challenges

MDPI

Article

In Silico Epitope-Based Vaccine Prediction against Fungal Infection Aspergillosis

Anamika Basu

Department of Biochemistry, Gurudas College, Kolkata 700054, India; anamika.biochem@gurudas.education

Abstract: Aspergillus fumigatus is a pathogenic microorganism that causes aspergillosis due to the presence of its allergenic proteins. During the last two years, a few clinical cases have been reported where allergic bronchopulmonary aspergillosis (ABPA) has been detected in COVID-19 patients. The administration of antifungal medicine did not provide satisfactory results. It is a challenging job for medical scientists to protect mankind by designing an epitope-based vaccine against the rare disease aspergillosis. Other than twenty-three allergenic proteins, this microorganism contains an extracellular cellulase CelA expansin protein (Afu5g08030), which is allergenic. To design a peptide vaccine against aspergillosis, the identification of B cell and T cell epitopes is state-of-the-art technology. In our latest research, probable T cell and B cell epitopes are predicted. Molecular docking analysis of these predicted epitopes with their receptors is performed. Here, the primary sequence of the expansin protein is extracted and analyzed. Then, its secondary and tertiary structures are predicted using a homology modeling method and validated. Considering the physicochemical properties of this antigenic protein, two short stretches of peptides, namely ^{80}KPQADEDPNASSSSSSS^{96} and ^{286}DGGKTWQGTTRTS^{298}, are predicted as linear B cell epitopes. Similarly, based on its contacts with the highest number of alleles, the peptide sequence ^{221}LDLFQNAFTQLADVS^{235} is chosen as the most possible T cell epitope for the protein present in Aspergillus fumigatus with the highest binding energy for MHC II allele HLA-DRB1* 01: 01. Considering the binding energy of the B cell epitope with IgE, the second epitope ^{286}DGGKTWQGTTRTS^{298} is designated as the most potential epitope of B cells for this protein. Docking studies were performed with the T cell epitope with the human ternary complex of T cell receptor, CD4 receptor, and peptide-MHC II molecule (PDB ID 3T0E) with a binding energy of −192 Kcal/mole. For peptide-based vaccines, the proposed B cell and T cell epitopes may be used against aspergillosis after further experimental analysis.

Keywords: aspergillosis; expansin protein; epitope-based vaccine design; B cell epitope; T cell epitope

check for updates

Citation: Basu, A. In Silico Epitope-Based Vaccine Prediction against Fungal Infection Aspergillosis. *Challenges* **2022**, *13*, 29. https://doi.org/10.3390/challe13020029

Academic Editor: Susan L. Prescott

Received: 10 May 2022
Accepted: 4 July 2022
Published: 6 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Aspergillus fumigatus is a pathogenic microorganism producing many major allergens, which, during exposure to the human respiratory tract, causes severe asthma and sinusitis. Aspergillus, the mold (a type of fungus), causes a fungal infection, aspergillosis. Aspergillosis has two types, e.g., allergic Aspergillus sinusitis and allergic bronchopulmonary aspergillosis (ABPA). Allergic bronchopulmonary aspergillosis (ABPA) affects 1 to 15% of cystic fibrosis patients. One study found that 2.5% of individuals who have asthma have ABPA, which is around 4.8 million people around the world [1]. Of the 4.8 million people with ABPA, around 400,000 still have chronic aspergillosis of the lungs. Several antifungal medications are marginally effective in some cases, so the understanding of host–microbe interactions between humans and fungi is very important. Moreover, the incidences of aspergillosis have increased for SARS-CoV 2 patients [2]. Several cases have been reported during the pandemic situation, including aspergillosis infection as an acute complication in COVID-19 patients [3]. Designing and exploring new treatment approaches using immunotherapeutic approaches is indispensable. Cytokine therapy, cellular immunotherapy,

and vaccine preparation are various strategies that are extensively discussed, in addition to antifungal medicines [4].

Expansin is a plant protein that plays an important role in the expansion of plant cell walls. Although expansin proteins are mostly present in the plant kingdom, they are also found in other organisms such as bacteria and fungi [5]. Several expansin proteins can cause hay fever and asthma in humans [6]. Numerous expansin genes are detected in the fungus Aspergillus with their allergenic activities in humans [7]. These proteins are responsible for allergic reactions when they come into contact with the wet surface of the human respiratory tract [8]. Among the twenty-three allergenic proteins from Aspergillus fumigatus, ten of them have less sequence similarity between humans and mice and are analyzed using immunoinformatics approaches [7]. The total genomic sequence of Aspergillus fumigatus [9] shows that an extracellular cellulase CelA expansin protein (Afu5g08030) is expressed as a fungal expansin. In the UniProtKB database (https://www.uniprot.org/uniprot/Q4WUD1, accessed on 22 September 2017) [10], the protein Q4WUD1 (Q4WUD1_ASPFU) is expressed from gene AFUA_5G08030 and is described as extracellular cellulase CelA/allergen Asp F7-like, putative protein. This protein can act as an allergen in humans during fungal infection [8]. However, no data for epitopes of this protein are available from immunological datasets. This protein has been examined in this analysis to predict informative epitopes that will assist in the design of possible vaccines. Information about B cell and T cell epitopes that can cause particular immune responses is used to recognize peptide vaccines [11,12]. These planned epitopes would provide cost-effective therapeutics of high quality against this type of pathogen. The use of online algorithms in immunoinformatics has identified and described effective immunogenic T and B cell epitopes for this expansin antigen to develop future peptide vaccines for *Aspergillus fumigatus*. Vaccinomics is an amalgamation of two subjects, immunogenetics and immunogenomics, with details of system biology and immune profiling to design infectious disease vaccines [13,14]. Here, bioinformatics approaches are used to develop next-generation vaccines as well as personalized medicine.

2. Materials and Methods

The epitope-based vaccine is executed by using the following methods:

a. Retrieval of Afu5g08030 protein sequence

The protein sequence of Q4WUD1 (Q4WUD1_ASPFU), which is an extracellulase CelA/allergen Asp F-7-like/putative expansin (Afu5g08030) from the organism *Aspergillus fumigatus* containing 344 amino acids, is retrieved from the UniProtKB database [10] in FASTA format. Post-translational modification data of this protein molecule show that the 1–22 position of protein sequence results in the signal peptide and 23–344 sequence forms the matured protein. The transmembrane region is predicted for the 266–272 region by using the MEMSATSVM: transmembrane helix prediction method [15] (shown in Figures S1 and S2).

b. Determination of protein antigenicity

VaxiJen v 2.0, the web prediction server, calculates the antigenicity of this protein [16].

c. Primary and secondary structure prediction of putative expansin protein (Afu5g08030) from organism *Aspergillus fumigatus*

The linear sequence of the interested protein has been analyzed using a Proteomics server, ExPASy Prot Param (www.expasy.org, accessed on 4 March 2018) [17]. Some physicochemical properties are investigated, such as the molecular weight, theoretical pI, amino acid composition, the total number of negatively and positively charged amino acid residues, extinction coefficients, atomic composition, estimated half-life, aliphatic index, instability index, and GRAVY, which stands for the grand average of hydropathicity. Various servers, such as NPS @ Network Protein Sequence Analysis with the SOMPA [18] algorithm, are used to analyze the alpha helix, beta-pleated sheets, turns, and coils of the interested protein. Thus, the number and location of the above-mentioned features of the

secondary structure of that protein are predicted, and finally, this prediction will help us to identify the most suitable epitopes for vaccine design.

d. Modeling and validation of 3D homology modeled structure

First of all, the sequence alignment technique is used to determine the best fit between other 3D structures of different proteins in the 3D simulation of a protein structure using the sequence homology approach. Centered on the target-template alignment, models are constructed using ProMod3 [19]. The coordinates that are preserved between the target and the template are copied to the model from the template. Using a fragment collection, insertions and deletions are remodeled. The side chains are then reconstructed. Finally, when using a force field, the geometry of the resulting model is regularized. An alternate model is constructed with PROMOD-III if loop modeling with ProMod3 fails [20].

