

Article

Prediction of African Swine Fever Virus Inhibitors by Molecular Docking-Driven Machine Learning Models

Jiwon Choi ^{1,2,*}, Jun Seop Yun ^{1,†}, Hyeon Song ¹, Yong-Keol Shin ³, Young-Hoon Kang ³, Palinda Ruvan Munashingha ³, Jeongyeon Yoon ³, Nam Hee Kim ¹, Hyun Sil Kim ¹, Jong In Yook ^{1,2}, Dongseob Tark ⁴, Yun-Sook Lim ^{5,*} and Soon B. Hwang ^{5,6}

¹ Department of Oral Pathology, Oral Cancer Research Institute, Yonsei University College of Dentistry, Seoul 03722, Korea; YJS8714@yuhs.ac (J.S.Y.); 0212HEA@yuhs.ac (H.S.); MIGO77@yuhs.ac (N.H.K.); KHS@yuhs.ac (H.S.K.); JIYOOK@yuhs.ac (J.I.Y.)

² Met Life Sciences Co. Ltd., Seoul 03722, Korea

³ Enzynomics Co. Ltd., Yuseong-gu, Daejeon 34050, Korea; ykshin@enzynomics.com (Y.-K.S.); yhkang@enzynomics.com (Y.-H.K.); mpruvan@enzynomics.com (P.R.M.); jyyoon@enzynomics.com (J.Y.)

⁴ Laboratory for Infectious Disease Prevention, Korea Zoonosis Research Institute, Jeonbuk National University, Iksan 54596, Korea; tarkds@jbnu.ac.kr

⁵ Laboratory of RNA Viral Diseases, Korea Zoonosis Research Institute, Jeonbuk National University, Iksan 54596, Korea; sbhwang@jbnu.ac.kr

⁶ Ilsong Institute of Life Science, Hallym University, Seoul 03722, Korea

* Correspondence: edccjw@gmail.com (J.C.); yslim@jbnu.ac.kr or yunsolim@hanmail.net (Y.-S.L.)

† Jiwon Choi and Jun Seop Yun contributed equally to this work.



Citation: Choi, J.; Yun, J.S.; Song, H.; Shin, Y.-K.; Kang, Y.-H.; Munashingha, P.R.; Yoon, J.; Kim, N.H.; Kim, H.S.; Yook, J.I.; et al. Prediction of African Swine Fever Virus Inhibitors by Molecular Docking-Driven Machine Learning Models. *Molecules* **2021**, *26*, 3592. <https://doi.org/10.3390/molecules26123592>

Academic Editor: Isabel Iriepa Canalda

Received: 17 April 2021

Accepted: 9 June 2021

Published: 11 June 2021

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



Copyright: © 2021 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 (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: African swine fever virus (ASFV) causes a highly contagious and severe hemorrhagic viral disease with high mortality in domestic pigs of all ages. Although the virus is harmless to humans, the ongoing ASFV epidemic could have severe economic consequences for global food security. Recent studies have found a few antiviral agents that can inhibit ASFV infections. However, currently, there are no vaccines or antiviral drugs. Hence, there is an urgent need to identify new drugs to treat ASFV. Based on the structural information data on the targets of ASFV, we used molecular docking and machine learning models to identify novel antiviral agents. We confirmed that compounds with high affinity present in the region of interest belonged to subsets in the chemical space using principal component analysis and *k*-means clustering in molecular docking studies of FDA-approved drugs. These methods predicted pentagastrin as a potential antiviral drug against ASFVs. Finally, it was also observed that the compound had an inhibitory effect on *Asfv*PolX activity. Results from the present study suggest that molecular docking and machine learning models can play an important role in identifying potential antiviral drugs against ASFVs.

Keywords: African swine fever virus; antiviral; molecular docking; machine learning

1. Introduction

African swine fever (ASFV) is an enveloped and double-stranded DNA virus belonging to the *Asfarviridae* family (genus *Asfivirus*), which contains a genome measuring 170–193 kbp, and replicates predominantly in the cytoplasm of macrophages [1]. The virus has a complex structure and a genome that mainly replicates in the cytoplasm of infected cells [2]. It has resulted in high mortality rates in pigs and is responsible for serious economic and production losses worldwide [3,4]. Recently, ASFV outbreaks reported in EU countries and Transcaucasia, especially in East Asian countries, have resulted in the culling of over one million pigs [5–7]. Vaccines and antiviral drugs can stop the spread of infection, but efforts to develop effective vaccines against ASFV have failed [8]. Therefore, there is a need for novel effective drugs and vaccines against ASFV. Currently, no vaccine or treatment is available against ASFV infection. Recent attempts to generate DNA vaccines, vectored vaccines, or live attenuated vaccines by defined gene deletions have yielded

encouraging data through in vitro assays [9–13]. However, there are still major challenges with respect to the safety of live-attenuated ASFV vaccines, and their implementation is highly limited. Several studies have attempted to identify multiple compounds that can inhibit ASFV infections. Antiviral drugs with identified targets and known mechanisms can be classified into five distinct categories as follows: nucleoside analogs, interferons, plant-derived compounds, antibiotics, small interfering RNA, and CRISPR/Cas9 [14–23]. Other antiviral drugs such as apigenin, resveratrol, and oxyresveratrol show potent, dose-dependent anti-ASFV activity in vitro, but they have unknown targets and mechanisms. To date, there have been no effective drugs for clinical trials against ASFV, and there is an urgent need to develop preventive and therapeutic reagents.

