Hemagglutinin in Databases. This step consists of exploring databases such as SabDab or PylgClassify to identify the best-known antibodies with affinity for the virus [14, 21].

Once the antibodies have been selected, three-dimensional structure prediction models will be compared to choose the most suitable one for antibodies and thus be used downstream. The models under analysis will be ImmuneBuilder, IgFold, AlphaFold 2 (local Colab fold version) and AlphaFold 3 (Beta).

ImmuneBuilder is a set of deep learning models trained to accurately predict the structure of antibodies (ABodyBuilder2), nanobodies (NanoBodyBuilder2) and T-cell receptors (TCRBuilder2). ImmuneBuilder was used in this project as it generates structures with high precision and is theoretically much faster than AlphaFold2 [28]. It is important to note that during the project, ABodyBuilder3 was launched, which has some improvements over its previous version, mainly in the refinement of the structures, improving the scalability and precision of the model [29], but it was not possible to compare it with the other models under study. IgFold uses representations acquired from the pre-trained AntiBERTy language model to directly predict 3D atomic

coordinates. These structures are accompanied by an accuracy estimate per residue, which provides information on the quality of the prediction [30]. Local Colab Fold, on the other hand, is a local version of AlphaFold 2 in which AlphaFold 2's homology search is replaced by MMseqs2 (Many-against-Many sequence searching), thus making it around 40 to 60 times faster without losing quality in the prediction [31]. Finally, AlphaFold 3, despite still being in a beta phase and therefore reduced to just 20 daily predictions, is a model capable of predicting complexes containing almost all types of molecules present in the PDB with high accuracy, outperforming even more specific models. This has been achieved through a substantial evolution of the AlphaFold 2 architecture and training procedure, both to accommodate more general chemical structures and to improve the efficiency of data learning [32].

Afterwards, to find out which model gives the best results, predictions of the selected antibodies will be performed. In this way, it is possible to compare the predictions made by the models with the three-dimensional structure already catalogued in databases. To do this, the Root mean square deviation of atomic positions (RMSDs) will be calculated.

After that, it is essential to characterize the epitopes, in other words, the regions of Hemagglutinin to which these antibodies bind, and specify which variant(s) those antibodies are effective with.

Then, we are going to use Antibody Sequences to generate multiple sequence alignments (MSA), an essential technique in the bioinformatic analysis of protein sequences. With this approach, we sought to identify conserved and variable patterns in the antibody sequences, as well as areas of interaction with hemagglutinin. This information is crucial to understanding the diversity and evolution of antibodies in response to the virus.

The next activity involves using ESM (Evolutionary Scale Modelling) to try to predict what kind of evolutionary mutations could occur in these antibodies, and which are the most favorable in the fight against influenza, ultimately obtaining their sequence to predict the structure of these antibodies with mutations.

Then, we will analyze the results. Once these sequences have been obtained, they need to be categorized using various processes, such as multiple alignment and clustering, to select the most interesting results. Finally, ProPythia will be used to extract the physicochemical properties, while the prediction model, tested previously, which shows the best accuracy values, will be used to predict its structural framework. Analyzing these properties will allow us to identify the sequences that meet specific criteria so that they can be tested in the laboratory.

4 Results

4.1 Selection of the antibodies

As mentioned in the methodology, after the literature review, the first step was to choose relevant antibodies that interact with the influenza hemagglutinin. To do this, the PDB and SabDab were simultaneously used to obtain all the information available

on them. The result of this selection can be found in Table 1, which contains the PDB entry and the link to the respective SabDaB page.

Table 1. PDB ID and respective SabDab link of the selected antibodies.