The 3D structure of this protein is also predicted by another method, BioSerf [21]. It is a fully automatic server for homology modeling. The procedure uses three prototype selection methods; PSI-BLAST [22] vs. PDB fasta [23], pGenTHREADER [24], and HH-Blits [25] vs. PDB. With conservative cut-offs, the top-scoring matches are then aggregated for the complete set of putative templates. All-by-all TM scores are then determined, and the matrix of scores is evaluated to exclude any conceivable outlying templates whose configuration is too distinct from the full set of templates. Finally, from the remaining templates, the best 10 templates are chosen and a model is created using MODELLER [26].

e. Estimation of the quality of modeled protein structure

The best model among the obtained models is calculated by QMEAN4 scores and Ramachandran plot using two kinds of selection methods. The consistency of the global and per-residue model was measured using the QMEAN scoring function [27]. The best model among the obtained models is calculated by the QMEAN4 scores and Ramachandran plot using two kinds of selection methods, i.e., PROCHECK [28] and RAMPAGE [29]. The consistency of the global and per-residue model is measured using the QMEAN scoring function [27]. Evaluation of the protein molecule's backbone conformation is assayed by the Ramachandran plot dividing the model's percentage of amino acid residues in the allowed and disallowed regions. Web server ERRAT [30] is used to predict the overall quality of the models, whereas energy minimization of the predicted model is subject to KoBaMIN, a protein structure refinement server [31].

f. Prediction of putative epitopes of B cells

To classify the positions of linear epitopes of the target protein, numerous physico-chemical properties such as flexibility, hydrophilicity, accessibility, exposed surface area, turns, polarity and the antigenic tendency of peptide chains are measured [32]. Various IEDB (www.iedb.org, accessed on 23 April 2018) [33] methods, including classical propensity scale methods such as the Kolaskar and Tongaonkar antigenicity scale [34], the Parker hydrophilicity prediction [35], the Emini surface accessibility prediction [36], Bepipred linear epitope prediction [37], Karplus and Schulz flexibility prediction [38], and Chou and Fashman prediction tool for beta-turn [39] are therefore used to estimate linear or continuous B cell epitopes for Aspergillus fumigatus expansin protein. The most likely B cell epitope of the specific protein has been identified with graphical results and prediction ratings.

g. Potential T-cell epitope Identification

Fungal antigens are processed in the human body with the help of MHC II molecules [40]. Therefore, linear T-cell epitopes for MHC-class II binding are reported for allergenic expansin proteins of fungal origin. The selected epitopes are documented with two consensus methods using an artificial neural network (ANN) [41] and stabilized matrix method (SMM) [42]. The findings are derived from the Immune Epitope Database (IEDB) tools for MHC-II binding prediction methods (www.iedb.org, accessed on 22 April 2018) [33]. This web server-based technique predicts the Major Histocompatibility Complex (MHC) class II

binding epitopes, considering binding scores to different MHC supertypes as a percentile rank.

h. Docking analysis of epitopes of B cells and T cells

T cell epitope ^{221}LDLFQNAFTQLADVS^{235} and B cell epitope ^{286}DGGKTWQGTTRTS^{298} are chosen and subjected to the 3D structure formation with the PEP-FOLD server [43,44] for docking studies. A ternary complex is formed when the MHC molecule, T cell epitope, and T cell receptor bind to each other during epitope recognition by T lymphocytes. T cells are capable of recognizing antigenic peptides by the CD4 receptor molecule, only after antigen processing and presenting by MHC molecules. Henceforth, stable ternary complex formation is an essential criterion for selected T cell epitope in designing a vaccine. To identify all molecular interactions with a whole ternary complex containing T cell receptor, a peptide with MHC molecule and CD4 receptor (PDB ID 3TOE), and the T cell epitope ^{221}LDLFQNAFTQLADVS^{235}, a molecular docking study is performed. Similarly, for B cell epitope ^{286}DGGKTWQGTTRTS^{298} and immunoglobulin E (PDB ID 4J4P), a docking study is executed using ClusPro [45,46], respectively.

3. Results & Discussion

A. Primary and secondary structure prediction of putative expansin protein (Afu5g08030) from organism *Aspergillus fumigatus*

The primary structure of this expansin protein is as follows:

>tr | Q4WUD1 | Q4WUD1_ASPFU Extracellular cellulase CelA/allergen Asp F7-like, putative OS = Neosartorya fumigata (strain ATCC MYA-4609/Af293/CBS 101,355/FGSC A1100) GN = AFUA_5G08030 PE = 4 SV = 1

MLYQRLTALGVAALVAASSVSASPMARGVKGRCRAKYNKAVSHVPTTHEATPTINL PVELPETQSQSPWPTVDEIVPVKKPQADEDPNASSSSSSSSSLSTTSALPSSPATQEQDTVS AAPAEPTTAAPETVNKAATTASSSSTTHSGKATFYGGNVSGGTCSFTGYTLPSGLFGTAYS GAAWNNAAECGACVSVTGPNGKTIKAMIVDQCPECEQDHLDLFQNAFTQLADVSKGI IPITWSFVPCGITTPLVLKNKEGTSPYWFSMQVMNANEPVAKLEVSTDGGKTWQGTTRTS YNFFEKSSGFGKDTVDVRVTGQSGATLTVTNVGTSSGSSVTAKSNL

The primary structure investigation shows that this fungal expansin Afu5g08030 protein has 344 amino acid residues and a 35.575 KD predicted molecular weight. Its estimated isoelectric pH is 5.84. The total number of negatively charged amino acid residues (Asp + Glu) and positively charged amino acid residues (Arg + Lys) are 26 and 23, respectively. Theoretically, the value of the isoelectric point and the number of charged residues of amino acids disclose the protein as being negatively charged. The estimated half-life of this protein in mammalian reticulocytes is 30 h and the computed instability index (II) is 39.10, which clarifies this protein as a stable compound. The aliphatic index is calculated as 39.10 and the amino acid M (Met) has appeared as the N-terminal amino acid of this antigenic protein. The grand average of hydropathicity is GRAVY of −0.237, which indicates that the protein is hydrophilic. The Grand Average Index of Hydropathicity (GRAVY) is used to depict the peptide hydrophobicity score, which measures the total of all amino acid hydropathy values divided by the length of the sequence. The amino acids Ala (A), Ser (S) and Thr (T) are found in higher proportion, i.e., 10.5%, 13.7%, and 12.2%, compared to the other amino acids in the protein. The amino acids His (H), Met (M), Trp (W) and Arg (R) are present in lower amounts, i.e., 1.7%, 1.5% and 1.7% in protein.

The secondary structure prediction by using SOMPA [18], with a threshold of 8.0, has shown that the presence of 69 (20.06%) α-helix, 19 (5.52%) β turn, 86 (25.00%) extended strand and 170 (49.42%) random coils (Figure 1). Positive GRAVY values indicate hydrophobic; negative values mean hydrophilic. The GRAVY value of the primary structure of the protein is predicted negative (−0.237), which suggests that most of the amino acid residues of the allergenic protein are hydrophilic. Similarly, this protein is stable. The threonine residue (T) 12.2% is present in the protein, which prefers to lie in the β sheet. The secondary structure analysis of the protein by the SOMPA algorithm shows that a 25% portion of the

protein remains in the β sheet. From the other researchers' works [47], the antigenic portion of a protein has been shown, probably belonging to the β sheet region of that protein. Among the two predicted B cell epitopes peptide 1 (80–96) and peptide 2 (286–298), the part of peptide 2, i.e., ^{291}WQGTTRT^{297}, is present in the β sheet, as shown in Figure 1. Similarly, predicted T cell epitope ^{237} GIIPITWSF^{245} is also situated in the β sheet structure.

