current issues in molecular biology

MDPI

Article

From Dietary Proteins to Antiviral Peptides: Machine Learning-Guided Discovery and Structural Evaluation Against H5N1 Influenza Polymerase

Layla A. Alahmari [1], Pascal Emmanuel owona [2], Paul Désiré Djomeni Dzeufiet [2], Nawal H. Siddig [3], Nawal Al-Hoshani [4], Fahad Al-Asmari [5], Khaled Alzhrani [6] and Tariq Aziz [7,8,*]

1 Department of Community Health, College of Applied Medical Sciences, Northern Border University, Arar, Saudi Arabia
2 Laboratory of Animal Physiology, Department of Animal Biology and Physiology, Faculty of Science, University of Yaoundé 1, Yaoundé P.O. Box 812, Cameroon
3 Department of Mathematical Sciences, College of Science, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia
4 Department of Biology, College of Science, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia
5 Department of Food and Nutrition Sciences, College of Agricultural and Food Sciences, King Faisal University, Al-Ahsa, Saudi Arabia
6 Department of Clinical Laboratory Sciences, College of Applied Medical Sciences, Jouf University, Sakaka, Saudi Arabia
7 Department of Biology, College of Science, University of Tabuk, Tabuk 71491, Saudi Arabia
8 Molecular Microbiology and Infectious Diseases Research Unit, University of Tabuk, Tabuk 71491, Saudi Arabia
* Correspondence: tariqckd@ut.edu.sa

Abstract

Food proteins are well-known bioactive peptides, but their inhibitory activities against the influenza H5N1 polymerase have not been widely studied. In this paper, we applied a computational pipeline to predict the antiviral activities of food-derived peptides against the *PB1* subunit of H5N1 polymerase. Nineteen proteins from soybean, cow milk, egg, pea, maize, rice, and cowpea were analyzed in terms of their physiochemical properties, allergenicity, and antimicrobial and hemolytic activities, and then digested in silico with pepsin and chymotrypsin. The peptides generated from each digestion process were ranked according to their antiviral activities using four different machine learning algorithms (SVM, Random Forest, IBk, and K-Star) and prioritized on the basis of consensus across multiple models as well as gastrointestinal release and non-hemolytic properties. Peptides chosen were docked to *PB1* protein on three independent docking servers (HawkDock, PatchDock, and pyDockWEB) and further analyzed in terms of interaction analysis and MD simulations with 200.4 ns. Some proteins exhibited a high score for antimicrobial activity, such as legumin A (0.970), β-conglycinin α subunit (0.941), and glutelin type-A 1 (0.911). Most proteins were similar to known allergens. The peptides that showed consistent results in terms of antiviral properties and docking include RIQQAIAAGILPLSPL, SSAVSSVCENPILQPY, and QPYRIQQAI-AAGILPL, even though the ranking of their efficacy by both platforms was different. The peptide SSAVSSVCENPILQPY is the most consistent candidate for further research and has 29 close contacts and four hydrogen bonds. In addition, the complex formed between it and *PB1* is stable throughout the simulation period. These results allow identifying peptide candidates for further investigation, but all of them are based only on computational results. They need to be proven experimentally.

Keywords: H5N1 influenza virus; antiviral peptides; food-derived proteins; molecular docking; molecular dynamics; machine learning; *PB1* subunit

Check for updates

Academic Editor: Haijun Gong

Received: 9 September 2026
Revised: 11 October 2026
Accepted: 17 October 2026
Published: 26 October 2026

Copyright: © 2026 by the authors. 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.

1. Introduction

Influenza A viruses have been a constant threat to the health of both humans and animals owing to the genetic variability of these viruses, their ability to adapt, and their capacity to cross species barriers [1]. Among the different subtypes of influenza A viruses, the subtype H5N1 deserves special attention owing to its widespread presence among avian species and the ability of the virus to infect mammals, including humans [2]. The adaptation and emergence of highly pathogenic avian influenza viruses in new hosts raise questions about the consequences that can result from such events for public health [3]. Highly pathogenic avian influenza A(H5N1) has become a zoonotic and multispecies disease worldwide after its evolution from an avian disease [4]. After the worldwide circulation of H5N1 viruses, especially clade 2.3.4.4b viruses, which have been circulating since 2021, there were reports of infection in domestic birds, wild birds, and several mammals [5]. The virus has extended its range from Europe and Asia, where the virus was originally found, to North and South America as well as other areas that were not previously affected by the subtype. For instance, in June 2026, Australia detected its first case of highly pathogenic H5N1 in a wild bird [6,7].

Infections in humans are rare and mainly occur due to exposure to infected animals or their environment, especially birds like poultry. Based on the World Health Organization (WHO), there have been about 900 lab-confirmed infections of H5N1 in humans since 2003, with half of them being fatal, but the number is a conservative estimate since it includes only identified infections [8]. Recently, infections in humans have still been occurring, mostly at the human–animal interface; for example, the WHO has recorded three human A(H5N1) infections in April and May 2026 [9]. Although some infections have been severe, no human-to-human transmission has been seen, and the WHO currently classifies H5N1 as an animal virus. However, H5N1 continues to circulate in different animal hosts, offering chances for genetic changes [8]. This is why FAO, WHO, and OIE advocate that continuous monitoring and the application of a One Health approach should be utilized for viral evolution and reducing transmission at the animal–human–environment interface [10]. It becomes evident that the changing epidemiological scenario calls for the need to come up with complementary antiviral strategies that would target essential mechanisms of H5N1 replication [11].

Knowledge of molecular mechanisms of viral replication and adaptation to the host cell is very important for the creation of effective antiviral strategies [12]. Replication of influenza A virus mostly takes place due to the work of the viral RNA polymerase complex that consists of three subunits. *PB1* is the catalytic subunit of the polymerase complex, carrying the RNA-dependent RNA polymerase active site and forming the structural core that links *PB2* and PA, which makes it an attractive structural target for peptide binding [13,14]. In the complex, RNA recognition is done by the *PB2* subunit, and *PA* provides the structure of the complex and the endonuclease activity required for cap-snatching during mRNA formation [15,16]. However, replication of the virus requires not only viral proteins, but also host cell factors. One of these factors is acidic nuclear phosphoprotein 32 (*ANP32*), which helps to activate the polymerase complex of influenza, and differences in the proteins from birds and mammals create a host range barrier [5,17]. Amongst the various *ANP32* family members, *ANP32B* has been found to be a host factor affecting the functions of influenza virus polymerase [17]. There have been studies regarding structural interaction between mammalian *ANP32B* and the H5N1 polymerase complex, with particular attention to the *PA* subunit [5]. It has been found that *ANP32B* works as a cofactor that binds to the polymerase complex; interacts with the complex through *PB1* and *PA* subunits; and stabilizes, assembles, and helps attain a favorable conformation for the complex [18]. In this way, *ANP32B* may help H5N1 virus replicate itself through influence on the polymerase

functions, while the differing degrees of compatibility between the viral polymerase proteins and *ANPB32* family members are important parts of the host range adaptation [19]. Thus, the interaction between the *PB2–PB1–PA* complex and *ANP32B* may influence RNA synthesis efficiency and cause differences in replication especially where an avian H5N1 polymerase acts in mammals [19,20].

The structural details of the polymerase–*ANP32B* complex have provided valuable information on the molecular interface between the viral polymerase and the host cell factor, making the complex a potential drug target [21]. As conventional drugs normally target viral factors, interfering with the polymerase–*ANP32B* interaction can be regarded as an alternate drug strategy [12,22]. Indeed, in general terms, such interference would affect viral replication at an early stage of the viral replication cycle [22] and investigating a host–virus interface would allow us to discover new antiviral drugs that operate with a mechanism other than the known mechanism of action of anti-influenza agents [23]. While strategies targeting viral enzymes or very variable surface proteins are more complicated, a strategy targeting a functional protein–protein interface necessarily involves disrupting a critical host–virus interaction [24], which has prompted scientists to propose the interface as an appealing target for new antiviral drugs [25,26] and a possible biological target for peptides [24]. *ANP32B* is an endogenous protein with biological functions in the cell. Thus, any compound targeting the interface between these proteins should be tested for its specificity, and none of the food-derived peptides, to the best of our knowledge, have ever been investigated using computational analysis of their effects on this polymerase complex. In that regard, the current study sought to explore whether peptides derived from food proteins could be selected as potential candidates for interactions with the H5N1 polymerase using purely computational methods. The H5N1 polymerase–*ANP32B* complex served as the biological rationale behind this study [12]. PB1 of the H5N1 polymerase was chosen as the structural target for docking, among several other models of the *PA*, *PB1*, and *PB2* subunits, based on the structural criteria. Whether the peptides' binding sites overlapped with those of *ANP32B* is still unknown; thus, the influence of the peptides on the protein–protein interaction is a hypothesis.

Concurrently, bioactive peptides isolated from foods have been the subject of much research interest in recent years because of their potential as naturally occurring substances with multiple biological activities [27,28]. Food-derived peptides can be formed by the hydrolysis of food proteins using enzymes, fermentation, or some other food processing technique [27]. According to their amino acid sequence and structure, food peptides may show different activities, such as antimicrobial activity, antioxidant activity, and immunomodulation [29]. The natural availability, diversity in peptide structures, and biological compatibility of food peptides make them an interesting source of novel therapeutic substances. Experimentally screening all of the potential peptides that may be present in foods is a lengthy and resource-heavy process [30]. Advances in artificial intelligence, machine learning, structural bioinformatics, and molecular modeling offer promising approaches for the faster identification of bioactive peptides [31]. In addition, artificial intelligence-based methods may be employed to predict biological activities and identify interesting peptide sequences, whereas computational techniques like molecular docking and molecular dynamics may help reveal the interactions of peptides with proteins and their stability [30]. Thus, the combination of these techniques allows the screening of large peptide libraries and optimization of peptide molecules before further experimental studies. Hence, the general objective of the current project is to design an AI-driven computational pipeline for the discovery and optimization of food-derived peptides interacting with H5N1 influenza polymerase.

2. Materials and Methods

2.1. Study Design

The present study was developed as a multiple step in silico pipeline for identification of antiviral peptides from foods potentially active against the H5N1 influenza A polymerase. A set of nineteen proteins from soybean, cow milk, egg, pea, maize, rice and cowpea have been searched in the UniProtKB and further analyzed in terms of their physicochemical parameters (ProtParam), allergenicity (AllerCatPro 2.0), and their possible antimicrobial and hemolytic activities (Macrel v1.4.0). These proteins were exposed to the simulated digestion in the presence of pepsin and chymotrypsin (MS-Digest, maximum one missed cleavage) and subsequently evaluated in terms of antiviral activity using the four machine learning algorithms of AVP-IC50Pred (SVM, RF, IBk, and K-Star). Candidates were selected using a multi-criteria selection process based on their potential gastrointestinal release, agreement with IC50 predictions across multiple models, and lack of hemolytic activity. Meanwhile, the structure of the avian H5N1 polymerase in complex with human *ANP32B* (PDB ID: 8R1J) was considered as the target template for prediction of the *PA*, *PB1*, and *PB2* subunits with SWISS-MODEL. Models were tested using the ProSA-web server and SAVES v6.1, and PB1 was chosen as the docking receptor since it is the only subunit where more than 90% of the residues occupy the favorable zones of the Ramachandran plot. The chosen peptides were docked to PB1 using three different docking software (HawkDock, PatchDock, and pyDockWEB). These results were compared using a consensus approach. The structures of the peptide–*PB1* complexes thus obtained were analyzed at the atomic level and the two-dimensional level using COCOMAPS 2.0 and LIGPLOT, respectively. The selected complexes were further analyzed in terms of their stability using 200.4 ns molecular dynamics simulations in Desmond (Schrödinger Suite) at 310 K in NPT ensemble (Figure 1).

2.2. Food-Derived Proteins Retrieval

A systematic search for food-derived protein sequences was conducted using the UniProt database (https://www.uniprot.org/, accessed on 1 August 2026). Proteins that have been known to exist in common food sources were chosen. These included proteins from soybean (*Glycine max*), cow milk (*Bos taurus*), egg (*Gallus gallus*), pea (*Pisum sativum*), maize (*Zea mays*), rice (*Oryza sativa*), and cowpea (*Vigna unguiculata*). The corresponding protein entries in UniProtKB were determined based on the UniProt accession numbers. In total, there were 19 protein sequences including storage proteins of soybean (glycinin G3, glycinin G4, and β-conglycinin α-subunit), bovine milk proteins (β-lactoglobulin, αS1-casein, αS2-casein, β-casein, and κ-casein), chicken egg protein ovalbumin, storage proteins of pea (vicilin and legumins), maize zein and glutelins, rice glutelins, and cowpea vicilin. The corresponding amino acid sequences were retrieved in FASTA format as shown in Table 1.

Table 1. Selected food-derived proteins.

Protein_ID	Protein_Name	Food_Source	Species	UniProt_ID	Length (aa)
P1	Glycinin G3	Soybean	*Glycine max*	P11828	481
P2	Glycinin G4	Soybean	*Glycine max*	P02858	563
P3	β-Conglycinin α subunit	Soybean	*Glycine max*	P0DO15	605
P4	β-Lactoglobulin	Cow milk	*Bos taurus*	P02754	178

Table 1. *Cont.*

Protein_ID	Protein_Name	Food_Source	Species	UniProt_ID	Length (aa)
P5	α-S1-casein	Cow milk	*Bos taurus*	P02662	214
P6	α-S2-casein	Cow milk	*Bos taurus*	P02663	222
P7	β-Casein	Cow milk	*Bos taurus*	P02666	224
P8	κ-Casein	Cow milk	*Bos taurus*	P02668	190
P9	Ovalbumin	Egg	*Gallus gallus*	P01012	385
P10	Vicilin	Pea	*Pisum sativum*	P13918	459
P11	Vicilin	Pea	*Pisum sativum*	Q702P0	415
P12	Zein-alpha GZ19AB11	Maize	*Zea mays*	P08416	234
P13	Glutelin-2	Maize	*Zea mays*	P04706	223
P14	Glutelin type-A 1	Rice	*Oryza sativa*	P07728	499
P15	Glutelin type-A 2	Rice	*Oryza sativa*	P07730	499
P16	Legumin A	Pea	*Pisum sativum*	P02857	517
P17	Legumin A2	Pea	*Pisum sativum*	P15838	520
P18	Legumin B	Pea	*Pisum sativum*	P14594	338
P19	Vicilin	Cowpea	*Vigna unguiculata*	A0A2U9K6L4	432

2.3. Physicochemical Characterization of the Selected Food-Derived Proteins

The protein sequences of the selected food proteins were acquired from the ExPASy ProtParam online database (https://web.expasy.org/protparam/, accessed on 1 August 2026) [32]. The length of the sequence, molecular weight, theoretical pI, amino acid composition, and other physicochemical properties were computed for each protein. The data collected were analyzed to determine the physicochemical properties of the selected proteins.

2.4. Allergenicity Assessment and Selection of Food-Derived Protein Candidates

The allergenicity was determined using AllerCatPro 2.0 (https://allercatpro.bii.a-star.edu.sg/, accessed on 1 August 2026) with the default server parameter settings to determine any possible similarity with known allergens to classify the allergenicity of the target proteins [33].

2.5. Antimicrobial and Hemolytic Properties Assessment of Food-Derived Protein Candidates

The antimicrobial and hemolytic nature of the chosen proteins was predicted using Macrel v1.4.0 (https://www.big-data-biology.org/software/macrel/, accessed on 1 August 2026), which is an open-source machine learning-based pipeline that is used for the detection and characterization of antimicrobial peptides [34]. The protein sequences were fed to the software in FASTA format, and then analysis was conducted using the peptide prediction module. The software classifies the sequences based on their antimicrobial nature and categorizes them into different families of AMPs using physicochemical and structural features. Each sequence had its AMP classification, AMP family, and AMP probability score. The hemolytic nature of the proteins was also predicted using the hemolysis classifier in the software.

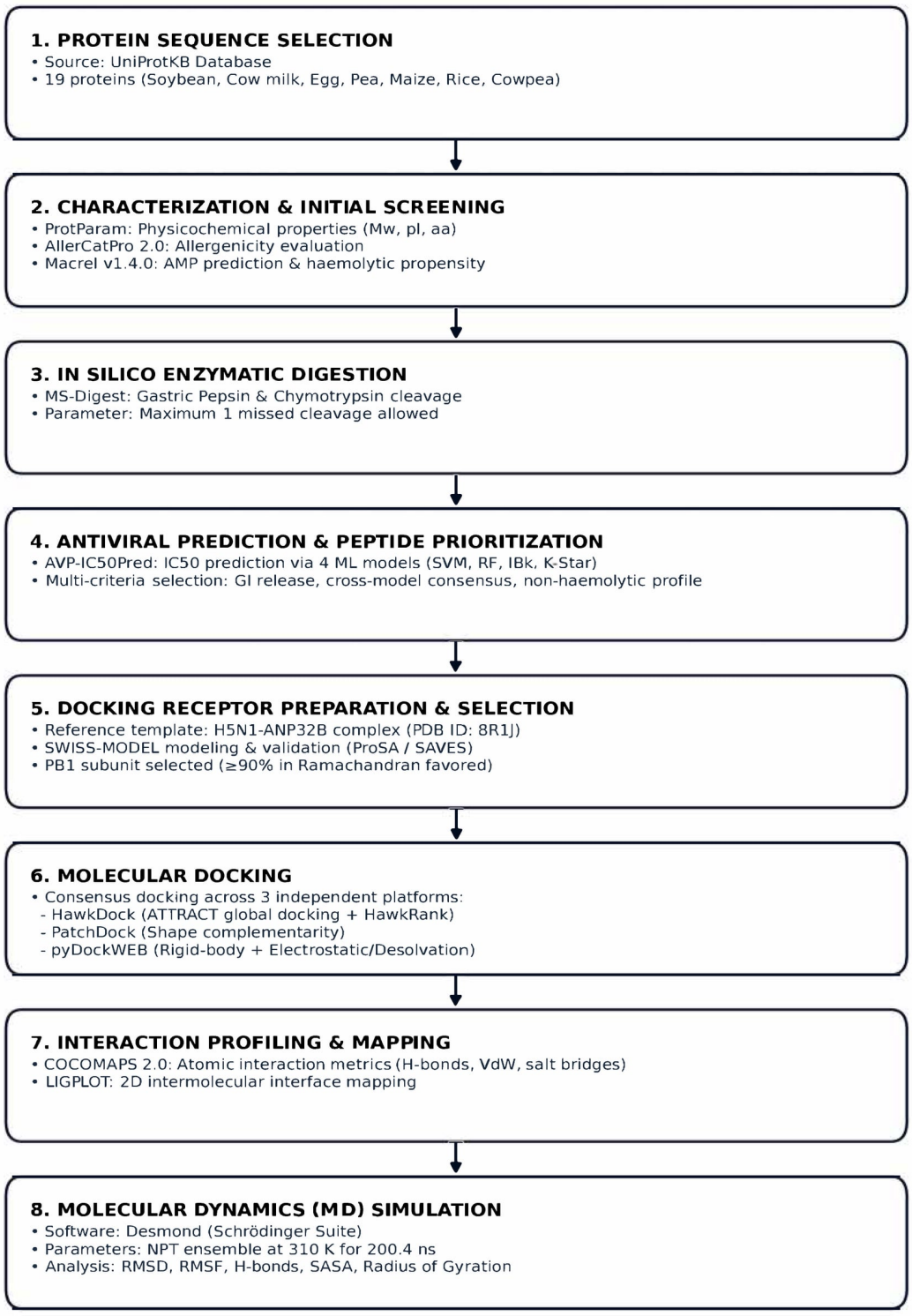


Figure 1. Description of the study protocol.

2.6. In Silico Digestion and Identification of Bioactive Peptide Candidates

The in silico enzymatic digestion of the selected protein sequences was performed using MS-Digest (https://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msdigest, accessed on 1 August 2026), available through the Protein Prospector platform of the UCSF Mass Spectrometry Facility [35]. The corresponding protein sequences were submitted in FASTA format, and pepsin gastric and chymotrypsin were selected as the proteolytic enzymes. The maximum number of missed cleavages was set to 1. The digestion parameters were applied according to the cleavage specificity implemented in MS-Digest.

For each resulting theoretical peptide, the software generated the peptide sequence, its position within the parent protein (start and end residues), the number of missed cleavages, and the calculated mass-to-charge ratio (m/z). The resulting peptide profiles were subsequently collected for further analysis of the peptides generated by the in silico chymotryptic and pepsin gastric digestions.

2.7. Prediction and Prioritization of Antiviral Peptide Candidates

The antiviral activity of peptides was predicted using the AVP-IC50Pred web server (http://crdd.osdd.net/servers/ic50avp, accessed on 2 August 2026) [36]. Peptide sequences were uploaded into the web-based server individually, and antiviral activity was predicted as the half-maximal inhibitory concentration (IC50) of the peptide in μM. The AVP-IC50Pred web-based server uses machine learning techniques with sequence information to predict IC50 values of antiviral peptides. Four machine learning algorithms were used for predicting the IC50 values of antiviral peptides in the server, and these were Support Vector Machine (SVM), Random Forest (RF), Instance-Based k-nearest neighbor (IBk), and K-Star. The predicted IC50 values of each of the peptides for all four classifiers were noted down for further analysis.

2.8. Selection and Prioritization of Peptide Candidates

After the digestion of the parent protein by pepsin and chymotrypsin in silico, the obtained peptides went through multi-criteria prioritization in order to select suitable candidates for analysis. First, the selected peptides were analyzed according to their likelihood of being generated by both pepsin and chymotrypsin, where both enzymes' proteolytic activity was considered [37]. The peptides were further analyzed regarding their predicted antiviral activity using four machine learning models, including Support Vector Machine (SVM), Random Forest (RF), IBk, and K-Star [38,39]. The peptides that exhibited positive prediction of antiviral activity by the mentioned models were selected, whereas contradictions between single-model predictions were considered. The predicted antimicrobial properties and hemolytic activity of each peptide were considered as an additional criterion, especially those peptides that were non-hemolytic [40,41]. Moreover, sequence correlations amongst peptides that were produced by the two proteases were taken into account, especially when a peptide found after pepsin digestion coincided with the sequence of a longer peptide found after chymotryptic digestion. From the combination of the above criteria, the final list of peptide candidates was selected. The selection process aimed at finding peptides with a combination of potential gastric secretion, antiviral activity, and favorable hemolytic activity prediction and was not only based on the lowest IC50 value prediction of a single computational model.

2.9. Structural Characterization of the H5N1 Polymerase Subunits

To predict the 2D and 3D structure of H5N1 influenza A polymerase dimer in complex with human ANP32B subunits, the corresponding PDB ID (8R1J) was entered into the database (https://www.rcsb.org/, accessed on 3 August 2026). The protein chains were previously identified in the macromolecules section, along with their characteristics, according to Table 2 below.

After this characterization, the secondary structure of each chain was assessed using PSIPRED (https://bioinf.cs.ucl.ac.uk/psipred/, accessed on 4 August 2026), following algorithms suggested by [42]. AlphaFold (https://alphafold.ebi.ac.uk/, accessed on 4 August 2026) was then used to predict the 3D structure. The prediction went through ProSA-web (https://prosa.services.came.sbg.ac.at/prosa.php) and SAVES v6.1 (https://saves.mbi.ucla.edu/, accessed on 4 August 2026) [43–46].

Table 2. Characteristics of the protein chains of H5N1 influenza A polymerase in complex with human ANP32B.