PDB ID	SabDab link
6UR5	https://opig.stats.ox.ac.uk/webapps/sabdab-sabpred/sabdab/structureviewer/?pdb=6ur5
5XKU	https://opig.stats.ox.ac.uk/webapps/sabdab-sabpred/sabdab/structureviewer/?pdb=5xku
4XNQ	https://opig.stats.ox.ac.uk/webapps/sabdab-sabpred/sabdab/structureviewer/?pdb=4xnq
6WHK	https://opig.stats.ox.ac.uk/webapps/sabdab-sabpred/sabdab/structureviewer/?pdb=6whk
6HJQ	https://opig.stats.ox.ac.uk/webapps/sabdab-sabpred/sabdab/structureviewer/?pdb=6hjq
5CJQ	https://opig.stats.ox.ac.uk/webapps/sabdab-sabpred/sabdab/structureviewer/?pdb=5cjq
4YK4	https://opig.stats.ox.ac.uk/webapps/sabdab-sabpred/sabdab/structureviewer/?pdb=4yk4
5W42	https://opig.stats.ox.ac.uk/webapps/sabdab-sabpred/sabdab/structureviewer/?pdb=5w42
3SDY	https://opig.stats.ox.ac.uk/webapps/sabdab-sabpred/sabdab/structureviewer/?pdb=3sdy
6A67	https://opig.stats.ox.ac.uk/webapps/sabdab-sabpred/sabdab/structureviewer/?pdb=6a67

Ten antibodies were selected to be used as the foundation for the rest of the project. After this selection, the chosen antibodies and their respective structure were analyzed using Pymol. Using the software, it was possible to identify the Heavy and Light Chain of each antibody, and then store the sequence corresponding to that fraction. With these sequences, it was possible to start testing the prediction models, as these regions with the most variability were used as input.

4.2 Selecting the structure prediction model

As already mentioned, four models were tested:

- ImmuneBuilder (https://github.com/oxpig/ImmuneBuilder);
- IgFold (https://github.com/Graylab/IgFold);
- AlphaFold 2 (Local Colab Fold - https://github.com/YoshitakaMo/localcolabfold);
- AlphaFold 3 (Beta - https://golgi.sandbox.google.com/).

After using the models, the corresponding structures were obtained. The figure below shows the results of the predictions obtained by the models for the 5CJQ antibody:

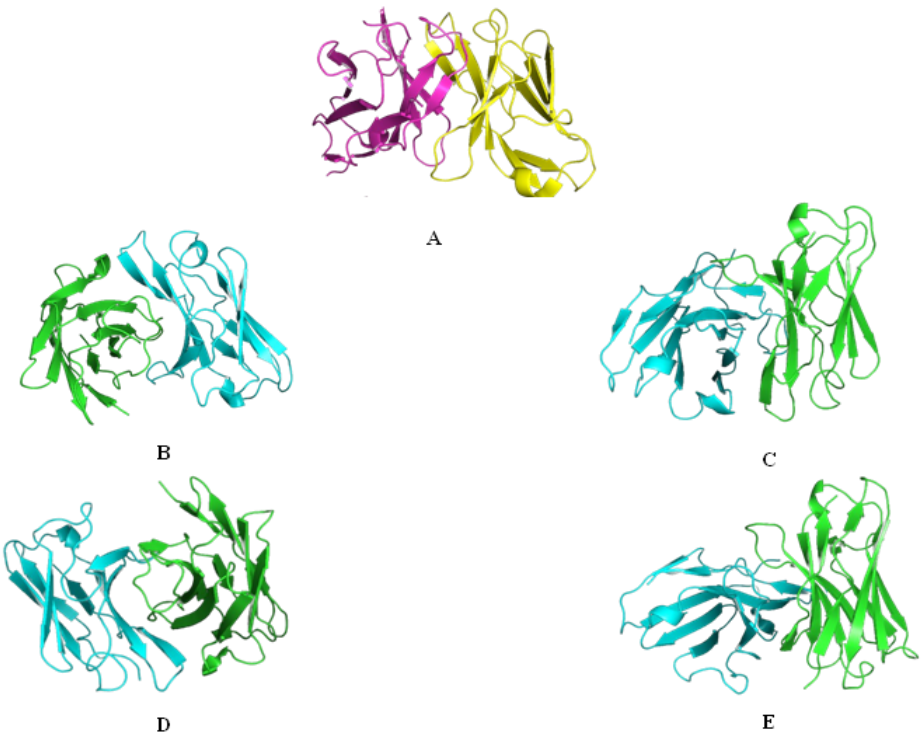

Fig. 4. 5CJQ antibody and the respective predictions obtained by the models mentioned above (A - Structure of the antibody determined experimentally; B - Prediction made by ImmuneBuilder; C - Prediction made by IgFold; D - Prediction made by AlphaFold 2; E - Prediction made by AlphaFold 3).