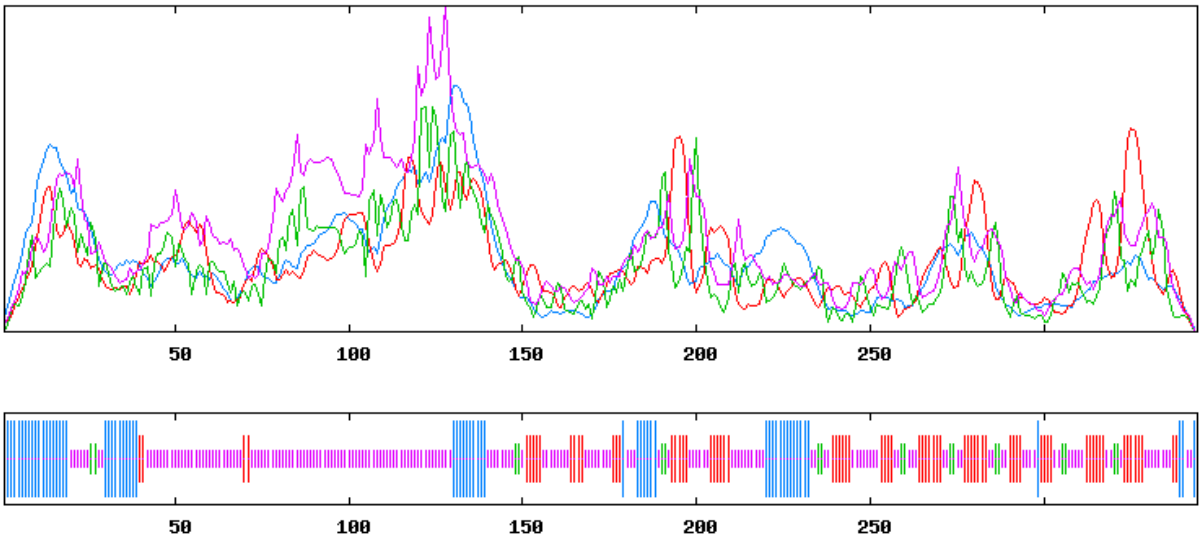


Figure 1. Secondary structure of the expansin protein obtained by using SOMPA. Here, helix, random coil, extended strand and beta-turn are blue, magenta, red and green, respectively. Also shown as a bar diagram.

The epitopes, KPQADEDPNASSSSSSS and DGGKTWQGTTRTS, are present in the coiled structure regions in the antigenic protein, whereas the predicted epitope LDLFQNAFTQLADVS is present in the helical secondary structure region, as shown in Figure 1.

B. Homology modeling and validation of 3D homology modeled structure

The three-dimensional structure of the fungal pollen protein has been modeled in model 1 and model 2 by using psipred [48] and SWISSMODEL [49] servers, respectively. In Table 1, the template 3D3O_A is identified for modeling the protein with the highest NetScore 72.732 and 197 alignment length (Table 1) for model 1. Similarly, this template 3D3O_A also shows the highest sequence identity (31.61) among three templates with a sequence similarity of 0.36 using the SWISSMODEL server (Table 2) for model 2.

Table 1. Template search results from psipred.

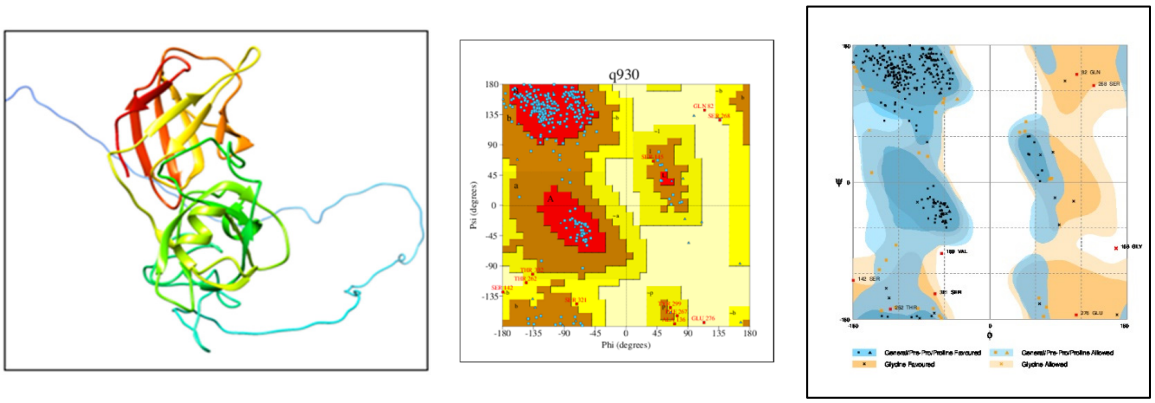
Template	NetScore	*p*-Value	PairE	SolvE	Alignment Score	Alignment Length	Structure Length	Sequence Length
3D3O-A	72.732	2×10^{-6}	−186.3	−15.0	340.0	197	208	344
1n10-A	58.550	6×10^{-5}	−92.5	−6.1	272.0	201	228	344
2HCZ-X	58.548	6×10^{-5}	−144.3	−7.8	256.0	213	242	344

The best 3D model selected by the QMEAN server [27] is model 2, obtained from the SWISSMODEL server [49] with its −1.73 QMEAN4 value (shown in Figure 2a). Model 1 has a −4.96 QMEAN value. The value of QMEAN4 is a linear combination of four possible mathematical terms, which can be translated to a Z ranking. The more desirable model is correlated with a higher Z ranking. For this model, Ramachandran diagrams are drawn using two web servers, e.g., RAMPAGE [29] and PROCHECK [28]. The high-resolution Ramachandran plot from RAMPAGE and PROCHECK web servers are shown

in Figures S3 and S4 in the Supplementary File. In Figure 2b,c, these diagrams are shown. In the Ramachandran map, the value of G-factors offers a measure of the uncommon or out-of-the-regular configuration in the predicted model. The total average value of G-factors is −0.35 for this model, which is not uncommon for dihedral angles and covalent main-chain forces. From the RAMPAGE web server [29], it is calculated that model 1 has 303 (88.6%) residues, present in favored regions. The 31 (9.1%) residues are present in the allowed region, and 8 (2.3%) residues are in the outlier region. Furthermore, for model 2, RAMPAGE has calculated that the 188 (96.4%) residues are present in the favored region. The Ramachandran plot for this modeled protein shows that 4 (2.1%) residues are in the allowed region and only three (1.5%) residues are present in the outlier region. In general, ~98.0% and ~2.0% residues should be expected in favored and allowed regions, respectively. The statistics for the Ramachandran plot from PROCHECK [28] for model 1 show that 233 (79.8%) residues are present in the most favored regions [A, B, L], 48 (16.4%) residues are present in the additional allowed regions [a, b, l, p], 8 (2.7%) residues are present in the generously allowed regions [~a, ~b, ~l, ~p] and 3 (1.0%) amino acid residues are present in the disallowed regions [X, X]. Similarly, Ramachandran plot statistics for model 2 have 146 residues (89.6%) among 197 present in the most favored regions [A, B, L], and 13 (8.0%) amino acid residues are present in the additional allowed regions [a, b, l, p], two residues are present (1.2%) in the generously allowed regions [~a, ~b, ~l, ~p] and two (1.2%) residues are present in the disallowed regions [X, X].