Entity ID	Molecule	Chains	Sequence Length	Organism	Gene Names
1	Polymerase acidic protein	A, D	716	Influenza A virus	*PA*
2	RNA-directed RNA polymerase catalytic subunit	B, E	757	Influenza A virus	*PB1*
3	Polymerase basic protein 2	C, F	898	Influenza A virus	*PB2*
4	Acidic leucine-rich nuclear phosphoprotein 32 family member B	G	251	Homo sapiens	*ANP32B*

2.10. Molecular Docking of Candidate Peptides to the PB1

2.10.1. Protein Structure Preparation for Molecular Docking

Structural alignment of the influenza A virus polymerase *PA–PB1* dimer in complex with human *ANP32B* protein was performed using crystal structure 8R1J (pdb_00008r1j). The protein sequences were determined from UniProt identifiers of the corresponding chains of the structure found in the PDB database, that is, A5Z236 (*PA*), A5Z231 (*PB1*), A5Z230 (*PB2*), and Q92688 (*ANP32B*). The respective FASTA sequences were downloaded and submitted to the SWISS-MODEL online server for generation of 3D models. For all the proteins, models showing 99–100% sequence identity to the corresponding reference sequence were chosen for further molecular docking.

2.10.2. Selection of the Docking Receptor

The three polymerase subunits *PA*, *PB1*, and *PB2* were analyzed using complementary structural quality parameters such as predicted aligned error (PAE), Z-score, knowledge based energy profile, and statistics derived from the Ramachandran plot [47,48]. The structure selected for molecular docking had to show mainly low PAE values, Z-scores, and knowledge-based energy profiles in addition to having a high number of residues in the favored regions of the Ramachandran plot [48]. A $\geq$90% minimum threshold of residues in the favored regions of the Ramachandran plot was another criteria used for the selection process. The protein structure with the best profile as per these criteria was considered to be the best receptor for docking. Using this procedure of selection, PB1 was selected as the docking receptor, as it was the only polymerase subunit with the required $\geq$90% in the favored regions of the Ramachandran plot and a good structural quality profile.

2.10.3. Docking Running

The molecular docking analyses were conducted in order to explore the putative interactions between the peptides and the *PB1* subunit. The peptide–*PB1* complexes were considered independently using three protein–peptide docking servers: HawkDock (https://cadd.zju.edu.cn/hawkdock/, accessed on 5 October 2026) [49], PatchDock (https://bio3d.cs.huji.ac.il/webserver/patchdock/about, accessed on 5 October 2026) [50], and pyDockWEB (https://life.bsc.es/pid/pydockweb, accessed on 5 October 2026) [51]. All the three docking servers were run using the default settings of the software. For HawkDock, the default protocol of HawkDock was applied using a global docking strategy with ATTRACT and then ranking with the HawkRank method without applying extra residue restraints by users. For PatchDock, the default docking parameters of protein–protein docking were utilized. For pyDockWEB, the default protocol for rigid body docking was performed using FTDock in order to calculate the docking orientation. Then, scoring was

made using the pyDock method taking into account electrostatic, desolvation, and to some extent van der Waals energy contributions.

2.11. Atomic and 2D Interaction Profiles of the Peptide–PB1

The docking interactions between the chosen peptides and *PB1* have been carried out using COCOMAPS 2.0 and LIGPLOT [52]. The peptide–protein docking complex was provided in PDB format in order to analyze and describe the interactions formed between the two molecules. COCOMAPS 2.0 (https://aocdweb.com/BioTools/cocomaps2/, accessed on 5 October 2026) was employed in order to quantify various kinds of interactions, such as proximal, van der Waals, hydrogen bonds, CH–O/N interactions, disulfide, salt bridges, and non-covalent interactions. In the second step, LIGPLOT was used for the generation of two-dimensional peptide–protein interfaces [46].

2.12. Molecular Dynamics Simulation and Structural Stability of the Peptide–PB1 Complexes

The MD simulation was conducted using the Desmond molecular dynamics simulation software suite (Schrödinger, LLC, New York, NY, USA). The protein–ligand complex was simulated for 200.4 ns in the NPT ensemble at 310 K temperature. Before conducting the MD simulation, the protein–ligand complex was solvated using explicit water molecules, while the charge was neutralized using counter ions. The final MD simulation system consisted of 12,076 protein atoms, 14,179 water molecules, corresponding ions, and ligand [53,54]. The analysis of the trajectories was done using the "Simulation Interaction Diagram" option in Schrödinger software. Stability of the protein–ligand complex was checked by monitoring the RMSD and RMSF values of protein and ligand structures along the simulation trajectory. The changes in the secondary structure of the protein were also monitored during the MD simulation. In addition, protein–ligand interactions such as hydrogen bonding, hydrophobic interaction, ionic interaction, and water bridge were analyzed during the MD simulation. Torsion profile studies and ligand properties like the radius of gyration, solvent accessible surface area, polar surface area, and intramolecular hydrogen bonds of the ligand were also carried out to assess the conformational dynamics of the ligand [53,54]. The generated trajectory was utilized to determine the conformational stability of the protein, interactions between ligand and protein, and the stability of protein–ligand complex throughout the simulation run up to 200 ns [53,54]. MM/PBSA studies were not carried out in this study owing to the large size of the protein–ligand complex. The receptor consists of 753 residues and roughly 12,076 protein atoms.

3. Results

3.1. Physicochemical Properties of Selected Food-Derived Proteins

The physicochemical parameters of the 19 proteins (Table 3) were estimated in order to characterize their size, charge, hydrophobicity, and stability since all these parameters are important for the solubility of the proteins, their behavior in gastrointestinal environment, and characteristics of the peptides produced as a result. The lengths of the proteins varied between 178 and 605 amino acids corresponding to the molecular weights of 19,883.25–70,306.21 Da. Such a variety included both small whey proteins like β-lactoglobulin and large multi-domain storage proteins like the β-conglycinin α subunit. That is why the set of proteins is characterized by various substrates that can be cleaved in the process of proteolysis. The theoretical pI varied from 4.93 (β-lactoglobulin) to 9.09 (glutelin type-A 1). A low pI corresponds to a negatively charged protein under physiological conditions, while a high pI corresponds to a positively charged protein. It is consistent with the numbers of charged residues: the TNNCR ranged from 1 (zein-alpha GZ19AB11) to 107 (β-conglycinin α subunit), and the TNPCR from 3 (zein-alpha

GZ19AB11) to 84 (β-conglycinin α subunit). Zein-alpha GZ19AB11 therefore has very few charged residues, while the β-conglycinin α subunit is highly charged, with an excess of negative charges.

Table 3. Physicochemical characteristics of the selected food-derived proteins.

Proteins	Number of AA	Theorical pI	MW	TNNCR	TNPCR	Ext. Coefficient	Formula	GRAVY	Stability Index
Glycinin G3	481	5.73	54,241.73	54	48	32,025	$C_{2377}H_{3703}N_{687}O_{737}S_{17}$	−0.617	64.84 (Unstable)
Glycinin G4	563	5.17	63,797.35	87	63	55,850	$C_{2765}H_{4325}N_{817}O_{902}S_{11}$	−0.952	59.32 (Unstable)
β-Conglycinin α subunit	605	5.12	70,306.21	107	84	33,600	$C_{3080}H_{4844}N_{892}O_{980}S_{8}$	−0.986	67.80 (Unstable)
β-Lactoglobulin	178	4.93	19,883.25	26	19	17,335	$C_{887}H_{1440}N_{224}O_{266}S_{12}$	−0.986	40.12 (Unstable)
α-S1-Casein	214	4.98	24,528.94	32	21	25,900	$C_{1108}H_{1719}N_{281}O_{333}S_{7}$	−0.481	56.03 (Unstable)
α-S2-Casein	222	8.55	26,018.69	28	31	28,880	$C_{1166}H_{1835}N_{303}O_{354}S_{8}$	−0.704	44.68 (Unstable)
β-Casein	224	5.26	25,107.33	23	16	11,460	$C_{1152}H_{1822}N_{284}O_{325}S_{8}$	−0.154	94.12 (Unstable)
κ-Casein	180	6.29	20,260.24	16	15	19,035	$C_{920}H_{1431}N_{233}O_{270}S_{6}$	−0.328	58.03 (Unstable)
Ovalbumin	386	5.19	42,881.24	47	35	31,400	$C_{1906}H_{3008}N_{500}O_{576}S_{23}$	−0.001	37.11 (Stable)
Vicilin	459	5.39	52,231.43	69	57	13,410	$C_{2294}H_{3647}N_{653}O_{731}S_{5}$	−0.693	49.92 (Unstable)
Vicilin	415	5.42	47,297.84	62	54	13,410	$C_{2082}H_{3315}N_{591}O_{665}S_{1}$	−0.735	51.73 (Unstable)
Zein-alpha GZ19AB11	234	8.52	25,439.65	1	3	11,920	$C_{1176}H_{1842}N_{296}O_{322}S_{5}$	0.382	75.02 (Unstable)
Glutelin-2	223	8.40	23,688.67	2	6	5960	$C_{1059}H_{1655}N_{303}O_{282}S_{17}$	−0.166	85.65 (Unstable)
Glutelin type-A 1	499	9.09	56,246.96	42	52	44,810	$C_{2460}H_{3856}N_{740}O_{752}S_{13}$	−0.539	51.36 (Unstable)
Glutelin type-A 2	499	8.93	56,306.15	42	50	50,310	$C_{2471}H_{3853}N_{735}O_{748}S_{15}$	−0.509	50.18 (Unstable)
Legumin A	517	6.16	58,805.35	76	71	35,870	$C_{2544}H_{4016}N_{782}O_{804}S_{12}$	−0.884	65.89 (Unstable)
Legumin A2	520	6.21	59,269.83	77	72	34,380	$C_{2562}H_{4044}N_{792}O_{809}S_{12}$	−0.894	67.35 (Unstable)
Legumin B	338	5.28	38,989.64	70	51	21,340	$C_{1669}H_{2657}N_{517}O_{562}S_{1}$	−1.199	68.01 (Unstable)
Vicilin	432	5.27	49,803	65	49	27,390	$C_{2200}H_{3428}N_{624}O_{691}S_{4}$	−0.747	56.54 (Unstable)

The extinction coefficient was from 5960 M^{-1} cm^{-1} (glutelin-2) to 55,850 M^{-1} cm^{-1} (glycinin G4). This is because it measures the quantity of tryptophan, tyrosine, and cysteine, meaning that it indicates the abundance of aromatic residues. This is important since chymotrypsin preferentially cuts after aromatic residues, and pepsin does the same, favoring hydrophobic and aromatic sites; hence, more peptides can be formed when there are more of these residues in the protein. The GRAVY value was from −1.199 (legumin B) to 0.382 (zein-alpha GZ19AB11). Hydrophilic proteins have negative values, and legumin B is the most hydrophilic among the listed proteins. The only hydrophobic protein was zein-alpha GZ19AB11, consistent with the fact that zein-alpha GZ19AB11 is an alcohol-soluble and hydrophobic maize prolamin. Most of the proteins were hydrophilic, with high chances of

enzyme accessibility. On the stability index, the range was from 37.11 (ovalbumin) to 94.12 (β-casein). The only protein whose stability was below the threshold of 40 was ovalbumin. Hence, it was expected to be stable in vitro, while all the others were unstable. In this case, having a higher instability index is an advantage since the objective of the study is to synthesize peptides using enzymatic hydrolysis.

Table number of negatively charged residues (Asp + Glu) (TNNCR), total number of positively charged residues (Arg + Lys) (TNPCR).

3.2. Allergenicity Profile of the Selected Food-Derived Proteins

Allergenicity of the 19 identified proteins was predicted using AllerCatPro 2.0 (Table 4). This program predicts the allergic potential of proteins on the basis of classification into categories of evidence for allergy based on their sequence similarity and 3D epitope similarity (3Depi) to known allergens. All proteins were classified as having "strong evidence" with exceptions of glycinin G1, β-conglycinin γ and δ-subunits, albumin, and serum albumin. In view of this, it is understandable that these proteins are considered major food allergens in soybean, milk, eggs, peas, and cereals. Evidence for allergenicity was provided by 3Depi >93% for several proteins; win80lin>30/35% for other proteins; and the presence of gluten-like Q-repeats with 3Depi 92–93% or greater for a few proteins. Glutelin-2 was assigned to this category on the basis of gluten-like Q-repeats together with win80lin above 35%. In contrast, legumin B was classified as "weak evidence" because its 3Depi value remained below 93%, and zein-alpha GZ19AB11 was classified as "no evidence", despite the presence of gluten-like Q-repeats in its sequence, which suggests that this motif alone was not sufficient to support an allergenic prediction.

Table 4. Predicted allergenicity of different food-derived proteins.

Proteins	Clinical Relevance of Most Similar Allergen		Similarity to Allergen and Resulting Predicted Evidence for Allergenicity	
	IgE Prevalence	Allergen Info	Result	Comment
Glycinin G3	1173	AG(3), food/insect allergen, anaphylaxis (high)	Strong evidence	Gluten-like Q-repeats and 3Depi > 93%
Glycinin G4	1173	AG(3), food/insect allergen, anaphylaxis (high)	Strong evidence	3Depi > 93%
β-Conglycinin α subunit	1229	AG(4), food/insect allergen, anaphylaxis (high)	Strong evidence	Gluten-like Q-repeats and 3Depi > 93%
β-Lactoglobulin	16,486	AG(4), food/insect allergen	Strong evidence	3Depi > 93%
α-S1-Casein	16,749	AG(6), food/insect allergen	Strong evidence	win80lin > 35%
α-S2-Casein	-	-	Strong evidence	win80lin > 35%
β-Casein	-	-	Strong evidence	win80lin > 35%

Table 4. *Cont.*

Proteins	Clinical Relevance of Most Similar Allergen		Similarity to Allergen and Resulting Predicted Evidence for Allergenicity	
	IgE Prevalence	Allergen Info	Result	Comment
κ-Casein	22,675	AG(6), food/insect allergen	Strong evidence	win80lin > 35%
Ovalbumin	18,144	AG(6), food/insect allergen	Strong evidence	3Depi > 93%
Vicilin	3	AG(2), food/insect allergen, anaphylaxis (high)	Strong evidence	3Depi > 93%
Vicilin	3	AG(2), food/insect allergen, anaphylaxis (high)	Strong evidence	Gluten-like Q-repeats and 3Depi > 93%
Zein-alpha GZ19AB11	-	-	No evidence	Gluten-like Q-repeats
Glutelin-2	-	AG(1)	Strong evidence	Gluten-like Q-repeats and win80lin > 35%
Glutelin type-A 1	-	AG(1)	Strong evidence	Gluten-like Q-repeats and 3Depi > 92%
Glutelin type-A 2	-	AG(1)	Strong evidence	Gluten-like Q-repeats and 3Depi > 92%
Legumin A	-	AG(1)	Strong evidence	3Depi > 93%
Legumin A2	-	AG(1)	Strong evidence	3Depi > 93%
Legumin B	-	-	Weak evidence	3Depi $\leq$ 93%
Vicilin	-	-	Strong evidence	Gluten-like Q-repeats and 3Depi > 93%

3.3. Antimicrobial and Antihemolytic Properties of the Investigated Proteins

The hemolytic and antimicrobial capabilities of 19 proteins were predicted with Macrel (see Table 5) to find the proteins capable of producing bioactive peptides with favorable safety profiles. Among the proteins, twelve were categorized as AMP positive and seven as non−AMP. The results imply that the antimicrobial capability is quite prevalent among the set of proteins but rather unevenly spread depending on their origin. Plant storage proteins proved to be the best candidates for such a capability, and the highest probabilities of AMP activity were registered for legumin A (0.970), β-conglycinin α subunit (0.941), glutelin type−A 1 (0.911), legumin A2 and glutelin type−A 2 (0.891 each), glycinin G3 (0.861), glycinin G4 (0.851), legumin B (0.842), and vicilin (0.822). A second entry of vicilin had probability of 0.772, while that of ovalbumin amounted to just 0.554–a borderline value considered positive. All the remaining proteins scored 0.000, including β-lactoglobulin, α-S1−casein, α-S2-casein, β−casein, κ−casein, and zein-alpha GZ19AB11, as well as 0.010 for glutelin−2, so none of these proteins could be considered AMP positive.

Table 5. Antimicrobial and antihemolytic properties.

Proteins	AMP_Family	Is_AMP	AMP_Probability	Hemolytic	Hemolytic_Probability
Glycinin G3	CDP	True	0.861	NonHemo	0.129
Glycinin G4	ADP	True	0.851	NonHemo	0.089
β-Conglycinin α subunit	ADP	True	0.941	NonHemo	0.089
β-Lactoglobulin	ADP	False	0.000	NonHemo	0.129
α-S1-Casein	ADP	False	0.000	NonHemo	0.069
α-S2-Casein	CDP	False	0.000	NonHemo	0.178
β-Casein	ADP	False	0.000	NonHemo	0.168
κ-Casein	CDP	False	0.000	NonHemo	0.089
Ovalbumin	ADP	True	0.554	NonHemo	0.139
Vicilin	ADP	True	0.822	NonHemo	0.059
Vicilin	ADP	False	0.386	NonHemo	0.069
Zein-alpha GZ19AB11	CDP	False	0.000	NonHemo	0.485
Glutelin-2	CDP	False	0.010	Hemo	0.545
Glutelin type-A 1	CDP	True	0.911	NonHemo	0.287
Glutelin type-A 2	CDP	True	0.891	NonHemo	0.307
Legumin A	CDP	True	0.970	NonHemo	0.069
Legumin A2	CDP	True	0.891	NonHemo	0.069
Legumin B	ADP	True	0.842	NonHemo	0.109
Vicilin	ALP	True	0.772	NonHemo	0.079

Regarding hemolytic activity, most of the proteins had been predicted to be non-hemolytic, having probabilities from 0.059 to 0.485. Vicilin had the lowest probability (0.059), followed by legumin A, legumin A2, vicilin, glycinin G4, the α subunit of the β-conglycinin, κ-casein, and legumin B. Probabilities from 0.129 to 0.168 were observed for glycinin G3, β-lactoglobulin, ovalbumin, β-casein, α-S2-casein (0.129 each), glutelin type-A 1 (0.139), glutelin type-A 2 (0.178), and zein-alpha GZ19AB11 (0.485), which were all below the classification threshold of 0.5. The only protein that had been classified as hemolytic was glutelin-2 (0.545). However, some of the highest ranking AMP candidates, such as legumin A, legumin A2, the α subunit of the β-conglycinin, glycinin G4, and legumin B, had low probabilities for hemolytic activity, which means that they not only possess good antimicrobial properties but also have low risks of adverse effects. In contrast, glutelin type-A 1 and 2 are both AMP positive, yet they have higher probabilities of being hemolytic than other candidates.

3.4. Results of Predicted In Silico Digestion and Identification of Bioactive Peptide Candidates

3.4.1. Chymotryptic Peptide Identification

The chymotryptic digestion gave rise to a number of peptide sequences that spanned through the entire protein sequence analyzed, having peptide sequence lengths between 8 and 21 amino acids and 0–1 missed cleavage sites (Table 6). The identified peptide sequences included the peptide fragments that span the residue numbers 1–39, 40–72, 82–119, 131–152, 156–184, and 193–234. Some of the identified peptide sequences include MAAK-IFCL, QQQQLPF, QQLPLVHL, TQPQLLPFY, SQLPAAYPQQFLPF, RIQQAIAAGILPLSPLF,

QPYRIQQAIAAGILPLS, and QQQQLPFSQLPAAYPQQF which have been identified as Rank 1 peptides. Multiple peptide sequences have been identified at different m/z values due to identification of the same peptide sequence in different measured conditions. The peptide sequences identified have m/z values ranging from 806.4 to 2120.3870.

Table 6. Chymotryptic digestion and peptide identification.

Rank	*m/z* (mi)	*m/z* (av)	Start	End	Missed Cleavages	Sequence
1	806.4043	806.8989	156	163	1	(L)AALNSPAY(L)
1	807.4433	808.0365	1	8	1	(-)MAAKIFCL(L)
1	811.3985	811.9182	218	224	1	(L)TNPAVFY(Q)
1	832.4563	832.9808	216	223	1	(L)ALTNPAVF(Y)
1	862.4669	863.0072	176	184	1	(L)AGVSPATFL(T)
1	871.4308	871.9740	132	138	0	(L)QQQQLPF(S)
1	879.4934	880.0378	225	233	0	(Y)QQPIIGGAL(F)
1	888.4574	889.0047	132	138	0	(L)QQQQLPF(S)
1	896.4732	897.1968	1	8	1	(-)MAAKIFCL(L)
1	896.5200	897.0685	225	233	0	(Y)QQPIIGGAL(F)
1	912.4682	913.1962	1	8	1	(-)MAAKIFCL(L)
1	930.5407	931.1295	82	89	1	(L)QQLPLVHL(L)
1	938.4838	939.2343	1	8	1	(-)MAAKIFCL(L)
1	939.4782	940.0471	15	24	0	(L)SASAATATIF(T)
1	943.5247	944.1253	185	192	1	(L)TQPQLLPF(Y)
1	947.5673	948.1602	82	89	1	(L)QQLPLVHL(L)
1	954.4787	955.2337	1	8	1	(-)MAAKIFCL(L)
1	984.5149	985.1345	165	172	1	(L)QQQQLLPF(S)
1	990.5003	991.0981	194	203	0	(Y)QHVAPNAGTL(L)
1	1001.5415	1002.1652	131	138	1	(Y)LQQQQLPF(S)
1	1001.5415	1002.1652	165	172	1	(L)QQQQLLPF(S)
1	1007.5269	1008.1287	194	203	0	(Y)QHVAPNAGTL(L)
1	1026.5619	1027.2157	225	234	1	(Y)QQPIIGGALF(-)
1	1041.5800	1042.1900	91	99	0	(L)AQNIRAQQL(Q)
1	1043.5884	1044.2463	225	234	1	(Y)QQPIIGGALF(-)
1	1059.5833	1060.2457	224	233	1	(F)YQQPIIGGAL(F)
1	1070.4901	1071.1412	108	116	1	(L)AAYSQQQQF(L)
1	1077.5575	1078.2174	173	183	1	(F)SQLAGVSPATF(L)
1	1118.5510	1119.2896	25	35	0	(F)TQCSQAPIASL(L)
1	1120.6109	1121.2892	194	204	1	(Y)QHVAPNAGTLL(Q)
1	1122.5578	1123.2612	111	119	1	(Y)SQQQQFLPF(N)
1	1154.6640	1155.3505	90	99	1	(L)LAQNIRAQQL(Q)
1	1170.5902	1171.3059	193	203	1	(F)YQHVAPNAGTL(L)
1	1249.6212	1250.4051	139	149	0	(F)SQLPAAYPQQF(L)

Table 6. *Cont.*

Rank	*m*/*z* (mi)	*m*/*z* (av)	Start	End	Missed Cleavages	Sequence
1	1305.6355	1306.4863	41	53	0	(L)SSAVSSVCENPIL(Q)
1	1363.8420	1364.6844	57	69	0	(Y)RIQQAIAAGILPL(S)
1	1410.7812	1411.6136	91	102	1	(L)AQNIRAQQLQQL(V)
1	1418.7196	1419.6468	40	53	1	(Y)LSSAVSSVCENPIL(Q)
1	1588.8040	1589.8622	25	39	1	(F)TQCSQAPIASLLPPY(L)
1	1606.8264	1607.8609	139	152	1	(F)SQLPAAYPQQFLPF(N)
1	1661.0109	1662.0410	57	72	1	(Y)RIQQAIAAGILPLSPL(F)
1	1693.8102	1694.9126	41	56	1	(L)SSAVSSVCENPILQPY(R)
1	1734.9901	1736.0800	54	69	1	(L)QPYRIQQAIAAGILPL(S)
1	1752.0167	1753.1106	54	69	1	(L)QPYRIQQAIAAGILPL(S)
1	2039.0114	2040.3139	15	35	1	(L)SASAATATIFTQCSQAPIASL(L)
1	2102.0342	2103.3564	132	149	1	(L)QQQQLPFSQLPAAYPQQF(L)
1	2119.0607	2120.3870	132	149	1	(L)QQQQLPFSQLPAAYPQQF(L)

3.4.2. Prediction of Pepsin Gastric Digestion and Peptide Identification

Pepsin gastric digestion and peptide identification is shown in Table 7. The in silico digestion of the protein with gastric pepsin resulted in eight peptide hits, where the lengths of the peptides are between 8 and 13 residues and their m/z values are between 847.3978 and 1192.5514. Peptide hits were located at residues 25–53 and 110–138 of the parent protein. The peptide hit AVSSVCENPI (residues 44–52) was identified by no missed cleavages, while others were characterized by one missed cleavage. Peptide sequences TQCSQAPI, QQQQLPF, and SQQQQFL had more than one m/z value; the last two were found at residues 132–138 and 111–117, respectively. The peptide SSAVSSVCENPI (residues 41–52) with m/z 1192.5514 is the longest identified peptide, while TQCSQAPI is the shortest.