ASFV is primarily replicated in the cytoplasm of swine macrophages, which is a major source of reactive oxygen species that cause constant damage to the ASFV genome [24–27]. To overcome potential DNA damage, such as apurinic/aprimidinic (AP) sites and/or single strand breaks, the virus has evolved its own DNA repair system, including AP endonuclease (*AsfvAP*), a repair DNA polymerase (*AsfvPolX*), and a DNA ligase (*AsfvLIG*) [28–30]. Unlike the homologous proteins, the fidelities of both *AsfvPolX* and *AsfvLIG* are very low. Therefore, in addition to their critical role in genome stability maintenance, these enzymes play an important role in the strategic mutagenesis and genotype formation of ASFV [28,31–33]. Owing to their functional importance, ASFV repair enzymes have been extensively studied [34,35].

In this study, we focused on an enzyme that helps synthesize DNA, the ASFV DNA polymerase X or *AsfvPolX*, to identify novel targets for the prevention and treatment of ASFV. *AsfvPolX* plays a role in the DNA repair process of the ASFV genome, and viral replication is partially dependent on the function of *AsfvPolX*. In addition, *AsfvPolX* has several unique structural features, including a 5'-phosphate (5'-P) binding pocket, a His115-Arg127 platform, and hydrophobic residues Val120 and Leu123, which can all affect the catalytic efficiency of *AsfvPolX*. It can be assumed that blocking the binding pocket with therapeutics may inhibit *AsfvPolX* activity and thereby impair the DNA repair process of the viral genome [36]. In the present study, we used the crystal structure of two different types of *AsfvPolX* protein structures (*AsfvPolX* apo and *AsfvPolX*/DNA structures) to identify potential antiviral drugs using a novel integrated strategy that combines molecular docking and machine learning models [14,37]. Artificial intelligence and machine learning have been successfully applied in bioinformatics [38–42]. To identify novel antiviral inhibitors, we first conducted molecular docking studies of FDA-approved drugs based on the structural information of the targets of ASFV and predicted ten top-ranked compounds with high docking scores using principal component analysis (PCA) and *k*-means clustering. Pentagastrin, which had the highest binding affinity in both the apo and DNA complex structures of *AsfvPolX*, was chosen as an antiviral candidate drug. We also identified the dose-dependent inhibitory effect of pentagastrin on *AsfvPolX* polymerase activity.

The use of this combination strategy can provide an alternative approach to identifying potential antiviral drugs against ASFVs from molecules that have already been tested or approved for other diseases. Furthermore, the region of interest that only appears in groups with high binding affinity to ASFV targets identified based on eight molecular descriptors used in this study can provide key information regarding the unique chemical characteristics of antiviral drugs against ASFVs. These results demonstrate that active compounds with regions of interest in the chemical space may be expected to provide insights into potential ASFV inhibitors.

2. Results and Discussion

2.1. Identification of Ligandable Pockets

The putative binding sites of the *AsfvPolX* protein were identified using SiteMap, with an applied force field of OPLS_2005 in the Schrödinger suite. SiteMap uses the hydrophobicity and accessibility of a detected binding site to assess the likelihood of a small-molecule inhibitor binding to it. We denoted the four top-ranking surface clefts

identified by the SiteMap module on the structure of *Asfv*PolX protein in complex with DNA (PDB ID: 5HRB) from site 1 to site 4 (Figure 1). Sites 1 and 3 were clearly distinct from the DNA binding sites; however, sites 2 and 4 were very close in the 5'-P binding pocket of *Asfv*PolX. We finally decided on sites 2 and 4 as the potential binding pockets of *Asfv*PolX to find antiviral drugs through docking simulations.