Table 2. Template search results from SWISSMODEL.

Template	Sequence Identity	Found by	Method	Resolution	Sequence Similarity	Coverage
3D3O-A	31.61	HHblits	X-ray	1.90 Å	0.36	0.56
1n10-A	18.78	HHblits	X-ray	2.90 Å	0.31	0.57
2HCZ-X	19.42	HHblits	X-ray	2.75 Å	0.37	0.56



(**a**) (**b**) (**c**)

Figure 2. (**a**) Modeled structure of Afu5g08030 (**b**,**c**) Ramachandran plot of the predicted protein.

Verify3D [50] is used to assess the quality of both model 1 and model 2. A total of 57.27% of the residues of model 1 have an average 3D/1D score $\geq$ 0.2 as shown in Figure 3. Similarly, 100% of the residues of model 2 have an average 3D/1D score $\geq$ 0.2, which is displayed in Figure 4. The plot for 3D/1D vs. average score for both model 1 and model 2 are shown in Figures S5 and S6 in Supplementary File S1. The overall quality factors for both models are 29.630 and 77.778. So, the validation of model 2 by ERRAT2 [30] is executed. Subsequently, model 2 is refined for the energy minimization process by using KoBaMIN [31], a web server for knowledge-based potential refinement for proteins.

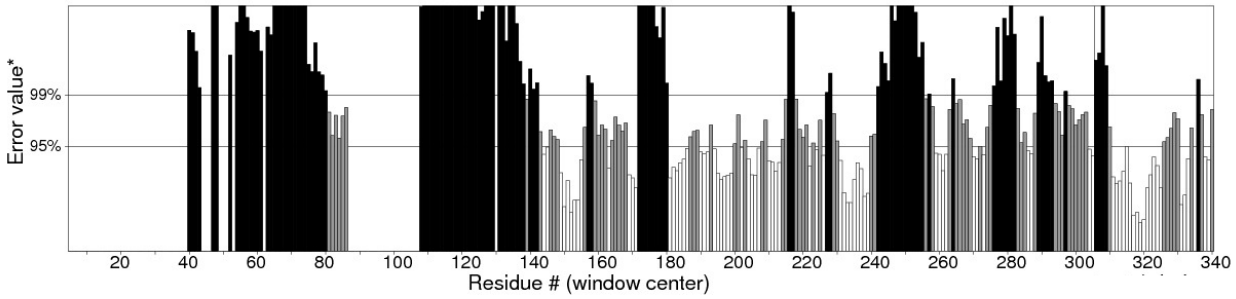


Figure 3. Quality estimation of modeled protein structure. * On the error axis, two lines are drawn to indicate the confidence with which it is possible to reject regions that exceed that error value.

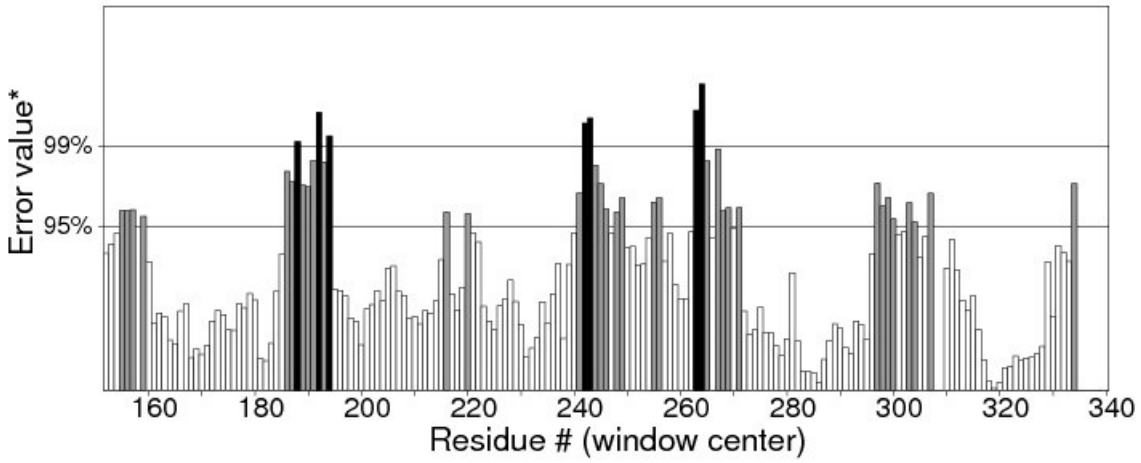


Figure 4. Quality estimation of model 2 protein structure. * On the error axis, two lines are drawn to indicate the confidence with which it is possible to reject regions that exceed that error value.

The minimized energy, RMSD values and time spent for minimization for both models are exhibited in Table 3. The energy minimized model 2 structure with its Z-score plot is shown in Figure 5. The Z-score plot describes the general quality of the modeled protein structure. The Z-score measures the deviation of the whole energy of the modeled structure concerning the energy.

Table 3. Minimized energy and other parameters from KoBaMIN.

Model No.	KB Energy (Kcal/mol)	RMSD (Initial)	RMSD (Reference)	GDT-TS (Reference)	GDT-HA (Reference)	Time Spent (s)
Model 1	−2708.3279	2.663	2.663	0.785	0.624	281
Model 2	−3235.077	0.379	0.379	1.000	0.971	50

C. Prediction of putative B cell epitope

The B cell epitope's characteristic features rely on the structure's hydrophilicity, flexibility, surface accessibility and beta-turn. Figures 6 and 7 show the prediction scores of the Chou and Fasman method for beta-turn [39], Emini surface accessibility [36], Kolaskar and Tangaonkar antigenicity [34], and Parker hydrophilicity [35]. Tables 4 and 5 outline the predicted linear B cell epitopes and the hydrophilicity, accessibility, flexibility, and beta-turn prediction score of each residue present in these epitopes.

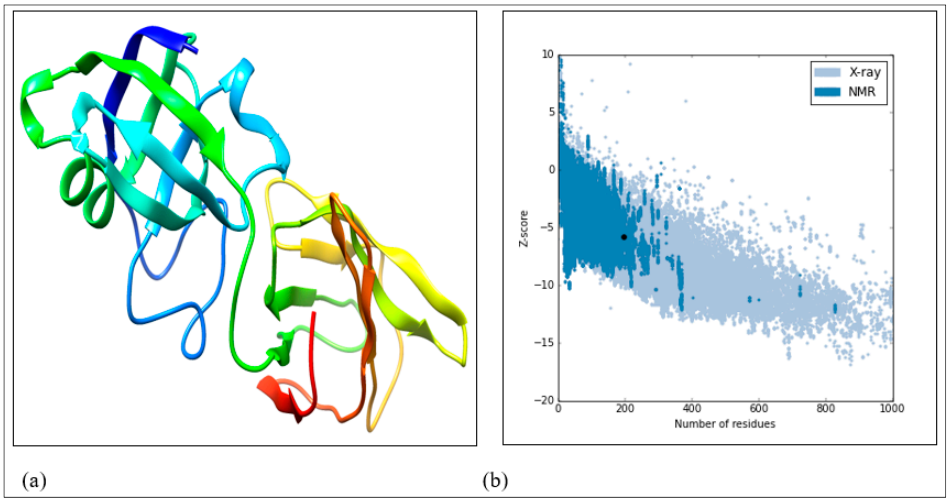


Figure 5. (**a**) Energy minimized structure with (**b**) its Z-score plot for model 2.