Table 7. Pepsin gastric digestion and peptide identification.

Rank	*m*/*z* (mi)	*m*/*z* (av)	Start	End	Missed Cleavages	Sequence
1	847.3978	847.9711	25	32	1	(F)TQCSQAPI(A)
1	871.4308	871.9740	132	138	1	(L)QQQQLPF(S)
1	878.4367	878.9659	111	117	1	(Y)SQQQQFL(P)
1	888.4574	889.0047	132	138	1	(L)QQQQLPF(S)
1	928.4159	928.9826	110	116	1	(A)YSQQQQF(L)
1	947.4503	948.0892	44	52	0	(A)VSSVCENPI(L)
1	1060.5343	1061.2497	44	53	1	(A)VSSVCENPIL(Q)
1	1192.5514	1193.3259	41	52	1	(L)SSAVSSVCENPI(L)

3.5. Predicted and Prioritized Antiviral, Antimicrobial, and Hemolytic Profiles of Peptide Candidates

3.5.1. Predicted Antiviral Activity of Chymotrypsin-Generated Peptides

IC50 values of the studied peptides were predicted in Table 8 in μM. One of the most promising peptides was QPYRIQQAIAAGILPL which showed favorable IC50 value predictions of 33.04, 30.69, 57.53, and 0.89 μM with SVM, RF, Ibk, and K-Star, respectively.

Another peptide showing favorable predictions was TQCSQAPIASLLPPY which exhibited relatively low IC50 values for SVM (33.26 μM), RF (41.10 μM), and Ibk (38.23 μM), while its K-Star prediction value was exceptionally low at 5.01 μM. The peptide SSAVSSVCENPILQPY showed also favorable profiles across four models, namely IC50 values of 36.36, 23.89, 26.71, and 8.51 μM, respectively. RIQQAIAAGILPLSPL was another peptide showing favorable IC50 value predictions, namely 32.08 μM (SVM), 23.01 μM (RF), 60.52 μM (Ibk), and 8.91 μM (K-Star). However, some peptides were characterized by noticeable inconsistencies between IC50 predictions in the models considered. For instance, peptide ALTNPAVF exhibited a very low IC50 prediction of 0.05 μM (K-Star) compared with other models. A similar pattern was observed for QQPIIGGAL, QQPIIGGALF, and RIQQAIAAGILPL, for which K-Star consistently predicted an IC50 of 0.89 μM, whereas the other algorithms produced considerably higher values.

Table 8. Prediction of the antiviral activity of peptides based on IC50 values.

Sequence	Length	SVM (μM)	RF (μM)	Ibk (μM)	K Star (μM)
AALNSPAY	8	70.08	85.23	88.87	100.01
MAAKIFCL	8	88.19	42.62	63.15	10.01
TNPAVFY	7	64.29	68.26	48.71	5.94
ALTNPAVF	8	60.64	62.08	67.91	0.05
AGVSPATFL	9	60.99	82.48	133.86	186.01
QQQQLPF	7	49.85	51.14	106.95	52.01
QQPIIGGAL	9	90.96	82.6	124.3	0.89
QQQQLPF	7	49.85	51.14	106.95	52.01
MAAKIFCL	8	88.19	42.62	63.15	10.01
QQPIIGGAL	9	90.96	82.6	124.3	0.89
MAAKIFCL	8	88.19	42.62	63.15	10.01
QQLPLVHL	8	51.33	43.22	93.03	200.01
MAAKIFCL	8	88.19	42.62	63.15	10.01
SASAATATIF	10	84.47	83.5	103.09	10.01
TQPQLLPF	8	41.38	63.19	62.99	52.01
QQLPLVHL	8	51.33	43.22	93.03	200.01
MAAKIFCL	8	88.19	42.62	63.15	10.01
QQQQLLPF	8	50.68	32.65	50.29	52.01
QHVAPNAGTL	10	64.14	108.87	176.22	61.41
LQQQQLPF	8	50.84	35.38	109.14	52.01
QQQQLLPF	8	50.68	32.65	50.29	52.01
QHVAPNAGTL	10	64.14	108.87	176.22	61.41
QQPIIGGALF	10	71.45	77.51	121.17	0.89
AQNIRAQQL	9	75.33	41.09	90.04	40.01
QQPIIGGALF	10	71.45	77.51	121.17	0.89
YQQPIIGGAL	10	97.62	88.93	121.17	200.01
AAYSQQQQF	9	51.54	65.22	65.74	13.86
SQLAGVSPATF	11	71.25	119.36	89.08	186.01

Table 8. *Cont.*

Sequence	Length	SVM (μM)	RF (μM)	Ibk (μM)	K Star (μM)
TQCSQAPIASL	11	74.04	85.51	78.62	13.86
QHVAPNAGTLL	11	56.6	99.07	170.71	193.01
SQQQQFLPF	9	45.42	66.26	30.48	52.01
LAQNIRAQQL	10	79.4	44.8	91.09	67.61
YQHVAPNAGTL	11	63.34	92.5	170.71	3.51
SQLPAAYPQQF	11	38.35	73.07	98.41	3.51
SSAVSSVCENPIL	13	66.36	31.67	25.74	8.51
RIQQAIAAGILPL	13	88.87	34.5	71.06	0.89
AQNIRAQQLQQL	12	57.49	31.14	108.03	0.35
LSSAVSSVCENPIL	14	60.85	18.68	30.44	8.51
TQCSQAPIASLLPPY	15	33.26	41.1	38.23	5.01
SQLPAAYPQQFLPF	14	38.71	52.57	73.98	52.01
RIQQAIAAGILPLSPL	16	32.08	23.01	60.52	8.91
SSAVSSVCENPILQPY	16	36.36	23.89	26.71	8.51
QPYRIQQAIAAGILPL	16	33.04	30.69	57.53	0.89
QPYRIQQAIAAGILPL	16	33.04	30.69	57.53	0.89
SASAATATIFTQCSQAPIASL	21	54.47	46.85	69.69	142.01
QQQQLPFSQLPAAYPQQF	18	39.91	55.74	96.63	52.01
QQQQLPFSQLPAAYPQQF	17	39.92	60.51	100.43	52.01

3.5.2. Predicted Results of Antiviral Activity of Gastric Pepsin-Generated Peptides

The predicted antiviral activity for the peptides resulting from the pepsin digestion of gastric samples is shown in Table 9. Eight entries of peptides were identified, which corresponded to seven distinct sequences, with peptide lengths varying between 7 to 11 amino acids. The predicted antiviral activity was assessed using four machine learning classifiers, including SVM, Random Forest (RF), IBk, and K-Star, where predicted values are given in μM. The values predicted using the SVM classification varied between 43.57 and 86.84 μM, while the predictions of RF varied between 51.14 and 95.34 μM. The predictions made using the IBk classifier varied between 7.27 and 106.95 μM, while the K-Star predictions varied between 0.22 and 285.01 μM. Of all the tested sequences, the YSQQQQF had the lowest SVM value (43.57 μM), while SQQQQFL had the lowest value predicted with the K-Star classifier (0.22 μM). On the other hand, the peptide VSSVCENPI had the following predicted values of 86.84, 73.96, 7.27, and 285.01 μM with the SVM, RF, IBk, and K-Star models, respectively. VSSVCENPIL, on the other hand, was found to have values of 85.15, 66.91, 17.34, and 285.01 μM. The longest peptide in the group, which was SSAVSSVCENPI, had predicted values of 63.65, 64.02, 18.88, and 42.01 μM with SVM, RF, IBk, and K-Star respectively.

Table 9. Predicted antiviral activity of gastric pepsin-generated peptides.

Sequence	Length	SVM (μM)	RF (μM)	Ibk (μM)	K Star (μM)
TQCSQAPI	8	80.61	95.34	76.9	263.01
QQQQLPF	7	49.85	51.14	106.95	52.01
SQQQQFL	7	44.49	58.53	26.78	0.22
QQQQLPF	7	49.85	51.14	106.95	52.01
YSQQQQF	7	43.57	57.32	47.61	28.01
VSSVCENPI	9	86.84	73.96	7.27	285.01
VSSVCENPIL	10	85.15	66.91	17.34	285.01
SSAVSSVCENPI	11	63.65	64.02	18.88	42.01

3.5.3. Predicted Antimicrobial and Hemolytic Properties of Candidate Peptides

Antimicrobial and hemolytic properties of the peptides were investigated to determine their predicted antimicrobial efficacy and safety (Table 10). It is worth mentioning that the majority of the peptides were predicted as having low antimicrobial activity since the majority of them were defined as non-antimicrobial (is_AMP = False) and belonging to the CLP family. In particular, AAKIFCL was the only peptide found with a high predicted AMP probability of 0.515, which made it antimicrobial and a member of the CDP family. Even though this peptide had the highest hemolytic probability among other candidates (0.436), it was still determined as being non-hemolytic, thus implying the possibility of its positive antimicrobial/hemolytic profile, which should be further experimentally validated. The second highest AMP probability was found for RIQQAIAAGILPL (0.406), even though it was also determined as non-antimicrobial and non-hemolytic. Other peptides were characterized by AMP probability between 0.030 and 0.337, demonstrating little antimicrobial activity. However, importantly, all peptides were found as having low hemolytic activity since the probability of hemolysis varied between 0.030 and 0.436.

Table 10. Antimicrobial and hemolytic profile of chymotrypsin-generated peptides.

Sequence	AMP_Family	Is_AMP	AMP_Probability	Hemolytic	Hemolytic Probability
AALNSPAY	CLP	False	0.050	NonHemo	0.178
AAKIFCL	CDP	True	0.515	NonHemo	0.436
TNPAVFY	CLP	False	0.129	NonHemo	0.168
ALTNPAVF	CLP	False	0.099	NonHemo	0.129
AGVSPATFL	CLP	False	0.129	NonHemo	0.158
QQQQLPF	CLP	False	0.248	NonHemo	0.089
QQPIIGGAL	CLP	False	0.119	NonHemo	0.168
QQQQLPF	CLP	False	0.248	NonHemo	0.089
AAKIFCL	CDP	True	0.515	NonHemo	0.436
QQPIIGGAL	CLP	False	0.119	NonHemo	0.168
AAKIFCL	CDP	True	0.515	NonHemo	0.436
QQLPLVHL	CLP	False	0.337	NonHemo	0.218
AAKIFCL	CDP	True	0.515	NonHemo	0.436

Table 10. *Cont.*

Sequence	AMP_Family	Is_AMP	AMP_Probability	Hemolytic	Hemolytic Probability
SASAATATIF	CLP	False	0.317	NonHemo	0.089
TQPQLLPF	CLP	False	0.099	NonHemo	0.139
QQLPLVHL	CLP	False	0.337	NonHemo	0.218
AAKIFCL	CDP	True	0.515	NonHemo	0.436
QQQQLLPF	CLP	False	0.198	NonHemo	0.089
QHVAPNAGTL	CLP	False	0.109	NonHemo	0.208
LQQQQLPF	CLP	False	0.218	NonHemo	0.109
QQQQLLPF	CLP	False	0.198	NonHemo	0.089
QHVAPNAGTL	CLP	False	0.109	NonHemo	0.208
AQNIRAQQL	CLP	False	0.257	NonHemo	0.129
QQPIIGGALF	CLP	False	0.139	NonHemo	0.149
YQQPIIGGAL	CLP	False	0.139	NonHemo	0.129
AAYSQQQQF	CLP	False	0.164	NonHemo	0.089
SQLAGVSPATF	CLP	False	0.208	NonHemo	0.129
TQCSQAPIASL	CDP	False	0.307	NonHemo	0.069
QHVAPNAGTLL	CLP	False	0.119	NonHemo	0.158
SQQQQFLPF	CLP	False	0.277	NonHemo	0.119
LAQNIRAQQL	CLP	False	0.297	NonHemo	0.198
YQHVAPNAGTL	CLP	False	0.257	NonHemo	0.168
SQLPAAYPQQF	CLP	False	0.129	NonHemo	0.218
SSAVSSVCENPIL	ADP	False	0.089	NonHemo	0.030
RIQQAIAAGILPL	CLP	False	0.406	NonHemo	0.416
AQNIRAQQLQQL	CLP	False	0.248	NonHemo	0.248
LSSAVSSVCENPIL	ADP	False	0.030	NonHemo	0.099
TQCSQAPIASLLPPY	CDP	False	0.158	NonHemo	0.188
SQLPAAYPQQFLPF	CLP	False	0.178	NonHemo	0.228
RIQQAIAAGILPLSPL	CLP	False	0.297	NonHemo	0.406
SSAVSSVCENPILQPY	ADP	False	0.139	NonHemo	0.069
QPYRIQQAIAAGILPL	CLP	False	0.099	NonHemo	0.228
QPYRIQQAIAAGILPL	CLP	False	0.099	NonHemo	0.228
SASAATATIFTQCSQAPIASL	CDP	False	0.277	NonHemo	0.198
QQQQLPFSQLPAAYPQQF	CLP	False	0.129	NonHemo	0.178
QQQQLPFSQLPAAYPQQF	CLP	False	0.129	NonHemo	0.178

3.5.4. Predicted Antimicrobial and Hemolytic Properties of Gastric Pepsin-Generated Peptides

The predicted antimicrobial and hemolytic profiles of the peptides generated following simulated gastric pepsin digestion are presented in Table 11. None of the identified peptides was predicted to exhibit antimicrobial activity, as all sequences were classified as non-antimicrobial (is_AMP = False), despite belonging to different predicted AMP families, including CLP, CDP, and ADP. The AMP probabilities remained relatively low, ranging from

0.069 to 0.248, with QQQQLPF displaying the highest predicted antimicrobial probability (0.248). However, this value was insufficient for classification as an AMP. Importantly, all peptides were predicted to be non-hemolytic, with hemolytic probabilities ranging from 0.030 to 0.158, indicating a low predicted propensity for erythrocyte membrane disruption. Among the peptides evaluated, VSSVCENPI exhibited the lowest hemolytic probability (0.030), whereas YSQQQQF showed the highest value (0.158).

Table 11. Antimicrobial and hemolytic profile of gastric pepsin-generated peptides.

Sequence	AMP_Family	Is_AMP	AMP_Probability	Hemolytic	Hemolytic_Probability
TQCSQAPI	CDP	False	0.188	NonHemo	0.059
QQQQLPF	CLP	False	0.248	NonHemo	0.089
SQQQQFL	CLP	False	0.198	NonHemo	0.109
QQQQLPF	CLP	False	0.248	NonHemo	0.089
YSQQQQF	CLP	False	0.198	NonHemo	0.158
VSSVCENPI	ADP	False	0.139	NonHemo	0.030
VSSVCENPIL	ADP	False	0.069	NonHemo	0.040
SSAVSSVCENPI	ADP	False	0.109	NonHemo	0.109

3.6. Predicted 3D Structure of the H5N1 Polymerase–ANP32B Target Complex

3.6.1. Results of Predicted Alignment Error Profiles of the H5N1 Polymerase Subunits

Figure 2 presents a comparative structure of the three main subunits of the H5N1 influenza A virus polymerase complex, i.e., PA (polymerase acidic protein), PB1 (RNA-directed RNA polymerase catalytic subunit), and PB2 (polymerase basic protein 2), together with their three-dimensional structures and predicted aligned error (PAE) maps. The subunit PA has an overall compact shape with α-helices and β-sheets, with its PAE map being mostly green colored (low predicted positional errors) (A). PB1 has a compact three-dimensional structure with a large number of secondary structures and mostly low predicted PAE (especially for the central regions) (B). PB2, which is the largest among the three subunits, has an extended modularity with some compact structures mixed with non-compact ones. Its PAE map has different levels of predicted errors (C).

3.6.2. Structural Quality and Knowledge-Based Energy Profiles of the H5N1 Polymerase Subunits

The quality of protein models utilized for docking analysis was estimated using ProSA-web (Figure 3). The total Z-scores were −8.29 for *PA*, −9.51 for *PB1*, −9.64 for *PB2*, and −7.84 for *ANP32B*. These scores lie within the usual range for structures solved experimentally with similar number of residues (500–750 residues for *PA*, *PB1* and *ANP32B*, shorter chain for *PB2*) because the black dots lie within the cluster of X-ray structures, which means that models have high quality and no serious problems. *PB2* (−9.64) and *PB1* (−9.51) obtained the best scores, followed by *PA* (−8.29) and *ANP32B* with the lowest but still acceptable score (−7.84). Local energy profiles confirm these results. For PB1 (residues 1–756) and *PB2* (residues 46–143 in the figure), knowledge-based energy was negative, around −1 to −3, over the almost entire length of the chains. PA (residues 1–716) displayed the same general trend, with a few short positive peaks (up to about +2 to +3) near the N-terminal region and at some isolated positions in the central part of the sequence, corresponding to local regions of lower reliability, probably loops, or flexible segments. *ANP32B* (residues 202–759 in the plotted window) was also mostly negative, with one main positive peak (about +3) in the first third of the plotted region and a few smaller ones.

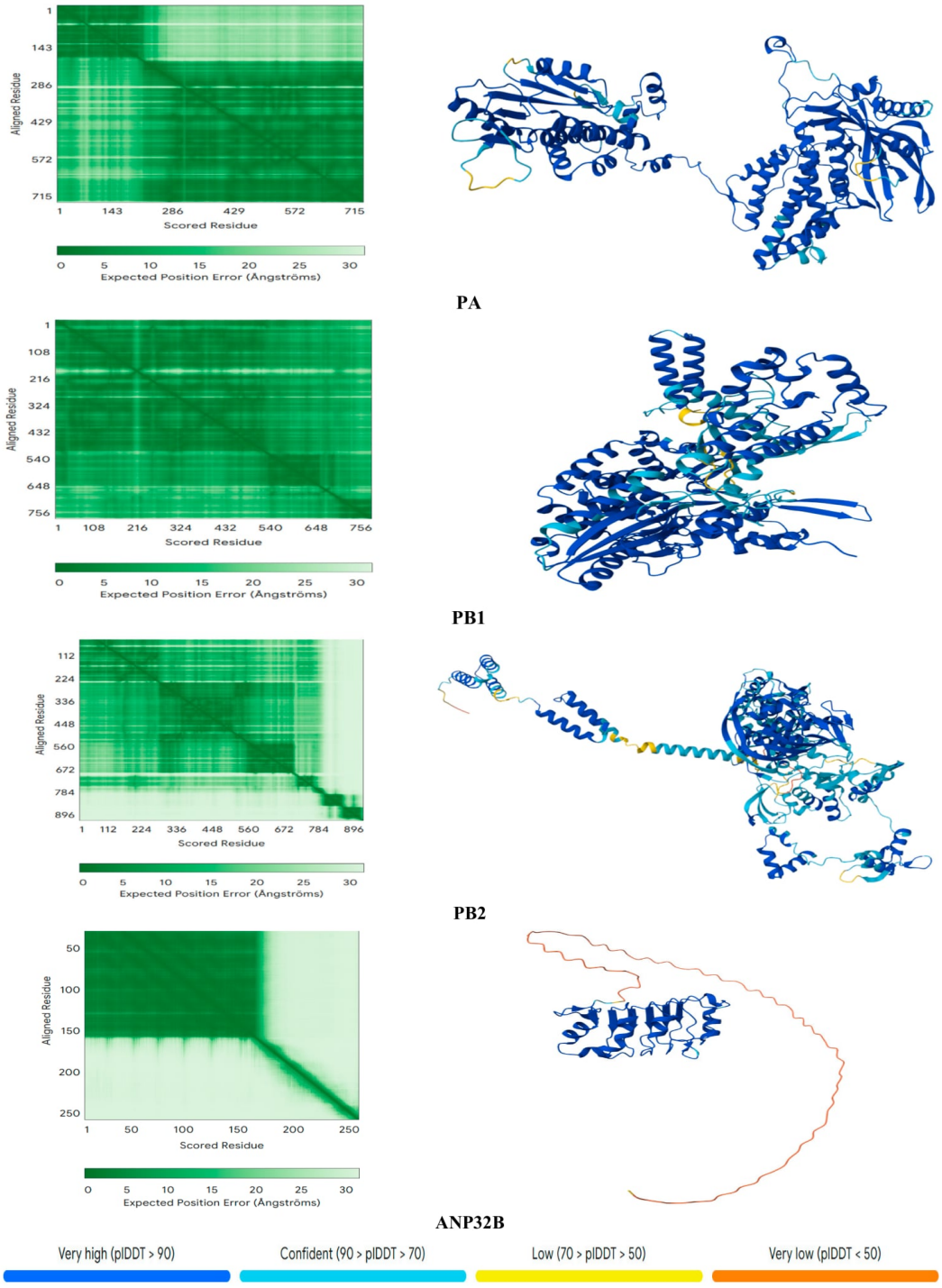


Figure 2. Predicted alignment error profiles of the H5N1 polymerase subunits.

3.6.3. Ramachandran Plots and of H5N1 Influenza Polymerase Subunits and *ANP32B*

The stereochemical quality of the *PA*, *PB1*, *PB2*, and *ANP32B* models was assessed using Ramachandran plot analysis (Figure 4). The percentage of residues in the favored regions was 90.8% for *PB1*, 89.1% for *PB2*, 87.8% for *ANP32B*, and 84.6% for *PA*, which corresponds to 9.2%, 10.9%, 12.2%, and 15.4% of residues in non−favored regions (additionally allowed, generously allowed, and disallowed), respectively. *PB1* was the only model above the 90% threshold set as a selection criterion for the docking receptor, whereas *PB2* was just below it (89.1%), so the difference between the two subunits is small. PA had the highest proportion of residues outside the favored regions (15.4%), and *ANP32B* was an intermediate one (12.2%).