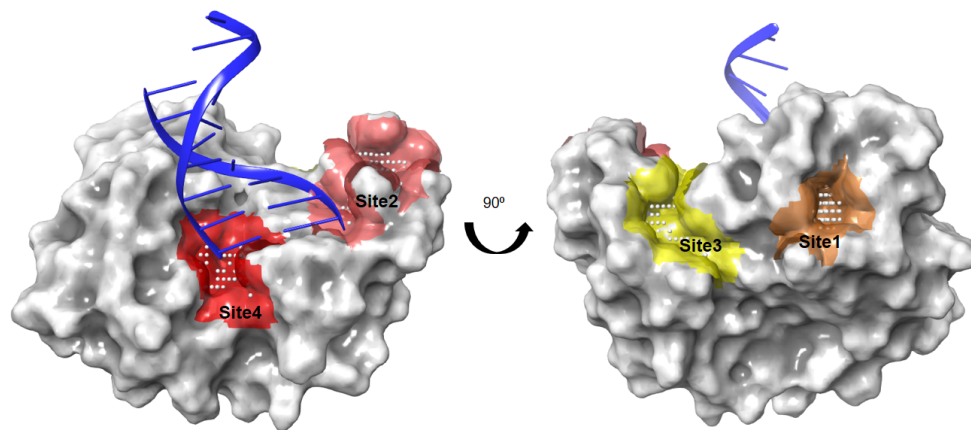


Figure 1. The four top-ranking surface pockets observed in the *Asfv*PolX protein identified by the SiteMap module of Schrödinger. Colored spaces represent the SiteMap predictions for four top-ranking surface clefts on the *Asfv*PolX protein (PDB ID: 5HRB, in grey surface representation): site 1: orange, 2: pink, 3: yellow, 4: red. The red and pink spaces correspond to the DNA binding site. Grey dots represent the cavity of the putative active site of the protein.

2.2. Docking Simulation to DNA Binding Site

To discover new antiviral drugs capable of inhibiting the activity of *Asfv*PolX, we first performed molecular docking using binding pockets identified through SiteMap. The preparation of datasets for molecular docking was performed with two different types of *Asfv*PolX protein structures (*Asfv*PolX apo and *Asfv*PolX/DNA structures), using FDA-approved drugs from the DrugBank database. The reported eight *Asfv*PolX complex structures demonstrated that *Asfv*PolX has one unique 5'-P binding pocket that can inhibit the activity of *Asfv*PolX and disrupt the DNA repair process of the ASFV genome [36]. Previous binding site analysis studies using SiteMap also identified the 5'-phosphate (5'-P) binding pocket of *Asfv*PolX as a potential binding pocket for antiviral inhibitors. To determine which conformations of the apo and DNA complex structures are structurally proximal conformations as template structures for virtual screening of antiviral compounds with high binding affinity, we implemented a molecular docking program using both the crystal structure of apo (PDB ID:1JAJ) and DNA complex structure (PDB ID:5HRB). We obtained 1729 drugs from the DrugBank database. After performing molecular docking, the 1JAJ and 5HRB datasets consisted of 1714 and 1712 compounds. Figure 2A,B show the histogram and boxplot distribution of the docking scores for the two structures. The distribution of docking score for the DNA complex structure was shifted to the left relative to that of the apo structure (Figure 2A). In addition, the mean docking scores of compounds in apo and DNA complex structures were -3.703 and -4.323 , respectively, which implied that compounds for DNA complex structure could be predicted to have higher binding affinity than compounds for apo structure (Figure 2B). These results indicated that the *Asfv*PolX-DNA complex structure is a suitable template for determining compounds with high binding affinity and can also provide important guidance for the design and discovery of potential antiviral inhibitors for *Asfv*PolX.

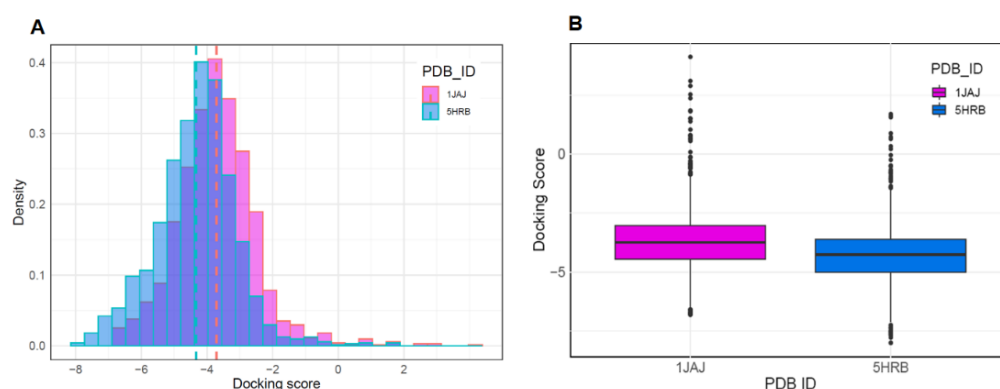


Figure 2. The DrugBank datasets were docked against the apo and DNA complex structure of *AsfvPolX* protein. The distributions of docking scores for each structure are shown as histogram (A) and box and whisker plots, with outliers indicated by black dots (B).

2.3. Identification of New Chemical Space against *AsfvPolX* Inhibitor

To investigate the distinct features of molecules based on docking results for apo and DNA complex structures, eight molecular descriptors of each molecule were calculated and applied to PCA (Figure 3). The 1JAJ and 5HRB datasets consisted of 1714 and 1712 compounds based on the docking score. We then constructed the 1JAJ and 5HRB datasets which consisted of 675 and 1035 compounds, respectively, with binding free energy ≤ -4.0 kcal/mol based on docking score. These datasets were selected for further PCA in this step. We also constructed a duplicated dataset consisting of 601 compounds with binding free energy ≤ -4.0 kcal/mol for both apo and DNA complex structures. In PCA visualizations of all the three datasets with high binding affinity, we determined that the datasets were generally well distributed throughout the chemical space, whereas the three datasets with binding affinities ≥ -4.0 kcal/mol were significantly concentrated in the center (Figure 3). Of the three datasets, 5HRB and duplicated datasets generated a lot of separately distributed points in the chemical space (Figure 3A). All the three datasets of compounds with low binding affinity show that the most densely packed regions were in the center (Figure 3B). PCA visualization of all three datasets containing compounds with binding affinity ≤ -4.0 kcal/mol or ≥ -4.0 kcal/mol (represented by “Non-binding” in Figure S2) shows that centralized distribution was clearly found in the non-binding dataset. From these results, we determined the region of interest present in 5HRB and duplicated the dataset with binding affinities ≤ -4.0 kcal/mol as a new chemical space for finding a potential inhibitor against ASFVs, which was then preferentially chosen to select the final compound for experimental testing.