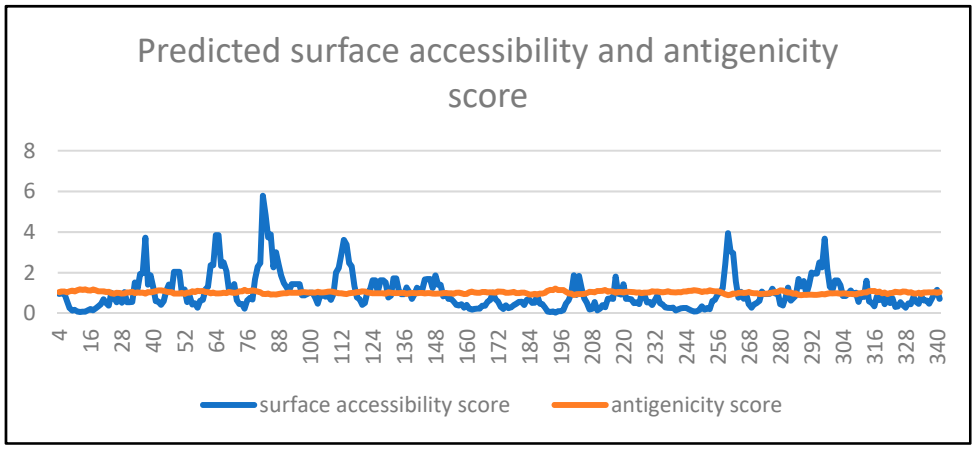


Figure 6. Prediction score for surface accessibility and antigenicity of the fungal protein.

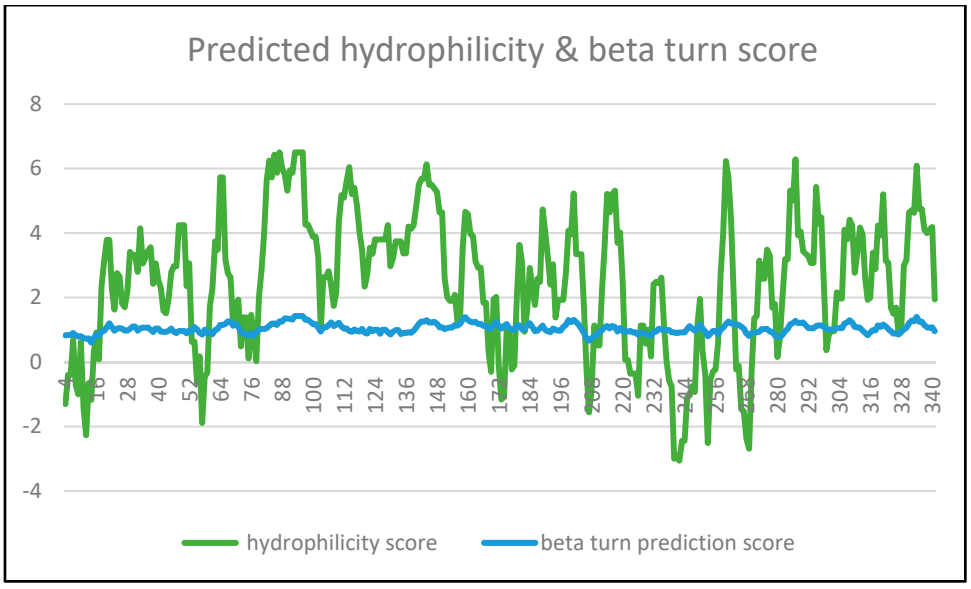


Figure 7. Prediction scores for hydrophilicity and beta-turn of the fungal protein for each amino acid residue.

Table 4. Different prediction scores for each amino acid in the first predicted B cell epitope.

PEPTIDE	Karplus and Schulz Flexibility Prediction (Threshold Value = 1.018)	Parker Hydrophilicity Prediction (Threshold Value = 2.505)	Emini Surface Accessibility Prediction (Threshold Value = 1.000)	Chou and Fasman Beta-Turn Prediction (Threshold Value = 1.039)
^{80}K	1.048	2.857	2.295	1.029
P	1.056	3.986	2.478	1.02
Q	1.059	5.629	5.783	1.054
A	1.059	6.243	4.829	1.119
D	1.062	5.729	3.734	1.191
E	1.06	6.429	3.883	1.197
D	1.053	5.871	2.265	1.151
P	1.046	6.5	3.005	1.261
N	1.046	6	2.411	1.257
A	1.061	5.814	1.866	1.356
S	1.087	5.314	1.497	1.351
S	1.12	5.943	1.298	1.339
S	1.146	5.871	1.081	1.32
S	1.161	6.5	1.435	1.43
S	1.169	6.5	1.435	1.43
S	1.169	6.5	1.435	1.43
^{96}S	1.61	6.5	1.435	1.43

Table 5. Different prediction scores for each amino acid in the second predicted B cell epitope.

PEPTIDE	Karplus and Schulz Flexibility Prediction (Threshold Value = 1.018)	Parker Hydrophilicity Prediction (Threshold Value = 2.505)	Emini Surface Accessibility Prediction (Threshold Value = 1.000)	Chou and Fasman Beta-Turn Prediction (Threshold Value = 1.039)
^{286}D	1.077	5.014	0.868	1.211
G	1.095	6.286	1.687	1.277
G	1.09	3.929	1.324	1.210
K	1.072	4.043	1.589	1.213
T	1.055	3.429	0.941	1.227
W	1.039	3.357	1.373	1.141
Q	1.045	3.286	2.002	1.056
G	1.060	3.071	1.961	1.047
T	1.067	3.071	1.961	1.047
T	1.071	5.429	2.499	1.114
R	1.058	4.300	2.261	1.137
T	1.038	4.486	3.674	1.137
^{298}S	1.013	2.429	2.204	1.086

For each residue of peptides ^{80}KPQADEDPNASSSSSS^{96} and ^{286}DGGKTWQGTTRTS^{298}, prediction scores not only for Parker hydrophilicity, Emini surface accessibility, Karplus and Schulz flexibility, but also Chou and Fashman beta-turn show that these small parts of selected protein will serve as linear B cell epitopes.

Kolaskar and Tongaonkar's antigenicity score and Chou and Fasman's beta-turn score are predicted for every amino acid present in the protein. These two scores remain almost unchanged for all the amino acids in the protein. Linear B cell epitopes are present on the surface of antigenic proteins. From Figure 7, it can be shown that the Emini surface accessibility score is highest in the 80–96 region, and for the 286–298 region, the surface accessibility score is very high. So, the amino acids that are present in these two regions are accessible to immunoglobulin molecules for immunological reactions. Similarly, the beta-turn score for these two regions is very high compared to the other regions. Since the antigenic part of a protein probably belongs to the β sheet region of that protein [47]. The Parker hydrophilicity predicted score shows that the amino acid side chains, which are present in regions 80–96 and 286–298, are hydrophilic. B cell epitopes are generally exposed to the surface of the protein sequence, due to the presence of the hydrophilic amino acids in the protein sequence. So, these two peptides are selected as predicted B cell epitopes of the antigenic protein.

D. Prediction of putative T cell epitope

The ^{221}LDLFQNAFTQLADVS^{235} peptide sequence is chosen among the five peptide sequences as the most likely T cell epitope for the specific protein present in Aspergillus fumigatus based on its connections with the largest number of alleles (Table 6) [41,42]. Not only is the IC50 value of this peptide the least for the MHC class II allele HLA-DRB1*01:01, but it also has the highest population coverage among the other peptides (Table 7).

Table 6. Result for interacting MHC II super type prediction.