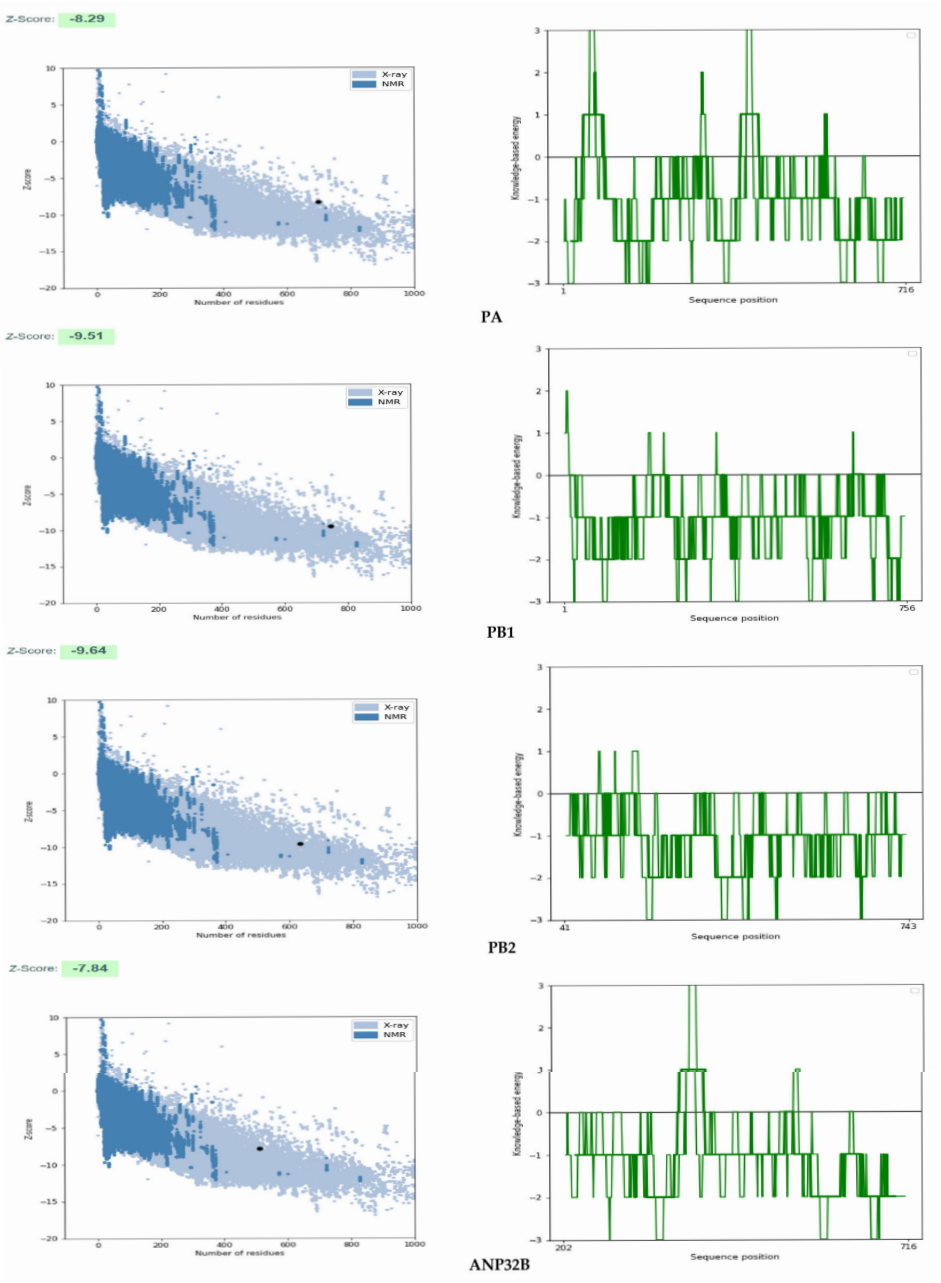


Figure 3. Structure quality assessment of the *PA*, *PB1*, *PB2*, and *ANP32B* protein models using ProSA-web. For each protein, the left panel shows the Z−score (black dot) in relation to the Z−scores of experimentally determined X−ray (light blue) and NMR (dark blue) structures of similar size, and the right panel shows the local knowledge−based energy profile along the sequence.

3.7. Predicted 2D Results of the Structure of the H5N1 Polymerase Subunits and ANP32B

Figure 5 The PSIPRED-predicted secondary structure of the H5N1 polymerase subunits *PA*, *PB1*, and *PB2*, along with *ANP32B*. Based on the predictions, all four proteins were seen to be mostly comprised of alpha helix and coils, with beta strands making up only a small part of the total protein structures. For the case of *PA* (757 residues), the structure consisted of 70% alpha helices, 20% coils, and 10% beta strands. The same case was observed in *PB1* (757 residues), where there were 70% alpha helices, 20% coils, and 10% beta strands. In the case of *PB2* (859 residues), 65% alpha helices, 22% coils, and 13% beta strands were predicted. With regards to *ANP32B* (251 residues), 55% alpha helices, 40% coils, and 5% beta strands were present in the predicted structure.

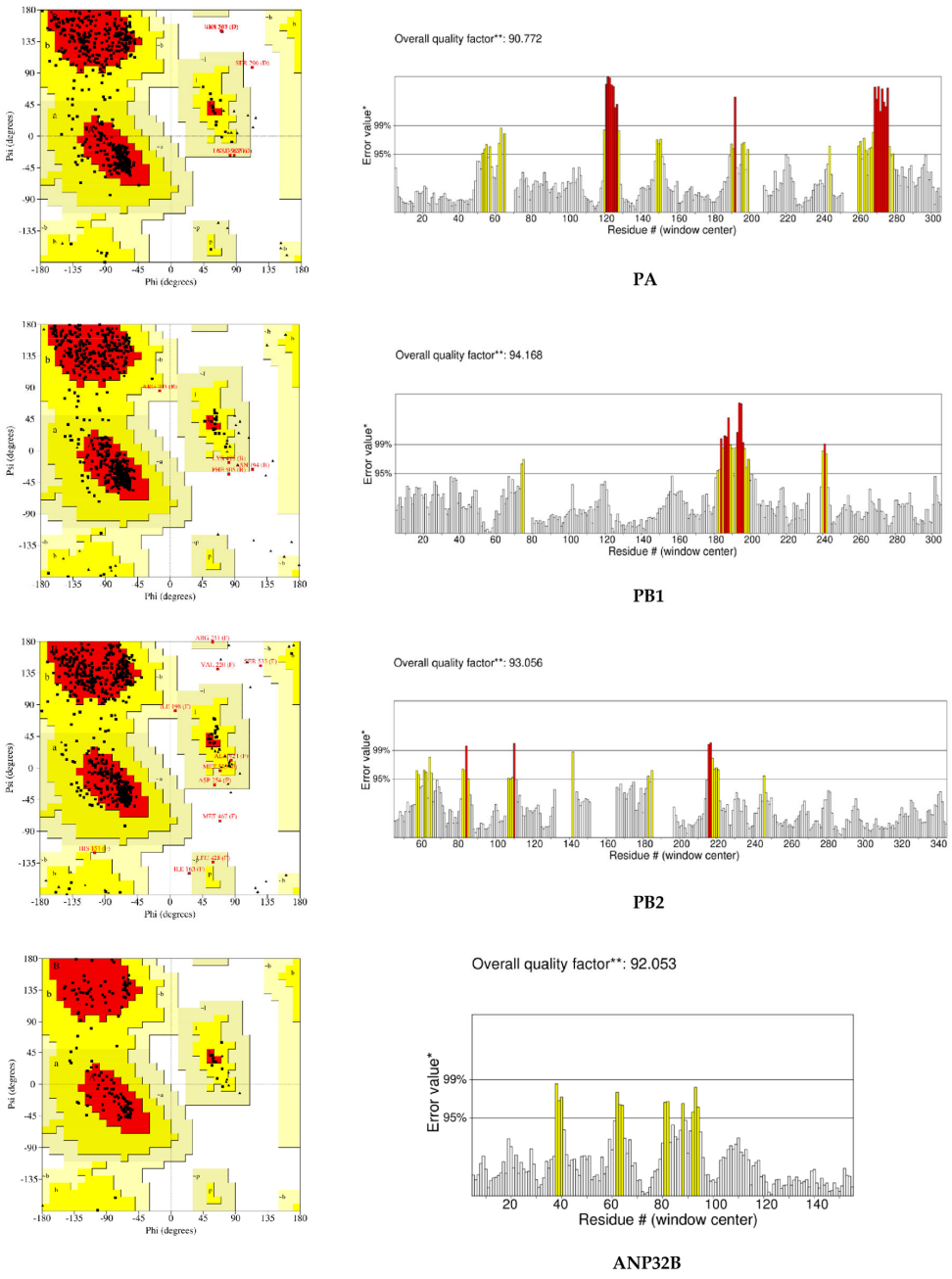


Figure 4. Structural and functional organization of the H5N1 polymerase subunits and ANP32B. * $p < 0.05$, ** $p < 0.01$.

3.8. Molecular Docking Data

3.8.1. 3D Interaction and Binding Score of the *PB1* Subunit and Selected Peptides Using HawkDock

The findings in Table 12 and Figure 6 show the binding energy profiles as well as 3D models of the interaction between the six peptides and *PB1* of the H5N1 influenza polymerase using the HawkDock server. The docking outcomes are described for the top ten ranked models for each peptide. The peptide that had the lowest binding energy score in rank one was RIQAAIAAGILPLSPL, with the score being −3663.38, then SSAVSSV-CENPILQPY with −3424.60, QPYRIQQAIAAGILPL with −3265.93, SSAVSSVCENPI with −3061.72, TQCSQPIASLLPPY with −2957.09, and TQCSQAPI with −2314.84. The ten ranked models for each peptide had varied binding energy scores. The three depict the docked peptide–PB1 complexes of the six peptides. In these structures, PB1 subunit is the

principal protein structure, while the docked peptide models are indicated in different colors depending on the number of the model.

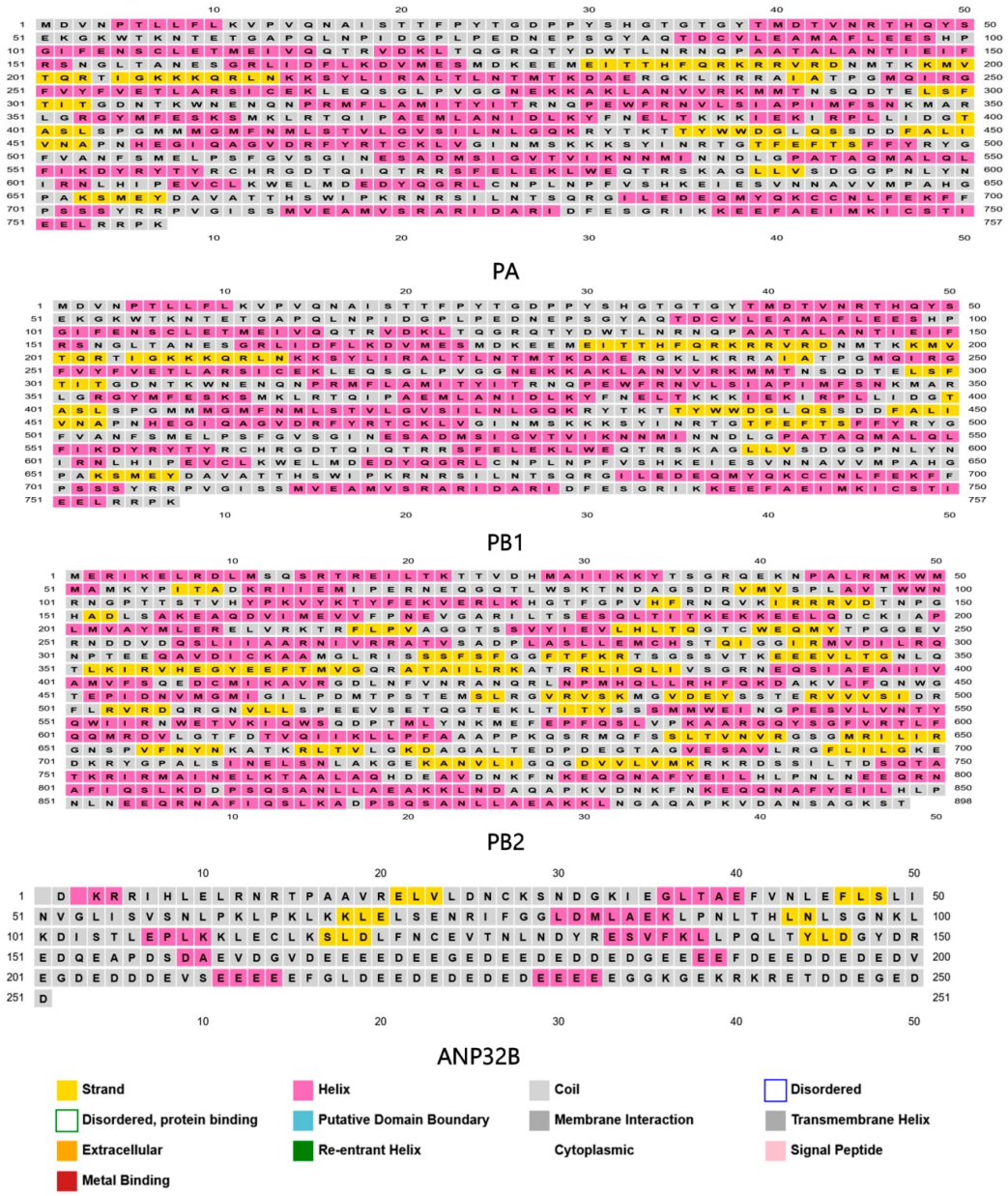


Figure 5. Predicted 2D structure of the H5N1 polymerase subunits and *ANP32B*.

Table 12. Binding energies between peptides and PB1 subunit of H5N1 influenza polymerase using the HawkDock server.

	QPYRIQQAIAAGILPL	TQCSQAPIASLLPPY	SSAVSSVCENPILQPY	RIQQAIAAGILPLSPL	SSAVSSVCENPI	TQCSQAPI
Rank			**Score**			
1	−3265.93	−2957.09	−3424.60	−3663.38	−3061.72	−2314.84
2	−3229.24	−2783.24	−3417.46	−3358.60	−3018.08	−2276.20
3	−3070.76	−2775.18	−3402.84	−3278.53	−2819.00	−2241.39
4	−2956.85	−2716.88	−3222.96	−3237.46	−2743.13	−2230.86
5	−2956.41	−2702.67	−3086.07	−3175.93	−2701.91	−2145.56
6	−2942.51	−2698.51	−2992.87	−3130.39	−2692.01	−2066.24
7	−2908.86	−2680.86	−2992.21	−3119.73	−2633.26	−2015.30

Table 12. *Cont.*

	QPYRIQQAIAAGILPL	TQCSQAPIASLLPPY	SSAVSSVCENPILQPY	RIQQAIAAGILPLSPL	SSAVSSVCENPI	TQCSQAPI
Rank			**Score**			
8	−2837.18	−2674.52	−2984.81	−2891.71	−2631.18	−1906.76
9	−2813.36	−2668.30	−2651.25	−2431.49	−2503.59	−1782.29
10	−2676.38	−2040.50	−2640.75	−2226.68	−2230.63	−1623.52

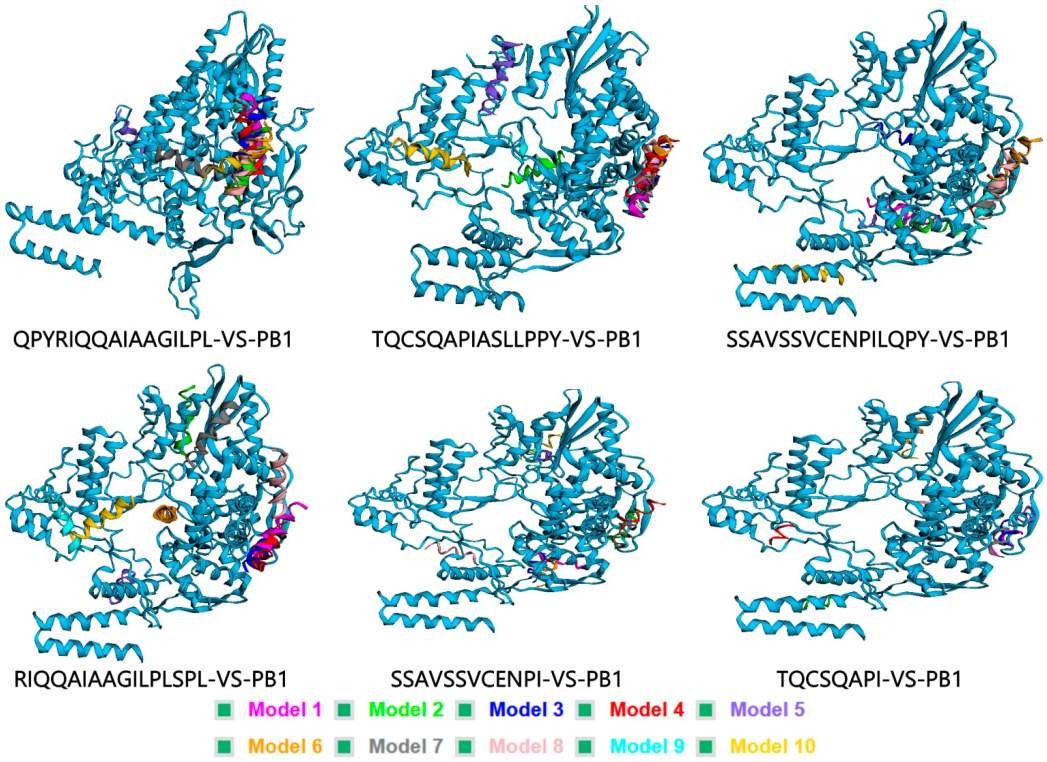


Figure 6. 3D interactions between peptides and *PB1* subunit of H5N1 influenza polymerase using the HawkDock server.

3.8.2. The 2D Interaction of Peptide–PB1 Obtained Using the HawkDock Server

Figure 7 Two-dimensional interaction of the six peptides with PB1 of the H5N1 influenza polymerase. The interaction models demonstrate the residues taking part in the peptide–PB1 interactions and corresponding non-covalent bonds found for the peptide. The hydrogen bond interactions are marked with the green dotted lines, whereas the semi-circle marks show the additional residue–residue interactions. The peptide QPYRIQQAIAAGILPLS has the interaction model displaying the contacts involving Arg4(B) and several residues of the opposite chain, including Arg316(A). The TQCSQAPIASLLPPY peptide has several contacts involving Gln2(B), Gln5(B), and Thr3(B) with additional interactions with residues Asn67(A), Ser345(A), and Arg316(A). The peptide SSAVSSVCENPI has several contacts involving Gln8(B), Ser5(B), and Ser6(B) together with residues Arg135(A), Arg350(A), Asn134(A), and Asp398(A). The peptide RIQQAIAAGILPLS has contacts involving Ser14(B) and Ile10(B) with additional interactions with Asn67(A). In addition to the contact of the model for the SSAVSSVCENPI peptide, there are contacts involving Ser6(B) and Ala3(B) together with Asn134(A) and Arg350(A).

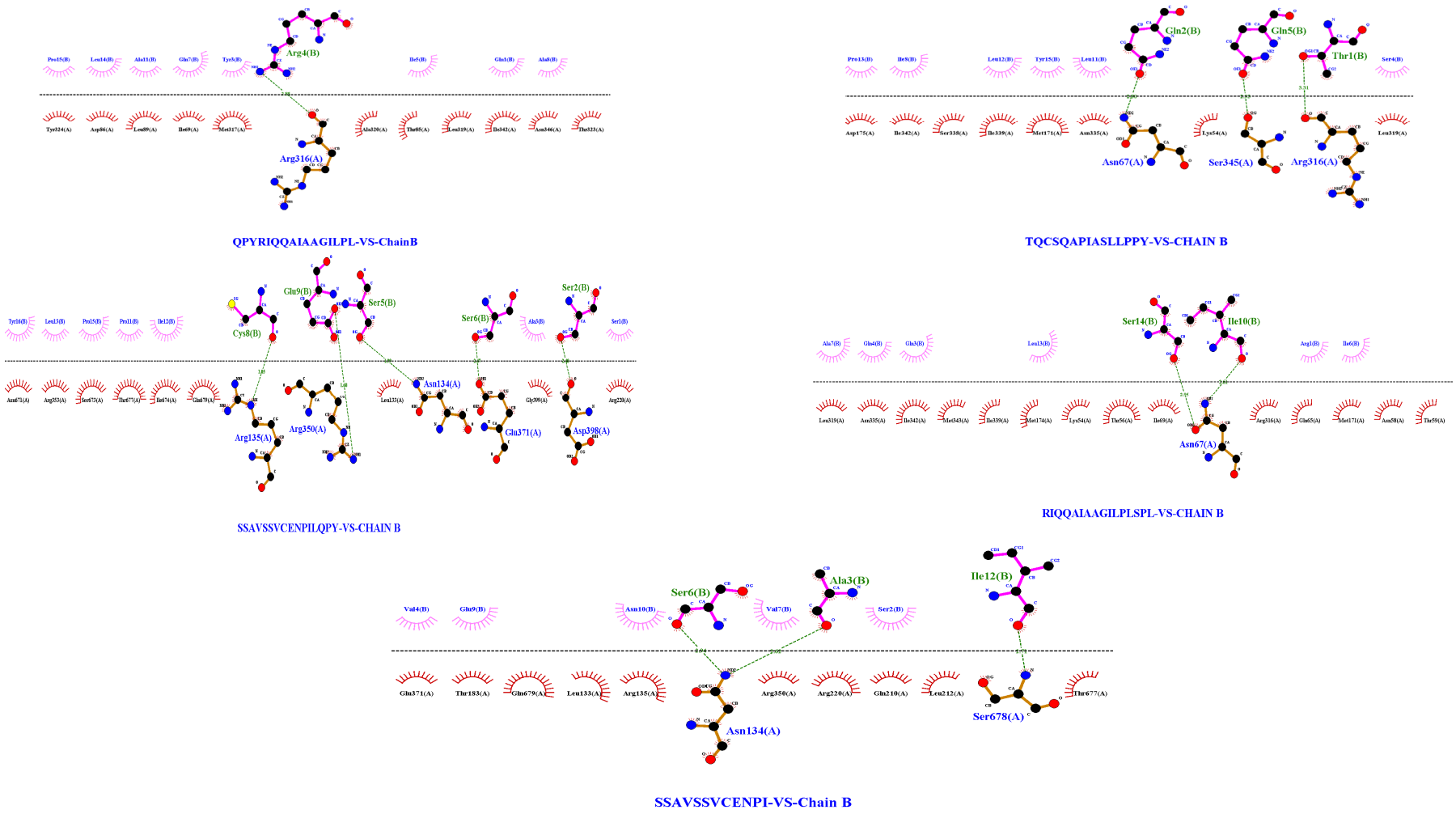


Figure 7. The 2D interactions between predicted peptides and the PB1 subunit.

3.8.3. Atomic Interactions Between Predicted Peptides and the *PB1* of H5N1 Polymerase Using the HawkDock Molecular Docking Tool

Analysis of molecular interactions indicated unique patterns of interaction among the peptides that are predicted and docked on the *PB1* subunit of the H5N1 influenza polymerase (Table 13 and Figure 8). Proximity contacts occurred within the range of 19 to 29 contacts, where the highest number of proximity contacts occurred in peptide D (SSAVSSVCENPILQPY), followed by peptides B and C (26 each) and F (24). Apolar van der Walls interactions occurred within the range of 12 to 19 contacts. The highest number of contacts occurred in peptide F (19), while peptide A (TQCSQAPI) had the highest number of CH-O/N (13). Hydrogen bonding interactions occurred for all peptides with the range of 2 to 4 bonds. Peptides C and D had the highest number (4 hydrogen bonds each), followed by peptide E (3), while peptides A, B, and F had 2 hydrogen bonds each. Disulfide (S-S) bonds occurred only for peptide D (1 interaction). Steric interactions occurred within the range of 3 to 6 interactions; peptide C had the highest number of interactions (6), followed by peptides D and E (5 each). In contrast, no salt bridges, water-mediated contacts, halogen bonds, metal-mediated contacts, π–π interactions, lp–π interactions, anion–π interactions, cation–π interactions, amino–π interactions, O/N/SH–π interactions, CH–π interactions, or polar van der Waals contacts were detected for any of the analyzed peptide–complex pairs.

Table 13. Atomic interactions between predicted peptides and the *PB1* subunit of H5N1 polymerase.

Atomic Interactions	A	B	C	D	E	F
Proximal contacts	20	26	26	29	19	24
Apolar vdW contacts	16	14	13	12	12	19
CH-O/N bonds	13	9	9	7	6	4
Clashes	3	3	6	5	5	4
H-bonds	2	2	4	4	3	2
S-S bonds	0	0	0	1	0	0
Salt bridges	0	0	0	0	0	0
Water mediated contacts	0	0	0	0	0	0
Halogen bonds	0	0	0	0	0	0
Metal mediated contacts	0	0	0	0	0	0
π–π interactions	0	0	0	0	0	0
lp–π interactions	0	0	0	0	0	0
Anion–π interactions	0	0	0	0	0	0
Cation–π interactions	0	0	0	0	0	0
Amino–π interactions	0	0	0	0	0	0
O/N/SH–π interactions	0	0	0	0	0	0
CH–π interactions	0	0	0	0	0	0
Polar vdW contacts	0	0	0	0	0	0

A: Interactions between TQCSQAPI and the *PB1* subunit of H5N1 influenza polymerase, **B**: interactions between TQCSQAPIASLLPPY and the *PB1* subunit of H5N1 influenza polymerase, **C**: interactions between SSAVSSVCENPI and the PB1 subunit of H5N1 influenza polymerase, **D**: interactions between SSAVSSVCENPILQPY and the *PB1* subunit of H5N1 influenza polymerase, **E**: interactions between RIQQAIAAGILPLSPL and the *PB1* subunit of H5N1 influenza polymerase, **F**: interactions between QPYRIQQAIAAGILPL and the *PB1* subunit of H5N1 influenza polymerase.