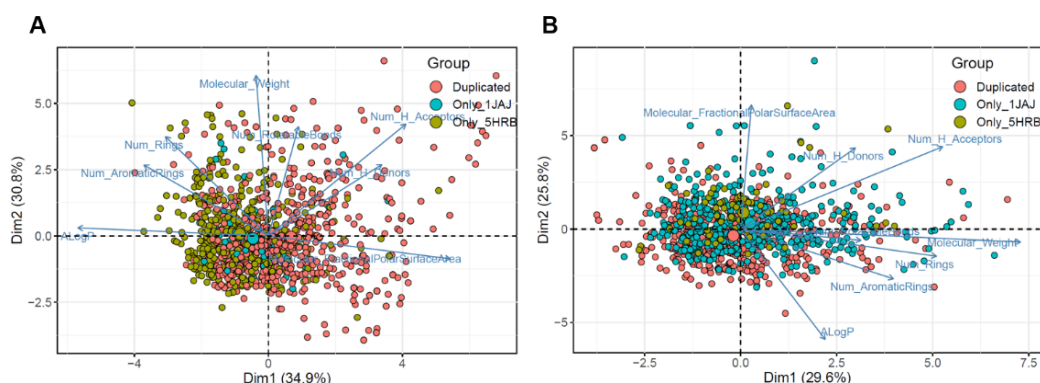


Figure 3. Visual representation of the chemical space of 1JAJ and 5HRB datasets. Distribution of the chemical space of the compounds in the 1JAJ and 5HRB datasets, according to principal component analysis. (A) Compounds with ≤ -4.0 kcal/mol binding free energies. (B) Compounds with ≥ -4.0 kcal/mol binding free energies. The loading plot vectors are represented by arrows for each physicochemical property.

2.4. Molecular Descriptor-Based Clustering

To investigate the region of interest in the chemical space, we used *k*-means clustering for three datasets consisting of compounds with binding affinities ≤ -4.0 kcal/mol based on eight molecular descriptors (Figure 4A). We expected that the diverse regions of 5HRB and duplicated dataset through PCA could provide several important chemical regions for finding compounds with higher binding affinities in the DNA-binding sites of ASFVs. For the datasets, we identified the optimal clustering consisting of groups of six clusters. These results indicated that physicochemical descriptor-based clustering can accurately provide a unique region for 5HRB and duplicated dataset containing compounds with high affinity, demonstrating that compounds with the best docking score in both ASFV protein structures are clustered together in the chemical space. For compounds in Clusters 1, 3, and 4, the mean docking scores were -5.27 , -4.63 , and -4.87 , respectively (Figure 4B). The mean docking scores of compounds in Clusters 2, 5, and 6 were -6.0 , -5.26 , and -5.44 , respectively, which implied that compounds of Cluster 2 could be predicted to have higher binding affinity than other clusters. This showed chemical diversity and regions of interest in chemical space. We showed that PCA-based clustering on the eight molecular descriptors used in this study could discriminate compounds with high binding affinity and provide valuable guidance on the design and discovery of potent antiviral drugs available for the prevention or treatment of ASFV infections.

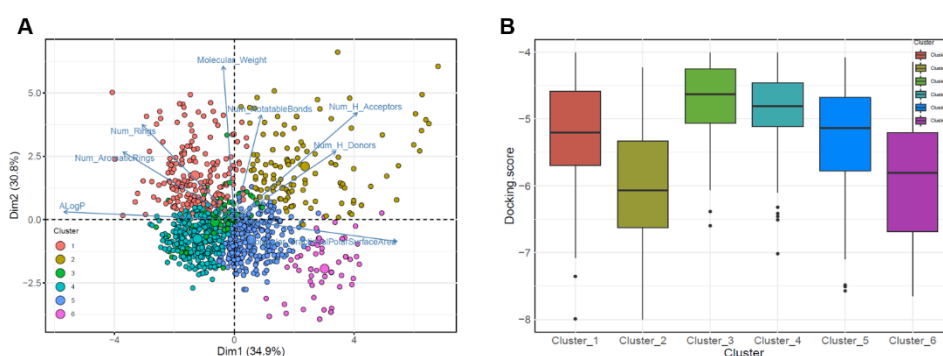


Figure 4. Principal component analysis clustering plot of the chemical space of prepared datasets. The loading plot vectors are represented by arrows for each physicochemical property. (A) The red, mustard, green, cyan, blue, and magenta dots correspond to clusters 1, 2, 3, 4, 5, and 6, respectively. Data points are color-coded by clusters of molecules. (B) Boxplots for SP Glide score distributions of prepared dataset for each cluster (Clusters 1–6).