Sl. No.	Peptide Sequence	Start Position	Interacting MHC II Alleles	Percentile Rank
1.	LDLFQNAFTQLADVS	221	HLA-DPA1*01:01	0.16
			HLA-DPA1*02:01	1
			HLA-DQA1*01:02	1.68
			HLA-DRB1*04:01	1.83
			HLA-DRB1*04:05	2.40
			HLA-DRB1*08:02	2.96
			HLA-DPA1*02:01	3.01
			HLA-DRB1*11:01	3.84
			HLA-DRB1*15:01	5.37
			HLA-DPA1*01:03	6.05
			HLA-DPB3*02:02	6.20
			HLA-DPA1*03:01	6.44
			HLA-DRB1*01:01	6.65
			HLA-DRB5*01:01	7.08
			HLA-DRB1*12:01	7.93
2.	TQEQDTVSAAPAEPT	112	HLA-DQA1*04:01	0.33
			HLA-DQA1*03:01	1.12
			HLA-DQA1*05:01	12.46
			HLA-DQA1*05:01	14.01
3.	WQGTRTSYNFFEKS	291	HLA-DPA1*01:01	0.73
			HLA-DPA1*02:01	5.63
			HLA-DPA1*02:01	7.77
			HLA-DRB1*11:01	15.78
4.	WPTVDEIVPVKKPQA	69	HLA-DRB1*11:01	0.61
			HLA-DRB1*08:02	1.92
			HLA-DRB1*09:01	5.40
			HLA-DRB5*01:01	11.11
			HLA-DRB1*03:01	11.43
5.	FGTAYSGAAWNNAAE	176	HLA-DRB1*09:01	0.66
			HLA-DRB1*04:01	8.36
			HLA-DRB5*01:01	10.53
			HLA-DQA1*05:01	13.58

Table 7. Binding affinity and population coverage for the selected peptide sequences from NetMHCI-Ipan.

Sl No.	Identified Peptide Sequences	Starting Position	Core Sequence	Interacting Allele	IC_{50} Value	Population Coverage
1.	LDLFQNAFTQLADVS	221	FQNAFTQLA	HLA-DRB1*01:01	18.70	99.03%
2.	TQEQDTVSAAPAEPT	112	VSAAPAEPT	HLA-DRB1*01:01	1148.22	75.22%
3.	WQGTRTSYNFFEKS	291	TRTSYNFFE	HLA-DRB1*04:05	1908.57	68.35%
4.	WPTVDEIVPVKKPQA	69	IVPVKKPQA	HLA-DRB1*11:01	931.14	29.43%
5.	FGTAYSGAAWNNAAE	176	YSGAAWNNA	HLA-DRB1*01:01	304.21	30.76%

The IC_{50} value indicates the binding efficiency of T cell epitopes with the MHC molecules. The lower the IC_{50} value, the higher the binding efficiency. For the world population, the population coverage for established T cell epitopes is calculated using the IEDB population coverage estimation tool [51]. This method calculates the percentage of people expected to respond to a given set of epitopes with known MHC restrictions.

E. Molecular docking results for the identified B cell and T cell epitopes

Docking experiments are conducted using the FFT-based algorithm [45] on the ClusPro 2.2 web server. Cluster ratings for the prediction of the lowest binding energy are calculated. Docking studies are performed with the T cell epitope with the human ternary complex of T cell receptor, peptide-MHC II molecule and CD4 receptor (PDB ID 3T0E) resulting in the binding energy of −192 Kcal/mole (as shown in Figure 8) [45,46]. The ternary complex is responsible for T cell recognition and antigen presentation. This molecular docking structure is stabilized by several hydrogen bonds (shown in Table 8).

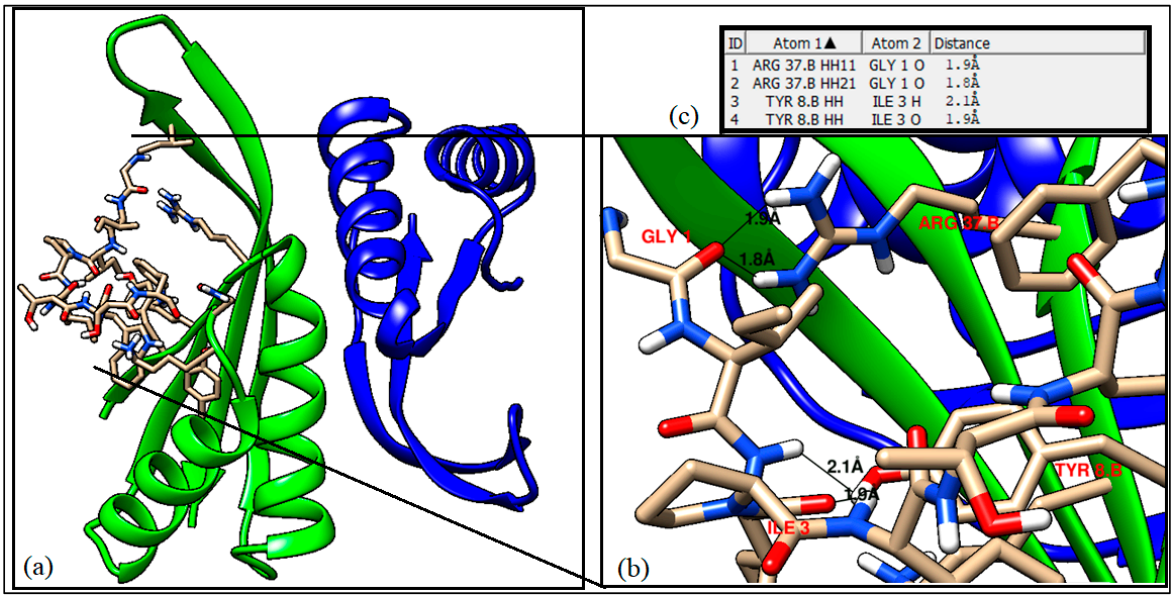


Figure 8. (**a**) Docking structure of predicted T cell epitope with ternary complex (**b**) Interacting amino acid residues in docking structure with hydrogen bonds (**c**) Atoms participate in hydrogen bonds with their bond distances.

Table 8. Details of H bonds between ternary structure and T cell epitope.

Sl No.	Atom 1	Atom 2	Distance (Å)
1.	ARG 37.B HH11	GLY 1 O	1.919
2.	ARG 37.B HH21	GLY 1 O	1.774
3.	TYR 8.B OH	ILE 3 H	2.086
4.	TYR 8.B HH	ILE 3 O	1.883

The molecular docking structure for T cell epitope LDLFQNAFTQLADVS with MHC II molecule is shown in Figure 8a, where the epitope is represented as a stick structure. The α and β chains of the MHC molecule are depicted as green and blue ribbons, respectively. Interacting atoms and bond lengths of H-bonds are marked in Figure 8b. H-bond distances are enlisted in Figure 8c and also shown in Table 8. The whole structure of the T cell receptor with the ternary complex is shown in Figure S13 in the Supplementary Material.

In contrast, a docking study is carried out with two predicted B cell epitopes with human immunoglobulin E (PDB ID 4J4P). Binding energies of the bound structures of B cell epitopes viz, ^{80}KPQADEDPNASSSSSSS^{96} and ^{286}DGGKTWQGTTRTS^{298} are −272.9 and −301.0, respectively. Considering the binding energy of the B cell epitope with IgE, the second epitope ^{286}DGGKTWQGTTRTS^{298} is designated as the most probable B cell epitope of this fungal protein and the interaction of this epitope with Ig E is shown in Figures S14 and S15. The components of binding energies for B cell epitope 1 and epitope 2, from the DockThor server [52], are shown in Table S6.