To compare the accuracy of the models, a visual evaluation of the structures is not considered a reliable metric, as it is impossible to categorise which of the models has the best results, as can be seen in Figure 4. As this is a protein structure prediction problem, it is important that other structural properties can be analysed. As this work uses a known target protein (i.e. the native structure has already been experimentally determined), the root mean square deviation of atomic positions (RMSDs) between the conformation obtained by the GA and the native structure can be used. In other words, the RMSD reveals how close one structure is (usually in Å) to the other and this closeness is greater the closer the value is to 0 [33]. The RMSDs were then calculated, and the following values were obtained for each prediction (Table 2):

Table 2. Calculation of RMSDs (Å) for the different models.

PDB ID	IMMUNEBUILDER	IGFOLD	AF2	AF3
6UR5	0,75	1,03	0,89	0,83
5XKU	11,42	12,05	11,32	0,60
4XNQ	20,18	20,30	20,28	20,36
6WHK	0,90	0,67	0,51	0,43
6HJQ	0,79	0,96	0,97	0,92
5CJQ	0,98	1,37	1,17	0,58
4YK4	19,73	19,76	19,75	0,44
5W42	0,62	0,97	0,84	0,58
3SDY	0,53	0,55	0,54	0,77
6A67	0,33	0,78	0,63	0,54
MAXIMUM	20,18	20,30	20,28	20,36
MINIMUM	0,33	0,55	0,51	0,43
MEDIAN	0,85	1,00	0,93	0,59
MEAN (SD)	5,622(±7,8)	5,8(±8,2)	5,6(±8,2)	2,6(±6,2)

The table shows that the models performed very well overall. It is worth highlighting the performance of AlphaFold 3, which despite still being in the beta phase, clearly stands out in terms of the accuracy of the structures it generated compared to the other models, with the lowest average values (2.60 Å ± 6.2), while the other models have very similar scores between them. The prediction with the highest accuracy belongs to ImmuneBuilder for the 6A67 antibody (0.33 Å), a very good result. However, it is worth highlighting the difficulty in predicting some of the antibodies, namely 4XNQ, where none of the models achieved a good result, and the 4YK4 and 5XKU antibodies where only AlphaFold 3 showed a reasonable result.

Besides accuracy, another important factor in predicting the three-dimensional structures of antibodies to be considered according to the literature is the model's execution time. For this metric, both ImmuneBuilder and AlphaFold 3 showed the best results, providing results in less than 1 minute on average. IgFold, on the other hand, took around 3 minutes per prediction, only beaten by AlphaFold 2 with an average of 5 minutes per prediction.

4.3 Simulating the evolution of antibodies

After selecting the prediction model that would be used later, we started preparing the ESM. To do this, the code in the repository (https://github.com/brianhie/efficient-evolution/tree/master) had to be analysed and interpreted to adapt it to the project's needs. However, although it was possible to successfully obtain 25 mutations for each antibody, to this day it has not been possible to adapt the python script to obtain the overall score of these mutations. It was only possible to obtain the logits, i.e. the values associated with each mutation by position in the sequence, which did not allow for the interpretation and categorisation of the mutations obtained to understand which would be the most useful.