2.5. Pentagastrin on *AsfvPolX* Polymerase Activity

Among the top-ranked ten ligands in Cluster 2, pentagastrin with the highest binding affinity in both apo and complex structures was chosen as a promising drug candidate. The list of the top-ranked ten ligands in Cluster 2 can be found in Table S1. To measure the effect of pentagastrin on *AsfvPolX* activity, BL21 (DE3)-RIL cells carrying pET28-*AsfvPolX* plasmid were cultured in the presence of isopropyl β -D-1-thiogalactopyranoside, and then the *AsfvPolX* protein was purified using a Ni^{2+} -nitrilotriacetic acid (Ni-NTA) column (Figure 5A). *AsfvPolX* protein was further purified using a glycerol gradient (Figure 5B). We first verified *AsfvPolX* polymerase activity. As shown in Figure 5C, *AsfvPolX* activity rapidly increased after the incorporation of nucleotides up to 200 ng of enzyme, and the activity gradually decreased thereafter. We then determined the effect of pentagastrin on *AsfvPolX* polymerase activity. As shown in Figure 5D, *AsfvPolX* activity was significantly decreased by pentagastrin in a dose-dependent manner. We showed that 100 μM of pentagastrin inhibited *AsfvPolX* activity by 60% (Figure 5D), indicating that pentagastrin may be a potent inhibitor of the ASFVs.

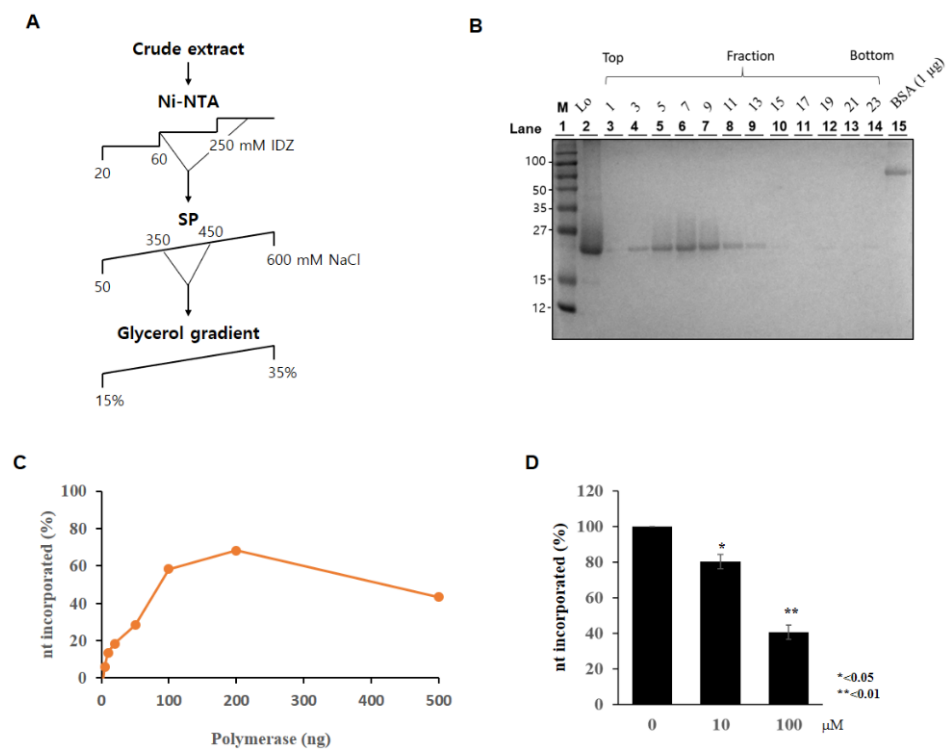


Figure 5. Pentagastrin inhibits *Asfv*PolX activity. **(A)** Schematic illustration of recombinant PolX purification. Ni-NTA, Ni²⁺-nitrilotriacetic acid affinity; SP, SP column chromatography; IDZ, imidazole. Glycerol gradient indicates the final purification procedure using glycerol gradient sedimentation. **(B)** Twelve fractions collected from glycerol gradient sedimentation were analyzed in a 15% SDS-PAGE and then the gel was stained with Coomassie Brilliant Blue R-250. Lo, Load. Molecular mass markers are indicated in kDa. **(C)** Polymerase activity was determined by measuring nt incorporation using indicated amounts of purified *Asfv*PolX protein. **(D)** Pentagastrin was mixed with 100 ng of ASFV PolX for 10 min and polymerase activities were measured as described in Materials and Methods.