4. Conclusions

A serious fungal infection known as COVID-19-associated pulmonary aspergillosis (CAPA) complicates the condition of severely ill COVID-19 patients. Numerous retrospective and prospective investigations have been carried out worldwide to better understand this deadly co-infection [53,54]. The mortality rate for COVID-19-associated pulmonary aspergillosis (CAPA) is 52% [55]. In CAPA, antifungal medication is advised, and in individuals who do not react, discontinuing or reducing concurrent corticosteroid therapy may be an option [56]. To overcome this challenging problem associated with COVID-19 patients, an epitope-based vaccine has been developed for protective purposes.

Since the allergenic protein is fungal, it can produce both B cell and T cell immune reactions. For B cell epitope prediction, several physical parameters are considered. In our study, the binding energy of B cell epitope with human immunoglobulin E reveals a thermodynamically stable complex structure. In the case of B cell immunity, the antigenic protein binds with immunoglobulin E molecules and triggers humoral immunity as a part of the adaptive immune system. Fungal infections, acquired via inhalation, are connected with T cell response. For *Aspergillus* spp., T cell activation is a decisive event for the development of protective immunity. The ternary complex of T cell receptor, peptide-MHC and CD4 molecule is used for docking analysis with predicted T cell epitope forming a stable docking structure.

From the various prediction scores, the most feasible B cell epitope is predicted as ^{286}DGGKTWQGTTRTS^{298}. Similarly,^{221}LDLFQNAFTQLADVS^{235} peptide is considered as the most credible T cell epitope considering the interactions with fifteen numbers of MHC alleles with the least IC_{50} value of 18.70 nm for HLA-DPA1*01:01allelic protein. Thus, the projected epitopes of B cells and T cells can be used after more experimental studies for peptide vaccine design.

Funding: This research received no external funding.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/challe13020029/s1 from Supplementary file S1.

Institutional Review Board Statement: This article does not contain any studies with human participants or animals performed by any of the authors. Hence, no formal consent is required.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data generated or analyzed during this study are included in this published article.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Denning, D.W.; Pleuvry, A.; Cole, D.C. Global burden of allergic bronchopulmonary aspergillosis with asthma and its complication chronic pulmonary aspergillosis in adults. *Med. Mycol.* **2013**, *51*, 361–370. [CrossRef] [PubMed]
2. Özdemir, Ö.; Pop, S.; Engin, M.M.N. COVID-19 in an Allergic Bronchopulmonary Aspergillosis Patient: A Case Report. *Turk. Arch. Pediatr.* **2021**, *56*, 396.
3. Arastehfar, A.; Carvalho, A.; van de Veerdonk, F.L.; Jenks, J.D.; Koehler, P.; Krause, R.; Cornely, O.A.; Perlin, D.S.; Lass-Flörl, C.; Hoenigl, M.; et al. COVID-19 associated pulmonary aspergillosis (CAPA)—From immunology to treatment. *J. Fungi* **2020**, *6*, 91. [CrossRef] [PubMed]
4. Feys, S.; Almyroudi, M.P.; Braspenning, R.; Lagrou, K.; Spriet, I.; Dimopoulos, G.; Wauters, J. A visual and comprehensive review on COVID-19-associated pulmonary aspergillosis (CAPA). *J. Fungi* **2021**, *7*, 1067. [CrossRef]
5. Georgelis, N.; Nikolaidis, N.; Cosgrove, D.J. Bacterial expansins and related proteins from the world of microbes. *Appl. Microbiol. Biotechnol.* **2015**, *99*, 3807–3823. [CrossRef]
6. Galli, S.J.; Tsai, M.; Piliponsky, A.M. The development of allergic inflammation. *Nature* **2008**, *454*, 445–454. [CrossRef]
7. Thakur, R.; Shankar, J. In silico Identification of Potential Peptides or Allergen Shot Candidates Against *Aspergillus fumigatus*. *BioResearch Open Access* **2016**, *5*, 330–341. [CrossRef]
8. Rabe, U. Are Fungal Spores Relevant Outdoor Allergens? *Laryngo-Rhino-Otologie* **2020**, *99*, 292.
9. Nierman, W.C.; Pain, A.; Anderson, M.J.; Wortman, J.R.; Kim, H.S.; Arroyo, J.; Berriman, M.; Abe, K.; Archer, D.B.; Bermejo, C.; et al. Genomic sequence of the pathogenic and allergenic filamentous fungus Aspergillus fumigatus. *Nature* **2005**, *438*, 1151. [CrossRef]
10. Available online: http://www.uniprot.org/uniprot/Q4WUD1 (accessed on 22 September 2017).
11. Da Silva, L.B.R.; Taborda, C.P.; Nosanchuk, J.D. Advances in Fungal Peptide Vaccines. *J. Fungi* **2020**, *6*, 119. [CrossRef]
12. Kardani, K.; Bolhassani, A.; Namvar, A. An overview of in silico vaccine design against different pathogens and cancer. *Expert Rev. Vaccines* **2020**, *19*, 699–726. [CrossRef] [PubMed]
13. Poland, G.A.; Ovsyannikova, I.G.; Kennedy, R.B. Personalized vaccinology: A review. *Vaccine* **2018**, *36*, 5350–5357. [CrossRef] [PubMed]
14. Whitaker, J.A.; Ovsyannikova, I.G.; Poland, G.A. Adversomics: A new paradigm for vaccine safety and design. *Expert Rev. Vaccines* **2015**, *14*, 935–947. [CrossRef] [PubMed]
15. Nugent, T.; Jones, D.T. Transmembrane protein topology prediction using support vector machines. *BMC Bioinform.* **2009**, *10*, 159. [CrossRef]
16. Doytchinova, I.A.; Flower, D.R. VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinform.* **2007**, *8*, 4. [CrossRef] [PubMed]
17. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Wilkins, M.R.; Appel, R.D.; Bairoch, A. Protein identification and analysis tools on the ExPASy server. In *The Proteomics Protocols Handbook*; Humana Press: Humana Totowa, NJ, USA, 2005; pp. 571–607.
18. Geourjon, C.; Deleage, G. SOPMA: Significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Bioinformatics* **1995**, *11*, 681–684. [CrossRef] [PubMed]
19. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F.T.; de Beer, T.A.P.; Rempfer, C.; Bordoli, L.; et al. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **2018**, *46*, W296–W303. [CrossRef]
20. Van Gunsteren, W.F.; Berendsen, H.J. Algorithms for macromolecular dynamics and constraint dynamics. *Mol. Phys.* **1977**, *34*, 1311–1327. [CrossRef]
21. Buchan, D.W.; Minneci, F.; Nugent, T.C.; Bryson, K.; Jones, D.T. Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res.* **2013**, *41*, W349–W357. [CrossRef]
22. Jones, D.T.; Swindells, M.B. Getting the most from PSI-BLAST. *Trends Biochem. Sci.* **2002**, *27*, 161–164. [CrossRef]
23. Martin, A.C. Mapping PDB chains to UniProtKB entries. *Bioinformatics* **2005**, *21*, 4297–4301. [CrossRef] [PubMed]
24. Lobley, A.; Sadowski, M.I.; Jones, D.T. pGenTHREADER and pDomTHREADER: New methods for improved protein fold recognition and superfamily discrimination. *Bioinformatics* **2009**, *25*, 1761–1767. [CrossRef] [PubMed]