5 Discussion

Throughout the work, it was possible to realize and identify the three-dimensional structure prediction models that worked better with antibodies, and the model with the best results was AlphaFold 3. This is due to its high precision and performance, even when predicting antibodies that other models had difficulty predicting. This performance can be justified by the great improvement in protein-ligand structure prediction, which shows that it is possible to deal with the great diversity of chemical space within a general deep learning framework and without resorting to an artificial separation between protein structure prediction and ligand docking. However, as can be seen, AlphaFold 3 also performed poorly on the 4XNQ antibody. This is because despite the great advance in modelling accuracy in AF3, there are still many targets for which accurate modelling can be a challenge. To obtain the highest accuracy, it may be necessary to generate many predictions and classify them, which entails additional computational cost [32]. The short prediction time obtained for the models, particularly AlphaFold 3 and ImmuneBuilder, was also predicted in the literature, as these models are usually faster than the others [28, 32].

With these results, it can already be seen that the application of Artificial Intelligence has had an incredible impact on the design and discovery of proteins, resulting in significant advances in the discovery, optimization and development of antibodies, saving months of expensive research in the laboratory. This makes AI a major point of interest for the pharmaceutical industry [34]. Currently, models can predict antibody affinity and non-specific binding using the dataset of antibody libraries for high and low levels of affinity and non-specific binding, allowing co-optimization of the affinity and specificity of therapeutic antibodies to accelerate drug development with highly potent antibodies [35] enabling a fully computational capability to generate new antibodies effective against targets [34]. However, despite the huge advances in the field, there are still challenges to be overcome, especially in antibody prediction, particularly in CDR loops (Complementarity-determining region). These regions show the greatest variability and are therefore often the area in which the greatest divergence is found [36].

6 Future work

A series of fundamental steps are still needed to fulfil the objectives initially proposed for this project. Firstly, it would be necessary to improve the Evolutionary Scale Modelling script so that it would be possible to see which mutations had the best score. This step is fundamental, and otherwise it will not be possible to proceed properly with the next activities. After this categorisation, it will be possible to extract the physicochemical properties using ProPythia and use the model previously defined as best for antibody prediction, in this case AlphaFold 3. By analysing both the three-dimensional structure and the properties of the mutations generated, we can identify whether any of the sequences meet specific criteria that make them relevant for testing in the laboratory.

7 References

1. Bridges, C.B., Kuehnert, M.J., Hall, C.B.: Transmission of Influenza: Implications for Control in Health Care Settings. Clinical Infectious Diseases. 37, 1094–1101 (2003). https://doi.org/10.1086/378292/2/37-8-1094-FIG005.GIF.
2. Kim, H., Webster, R.G., Webby, R.J.: Influenza Virus: Dealing with a Drifting and Shifting Pathogen. Viral Immunol. 31, 174–183 (2018). https://doi.org/10.1089/VIM.2017.0141/ASSET/IMAGES/LARGE/FIGURE2.JPEG.
3. Clohisey, S., Baillie, J.K.: Host susceptibility to severe influenza A virus infection. Crit Care. 23, 1–10 (2019). https://doi.org/10.1186/S13054-019-2566-7/TABLES/1.
4. Xu, X., Blanton, L., Elal, A.I.A., Alabi, N., Barnes, J., Biggerstaff, M., Brammer, L., Budd, A.P., Burns, E., Cummings, C.N., Garg, S., Kondor, R., Gubareva, L., Kniss, K., Nyanseor, S., O'Halloran, A., Rolfes, M., Sessions, W., Dugan, V.G., Fry, A.M., Wentworth, D.E., Stevens, J., Jernigan, D.: Update: Influenza Activity in the United States During the 2018–19 Season and Composition of the 2019–20 Influenza Vaccine. MMWR Morb Mortal Wkly Rep. 68, 544–551 (2019). https://doi.org/10.15585/MMWR.MM6824A3.
5. Leekha, S., Zitterkopf, N.L., Espy, M.J., Smith, T.F., Thompson, R.L., Sampathkumar, P.: Duration of Influenza A Virus Shedding in Hospitalized Patients and Implications for Infection Control. Infect Control Hosp Epidemiol. 28, 1071–1076 (2007). https://doi.org/10.1086/520101.
6. Gaitonde, D.Y., Moore, F.C., Morgan, M.K.: Influenza: Diagnosis and Treatment. Am Fam Physician. 100, 751–758 (2019).
7. Javanian, M., Barary, M., Ghebrehewet, S., Koppolu, V., Vasigala, V.K.R., Ebrahimpour, S.: A brief review of influenza virus infection, (2021). https://doi.org/10.1002/jmv.26990.
8. Te Velthuis, A.J.W., Fodor, E.: Influenza virus RNA polymerase: insights into the mechanisms of viral RNA synthesis. Nature Reviews Microbiology 2016 14:8. 14, 479–493 (2016). https://doi.org/10.1038/nrmicro.2016.87.