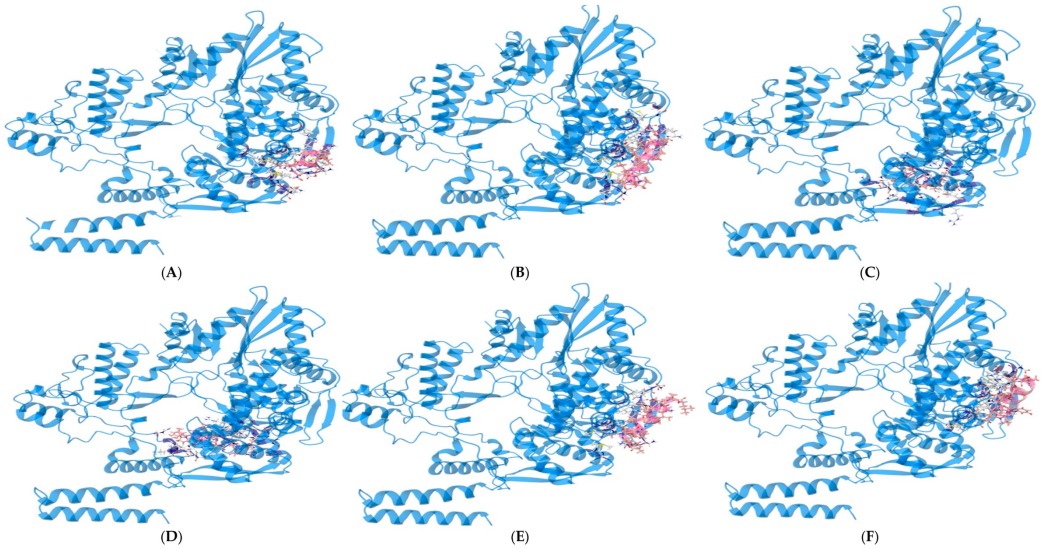


Figure 8. The three-dimensional structure atomic interactions between predicted peptides and the H5N1 polymerase using the HawkDock based docking approach. (**A**): Interactions between TQCSQAPI and the *PB1* subunit of H5N1 influenza polymerase, (**B**): interactions between TQCSQAPIASLLPPY and the PB1 subunit of H5N1 influenza polymerase, (**C**): interactions between SSAVSSVCENPI and the *PB1* subunit of H5N1 influenza polymerase, (**D**): interactions between SSAVSSVCENPILQPY and the PB1 subunit of H5N1 influenza polymerase, (**E**): interactions between RIQQAIAAGILPLSPL and the PB1 subunit of H5N1 influenza polymerase, (**F**): interactions between QPYRIQQAIAAGILPL and the *PB1* subunit of H5N1 influenza polymerase.

3.8.4. Molecular Docking Results Using pyDockWEB

Results from the pyDockWEB analysis revealed that the total docking energy was different across the ranked solutions for all peptide–protein complexes. The total energy of SSAVSSVCENPILQPY had variations from −29.537 to −24.237. TQCSQAPIASLLPPY had variations from −26.786 to −23.736, and QPYRIQQAIAAGILPL had variations from −26.701 to −22.737. The other variations in energy for RIQQAIAAGILPLSPL were from −29.537 to −24.237. For the smaller peptide sequences, TQCSQAPI had variations in energy of −16.982 to −14.617, and SSAVSSVCENP had variations from −25.897 to −20.953. For all reported complexes, the most negative energy values occurred for SSAVSSVCENPILQPY and RIQQAIAAGILPLSPL (−29.537), followed by TQCSQAPIASLLPPY (−26.786) and QPYRIQQAIAAGILPL (−26.701), while the other complexes had less negative energy values. See full ranked data in Table 14 and Figure 9.

Table 14. Binding scores between candidate peptides and *PB1* using pyDockWEB.

Rank	SSAVSSVCENPILQPY	TQCSQAPIASLLPPY	QPYRIQQAIAAGILPL	RIQQAIAAGILPLSPL	TQCSQAPI	SSAVSSVCENP
	Total Energy	**Total Energy**	**Total Energy**	**Total Energy**	**Total Energy**	**Total Energy**
1	−29.537	−26.786	−26.701	−29.537	−16.982	−25.897
2	−26.708	−26.586	−26.344	−26.708	−16.976	−25.247
3	−25.372	−26.294	−25.641	−25.372	−16.817	−24.260
4	−25.239	−25.673	−24.785	−25.239	−16.486	−23.399
5	−25.088	−24.698	−24.458	−25.088	−16.401	−23.175
6	−24.955	−24.672	−23.871	−24.955	−16.353	−22.476
7	−24.608	−24.297	−23.544	−24.608	−14.966	−22.088
8	−24.569	−24.188	−23.026	−24.569	−14.926	−21.695
9	−24.456	−23.784	−22.964	−24.456	−14.901	−21.493
10	−24.237	−23.736	−22.737	−24.237	−14.617	−20.953

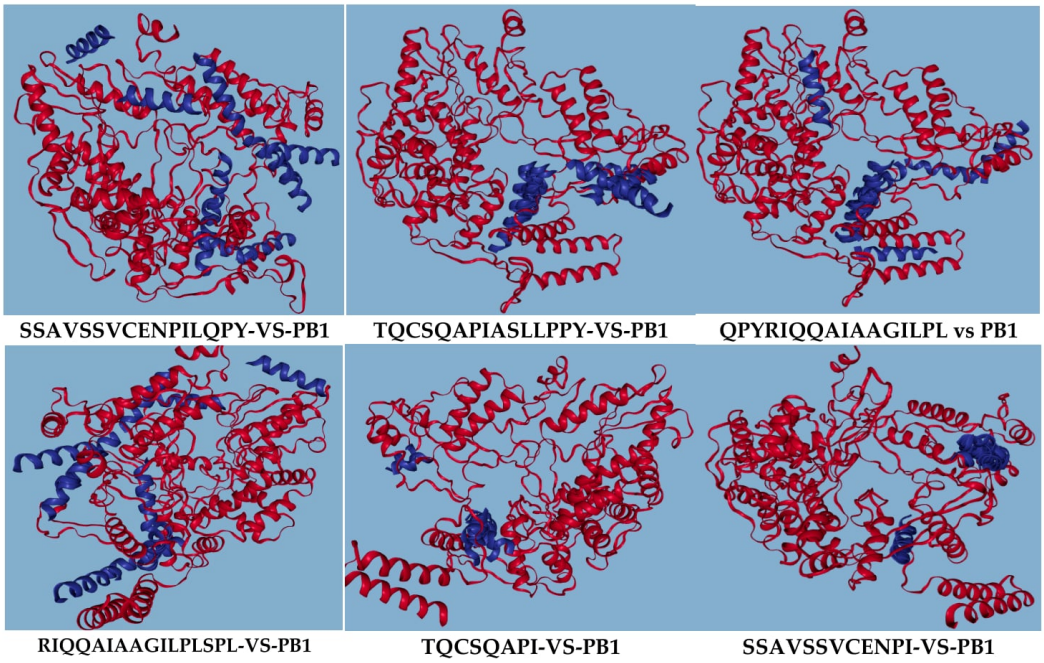


Figure 9. The 3D complexes between candidate peptides and *PB1* using pyDockWEB.

3.8.5. Molecular Docking Results Using PatchDock

Atomic contact energy (ACE) scores were obtained for the six protein–peptide complex pairs tested using PatchDock analysis, covering ten ranked docking solutions for each pair. The ACE scores varied from −338.73 to −94.03 for QPYRIQQAIAAGILPL, −448.58 to −214.49 for TQCSQAPIASLLPPY, −336.48 to −211.98 for SSAVSSVCENPILQPY, −286.97 to −116.02 for RIQQAIAAGILPLSPL, −254.82 to −133.47 for SSAVSSVCENPI, and −198.46 to −101.38 for TQCSQAPI. Out of the ten ranked docking solutions, the most negative ACE score was observed for TQCSQAPIASLLPPY (−448.58), followed by QPYRIQQAIAAGILPL (−338.73), SSAVSSVCENPILQPY (−336.48), RIQQAIAAGILPLSPL (−286.97), SSAVSSVCENPI (−254.82), and TQCSQAPI (−198.46). The PatchDock ACE scores are shown in Table 15 and Figure 10.

Table 15. PatchDock atomic contact energy (ACE) scores of the protein–peptide docking complexes.

Rank	QPYRIQQAIAAGILPL	TQCSQAPIASLLPPY	SSAVSSVCENPILQPY	RIQQAIAAGILPLSPL	SSAVSSVCENPI	TQCSQAPI
	ACE	ACE	ACE	ACE	ACE	ACE
1	−338.73	−448.58	−336.48	−286.97	−254.82	−198.46
2	−278.74	−335.84	−332.51	−220.56	−231.74	−184.37
3	−218.51	−314.65	−269.41	−206.35	−218.63	−171.25
4	−209.69	−307.08	−261.41	−192.86	−205.91	−159.82
5	−206.72	−302.22	−250.39	−183.91	−194.76	−148.63
6	−195.38	−300.84	−243.61	−149.48	−181.52	−137.94
7	−130.73	−229.03	−237.26	−146.61	−169.38	−126.71
8	−116.13	−214.49	−222.64	−145.87	−157.26	−118.54
9	−113.90	−297.43	−217.72	−120.06	−145.83	−109.62
10	−94.03	−294.82	−211.98	−116.02	−133.47	−101.38

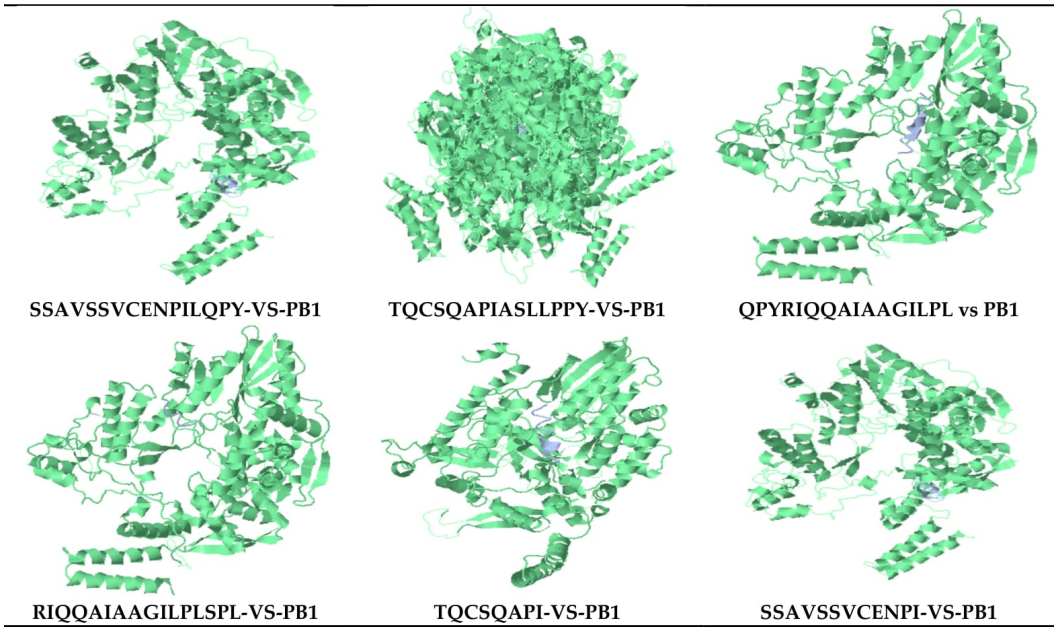


Figure 10. The 3D complexes between candidate peptides and *PB1* using PatchDock.

3.9. Dynamic Simulation Parameters

3.9.1. Dynamic Simulation of Docking Complexes from HawkDock

Figures 11–13 present the dynamic simulation parameters of the *PB1* subunit and selected peptides. For the complex with the peptide RIQQAIAAGILPSPL, the total energy shows marked fluctuations during the initial and intermediate stages before becoming more stable toward the end of the simulation. The number of hydrogen bonds generally fluctuates between 4 and 7 interactions, with occasional increases, suggesting that the peptide maintains persistent non-covalent interactions with the protein. The radius of gyration (Rg) varies approximately between 8 and 10 Å, followed by stabilization around 8.5–9 Å, indicating a slight structural rearrangement and subsequent maintenance of a relatively compact conformation. The solvent-accessible surface area (SASA) also fluctuates, reflecting changes in the exposure of certain regions to the solvent. The RMSD increases rapidly at the beginning and then reaches a plateau, suggesting an initial structural adaptation followed by a relatively stable dynamic conformation. RMSF analysis reveals several highly flexible regions, probably corresponding to loops or peripheral regions of the polymerase. For the TQCSQAPIASLLPPY complex, the analyzed parameters similarly indicate progressive stabilization. The number of hydrogen bonds generally remains around three to five, with transient increases, suggesting sustained but somewhat variable peptide–protein interactions. The radius of gyration stabilizes around 7.5–8 Å after structural rearrangement, indicating a relatively compact complex. The RMSD increases rapidly during the early stage and subsequently reaches a plateau around 6–7 Å, although noticeable fluctuations remain, particularly for one of the components represented in red, suggesting somewhat higher structural dynamics. The RMSF profile also shows several highly flexible regions distributed along the sequence, which may reflect the intrinsic mobility of loops and terminal regions rather than global instability.

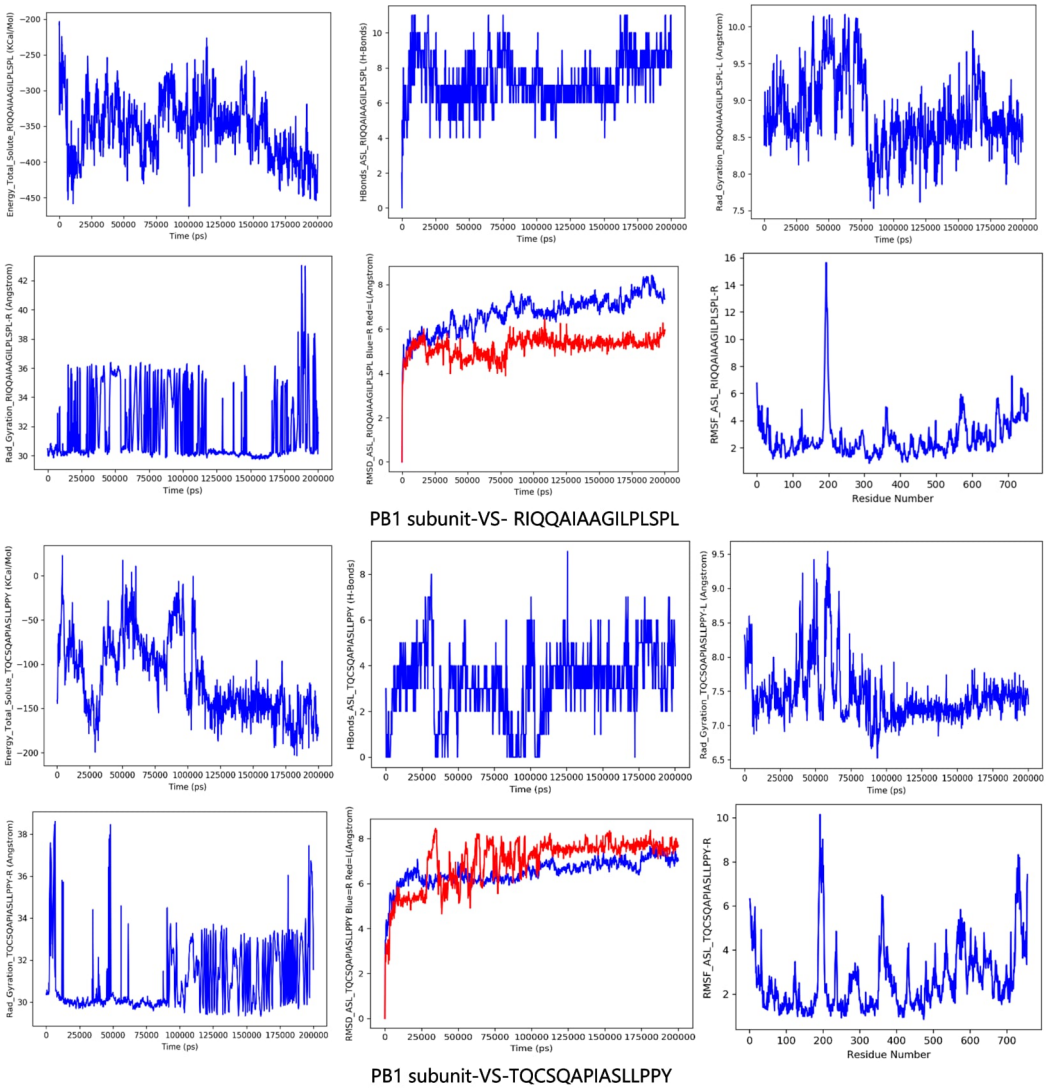


Figure 11. Dynamic simulation parameters of *PB1* subunit–TQCSQAPIASLLPPY/SSAVSSVCENPI HawkDock complexes.

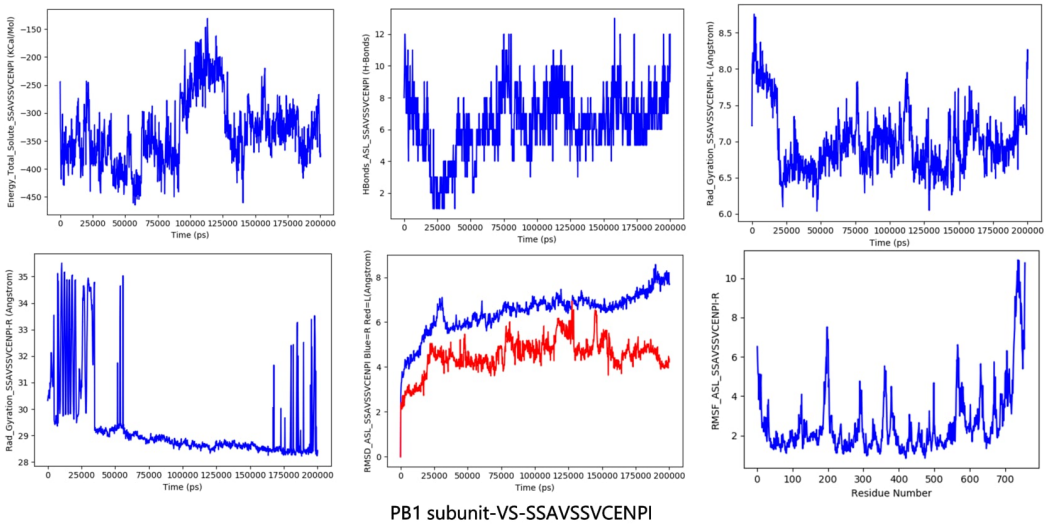


Figure 12. *Cont.*

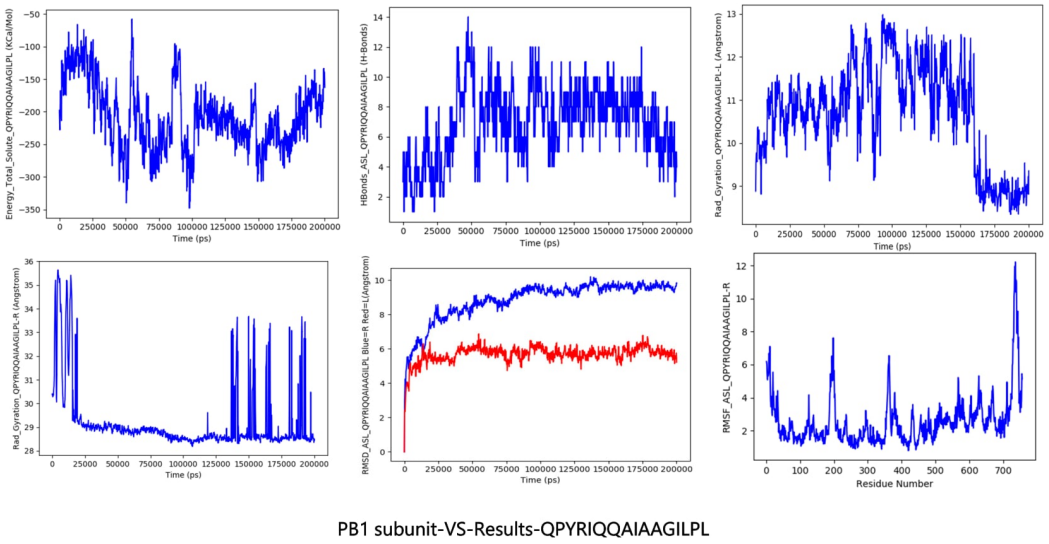


Figure 12. Dynamic simulation parameters of *PB1* subunit–QPYRIQQAIAAGILPL/RIQQAI–AGILPLSPL HawkDock complexes.

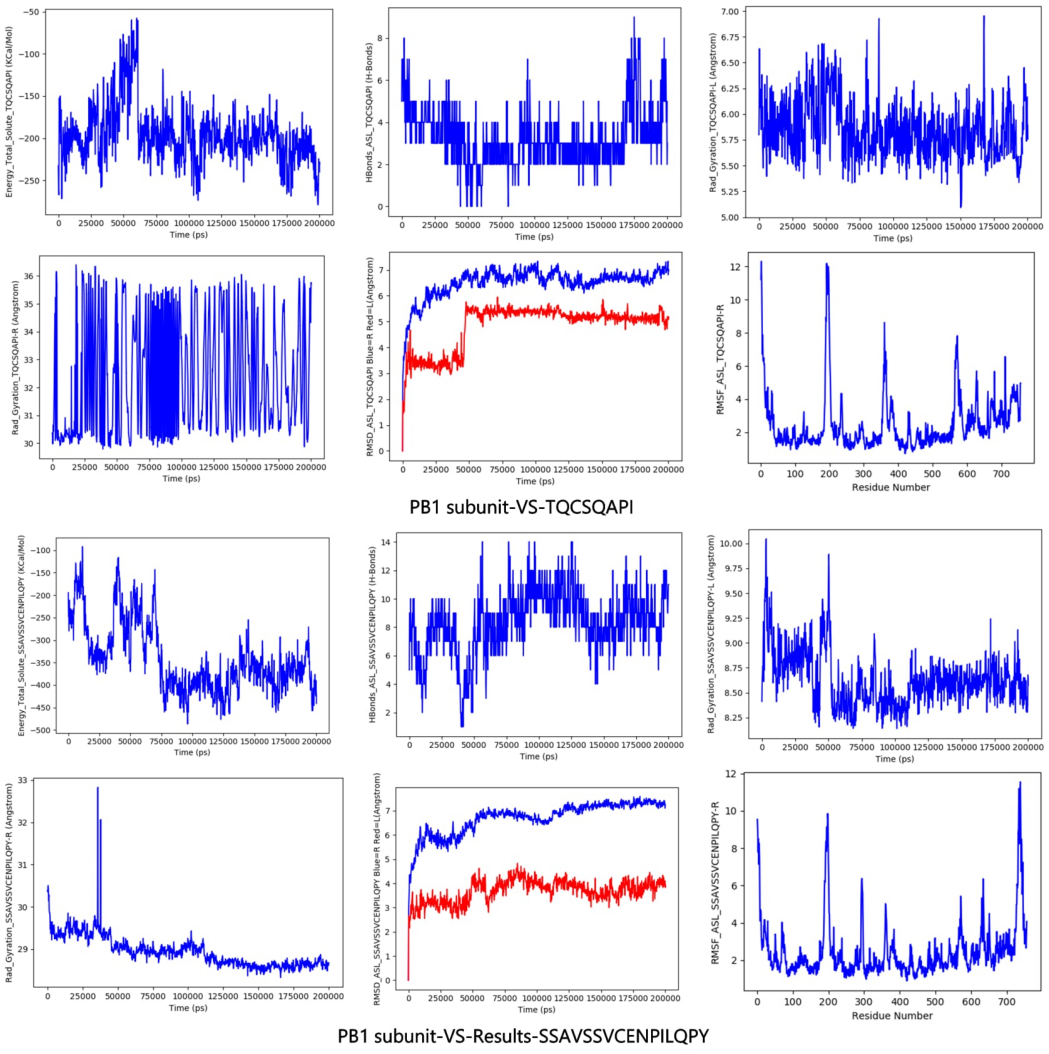


Figure 13. Dynamic simulation parameters of *PB1* subunit–SSAVSSVCENPILQPY/TQCSQAPI complexes.