25. Remmert, M.; Biegert, A.; Hauser, A.; Söding, J. HHblits: Lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat. Methods* **2012**, *9*, 173–175. [CrossRef]
26. Webb, B.; Sali, A. Comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinform.* **2016**, *54*, 5–6. [CrossRef] [PubMed]
27. Benkert, P.; Künzli, M.; Schwede, T. QMEAN server for protein model quality estimation. *Nucleic Acids Res.* **2009**, *37* (Suppl. S2), W510–W514. [CrossRef] [PubMed]
28. Laskowski, R.A.; MacArthur, M.W.; Thornton, J.M. (2006). PROCHECK: Validation of protein-structure coordinates. *Int. Tables Crystallogr.* **2006**, *F*, 722–725.
29. Lovell, S.C.; Davis, I.W.; Arendall, W.B.; de Bakker PI, W.; Word, J.M.; Prisant, M.G.; Richardson, J.S.; Richardson, D.C. Structure validation by Calpha geometry: Phi, psi and Cbeta deviation. *Proteins* **2003**, *50*, 437–450. [CrossRef]
30. Lengths, M.; Angles, M. Limitations of structure evaluation tools errata. *Quick Guidel. Comput. Drug Des.* **2018**, *16*, 75.
31. Rodrigues, J.P.; Levitt, M.; Chopra, G. KoBaMIN: A knowledge-based minimization web server for protein structure refinement. *Nucleic Acids Res.* **2012**, *40*, W323–W328. [CrossRef]
32. Fieser, T.M.; Tainer, J.A.; Geysen, H.M.; Houghten, R.A.; Lerner, R.A. Influence of protein flexibility and peptide conformation on reactivity of monoclonal anti-peptide antibodies with a protein alpha-helix. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 8568–8572. [CrossRef]
33. Dhanda, S.K.; Mahajan, S.; Paul, S.; Yan, Z.; Kim, H.; Jespersen, M.C.; Jurtz, V.; Andreatta, M.; Greenbaum, J.A.; Marcatili, P.; et al. IEDB-AR: Immune epitope database—Analysis resource in 2019. *Nucleic Acids Res.* **2019**, *47*, W502–W506. [CrossRef] [PubMed]
34. Kolaskar, A.S.; Tongaonkar, P.C. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett.* **1990**, *276*, 172–174. [CrossRef]
35. Parker, J.M.R.; Guo, D.; Hodges, R.S. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: Correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. *Biochemistry* **1986**, *25*, 5425–5432. [CrossRef] [PubMed]
36. Emini, E.A.; Hughes, J.V.; Perlow, D.; Boger, J. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *J. Virol.* **1985**, *55*, 836–839. [CrossRef]
37. Larsen JE, P.; Lund, O.; Nielsen, M. Improved method for predicting linear B-cell epitopes. *Immunome Res.* **2006**, *2*, 1–7. [CrossRef] [PubMed]
38. Karplus, P.A.; Schulz, G.E. Prediction of chain flexibility in proteins. *NW* **1985**, *72*, 212–213. [CrossRef]
39. Chou, P.Y.; Fasman, G.D. Prediction of the secondary structure of proteins from their amino acid sequence. *Adv. Enzym.* **1978**, *47*, 45–147.
40. Larsen, M.V.; Lundegaard, C.; Lamberth, K.; Buus, S.; Lund, O.; Nielsen, M. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinform.* **2007**, *8*, 424. [CrossRef]
41. Nielsen, M.; Lundegaard, C.; Worning, P.; Lauemøller, S.L.; Lamberth, K.; Buus, S.; Brunak, S.; Lund, O. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Sci.* **2003**, *12*, 1007–1017. [CrossRef]
42. Peters, B.; Sette, A. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC Bioinform.* **2005**, *6*, 132. [CrossRef]
43. Maupetit, J.; Derreumaux, P.; Tuffery, P. PEP-FOLD: An online resource for de novo peptide structure prediction. *Nucleic Acids Res.* **2009**, *37* (Suppl. S2), W498–W503. [CrossRef] [PubMed]
44. Shen, Y.; Maupetit, J.; Derreumaux, P.; Tufféry, P. Improved PEP-FOLD approach for peptide and mini protein structure prediction. *J. Chem. Theory Comput.* **2014**, *10*, 4745–4758. [CrossRef] [PubMed]
45. Kozakov, D.; Hall, D.R.; Xia, B.; Porter, K.A.; Padhorny, D.; Yueh, C.; Beglov, D.; Vajda, S. The ClusPro web server for protein-protein docking. *Nat. Protoc.* **2017**, *12*, 255. [CrossRef] [PubMed]
46. Kozakov, D.; Hall, D.R.; Beglov, D.; Brenke, R.; Comeau, S.R.; Shen, Y.; Li, K.; Zheng, J.; Vakili, P.; Paschalidis, I.C.; et al. Achieving reliability and high accuracy in automated protein docking: ClusPro, PIPER, SDU, and stability analysis in CAPRI rounds 13–19. *Proteins Struct. Funct. Bioinform.* **2010**, *78*, 3124–3130. [CrossRef] [PubMed]
47. Janeway, C.A., Jr.; Travers, P.; Walport, M.; Shlomchik, M.J. The interaction of the antibody molecule with specific antigen. In *Immunobiology: The Immune System in Health and Disease*, 5th ed.; Garland Science: Bethesda, MD, USA, 2001.
48. Buchan, D.W.; Jones, D.T. The PSIPRED protein analysis workbench: 20 years on. *Nucleic Acids Res.* **2019**, *47*, W402–W407. [CrossRef]
49. Bienert, S.; Waterhouse, A.; De Beer, T.A.; Tauriello, G.; Studer, G.; Bordoli, L.; Schwede, T. The SWISS-MODEL Repository—new features and functionality. *Nucleic Acids Res.* **2017**, *45*, D313–D319. [CrossRef]
50. Eisenberg, D.; Lüthy, R.; Bowie, J.U. [20] VERIFY3D: Assessment of protein models with three-dimensional profiles. *Methods Enzymol.* **1997**, *277*, 396–404.
51. Bui, H.H.; Sidney, J.; Dinh, K.; Southwood, S.; Newman, M.J.; Sette, A. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinform.* **2006**, *7*, 153. [CrossRef]
52. Santos, K.B.; Guedes, I.A.; Karl, A.L.; Dardenne, L.E. Highly Flexible Ligand Docking: Benchmarking of the DockThor Program on the LEADS-PEP Protein–Peptide Data Set. *J. Chem. Inf. Modeling* **2020**, *60*, 667–683. [CrossRef]
53. Sharma, A.; Hofmeyr, A.; Bansal, A.; Thakkar, D.; Lam, L.; Harrington, Z.; Bhonagiri, D. COVID-19 associated pulmonary aspergillosis (CAPA): An Australian case report. *Med. Mycol. Case Rep.* **2021**, *31*, 6–10. [CrossRef]

54. Salmanton-García, J.; Sprute, R.; Stemler, J.; Bartoletti, M.; Dupont, D.; Valerio, M.; Garcia-Vidal, C.; Falces-Romero, I.; Machado, M.; de la Villa, S.; et al. COVID-19–associated pulmonary aspergillosis, March–August 2020. *Emerg. Infect. Dis.* **2021**, *27*, 1077. [CrossRef] [PubMed]
55. Skok, K.; Vander, K.; Setaffy, L.; Kessler, H.H.; Aberle, S.; Bargfrieder, U.; Trauner, M.; Lax, S.F. COVID-19 autopsies: Procedure, technical aspects and cause of fatal course. Experiences from a single center. *Pathol.-Res. Pract.* **2021**, *217*, 153305. [CrossRef] [PubMed]
56. Verweij, P.E.; Brüggemann, R.J.; Azoulay, E.; Bassetti, M.; Blot, S.; Buil, J.B.; Calandra, T.; Chiller, T.; Clancy, C.J.; Cornely, O.A.; et al. Taskforce report on the diagnosis and clinical management of COVID-19-associated pulmonary aspergillosis. *Intensive Care Med.* **2021**, *47*, 819–834. [CrossRef] [PubMed]