2.6. Binding Interaction Patterns of Pentagastrin

The binding interactions between pentagastrin and *Asfv*PolX protein, identified using Glide (Schrödinger, LLC, New York, NY, USA, 2020), showed that pentagastrin was fitted into a groove at the same DNA binding pocket, and that it had a similar conformation as that of 1nt-gap DNA1 (Figure 6A). In the docking model of pentagastrin, the phenyl rings docked into a hydrophobic cavity formed by residues Val120, Leu123, and Ile124 of *Asfv*PolX protein. The hydroxyl group of pentagastrin formed hydrogen bonds with Lys85 and Lys136 at the binding site (Figure 6B). The prediction of the binding poses of the other nine candidate ligands through molecular docking simulations is represented in Figure S3. Docking studies demonstrated that these residues contribute to the binding interaction of pentagastrin with the *Asfv*PolX protein and could be important for the design of new ligands.

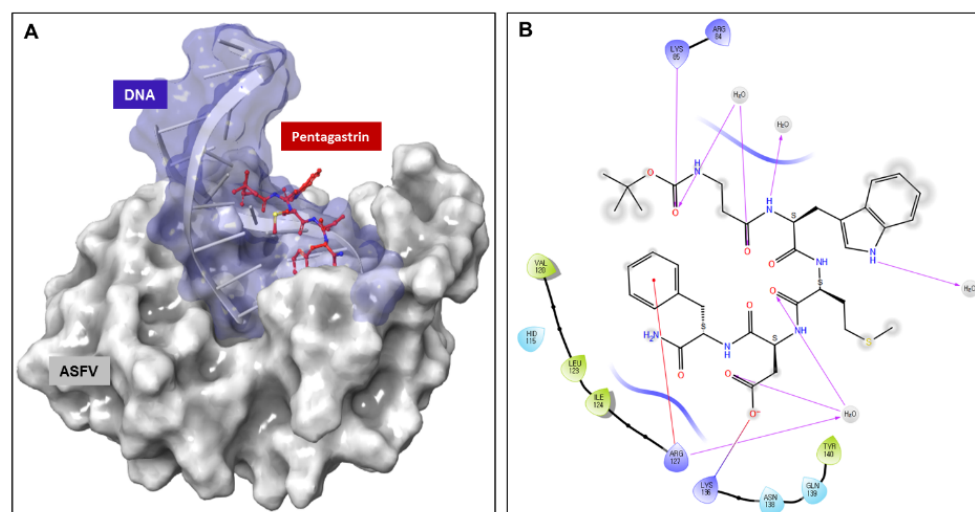


Figure 6. Molecular docking of pentagastrin to AsfvPolX protein. (A) Three-dimensional ligand interaction of pentagastrin generated by using Schrödinger software. The binding site in pentagastrin is shown as a surface model (B). Two-dimensional ligand interaction diagram of pentagastrin with the essential amino acid residues at the DNA binding site are represented in circles.

3. Materials and Methods

3.1. Binding Site Analysis

The binding site was characterized using the SiteMap module of the Schrödinger software package (Schrödinger, New York, NY, USA) using the high-resolution crystal structure, *AsfvPolX*, in complex with DNA (Protein Data Bank [PDB] ID, 5HRB) [43]. The SiteMap module can be used to identify and clarify potential binding sites (site points) that are most likely contribute to tight protein–ligand or protein–protein binding. The following physicochemical properties of the sites were calculated using the SiteMap program: size, volume, degree of enclosure/exposure, degree of contact, hydrophobic/hydrophilic character, hydrophobic/hydrophilic balance, and hydrogen-bonding possibilities (acceptors/donors). The algorithm of SiteMap determines the likeliness of a site point whose contribution depends on how close the site points are to the protein surface and how site points are sheltered from the solvent. These site points, determined by SiteMap, can be used to define the active site of a protein.

3.2. Docking Simulation

To examine the binding interactions of the *AsfvPolX*-ligand complexes, molecular docking studies were performed using Glide software (Schrödinger, New York, NY, USA), which uses an optimized potential for liquid simulations (OPLS)-2005 force field, and refinement was carried out according to the recommendations of the Schrödinger Protein Preparation Wizard. The ligand database consists of 2413 FDA-approved drugs obtained from DrugBank. Prior to screening, the database was filtered to select compounds with molecular weights >200 Da and 1729 compounds were obtained. LigPrep (Schrödinger, New York, NY, USA) was used to generate three-dimensional ligand structures. The active grid was generated using the Receptor grid application in the Glide module. On a defined receptor grid, flexible docking was performed using the standard precision mode of Glide [44,45]. The datasets of apo (PDB ID:1JAJ) and DNA complex structures (PDB ID:5HRB) contained 1714 and 1712 compounds, respectively, based on the Glide score.

3.3. Dataset Preparation

The dataset was generated from the DrugBank database (<https://go.drugbank.com> (DrugBank Release Version 5.1.4, 2 July 2019)), consisting of more than 11,000 drugs. Drugs in this database are classified into different types as follows: trial stage drugs, approved drugs, withdrawn drugs, and investigational drugs. In this study, we focused only on FDA-

approved drugs (around 2413 compounds) (DrugBank Release Version 5.1.4, 2 July 2019). These compounds were downloaded in the form of a structure data file from the DrugBank database. We removed compounds with a molecular weight less than approximately 200, and thus obtained a dataset comprising 1729 drugs. To perform molecular docking, we used apo and DNA complex structures of the *Asfo*PolX target to obtain sets of apo (PDB ID:1JAJ) and DNA complex structures (PDB ID:5HRB), which consisted of 1714 and 1712 compounds, respectively, based on the docking score. Finally, 1JAJ and 5HRB datasets consisting of 675 and 1035 compounds with binding affinities ≤ -4.0 kcal/mol based on docking score were subjected to further PCA analysis.