For the complex with the peptide VS-SSAVSSVCENPI, the total energy shows pronounced fluctuations during the initial and intermediate stages before reaching a compar-

atively stable region toward the end of the simulation. The number of hydrogen bonds generally fluctuates between approximately three and eight interactions, with several transient increases, suggesting that the peptide maintains persistent non-covalent interactions with the polymerase. The radius of gyration (Rg) initially decreases from approximately 8.5 Å to around 6.5 Å and subsequently fluctuates mainly between 6.5 and 7.2 Å, indicating an initial structural rearrangement followed by the maintenance of a relatively compact conformation. The SASA profile also shows a pronounced decrease during the early stage followed by a more stable pattern, suggesting an initial reduction in solvent exposure and subsequent structural stabilization. The RMSD increases rapidly at the beginning of the simulation and then reaches a relatively stable plateau, with the protein component remaining mainly around 6–7 Å and the second component around 4–5 Å. However, local fluctuations are observed. RMSF analysis reveals several flexible regions with pronounced peaks, particularly toward the terminal and loop regions, while most residues display moderate fluctuations, indicating localized flexibility rather than global instability. For the QPYRIQQAIAAGILPL complex, the analyzed parameters also indicate persistent peptide–polymerase interactions but reveal more pronounced structural rearrangements. The total energy fluctuates considerably throughout the trajectory, with several transitions between lower and higher energy states, before showing a more stable behavior toward the final stages. The number of hydrogen bonds generally remains around 4–9 interactions, with frequent transient increases, indicating sustained but variable non-covalent interactions between the peptide and polymerase. The radius of gyration initially increases from approximately 30 to 34 Å and then gradually decreases, with a marked reduction toward the end of the simulation, suggesting significant structural rearrangement followed by increased compactness. The SASA profile also displays considerable fluctuations, with an overall decrease toward the final part of the trajectory, indicating changes in solvent accessibility associated with the conformational reorganization of the complex. The RMSD reaches approximately 9–10 Å for the blue component and around 5–6 Å for the red component, with noticeable fluctuations throughout the simulation, suggesting greater conformational dynamics compared with VS-SSAVSSVCENPI. The RMSF profile similarly shows several pronounced peaks distributed along the sequence.

The simulations indicate that both RNA-directed RNA polymerase catalytic subunit–peptide complexes involving chains B and E maintain persistent interactions throughout the 200,000 ps (200 ns) trajectories, although distinct differences in their dynamic behavior are observed. For the complex with the peptide TQCSQAPI, the total energy initially fluctuates strongly and reaches higher energy values during the first part of the simulation before decreasing and remaining relatively stable around −200 to −230 kcal/mol during the later stages. The number of hydrogen bonds generally fluctuates between approximately two and six interactions, with a temporary decrease during the middle part of the trajectory followed by an increase toward the end, indicating dynamic but persistent non-covalent interactions between the peptide and polymerase. The radius of gyration (Rg) remains mainly between approximately 5.5 and 6.5 Å, with occasional peaks, suggesting that the complex maintains a relatively compact structure despite local conformational fluctuations. The SASA profile exhibits substantial variations throughout the simulation, with values mainly around 30–36 $Å^2$, indicating changes in solvent exposure and structural rearrangements of the complex. The RMSD increases rapidly during the initial stage and subsequently reaches a relatively stable plateau, with the blue component remaining around 6–7 Å and the red component increasing to approximately 4.5–5 Å, suggesting an initial adaptation followed by a comparatively stable dynamic conformation. RMSF analysis reveals several pronounced peaks, particularly around specific residues, indicating localized flexibility in loop or terminal regions rather than global destabilization. For the second complex, con-

taining the peptide SSAVSSVCENPILQPY, the total energy shows larger variations during the first half of the trajectory, followed by a more stable region toward the end, although fluctuations remain. The hydrogen bond profile generally ranges from approximately five to 12 interactions, with several periods of increased hydrogen bond formation, suggesting a strong and persistent network of non-covalent interactions between the peptide and polymerase. The radius of gyration initially decreases from approximately 9.0–10.0 Å to around 8.3–8.6 Å and then remains relatively stable, indicating an initial structural rearrangement followed by stabilization in a more compact conformation. The SASA profile also decreases during the early stages and subsequently fluctuates around approximately 28–30 $Å^2$, suggesting reduced solvent exposure and maintenance of a relatively compact structure. The RMSD increases rapidly at the beginning and then reaches a plateau, with the blue component gradually stabilizing around 7–7.5 Å and the red component remaining mainly around 3.5–4.5 Å, indicating that the complex undergoes an initial conformational adaptation followed by a relatively stable dynamic state.

3.9.2. Dynamic Simulation of Docking Complexes from PatchDock and pyDockWEB

2D *PB1*–Peptide Interaction Obtained from PatchDock and pyDockWEB

The binding modes of the six peptides were analyzed through two-dimensional interaction diagrams of the PatchDock and pyDockWEB complexes (Figure 14). The two tools gave comparable interfaces for the long peptides but clearly different ones for several others. QPYRIQQAIAAGILPL was bound in an extended conformation by both tools, with PatchDock concentrating contacts along its central and C-terminal parts through several long polar contacts, whereas pyDock distributed contacts along the whole peptide with a larger number of hydrophobic and charged residues at both ends. For RIQQAIAAGILPLSPL, PatchDock produced a curved peptide surrounded by a moderate number of polar and hydrophobic residues, while pyDock gave a peptide with a folded C-terminal region and few contacting residues, mostly around its central segment. For SSAVSSVCENPILQPY, the PatchDock pose showed contacts mainly with the N-terminal half and a C-terminal end bending toward the receptor, whereas the pyDock pose lay flat on the receptor surface with dense polar and charged contacts along its entire length, giving the more extensive interface. The short peptide SSAVSSVCENPI showed very few contacting residues with PatchDock and a largely solvent-exposed backbone, which indicates a weak, localized interface consistent with the low hydrogen bond count and the single dominant residue observed in the dynamics. TQCSQAPIASLLPPY displayed the weakest interface, with almost no contacts in the PatchDock pose and only a few additional polar contacts in the pyDock pose, so both tools suggest a weak binding mode with a slight advantage for pyDock. Finally, the short peptide TQCSQAPI formed a compact interface in both poses, but the PatchDock pose was better defined, with about ten surrounding residues and numerous polar contacts around its central and N-terminal parts, against fewer and more dispersed contacts with pyDock. Overall, neither tool systematically produced the denser interface. PatchDock was favorable for QPYRIQQAIAAGILPL and TQCSQAPI, and pyDock was favorable for SSAVSSVCENPILQPY and TQCSQAPIASLLPPY. The longer peptides (15–16 residues) generally established more contacts than the shorter ones, depending on a few anchoring residues. These static diagrams should be interpreted together with the molecular dynamics results and completed by binding free energy calculations.

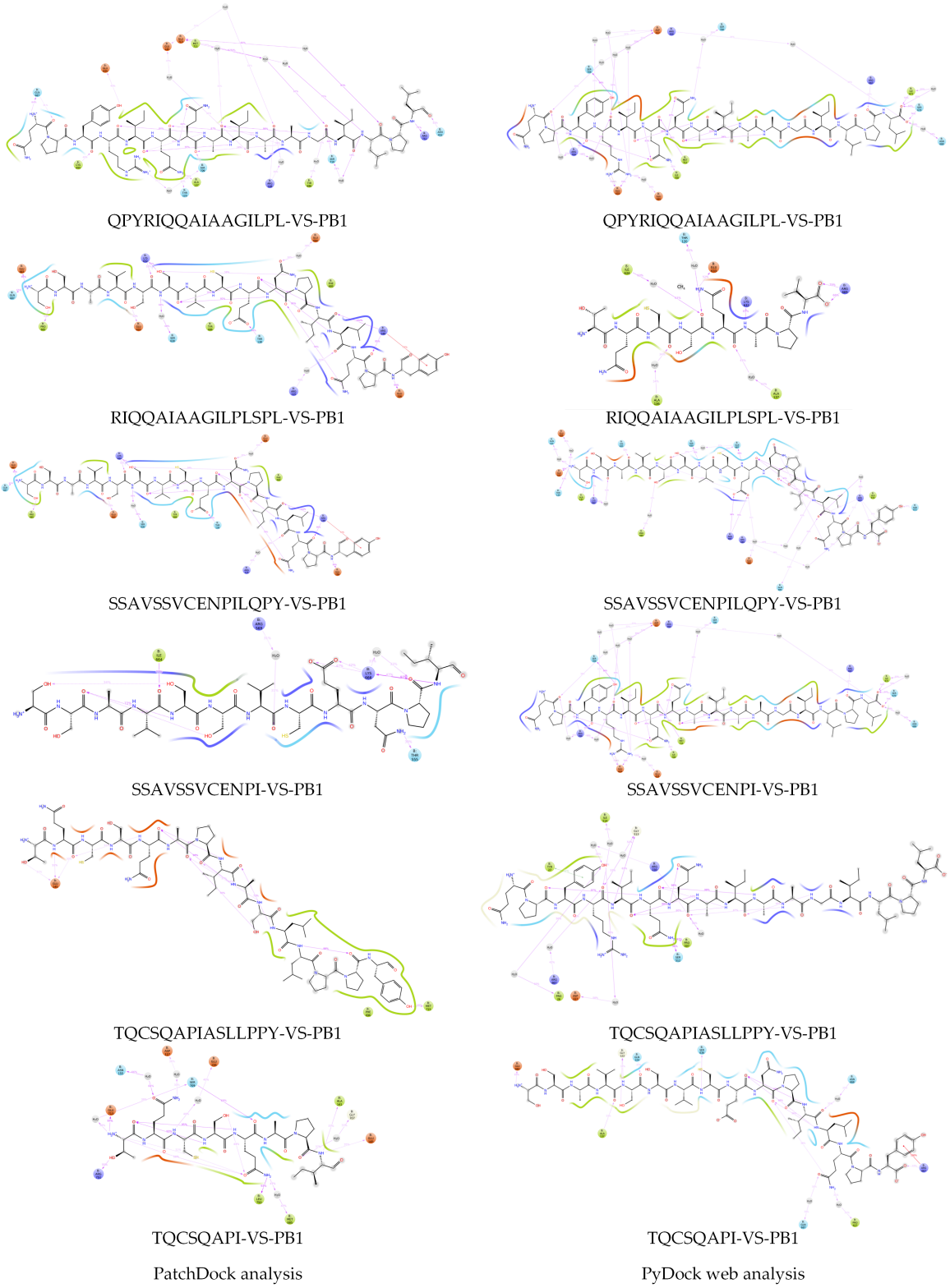


Figure 14. Two-dimensional protein–peptide interaction diagrams of the six peptide–PB1 complexes obtained from PatchDock (**left**) and pyDockWEB (**right**) poses. Dashed arrows indicate hydrogen bonds, pointing from the donor to the acceptor, with the percentage showing the frequency of the interaction. The solid red line indicates a cation–π interaction. Residue bubbles are coloured by residue type: orange, negatively charged; purple/blue, positively charged; cyan, polar uncharged; green, hydrophobic. Grey "H_2O" bubbles indicate water-mediated contacts.

MD of QPYRIQQAIAAGILPL–PB1 and RIQQAIAAGILPLSPL–*PB1* Obtained with PatchDock and pyDockWEB

Figure 15 illustrates the parameters of molecular dynamics simulations of two peptide–PB1 complexes, QPYRIQQAIAAGILPL–*PB1* and RIQQAIAAGILPLSPL–*PB1*, obtained using PatchDock and pyDockWEB docking models. As shown in Figure 15A, the QPYRIQQAIAAGILPL–PB1 complex generated using PatchDock had roughly 8–11 hydrogen bonds throughout the course of the simulation, reaching up to 16 at its peak while falling sometimes to 4–6 hydrogen bonds. In Figure 15B, the pyDockWEB generated complex mostly had 6–9 hydrogen bonds, occasionally spiking to 12–14 hydrogen bonds, showing slightly fewer typical number of hydrogen bonds compared to the PatchDock-derived model. On the other hand, RIQQAIAAGILPLSPL–*PB1* complex had the trend of increased hydrogen bond numbers in Figure 15C; the number of hydrogen bonds varied from 3–5 in early simulation to 6–8 in later simulation stages, with occasional spikes up to 10–11 hydrogen bonds. As shown in Figure 15D, RIQQAIAAGILPLSPL–*PB1* complex generated through pyDockWEB model had a larger number of hydrogen bonds, which ranged from 9 to 12 mostly, with occasional spikes up to 14–15 hydrogen bonds. Figure 15E illustrates the QPYRIQQAIAAGILPL–*PB1* complex constructed by PatchDock, with its fluctuation range initially being approximately 8.5–9.7 Å and then gradually becoming stable at around 8.7–9.0 Å. Figure 15F illustrates the corresponding pyDockWEB model, which had its Rg value initially fluctuating around 9.6–10.0 Å and sometimes even going up to 10.2–10.4 Å. These larger Rg values mean that there is greater spatial distribution in this particular model, while interpretation of these values also depends on the structures and atom selections of the calculation. For the RIQQAIAAGILPLSPL–*PB1* complex, Figure 15G demonstrates significant fluctuation in its Rg value, ranging between approximately 7 and 10 Å before declining dramatically to 6.8–7.2 Å at approximately 160–180 ns and then regaining 8–9 Å. This significant change might be an indication of some conformational change within this simulation. Figure 15H illustrates the corresponding pyDockWEB complex with an initial increase in Rg value to around 8.4–8.6 Å followed by its fluctuation at approximately 8.2–8.4 Å. RMSD profiles also allowed us to evaluate structural adaptation of peptides. As can be seen in Figure 15I, the complex of QPYRIQQAIAAGILPL–*PB1* created using PatchDock demonstrated the growth of protein RMSD up to approximately 3.0–3.4 Å at first, then the RMSD of the peptide began to grow steadily and after that both RMSDs reached approximately equivalent levels in the course of the later simulation. In turn, Figure 15J shows that the corresponding complex created using pyDockWEB experienced fast initial adaptation, then fluctuations and transitions around 70–100 ns, followed by recovery of the trajectories and relatively constant fluctuation. Regarding RIQQAIAAGILPLSPL–*PB1* complex, as can be seen from Figure 15K, it demonstrated the growth of protein RMSD up to approximately 3.0–3.7 Å, while peptide RMSD fluctuated much more than it, thus, indicating more variability of the peptide's position or structure relative to the protein. In its turn, Figure 15L demonstrates that the corresponding complex created using pyDockWEB had initially grown RMSD, then had relatively stable late-stage profile, where protein RMSD was approximately 3.0–3.5 Å and peptide trajectory had similar profile. The RIQQAIAAGILPLSPL–*PB1* complex predicted using pyDockWEB had a higher average number of hydrogen bonds than its PatchDock version. However, for the QPYRIQQAIAAGILPL–*PB1* complex, the reverse was the case for their respective hydrogen bond graphs. In addition, the RMSD graph indicated that both systems went through an initial conformational adjustment phase followed by stability, although the peptide had higher fluctuations in Figure 15K. This shows some preliminary computational proof of the different dynamics of the two complexes; however, the average number of

hydrogen bonds, Rg, and RMSD can only offer some clues to the relative binding affinity and inhibiting capabilities of the peptides.

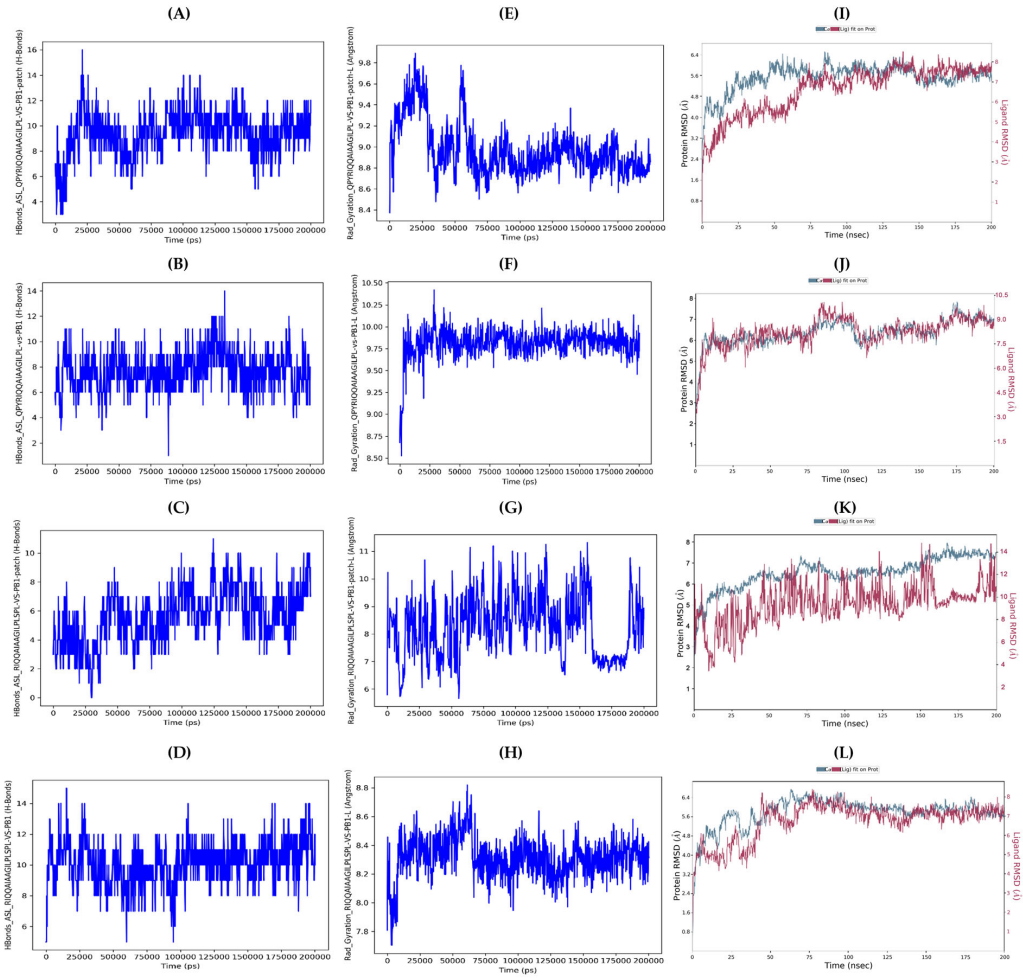


Figure 15. Simulation profile of two peptide–*PB1* complexes modeled using PatchDock and pyDockWEB programs. (**A**,**E**,**I**) Hydrogen bonding profile, radius of gyration (Rg), and root mean square deviation (RMSD), respectively, for the QPYRIQQAIAAGILPL–*PB1* complex modeled using PatchDock; (**B**,**F**,**J**) hydrogen bonding profile, Rg, and RMSD, respectively, for the same complex modeled using pyDockWEB; (**C**,**G**,**K**) hydrogen bonding profile, Rg, and RMSD, respectively, for the RIQQAIAAGILPLSPL–*PB1* complex modeled using PatchDock; and (**D**,**H**,**L**) hydrogen bonding profile, Rg, and RMSD, respectively, for the same complex modeled using pyDockWEB. The simulation was conducted for around 200 ns.

MD of SSAVSSVCENPILPQY–PB1 and SSAVSSVCENPI–*PB1* Obtained with PatchDock and pyDockWEB

Figure 16 illustrates the molecular dynamics simulation parameters of the peptide–*PB1* complexes involving SSAVSSVCENPILPQY and SSAVSSVCENPI, generated using PatchDock and pyDockWEB docking models. As observed in Figure 16A, the resulting peptide–*PB1* complex formed using PatchDock showed many variations in the number of hydrogen bonds during the simulation process. There was generally a range of about 8 to 13 hydrogen bonds, while there were also peaks of about 16–18 as well as drops to about 5–7 hydrogen bonds. These variations indicated that there was a dynamic hydrogen bonding interaction network at the peptide–protein interface. In Figure 16B, the resulting pyDockWEB complex showed a fairly constant number of hydrogen bonds ranging mostly from 9 to 12, with peaks of about 13–15 and drops to about 5–7. This trend implied that the number of hydrogen bonds was quite stable despite the local variations. As for the

SSAVSSVCENPI–*PB1* complex, Figure 16C shows an overall increase in the number of hydrogen bonds throughout the initial and intermediate periods of the simulation. It generally varied within a range of about 2–5 before rising to 6–9, and even occasionally reaching peaks of 10–11 hydrogen bonds. Figure 16D, on the other hand, reveals that the corresponding pyDockWEB-derived complex maintained about 4 to 6 hydrogen bonds, with rare occurrences of 7 to 9 and 1 to 3 hydrogen bonds. This suggests that, compared to the PatchDock-derived complex, the pyDockWEB structure had fewer hydrogen bonds on average, implying that there were differences in the hydrogen bonding patterns of the two docking poses. The Rg values of the complexes offer additional data regarding their stability. In the case of the SSAVSSVCENPILPQY–*PB1* complex created using PatchDock, as depicted in Figure 16E, Rg values started at 9.5 to 10.0 Å and declined quite rapidly to 8.4 to 8.7 Å. Thereafter, the Rg values oscillated between 8.4 and 8.9 Å for the rest of the simulation period.