9. Hutchinson, E.C., Charles, P.D., Hester, S.S., Thomas, B., Trudgian, D., Martínez-Alonso, M., Fodor, E.: Conserved and host-specific features of influenza virion architecture. Nature Communications 2014 5:1. 5, 1–11 (2014). https://doi.org/10.1038/ncomms5816.
10. Maeda, T., Ohnishi, S. ichi: Activation of influenza virus by acidic media causes hemolysis and fusion of erythrocytes. FEBS Lett. 122, 283–287 (1980). https://doi.org/10.1016/0014-5793(80)80457-1.
11. Luo, M.: Influenza virus entry. Adv Exp Med Biol. 726, 201–221 (2012). https://doi.org/10.1007/978-1-4614-0980-9_9.
12. Gamblin, S.J., Vachieri, S.G., Xiong, X., Zhang, J., Martin, S.R., Skehel, J.J.: Hemagglutinin Structure and Activities. Cold Spring Harb Perspect Med. 11, a038638 (2021). https://doi.org/10.1101/CSHPERSPECT.A038638.
13. Wu, N.C., Wilson, I.A.: Influenza Hemagglutinin Structures and Antibody Recognition. (2020). https://doi.org/10.1101/cshperspect.a038778.
14. Dunbar, J., Krawczyk, K., Leem, J., Baker, T., Fuchs, A., Georges, G., Shi, J., Deane, C.M.: SAbDab: the structural antibody database. Nucleic Acids Res. 42, D1140–D1146 (2014). https://doi.org/10.1093/NAR/GKT1043.
15. Walsh, G.: Biopharmaceutical benchmarks 2010. Nature Biotechnology 2010 28:9. 28, 917–924 (2010). https://doi.org/10.1038/nbt0910-917.
16. Gao, R., Sheng, Z., Sreenivasan, C.C., Wang, D., Li, F.: Influenza A Virus Antibodies with Antibody-Dependent Cellular Cytotoxicity Function. Viruses 2020, Vol. 12, Page 276. 12, 276 (2020). https://doi.org/10.3390/V12030276.
17. Von Holle, T.A., Anthony Moody, M.: Influenza and antibody-dependent cellular cytotoxicity. Front Immunol. 10, 457028 (2019). https://doi.org/10.3389/FIMMU.2019.01457/BIBTEX.
18. Padilla-Quirarte, H.O., Lopez-Guerrero, D. V., Gutierrez-Xicotencatl, L., Esquivel-Guadarrama, F.: Protective antibodies against influenza proteins. Front Immunol. 10, 461698 (2019). https://doi.org/10.3389/FIMMU.2019.01677/BIBTEX.
19. Choi, Y., Deane, C.M.: Predicting antibody complementarity determining region structures without classification. Mol Biosyst. 7, 3327–3334 (2011). https://doi.org/10.1039/C1MB05223C.
20. Hoogland, M.A.: How Medical Students Discover and Use Medical Information Tools. Med Ref Serv Q. 38, 347–357 (2019). https://doi.org/10.1080/02763869.2019.1661197.
21. Adolf-Bryfogle, J., Xu, Q., North, B., Lehmann, A., Dunbrack, R.L.: PyigClassify: A database of antibody CDR structural classifications. Nucleic Acids Res. 43, D432–D438 (2015). https://doi.org/10.1093/nar/gku1106.
22. Burley, S.K., Berman, H.M., Kleywegt, G.J., Markley, J.L., Nakamura, H., Velankar, S.: Protein Data Bank (PDB): The single global macromolecular structure archive. In: Methods in Molecular Biology. pp. 627–641. Humana Press Inc. (2017). https://doi.org/10.1007/978-1-4939-7000-1_26.
23. Sgarbossa, D., Lupo, U., Bitbol, A.F.: Generative power of a protein language model trained on multiple sequence alignments. Elife. 12, (2023). https://doi.org/10.7554/eLife.79854.