3.4. Principal Component Analysis and Cluster Analysis

A molecular descriptor is a numerical description that represents the physical and chemical information of a compound. The descriptor parameters for the datasets were generated using the molecular descriptor calculator included in the QikProp module of Schrödinger (Maestro, Schrödinger, LLC, New York, NY, USA, 2020). The calculated descriptors included molecular weight, number of hydrogen bond acceptors, number of hydrogen bond donors, ALogP, number of rotatable bonds, number of aromatic rings, number of rings, and polar surface area. The frequency distribution of physicochemical properties for the datasets is shown in Figure S1. To investigate and compare the distinct features of molecules based on the docking results for apo and DNA complex structures, the eight aforementioned molecular descriptors of each molecule were calculated. PCA is a multivariate statistical method used for exploratory data analysis. It allows the representation of the property space by projection into the principal component plane (PC1, PC2), which is the mathematical equivalent of taking a picture from the most favorable angle [46–48]. To extract the most important information from the dataset, PCA was employed to explore the chemical space of antiviral compounds as a function of the eight molecular descriptors, using the FactoMineR package in R [49,50]. PCA clustering was calculated for each dataset using the *k*-means clustering method within the Factoextra R package. All histograms and scatter plots were generated using the R software.

3.5. Plasmid Construction

To confirm the *in vitro* antiviral activity for the candidate compounds identified in this study, the full-length PolX gene (*O174L*) was manually synthesized (BIONICS) and cloned into the BamHI and HindIII sites of the pET28 vector.

3.6. Purification of PolX Proteins

The pET28-PolX plasmid was transformed into *Escherichia coli* BL21 (DE3)-RIL, and cells were grown at 37 °C until an A_{600} value of 0.5 was reached. Cells were then induced with 1 mM isopropyl β -D-1-thiogalactopyranoside for 3 h and collected by centrifugation. The cell pellet was resuspended in buffer T₃₀₀ (50 mM Tris-HCl [pH 7.5], 300 mM NaCl, 0.1% Nonidet P-40, 10% glycerol, and 0.1 mM phenylmethylsulfonyl fluoride). The subscript number in T₃₀₀ indicates the concentration of NaCl (mM). Cells were then lysed by sonication, and the cell extract was centrifuged for 30 min at 13,000 rpm. The supernatant was loaded onto a Ni-NTA column (Qiagen, Valencia, CA, USA) column and washed successively with low concentrations of imidazole (20 mM and 60 mM) in T₃₀₀ until no protein was eluted, and the protein was eluted with 250 mM imidazole in the same buffer. The peak protein fractions were pooled and diluted in 50 mM NaCl. The pooled protein was then loaded onto a HiTrap SP HP column (GE Healthcare, Chicago, IL, USA) that was eluted with a linear gradient (50–600 mM) of NaCl in buffer T. The eluted fractions containing the highest concentration of protein were collected (eluting between 350 and 450 mM of NaCl), and an aliquot was subjected to glycerol gradient centrifugation (15–35% glycerol) at 54,000 rpm for 26 h in a Beckman SW55Ti rotor. Gradient fractions containing active PolX were stored at -70 °C until use.

3.7. Measurement of Polymerase Activity

Polymerase activity of PolX was measured using the polymerase activity assay kit (Enzynomics, Daejeon, Korea), according to the manufacturer's instructions, with minor modifications. Briefly, PolX protein was mixed with pentagastrin (Sigma-Aldrich, St Louis, MO, USA) for 10 min at 4 °C and then further mixed with 2× reaction buffer and primed substrate. For PolX induction, the mixture was incubated for 30 min at 37 °C. Following inactivation of polymerase activity, the synthesized DNA was denatured at 95 °C for 20 min. For annealing DNA, the temperature was decreased at 65 °C for 10 min, 45 °C for 5 min, and 25 °C for 5 min. One hundred cycles of 5 s, with 0.5 °C temperature increments from 45 °C to 95 °C, were used for the melting curves to measure synthesized DNA contents using a CFX Connect real-time system (Bio-Rad Laboratories). The 2× reaction mixture was incubated with a standard substrate (S1, S2, S3) to create a standard curve.