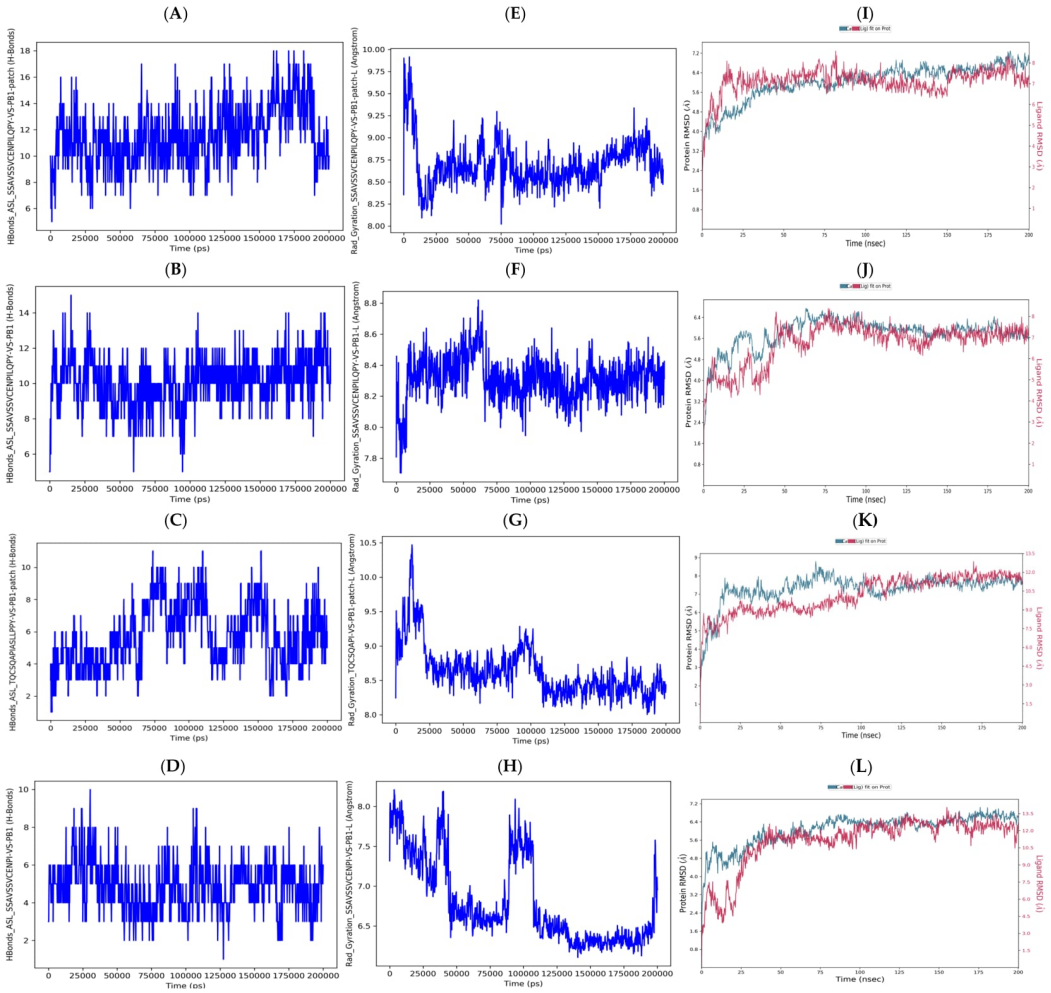


Figure 16. Simulation profile of two peptide–PB1 complexes modeled using PatchDock and pyDockWEB programs. (**A**,**E**,**I**) Hydrogen bonding profile, radius of gyration (Rg), and root mean square deviation (RMSD), respectively, for the SSAVSSVCENPILPQY–*PB1* complex modeled using PatchDock; (**B**,**F**,**J**) hydrogen bonding profile, Rg, and RMSD, respectively, for the same complex modeled using pyDockWEB; (**C**,**G**,**K**) hydrogen bonding profile, Rg, and RMSD, respectively, for the SSAVSSVCENPI–*PB1* complex modeled using PatchDock; and (**D**,**H**,**L**) hydrogen bonding profile, Rg, and RMSD, respectively, for the same complex modeled using pyDockWEB. The simulation was conducted for around 200 ns.

Figure 16F depicts that the Rg value for the corresponding pyDockWEB model initially equaled about 7.8–8.2 Å, then oscillated mostly within the range of 8.2–8.5 Å, sometimes

reaching higher values of 8.6–8.7 Å. This fairly uniform behavior implies that the general compactness of this complex did not change significantly over time throughout most of the simulation. In the case of the SSAVSSVCENPI–*PB1* complex, Figure 16G illustrates significant changes in the value of Rg in the first phase of the simulation. It initially grew up to the value of 9.5–9.8 Å and then dropped and oscillated mainly within 8.7–9.0 Å for much of the rest of the trajectory. This behavior could be indicative of the initial conformational rearrangement, and stabilization of a more compact form thereafter. Finally, as seen in Figure 16H, the value of Rg of the corresponding pyDockWEB-generated complex gradually decreased from approximately 7.5–8.0 Å to 4.2–4.5 Å in the last phases of the simulation, except for a temporary rise to 7.0–7.6 Å.

Such high variability indicates a significant difference in the spatial arrangement of the atoms used in the analysis. This kind of activity could be caused by conformational changes and modifications in the density of the structure; nevertheless, it would be wise to compare the final spike with the trajectory and simulation results to confirm its physical reality or artificial character. The RMSD graphs demonstrate how the peptide–PB1 complexes adapt structurally during the simulation process. As seen from the graph presented in Figure 16I, the SSAVSSVCENPILPQY–*PB1* complex built with PatchDock demonstrates an initial spike of protein and peptide RMSD values and then fluctuations, stabilizing at 3.0–3.5 Å for the protein and about 3.0–3.5 Å for the peptide throughout most of the later simulation period. The combination of the two paths indicates that there were some structural modifications done prior to the protein and peptide maintaining a fairly constant deviation from their reference structures. As can be seen in Figure 16J, the corresponding pyDockWEB model showed an increasing trend in the RMSD values at first, followed by some fluctuation and transition phases during the early to middle part of the path.

For the case of SSAVSSVCENPI–*PB1*, Figure 16K demonstrates the increase in both protein and peptide RMSD throughout the simulation followed by the decrease in the rate of change of the values, which become rather stable throughout the latter stage of the simulation. The protein RMSD becomes approximately equal to 3.0–3.5 Å, while the peptide RMSD demonstrates the similar increasing trend. It means that the complex was adapting its structure before reaching the relatively stable conformational state. Figure 16L represents the pyDockWEB model for the complex, where both protein and peptide RMSDs increase at the beginning of the simulation and then demonstrate stable fluctuations. The protein RMSD is approximately 3.0–3.5 Å during the latter simulation, while the peptide RMSD demonstrates a similar behavior.

4. Discussion

In the present study, proteins from foods have been regarded as potential sources of peptides exhibiting antiviral activity toward PB1 protein of H5N1 influenza polymerase complex using a computational process including protein characterization, simulated digestion, machine learning prediction, docking, interaction analysis, and molecular dynamics (MD) simulation. In particular, the role of the polymerase interacting activity of *ANP32B* was determined based on the structural aspect that was examined in this study. Although *ANP32B* can perform multiple physiological roles, the present analysis has predicted the interaction between *ANP32B* and H5N1 polymerase as a potential target for the viral polymerase since interference with the interaction might affect the activity of the viral polymerase [21]. Such an implication is not expected to affect the other physiological roles of *ANP32B*, which need further investigation experimentally.

New developments in peptide discovery using in silico methods have shown the promise that can be achieved by coupling in silico enzymatic digestion, bioactivity prediction, physiochemical characterization, molecular docking, and molecular dynamics

simulation to identify bioactive food-derived peptides [30,55]. A recent review published in 2025 has pointed out the increasing use of computational and molecular docking approaches in the identification and characterization of food-derived bioactive peptides but also highlighted the requirement for an integrated approach to prioritize such peptides [56]. In relation to influenza A virus, several computational studies have investigated natural compounds against H5N1 viral proteins using molecular modeling, docking and pharmacokinetics, or ADMET analysis; however, these studies have mostly involved small molecule phytochemicals and not peptides. Zekri et al. [57] investigated natural compounds against H5N1 proteins using an integrated approach involving molecular modeling and ADMET analysis. Unlike the previous studies, the current study focuses on peptides obtained from food proteins and uses in silico enzymatic digestion, antimicrobial and antiviral peptide prediction, allergenicity assessment, physicochemical characterization, molecular docking, and molecular dynamics simulations to systematically screen for peptides targeting the PB1 subunit of the H5N1 polymerase. Thus, rather than considering the individual computational tools as novel, our approach combines established methods within a food-derived peptide discovery framework specifically designed for the identification and structural prioritization of potential H5N1 polymerase-targeting peptides. Because all results below are predictions without experimental validation, they are discussed as indicators for candidate prioritization and not as evidence of activity.

Step one of the pipeline revealed that the selected 19 proteins are characterized by large variability in terms of physicochemical features, varying in length from 178 to 605 residues and molecular mass equal to approximately 19.9 to 70.3 kDa. Generally, the proteins were hydrophilic, judging by the negative GRAVY values, but there is a hydrophobic protein, namely zein-alpha GZ19AB11. This variability should be taken into account in relation to the search for peptides because such properties as hydrophobicity, net charge, and amino acid composition influence solubility of the peptides and their ability to interact with cell membrane and proteins [58,59]. In general, the presence of hydrophilic proteins in the current dataset is consistent with the properties of different food proteins used in previous studies as precursors for peptides [60–62]. However, the high instability indices, found in most proteins, should be analyzed carefully because this characteristic is determined by the sequence of amino acids, but not the stability of protein in physiological conditions. Therefore, these properties are best interpreted as a description of the diversity of the sequence space and not as a selection criterion. Having established this diversity, the next question was which of these proteins is most likely to be sources of bioactive peptides.

The biological relevance of selected proteins was substantiated by earlier studies revealing the ability of food proteins to form peptides possessing antimicrobial, antiviral, and immunomodulating effects via enzymatic hydrolysis [63,64]. Milk proteins such as caseins and β-lactoglobulin are well characterized sources of bioactive peptides, and some of the caseins were found to be capable of antiviral activity against influenza viruses [65,66]. It is known that soybean and cereal storage proteins, such as glycinin and β-conglycinin, can serve as antimicrobial peptides. It was hypothesized that these proteins might have antiviral activity as well [67]. In the current study, legumin A (0.970), β-conglycinin α subunit (0.941), glutelin type-A 1 (0.911), legumin A2 (0.891), and glutelin type-A 2 (0.891) demonstrated the highest probabilities of antimicrobial activity. The obtained results confirm previous data about high bioactivity of peptides derived from soybean and cereal storage proteins [68]. On the contrary, the scores for all five milk proteins were equal to 0.000; however, they were described as biologically active previously. This may be due to the fact that the prediction was carried out on intact proteins while activity in dairy proteins is generally expressed as a result of hydrolysis and highlights the importance of studying the digestion products rather than the intact proteins. It is also important to consider that

having a high AMP probability does not imply that the protein or peptide is antiviral in nature. Antimicrobial activity prediction algorithms generally predict the presence of antimicrobial activity without discriminating among the activity being antibacterial, antifungal, or antiviral.

Moreover, the identified allergenicity indicated the necessity of application of several criteria by choosing peptides. The majority of the parent proteins possessed significant predicted similarity to known allergens, with the exception of Zein-alpha GZ19AB11, which possessed no allergenicity signal, whereas the allergenicity signal of Legumin B was slightly lower. The obtained results are mainly in agreement with the allergenicity of some milk, soybean, and egg proteins that are contained in the dataset [69]. At the same time, it should be noted that the predicted allergenicity of the parent protein should not be immediately applied to all its peptide derivatives. Proteolytic cleavage leads to the destruction of the required structures for IgE binding, whereas some peptide sequences preserve the needed structures (EFSA Panel on Genetically Modified Organisms) [70]. Concerning the hemolytic potential, the majority of the proteins had a low probability of being hemolytic. However, Glutelin-2 was the only protein that was predicted as hemolytic, while Zein-alpha GZ19AB11 was the most probable among the non-hemolytic proteins. The results are in agreement with the application of hemolysis as the first step in the computational peptides selection [40]. The above predictions cannot confirm the safety of the peptides that have been developed through the current research because the length, charge, hydrophobicity, and amphipathicity of the peptides are important considerations in this context. The prediction of the AMP, allergenicity, and hemolysis profiles led to the selection of the parent proteins for extraction of the peptides.

The simulation of proteolysis was used to connect parent proteins with possible peptides. Chymotrypsin cleavage simulation yielded several overlapping peptides thar were 8–21 amino acid residues long. In turn, the simulation of pepsin cleavage yielded only seven peptides. The appearance of peptide families is in agreement with previous observations that enzymatic hydrolysis of food proteins yields peptides with different length and biological characteristics [71]. Of special note, VSSVCENPI and SSAVSSVCENPI peptides were obtained via simulation of pepsin cleavage and further analyzed for their antiviral and structural features. The appearance of the latter two peptides in the result of simulation of pepsin cleavage implies that at least some peptides selected computationally can be formed during proteolysis resembling the gastrointestinal one. However, this observation does not mean that the peptides are resistant to gastrointestinal digestion or bioavailable since the in silico digestion does not reflect all the complexity of the real process. The previous investigations have demonstrated that peptides predicted or detected to have biological activity could continue to be processed via proteolysis, which affects their activity [72,73]. Once the pool of peptides was defined, the next task was to rank them according to predicted antiviral activity.

The machine learning approach allowed obtaining a quantitative score for peptide ranking. A number of sequences such as QPYRIQQAIAAGILPL, TQCSQAPIASLLPPY, SSAVSSVCENPILQPY, and RIQQAIAAGILPLSPL demonstrated relatively good predicted IC50 values obtained using various algorithms. Algorithms yielded dissimilar values from time to time. For example, the K-Star algorithm resulted in a very low value of the predicted IC50 equal to 0.05 μM for the sequence ALTNPAVF; however, the predicted values yielded using SVM, Random Forest, and IBk were much higher. The reason behind this lies in the high sensitivity of antiviral peptide predictions to the algorithm chosen and the training dataset as well as the features used by each algorithm [74,75]. Qureshi et al. [36] obtained Pearson correlation coefficients of 0.66, 0.64, 0.56, and 0.55 for SVM, Random Forest, IBk, and K-Star algorithms, respectively. It is not surprising that there are variations in the

current predictions; therefore, a consensus approach is vital. Moreover, other studies have reported that the prediction of antiviral peptides is improved when both the physical and evolutionary features of the sequence are used simultaneously, despite the impossibility of computational predictions to substitute the experimental assay [75–77]. Peptides supported by multiple models are therefore stronger computational candidates than peptides predicted by a single algorithm.

The platforms agreed on some features and disagreed on other features. Due to the differences in the methods of calculation of scores, which involve different scoring functions (HawkRank scoring after ATTRACT docking, electrostatic and desolvation energy for pyDockWEB, and shape complementarity along with ACE for PatchDock), the results are incomparable by absolute values but comparable by peptide ranking. First, the longer peptides were generally ranked higher than the shorter ones. The peptide TQCSQAPI received the lowest scores in any platform, and the peptide SSAVSSVCENPI had worse scores than its longer analog SSAVSSVCENPILQPY in PatchDock. Second, the peptides RIQQAIAAGILPLSPL and SSAVSSVCENPILQPY were ranked the highest in both HawkDock and pyDockWEB because they had the lowest energy value (−29.537). However, in PatchDock, TQCSQAPIASLLPPY was the best peptide, whereas RIQQAIAAGILPLSPL had the fourth score out of six peptides. None of the peptides scored first on all three platforms; however, SSAVSSVCENPILQPY was always in the top three of the four longer peptides on each of them. Together with its largest number of close contacts (29) and hydrogen bonds (4) presented below, it becomes the most reliable choice at the docking stage. Notably, peptides showing the best docking results turned out to have favorable predictions as well. This is consistent with the previously described computational methods, where activity predictions were combined with molecular docking for reduction of candidates for experiments [78]. This provides an additional level of validation of priority, even though such methods are not completely independent from each other. The better scores obtained for the longer peptides may be associated with the increased number of residues that can form intermolecular contacts. However, the peptide length cannot be regarded as the only criterion of binding affinity, since the longer peptides are highly flexible. Lastly, the intervals of scores for each peptide prove that the score is calculated based on particular docking result.

An analysis of the expected intermolecular interactions made it possible to more accurately differentiate the docked complexes. Complex SSAVSSVCENPILQPY showed the maximum number of close contacts (29) and four hydrogen bonds, while SSAVSSVCENPI and TQCSQAPIASLLPPY showed 26 close contacts each. Generally speaking, a great number of hydrogen bonds along with close polar and non-polar contacts is characteristic of peptide–protein complexes [79]. It should be noted that the peptide with the maximum number of close contacts was not characterized by the best docking score in HawkDock. Therefore, the number of close contacts does not influence the docking energy of the peptide, and it is necessary to conduct more detailed calculations to distinguish between the alternatives. Some steric clashes in the predicted complexes, especially in complex SSAVSSVCENPI, proved the need to check for geometric correctness of the predicted complex structure. Since docking and contact analysis describe the static pose of the molecule, molecular dynamics simulation was used to check whether these structures are stable overtime.

The results from the MD simulations (200.4 ns) gave an insight into the dynamic nature of the complexes simulated. Some of the complexes (SSAVSSVCENPI and TQCSQAPIASLLPPY) underwent an initial conformational adjustment, after which they experienced relatively stable dynamics in terms of RMSD, radius of gyration (Rg), and solvent-accessible surface area (SASA), whereas SSAVSSVCENPILQPY exhibited a stable hydrogen bonding

pattern, Rg, and SASA in the latter parts of the simulation. These findings agree with previous studies that stable hydrogen bonding patterns, stable molecular compactness, and small structural fluctuation indicate stable interaction between peptides and proteins [40]. However, fluctuations in RMSD and RMSF do not automatically suggest instability given that conformational changes within proteins and peptides naturally take place during the simulation [80]. The stability is only valid for the simulation period and without MM/PBSA calculation, it is not possible to determine any binding free energy.

In terms of predictions, the selected peptides could, in theory, block the interaction between *ANP32B* and the polymerase through their binding at the interaction site, thus preventing the viral polymerase from using this host protein. Nevertheless, since ANP32B is a host protein taking part in cell functioning, the peptides can influence the physiologically normal protein–protein interactions with *ANP32B* [81]. In other words, the current computational strategy does not provide selective inhibition of *ANP32B* functionality but rather predicts the possibility that the interaction between the peptides and the interaction site of the host–virus can disturb this particular virus–host interaction. Therefore, the predicted binding affinity and interactions are indicative and require further validation.

5. Conclusions

The study investigated 19 proteins obtained from soybean, cow milk, egg, pea, maize, rice, and cowpea as peptide sources from foods that can target the *PB1* subunit of the H5N1 influenza virus polymerase. Protein-level analysis indicated that the majority of the proteins tested were hydrophilic and predicted to have high allergenicity signals, except for Legumin B (weak prediction) and Zein-alpha GZ19AB11 (no prediction). Twelve out of 19 proteins were predicted to have antimicrobial activity, with the top predictions belonging to Legumin A (0.970) and the β-conglycinin α subunit (0.941) and the milk proteins having zero predicted antimicrobial activity. The only protein predicted to have hemolytic properties was Glutelin-2 (0.545). Simulated proteolysis using pepsin and chymotrypsin yielded overlapping peptides consisting of eight to 21 residues. These peptides were ranked using four different machine learning algorithms and prioritized according to their consensus ranking, probable gastrointestinal release, and non-hemolytic profile. *PB1* was selected as the docking receptor because it was the only polymerase subunit with 90% or more of its residues in the favored regions of the Ramachandran plot. The main result is a multi-step approach that reduced 19 proteins to a few candidate peptides. RIQQAIAAGILPLSPL, SSAVSSVCENPILQPY, and QPYRIQQAIAAGILPL appeared to be the peptides with the most convincing computational data (good predicted IC50 values and good docking to *PB1*). The three platforms did not give the same ranking of the peptides; thus, none of the peptides was superior on all three platforms. Among the three, SSAVSSVCENPILQPY showed the most consistency. It made the largest number of close contacts (29) and formed four hydrogen bonds in the docked pose; moreover, its hydrogen bond network, radius of gyration, and solvent-accessible surface area remained stable during the latter part of the simulation. All results presented are predictions only, and they must be taken as hypotheses. It should be mentioned that the present study is entirely based on computational approaches, and there was no experimental validation of antiviral activity, binding affinity, toxicity, hemolysis, allergenicity, or gastrointestinal stability of the obtained peptides. Prediction models and docking scores have limitations, and the molecular dynamics simulations were carried out under computational conditions and without MM/PBSA free energy calculation. Synthesis and in vitro testing of these peptides are the next steps needed.

Author Contributions: Conceptualization, T.A.; methodology, P.E.o.; software, P.D.D.D.; validation, N.H.S.; formal analysis, N.A.-H.; investigation, L.A.A.; resources, T.A.; data curation, K.A.;

writing—original draft preparation, L.A.A.; writing—review and editing, P.E.o. and P.D.D.D.; visualization, F.A.-A.; supervision, T.A.; project administration, T.A.; funding acquisition, T.A. All authors have read and agreed to the published version of the manuscript.

Funding: The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA for funding this research work through the project number NBU-FFR-2026-2273-11. The authors are thankful to Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2026R924), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia (Grant No. KFU265070).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

AMP	Antimicrobial Peptide
ANP32B	Acidic Leucine-Rich Nuclear Phosphoprotein 32 Family Member B
AVP	Antiviral Peptide
CHARMM	Chemistry at HARvard Macromolecular Mechanics
COCOMAPS	Contacts of Structural Units in Proteins Mapped by an Analytical Tool
GRAVY	Grand Average of Hydropathy
H5N1	Hemagglutinin 5/Neuraminidase 1 Influenza A Virus
IC50	Half-Maximal Inhibitory Concentration
MD	Molecular Dynamics
MM/PBSA	Molecular Mechanics/Poisson–Boltzmann Surface Area
PDB	Protein Data Bank
PA	Polymerase Acidic Protein
PB1	Polymerase Basic Protein 1
PB2	Polymerase Basic Protein 2
RMSD	Root Mean Square Deviation

References

1. Yoo, S.J.; Kwon, T.; Lyoo, Y.S. Challenges of Influenza A Viruses in Humans and Animals and Current Animal Vaccines as an Effective Control Measure. *Clin. Exp. Vaccine Res.* **2018**, *7*, 1–15. [CrossRef] [PubMed]
2. de Jong, M.D.; Hien, T.T. Avian Influenza A (H5N1). *J. Clin. Virol.* **2006**, *35*, 2–13. [CrossRef] [PubMed]
3. Kamel, M.; Aleya, S.; Almagharbeh, W.T.; Aleya, L.; Abdel-Daim, M.M. The Emergence of Highly Pathogenic Avian Influenza H5N1 in Dairy Cattle: Implications for Public Health, Animal Health, and Pandemic Preparedness. *Eur. J. Clin. Microbiol. Infect. Dis.* **2025**, *44*, 1817–1833. [CrossRef] [PubMed]
4. Simancas-Racines, A.; Reytor-González, C.; Toral, M.; Simancas-Racines, D. H5N1 Avian Influenza: A Narrative Review of Scientific Advances and Global Policy Challenges. *Viruses* **2025**, *17*, 927. [CrossRef] [PubMed]
5. Zhang, L.; Lai, Y.; Cui, Y.; Yang, Q.; Shao, Y.; Ding, S.; Wang, H.; Wang, L.; Gao, G.F.; Deng, T. Emergence of Mammalian-Adaptive PB2 Mutations Enhances Polymerase Activity and Pathogenicity of Cattle-Derived H5N1 Influenza A Virus. *Nat. Commun.* **2025**, *17*, 1011. [CrossRef] [PubMed]
6. Bruno, L.; Nappo, M.A.; Frontoso, R.; Montinaro, S.; Di Lecce, R.; Guarnieri, C.; Ferrari, L.; Corradi, A. Avian Influenza Viruses: Global Panzootic, Host Range Expansion and Emerging One-Health Threats. *Vet. Sci.* **2026**, *13*, 67. [CrossRef] [PubMed]
7. Sánchez-Rodríguez, F.; Diaz-Gavidia, C.; Ruíz, S.; Jimenez-Bluhm, P. Highly Pathogenic Avian Influenza H5N1 in South America, 2022–2025: Spread, Affected Species, and Southward Expansion into the Antarctic Region. *Viruses* **2026**, *18*, 764. [CrossRef] [PubMed]