24. Rives, A., Meier, J., Sercu, T., Goyal, S., Lin, Z., Liu, J., Guo, D., Ott, M., Zitnick, C.L., Ma, J., Fergus, R.: Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. Proc Natl Acad Sci U S A. 118, e2016239118 (2021). https://doi.org/10.1073/PNAS.2016239118/SUPPL_FILE/PNAS.2016239118.SAPP.PDF.
25. Rao, R., Meier, J., Sercu, T., Ovchinnikov, S., Rives, A.: Transformer protein language models are unsupervised structure learners. bioRxiv. 2020.12.15.422761 (2020). https://doi.org/10.1101/2020.12.15.422761.
26. Hie, B.L., Shanker, V.R., Xu, D., Bruun, T.U.J., Weidenbacher, P.A., Tang, S., Wu, W., Pak, J.E., Kim, P.S.: Efficient evolution of human antibodies from general protein language models. Nat Biotechnol. 42, 275–283 (2024). https://doi.org/10.1038/s41587-023-01763-2.
27. Sequeira, A.M., Lousa, D., Rocha, M.: ProPythia: A Python package for protein classification based on machine and deep learning. Neurocomputing. 484, 172–182 (2022). https://doi.org/10.1016/j.neucom.2021.07.102.
28. Abanades, B., Wong, W.K., Boyles, F., Georges, G., Bujotzek, A., Deane, C.M.: ImmuneBuilder: Deep-Learning models for predicting the structures of immune proteins. Communications Biology 2023 6:1. 6, 1–8 (2023). https://doi.org/10.1038/s42003-023-04927-7.
29. Kenlay, H., Dreyer, F.A., Cutting, D., Nissley, D., Deane, C.M.: ABody-Builder3: Improved and scalable antibody structure predictions.
30. Ruffolo, J.A., Chu, L.S., Mahajan, S.P., Gray, J.J.: Fast, accurate antibody structure prediction from deep learning on massive set of natural antibodies. Nature Communications 2023 14:1. 14, 1–13 (2023). https://doi.org/10.1038/s41467-023-38063-x.
31. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., Steinegger, M.: ColabFold: making protein folding accessible to all. Nature Methods 2022 19:6. 19, 679–682 (2022). https://doi.org/10.1038/s41592-022-01488-1.
32. 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. 630, 493–500 (2024). https://doi.org/10.1038/s41586-024-07487-w.
33. Henrique Ishivatari, L.U., Laura de Oliveira, L., Luís Barroso da Silva, F., Tinós, R.: Algoritmos Genéticos Com Função De Avaliação Dinâmica Para O Problema De Predição De Estruturas De Proteínas. (2017). https://doi.org/10.21528/CBIC2011-25.3.

34. Bai, G., Sun, C., Guo, Z., Wang, Y., Zeng, X., Su, Y., Zhao, Q., Ma, B.: Accelerating antibody discovery and design with artificial intelligence: Recent advances and prospects. Semin Cancer Biol. 95, 13–24 (2023). https://doi.org/10.1016/J.SEMCANCER.2023.06.005.
35. Goulet, D.R., Atkins, W.M.: Considerations for the Design of Antibody-Based Therapeutics. J Pharm Sci. 109, 74–103 (2020). https://doi.org/10.1016/J.XPHS.2019.05.031.
36. Fernández-Quintero, M.L., Kokot, J., Waibl, F., Fischer, A.L.M., Quoika, P.K., Deane, C.M., Liedl, K.R.: Challenges in antibody structure prediction. MAbs. 15, (2023). https://doi.org/10.1080/19420862.2023.2175319.