8. Galli, M.; Giacomelli, A.; Lai, A.; Zehender, G. H5N1 Influenza A Virus: Lessons from Past Outbreaks and Emerging Threats. *Infez. Med.* **2025**, *33*, 76–89. [CrossRef] [PubMed]
9. Billias, N.; D'Alessandro, V.; Pouliopoulou, D.V.; Wong, J.J.; Miller, E.; Hopkins, J.P.; Boutsikari, E.C.; Cipriano, L.E.; da Veiga Pereira, T.; Johnstone, J.; et al. Mapping Reported Modes of Transmission of Highly Pathogenic Avian Influenza A (H5N1) to Humans: A Scoping Review. *One Health* **2026**, *23*, 101492. [CrossRef] [PubMed]
10. Milazzo, A.; Liu, J.; Multani, P.; Steele, S.; Hoon, E.; Chaber, A.L. One Health Implementation: A Systematic Scoping Review Using the Quadripartite One Health Joint Plan of Action. *One Health* **2025**, *20*, 101008. Correction in *One Health* **2026**, *22*, 101423. [CrossRef] [PubMed]
11. El-Bidawy, M.H.; Mohammad, I.; Ansari, M.R.; Hajelbashir, M.I.; Khan, M.S.; Poyil, M.M.; Bari, M.N.; Arafah, A.M.R.; Kamal, M.A.; Ahsan, S.T.M. Highly Pathogenic Avian Influenza: Tracking the Progression from IAV (H5N1) to IAV (H7N9) and Preparing for Emerging Challenges. *Microorganisms* **2026**, *14*, 12. [CrossRef] [PubMed]
12. Hui, X.; Ding, S.; Gao, S.; Xu, S.; Zhao, T.; Tian, X.; Wang, H. Host-Directed Antiviral Strategies Against Influenza Viruses: Host Targets, Multi-Omics Approaches and AI-Assisted Discovery. *Vet. Sci.* **2026**, *13*, 626. [CrossRef] [PubMed]
13. Kawaguchi, A.; Naito, T.; Nagata, K. Involvement of Influenza Virus PA Subunit in Assembly of Functional RNA Polymerase Complexes. *J. Virol.* **2005**, *79*, 732–744. [CrossRef] [PubMed]
14. Carter, T.; Iqbal, M. The Influenza A Virus Replication Cycle: A Comprehensive Review. *Viruses* **2024**, *16*, 316. [CrossRef] [PubMed]
15. Fislová, T.; Thomas, B.; Graef, K.M.; Fodor, E. Association of the Influenza Virus RNA Polymerase Subunit PB2 with the Host Chaperonin CCT. *J. Virol.* **2010**, *84*, 8691–8699. [CrossRef] [PubMed]
16. Llompart, C.M.; Nieto, A.; Rodriguez-Frandsen, A. Specific Residues of PB2 and PA Influenza Virus Polymerase Subunits Confer the Ability for RNA Polymerase II Degradation and Virus Pathogenicity in Mice. *J. Virol.* **2014**, *88*, 3455–3463. [CrossRef] [PubMed]
17. Zhang, H.; Zhang, Z.; Wang, Y.; Wang, M.; Wang, X.; Zhang, X.; Ji, S.; Du, C.; Chen, H.; Wang, X. Fundamental Contribution and Host Range Determination of ANP32A and ANP32B in Influenza A Virus Polymerase Activity. *J. Virol.* **2019**, *93*, e00174-19. [CrossRef] [PubMed]
18. Peacock, T.P.; Sheppard, C.M.; Lister, M.G.; Staller, E.; Frise, R.; Swann, O.C.; Goldhill, D.H.; Long, J.S.; Barclay, W.S. Mammalian ANP32A and ANP32B Proteins Drive Differential Polymerase Adaptations in Avian Influenza Virus. *J. Virol.* **2023**, *97*, e0021323. [CrossRef] [PubMed]
19. Budt, M.; Barac, I.; Kohs, J.; Krischuns, T.; Naffakh, N.; Wolff, T. Historical Pandemic and Contemporary Influenza A Viruses Reveal PB2 M631L as a Convergent Adaptation to Human ANP32. *Microorganisms* **2026**, *14*, 859. [CrossRef] [PubMed]
20. Cui, S.; Hou, X.; Pu, S.; Luo, J.; Zhu, H.; He, P.; Gao, C.; Gao, R.; Han, J. Identification and Characterization of PB2 Mutations Associated with Mammalian Adaptation of Highly Pathogenic H5N1 Avian Influenza Viruses. *Front. Microbiol.* **2026**, *17*, 1867604. [CrossRef] [PubMed]
21. Staller, E.; Carrique, L.; Swann, O.C.; Fan, H.; Keown, J.R.; Sheppard, C.M.; Barclay, W.S.; Grimes, J.M.; Fodor, E. Structures of H5N1 Influenza Polymerase with ANP32B Reveal Mechanisms of Genome Replication and Host Adaptation. *Nat. Commun.* **2024**, *15*, 4123. [CrossRef] [PubMed]
22. Bi, Z. Adaptation Differences and Mechanisms of Influenza Viruses to ANP32 Proteins Across Species. *J. Virol.* **2026**, *100*, e0190025. [CrossRef] [PubMed]
23. Pleschka, S. Special Issue "Host Targeted Therapeutics Against Virus Infections". *Viruses* **2024**, *16*, 1825. [CrossRef] [PubMed]
24. Idrees, S.; Chen, H.; Panth, N.; Paudel, K.R.; Hansbro, P.M. Exploring Viral-Host Protein Interactions as Antiviral Therapies: A Computational Perspective. *Microorganisms* **2024**, *12*, 630. [CrossRef] [PubMed]
25. Cao, R.; Guo, F.; Huang, T.; Zhang, Y.; Chen, Y.; Yuan, J.; Fu, T.; Wang, Z.; Liu, D. Ubiquitin-Dependent Regulation of Influenza A Virus Polymerase and vRNP Function: Mechanisms and Therapeutic Opportunities. *Microorganisms* **2026**, *14*, 1684. [CrossRef] [PubMed]
26. Patel, H.; Kukol, A. Harnessing Viral Internal Proteins to Combat Flu and Beyond. *Virology* **2025**, *604*, 110414. [CrossRef] [PubMed]
27. Chen, H.; Sun, J.; Fang, H.; Lin, Y.; Wu, H.; Lin, D.; Yang, Z.; Zhou, Q.; Zhao, B.; Zhou, T.; et al. Food-Derived Bioactive Peptides: Health Benefits, Structure–Activity Relationships, and Translational Prospects. *J. Zhejiang Univ. Sci. B* **2025**, *26*, 1037–1058. [CrossRef] [PubMed]
28. Machado, M.; Bautista-Hérnandez, I.; Gómez-García, R.; Silva, S.; Costa, E.M. Bioactive Food Proteins: Bridging Nutritional and Functional Benefits with Sustainable Protein Sources. *Foods* **2025**, *14*, 3035. [CrossRef] [PubMed]
29. Karami, Z.; Akbari-Adergani, B. Bioactive Food Derived Peptides: A Review on Correlation between Structure of Bioactive Peptides and Their Functional Properties. *J. Food Sci. Technol.* **2019**, *56*, 535–547. [CrossRef] [PubMed]
30. Agoni, C.; Fernández-Díaz, R.; Timmons, P.B.; Adelfio, A.; Gómez, H.; Shields, D.C. Molecular Modelling in Bioactive Peptide Discovery and Characterisation. *Biomolecules* **2025**, *15*, 524. [CrossRef] [PubMed]

31. Liu, X.; Guan, F.; Luo, H.; Yao, B.; Tian, J. Artificial Intelligence-Driven Discovery of Bioactive Peptides: Computational Approaches and Future Perspectives. *aBIOTECH* **2025**, *7*, 100014. [CrossRef] [PubMed]
32. Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, I.; Appel, R.D.; Bairoch, A. ExPASy: The Proteomics Server for In-Depth Protein Knowledge and Analysis. *Nucleic Acids Res.* **2003**, *31*, 3784–3788. [CrossRef] [PubMed]
33. Maurer-Stroh, S.; Krutz, N.L.; Kern, P.S.; Gunalan, V.; Nguyen, M.N.; Limviphuvadh, V.; Eisenhaber, F.; Gerberick, G.F. AllerCatPro—Prediction of Protein Allergenicity Potential from the Protein Sequence. *Bioinformatics* **2019**, *35*, 3020–3027. [CrossRef] [PubMed]
34. Santos-Júnior, C.D.; Pan, S.; Zhao, X.M.; Coelho, L.P. Macrel: Antimicrobial Peptide Screening in Genomes and Metagenomes. *PeerJ* **2020**, *8*, e10555. [CrossRef] [PubMed]
35. Sajjad, A.; Haq, I.; Syed, R.; Anwar, F.; Hamza, M.; Musharaf, M.; Kiani, T.; Nouroz, F. In-Silico Molecular Analysis and Blocking of the Viral G Protein of Nipah Virus Interacting with Ephrin B2 and B3 Receptor by Using Peptide Mass Fingerprinting. *Front. Bioinform.* **2025**, *5*, 1526566. [CrossRef] [PubMed]
36. Qureshi, A.; Tandon, H.; Kumar, M. AVP-IC50 Pred: Multiple Machine Learning Techniques-Based Prediction of Peptide Antiviral Activity in Terms of Half Maximal Inhibitory Concentration (IC50). *Biopolymers* **2015**, *104*, 753–763. [CrossRef] [PubMed]
37. Dave, L.A.; Montoya, C.A.; Rutherfurd, S.M.; Moughan, P.J. Gastrointestinal Endogenous Proteins as a Source of Bioactive Peptides—An In Silico Study. *PLoS ONE* **2014**, *9*, e98922. [CrossRef] [PubMed]
38. Charoenkwan, P.; Anuwongcharoen, N.; Nantasenamat, C.; Hasan, M.M.; Shoombuatong, W. In Silico Approaches for the Prediction and Analysis of Antiviral Peptides: A Review. *Curr. Pharm. Des.* **2021**, *27*, 2180–2188. [CrossRef] [PubMed]
39. Lefin, N.; Herrera-Belén, L.; Farias, J.G.; Beltrán, J.F. Review and Perspective on Bioinformatics Tools Using Machine Learning and Deep Learning for Predicting Antiviral Peptides. *Mol. Divers.* **2024**, *28*, 2365–2374. [CrossRef] [PubMed]
40. Rathore, A.S.; Kumar, N.; Choudhury, S.; Mehta, N.K.; Raghava, G.P.S. Prediction of Hemolytic Peptides and Their Hemolytic Concentration. *Commun. Biol.* **2025**, *8*, 176. [CrossRef] [PubMed]
41. Hasan, M.M.; Schaduangrat, N.; Basith, S.; Lee, G.; Shoombuatong, W.; Manavalan, B. HLPpred-Fuse: Improved and Robust Prediction of Hemolytic Peptide and Its Activity by Fusing Multiple Feature Representation. *Bioinformatics* **2020**, *36*, 3350–3356. [CrossRef] [PubMed]
42. Buchan, D.W.A.; Moffat, L.; Lau, A.; Kandathil, S.M.; Jones, D.T. Deep Learning for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res.* **2024**, *52*, W287–W293. [CrossRef] [PubMed]
43. Wiederstein, M.; Sippl, M.J. ProSA-Web: Interactive Web Service for the Recognition of Errors in Three-Dimensional Structures of Proteins. *Nucleic Acids Res.* **2007**, *35*, W407–W410. [CrossRef] [PubMed]
44. Owona, P.E.; Mengue Ngadena, Y.S.; Bidingha, A.G.R.; Bilanda, D.C.; Ngo Song, M.C.; Dzeufiet Djomeni, P.D.D.; Rania Ali El Hadi, M.; Alwethaynani, M.S.; Al-Hoshani, N.; Binsaleh, A.Y.; et al. Deciphering the Effects of *Pterocarpus mildbraedii* and *Pterocarpus soyauxii* against SARS-CoV-2 and HIV Coinfection through the Integration of ADMET, Network Pharmacology, In Silico Molecular Docking, Pharmacophore Analysis and Computational Inhibition of CDK2 and HIV-1 Viral Proteins. *J. Comput. Biophys. Chem.* **2026**, *25*, 2855–2896. [CrossRef]
45. Owona, P.E.; Mengue Ngadena, Y.S.; Melogmo Dongmo, Y.K.; Bidingha, R.A.G.; Tamafo Fouegue, A.D.; Ambani Omgba, J.J.; Bilanda, D.C.; Dzeufiet Djomeni, P.; Binsaleh, A.Y.; Al-Hoshani, N.; et al. Phytoestrogens 7-O-Acetylformonetin and Khrinone A from *Pterocarpus soyauxii* Taub. Inhibit the Human Papillomavirus Replication: An Integrative Immunological, Chemical, and Computational Approach to Cervical Cancer Prevention. *Chin. J. Anal. Chem.* **2026**, *54*, 100678. [CrossRef]
46. Emmanuel, O.P.; Sandrine, M.N.Y.; Claude, B.D.; Ogunlakin, A.D.; Ronald, B.A.G.; Desire, D.D.P.; Aziz, T.; Aljumaa, M.A.; Alkhatib, S.N.; Sagini, H.A. Integrative Immunoinformatics and Structural Modeling for the Rational Design of a Multi-Epitope Vaccine Candidate against Human Cytomegalovirus. *Sci. Rep.* **2026**, *16*, 22402. [CrossRef] [PubMed]
47. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [CrossRef] [PubMed]
48. Chen, V.B.; Arendall, W.B., III; Headd, J.J.; Keedy, D.A.; Immormino, R.M.; Kapral, G.J.; Murray, L.W.; Richardson, J.S.; Richardson, D.C. MolProbity: All-Atom Structure Validation for Macromolecular Crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 12–21. [CrossRef] [PubMed]
49. Weng, G.; Wang, E.; Wang, Z.; Liu, H.; Zhu, F.; Li, D.; Hou, T. HawkDock: A Web Server to Predict and Analyze the Protein–Protein Complex Based on Computational Docking and MM/GBSA. *Nucleic Acids Res.* **2019**, *47*, W322–W330. [CrossRef] [PubMed]
50. Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R.; Wolfson, H.J. PatchDock and SymmDock: Servers for Rigid and Symmetric Docking. *Nucleic Acids Res.* **2005**, *33*, W363–W367. [CrossRef] [PubMed]
51. Jiménez-García, B.; Pons, C.; Fernández-Recio, J. pyDockWEB: A Web Server for Rigid-Body Protein–Protein Docking Using Electrostatics and Desolvation Energies. *Bioinformatics* **2013**, *29*, 1698–1699. [CrossRef] [PubMed]
52. Chawla, M.; Kalra, U.; Petta, A.; Sharma, S.; Shaikh, A.R.; Cavallo, L.; Oliva, R. COCOMAPS 2.0: A Web Server for Identifying, Analyzing, and Visualizing Atomic Interactions at the Interface of Biomolecular Complexes. *Bioinformatics* **2025**, *41*, btaf606. [CrossRef] [PubMed]

53. Alghamdi, A.; Abouzied, A.S.; Alamri, A.; Anwar, S.; Ansari, M.; Khadra, I.; Zaki, Y.H.; Gomha, S.M. Synthesis, Molecular Docking, and Dynamic Simulation Targeting Main Protease (Mpro) of New, Thiazole Clubbed Pyridine Scaffolds as Potential COVID-19 Inhibitors. *Curr. Issues Mol. Biol.* **2023**, *45*, 1422–1442. [CrossRef] [PubMed]
54. Alzain, A.A.; Almogaddam, M.A.; Yousif, R.; Alqarni, M.H.; Foudah, A.I.; Osman, W.; Elamin, K.M.; Mohamed, H.M.; Moglad, E.; Ashour, A.; et al. Molecular Docking, Molecular Dynamics Simulation, and Pharmacophore-Based Virtual Screening Unveil Natural Compounds with TIM-3 Inhibitory Activity. *J. Pharm. Bioallied Sci.* **2025**, *17*, S1882–S1887. [CrossRef] [PubMed]
55. Tan, J.; Huang, S.; Wu, D.; Zhao, Z.; Zhao, Y.; Fu, Y.; Wu, W. AI-Driven Rapid Screening and Characterization of Dipeptidyl Peptidase-IV (DPP-IV) Inhibitory Peptides from Goat Blood Proteins: An Integrative In Silico and Experimental Strategy. *Foods* **2026**, *15*, 398. [CrossRef] [PubMed]
56. Singh, B.P.; Paul, S.; Ferreira-Santos, P.; Bhushan, B.; Hati, S.; Goel, G.; Udenigwe, C. In Silico and Molecular Docking Approaches in Food-Derived Bioactive Peptide Discovery: Trends, Challenges, and Prospects. *Food Res. Int.* **2025**, *222*, 117637. [CrossRef] [PubMed]
57. Zekri, A.; Ouassaf, M.; Khan, S.U.; Rengasamy, K.R.R.; Alhatlani, B.Y. Discovery of Novel Natural Inhibitors of H5N1 Neuraminidase Using Integrated Molecular Modeling and ADMET Prediction. *Bioengineering* **2025**, *12*, 622. [CrossRef] [PubMed]
58. Kovacs, J.M.; Mant, C.T.; Hodges, R.S. Determination of Intrinsic Hydrophilicity/Hydrophobicity of Amino Acid Side Chains in Peptides in the Absence of Nearest-Neighbor or Conformational Effects. *Biopolymers* **2006**, *84*, 283–297. [CrossRef] [PubMed]
59. Terziyska, M.; Terziyski, Z.; Ilieva, I.; Bozhkov, S.; Vladev, V. Knowledge Discovery from Bioactive Peptide Data in the PepLab Database Through Quantitative Analysis and Machine Learning. *Sci* **2025**, *7*, 122. [CrossRef]
60. Park, Y.W.; Nam, M.S. Bioactive Peptides in Milk and Dairy Products: A Review. *Korean J. Food Sci. Anim. Resour.* **2015**, *35*, 831–840. [CrossRef] [PubMed]
61. Peres Fabbri, L.; Cavallero, A.; Vidotto, F.; Gabriele, M. Bioactive Peptides from Fermented Foods: Production Approaches, Sources, and Potential Health Benefits. *Foods* **2024**, *13*, 3369. [CrossRef] [PubMed]
62. Magliery, T.J. Protein Stability: Computation, Sequence Statistics, and New Experimental Methods. *Curr. Opin. Struct. Biol.* **2015**, *33*, 161–168. [CrossRef] [PubMed]
63. Corrêa, J.A.F.; de Melo Nazareth, T.; Rocha, G.F.D.; Luciano, F.B. Bioactive Antimicrobial Peptides from Food Proteins: Perspectives and Challenges for Controlling Foodborne Pathogens. *Pathogens* **2023**, *12*, 477. [CrossRef] [PubMed]
64. Colón-Sandoval, A.; Contreras-Angulo, L.A.; Cabanillas-Bojórquez, L.A.; Gutiérrez-Grijalva, E.P.; León-Félix, J.; Leyva-López, N.; López-Martínez, L.X.; García-Cebreros, M.D.; Heredia, J.B. Peptides from General By-Products: Unveiling Their Potential Biological Activities in Human Health. *Molecules* **2025**, *30*, 4821. [CrossRef] [PubMed]
65. Santos, I.; Silva, M.; Grácio, M.; Pedroso, L.; Lima, A. Milk Antiviral Proteins and Derived Peptides against Zoonoses. *Int. J. Mol. Sci.* **2024**, *25*, 1842. [CrossRef] [PubMed]
66. Meleti, E.; Koureas, M.; Manouras, A.; Giannouli, P.; Malissiova, E. Bioactive Peptides from Dairy Products: A Systematic Review of Advances, Mechanisms, Benefits, and Functional Potential. *Dairy* **2025**, *6*, 65. [CrossRef]
67. Zhu, Y.; Chen, G.; Diao, J.; Wang, C. Recent Advances in Exploring and Exploiting Soybean Functional Peptides—A Review. *Front. Nutr.* **2023**, *10*, 1185047. [CrossRef] [PubMed]
68. Chatterjee, C.; Gleddie, S.; Xiao, C.W. Soybean Bioactive Peptides and Their Functional Properties. *Nutrients* **2018**, *10*, 1211. [CrossRef] [PubMed]
69. Dos Santos, A.L.S.; Dos Santos, P.P.B.; de Almeida Amaral, G.; Soares, E.C.; de Oliveira e Silva, C.A.; de Souza, S.V.C. Effect of Thermal Processing on the Antigenicity of Allergenic Milk, Egg and Soy Proteins. *J. Food Sci. Technol.* **2022**, *59*, 2617–2628. [CrossRef] [PubMed]
70. EFSA Panel on Genetically Modified Organisms (GMO); Mullins, E.; Bresson, J.L.; Dalmay, T.; Dewhurst, I.C.; Epstein, M.M.; George Firbank, L.; Guerche, P.; Hejatko, J.; Naegeli, H.; et al. Scientific Opinion on Development Needs for the Allergenicity and Protein Safety Assessment of Food and Feed Products Derived from Biotechnology. *EFSA J.* **2022**, *20*, e07044. [CrossRef] [PubMed]
71. Akbarian, M.; Khani, A.; Eghbalpour, S.; Uversky, V.N. Bioactive Peptides: Synthesis, Sources, Applications, and Proposed Mechanisms of Action. *Int. J. Mol. Sci.* **2022**, *23*, 1445. [CrossRef] [PubMed]
72. Miguéns-Gómez, A.; Grau-Bové, C.; Sierra-Cruz, M.; Jorba-Martín, R.; Caro, A.; Rodríguez-Gallego, E.; Beltrán-Debón, R.; Blay, M.T.; Terra, X.; Ardévol, A.; et al. Gastrointestinally Digested Protein from the Insect *Alphitobius diaperinus* Stimulates a Different Intestinal Secretome than Beef or Almond, Producing a Differential Response in Food Intake in Rats. *Nutrients* **2020**, *12*, 2366. [CrossRef] [PubMed]
73. Pei, J.; Gao, X.; Pan, D.; Hua, Y.; He, J.; Liu, Z.; Dang, Y. Advances in the Stability Challenges of Bioactive Peptides and Improvement Strategies. *Curr. Res. Food Sci.* **2022**, *5*, 2162–2170. [CrossRef] [PubMed]
74. Ali, F.; Kumar, H.; Alghamdi, W.; Kateb, F.A.; Alarfaj, F.K. Recent Advances in Machine Learning-Based Models for Prediction of Antiviral Peptides. *Arch. Comput. Methods Eng.* **2023**, *30*, 4033–4044. [CrossRef] [PubMed]

75. Hargrove, H.; Tong, B.; Elkabanny, A.H.; Zhang, X.F. Machine Learning in Preclinical Development of Antiviral Peptide Candidates: A Review of the Current Landscape. *Viruses* **2026**, *18*, 260. [CrossRef] [PubMed]
76. Chowdhury, A.S.; Reehl, S.M.; Kehn-Hall, K.; Bishop, B.; Webb-Robertson, B.M. Better Understanding and Prediction of Antiviral Peptides through Primary and Secondary Structure Feature Importance. *Sci. Rep.* **2020**, *10*, 19260. [CrossRef] [PubMed]
77. Yousaf, W.; Haseeb, A.; Shen, Y.; Li, H.; Fan, K.; Sun, N.; Sun, P.; Sun, Y.; Yang, H.; Yin, W.; et al. Data-Driven Discovery of Antiviral Peptides against PRRSV Using Multiple Machine Learning Models. *Front. Vet. Sci.* **2025**, *12*, 1681083. [CrossRef] [PubMed]
78. Rahmati, S.; Zandi, F.; Ahmadi, K.; Adeli, A.; Rastegarpanah, N.; Amanlou, M.; Vaziri, B. Computational Structure-Based Design of Antiviral Peptides as Potential Protein-Protein Interaction Inhibitors of Rabies Virus Phosphoprotein and Human LC8. *Heliyon* **2024**, *11*, e41520. [CrossRef] [PubMed]
79. Ferreira de Freitas, R.; Schapira, M. A Systematic Analysis of Atomic Protein–Ligand Interactions in the PDB. *MedChemComm* **2017**, *8*, 1970–1981. [CrossRef] [PubMed]
80. Wolff, A.M.; Nango, E.; Young, I.D.; Brewster, A.S.; Kubo, M.; Nomura, T.; Sugahara, M.; Owada, S.; Barad, B.A.; Ito, K.; et al. Mapping Protein Dynamics at High Spatial Resolution with Temperature-Jump X-Ray Crystallography. *Nat. Chem.* **2023**, *15*, 1549–1558. [CrossRef] [PubMed]
81. Yang, S.; Zhou, L.; Reilly, P.T.; Shen, S.M.; He, P.; Zhu, X.N.; Li, C.X.; Wang, L.S.; Mak, T.W.; Chen, G.Q.; et al. ANP32B Deficiency Impairs Proliferation and Suppresses Tumor Progression by Regulating AKT Phosphorylation. *Cell Death Dis.* **2016**, *7*, e2082. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.