4. Conclusions

In this study, we applied molecular docking and machine learning to build models from publicly available screening data (DrugBank) to predict antiviral drugs that could potentially inhibit ASFV. To determine which conformations of the apo and DNA complex structures are structurally proximal conformations as template structures for virtual screening, we first conducted a docking simulation using both structures. To investigate and compare molecules based on the docking scores for apo and DNA complex structures, eight molecular descriptors of each molecule were calculated. From molecular descriptor-based PCA, we determined that 5HRB and a duplicated dataset containing compounds with high binding affinity were generally well distributed throughout the chemical space. Conversely, the three datasets of compounds with low binding affinity were distributed concentrically around the center of the chemical space. In this work, the *k*-means clustering method was used to establish classification models for predicting whether compounds with high binding affinity are inhibitors of ASFVs. For the datasets, we identified the optimal clustering consisting of six clusters. In particular, compounds in Cluster 2 were found to have a much higher binding affinity on average compared to the other clusters. Therefore, we selected the top-ranked (with the top score) pentagastrin on both apo and DNA complex structures for further in vitro studies. Pentagastrin inhibited *Asfv*PolX activity by 60% at 100 μM concentrations, indicating that the strategy used in this study can reduce the processing time and cost required to identify potential antiviral drugs against ASFVs from molecules already tested or approved for other diseases. These studies to identify drugs with antiviral activity against ASFV have helped in determining active antiviral drugs available for the treatment of ASFV infections. The present study, performed through molecular docking and machine learning models, can play an important role in identifying potential antiviral drugs against ASFVs.

Supplementary Materials: The following are available online, Figure S1: Chemical properties of the compounds from the dataset are compared using histograms for eight molecular descriptors. The dotted lines represent mean values and all histograms were generated with the R software. Figure S2: The chemical space distributions based on the principal component analysis. Data points are color-coded by each dataset. The loading plot vectors are represented by arrows for each physicochemical property; Figure S3: The predicted binding pose of top-ranked nine ligands to *Asfv*PolX protein; Table S1: The list of top-ranked ten ligands in Cluster 2.

Author Contributions: Conceptualization, J.C. and Y.-S.L.; methodology, software, and experiments, J.C., J.S.Y., H.S., N.H.K., Y.-K.S., Y.-H.K., P.R.M., J.I.Y. and Y.-S.L.; writing the manuscript, J.C. and Y.-S.L.; review and editing, J.C., D.T., S.B.H. and Y.-S.L.; project administration and supervision, J.C., H.S.K., J.I.Y., S.B.H. and Y.-S.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Research Foundation of Korea grants (NRF-2017R1A2B3002241, NRF-2018R1D1A1B07050744, NRF-2019R1A2C2084535, NRF-2019R1A2C1086914, and NRF-2021R1A2C3003496) funded by the Korean government (MSIT) and the Korea Institute of Planning and Evaluation for Technology in Food, Agriculture and Forestry (IPET) through the Animal Disease Management Technology Development Program. It was also funded by the Ministry of Agriculture, Food and Rural Affairs (MAFRA) (grant no. 320053-2 to D.T.).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing not applicable.

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

Sample Availability: Samples of the compounds are not available.

References

1. Sánchez-Cordón, P.J.; Montoya, M.; Reis, A.L.; Dixon, L.K. African swine fever: A re-emerging viral disease threatening the global pig industry. *Vet. J.* **2018**, *233*, 41–48. [[CrossRef](#)]
2. Dixon, L.K.; Chapman, D.A.; Netherton, C.L.; Upton, C. African swine fever virus replication and genomics. *Virus Res.* **2013**, *173*, 3–14. [[CrossRef](#)] [[PubMed](#)]
3. Parker, J.; Plowright, W.; Pierce, M. The epizootiology of African swine fever in Africa. *Veter. Rec.* **1969**, *85*, 668–674.
4. Thomson, G.; Gainaru, M.; van Dellen, A. Experimental infection of warthog (*Phacochoerus aethiopicus*) with African swine fever virus. *Onderstepoort J. Vet. Res.* **1980**, *1*, 4719–4722.
5. Plotkin, A.S.; Plotkin, S.L. The development of vaccines: How the past led to the future. *Nat. Rev. Microbiol.* **2011**, *9*, 889–893. [[CrossRef](#)] [[PubMed](#)]
6. Sánchez-Vizcaíno, J.; Mur, L.; Gomez-Villamandos, J.; Carrasco, L. An Update on the Epidemiology and Pathology of African Swine Fever. *J. Comp. Pathol.* **2015**, *152*, 9–21. [[CrossRef](#)]
7. Wang, T.; Sun, Y.; Qiu, H.-J. African swine fever: An unprecedented disaster and challenge to China. *Infect. Dis. Poverty* **2018**, *7*, 111. [[CrossRef](#)] [[PubMed](#)]
8. Sánchez, E.G.; Pérez-Núñez, D.; Revilla, Y. Development of vaccines against African swine fever virus. *Virus Res.* **2019**, *265*, 150–155. [[CrossRef](#)]
9. Hübner, A.; Petersen, B.; Keil, G.M.; Niemann, H.; Mettenleiter, T.C.; Fuchs, W. Efficient inhibition of African swine fever virus replication by CRISPR/Cas9 targeting of the viral p30 gene (CP204L). *Sci. Rep.* **2018**, *8*, 1449. [[CrossRef](#)]
10. Lacasta, A.; Ballester, M.; Monteagudo, P.L.; Rodriguez, J.M.; Salas, M.L.; Accensi, F.; Pina-Pedrero, S.; Bensaid, A.; Argilagué, J.; Lopez-Soria, S.; et al. Expression Library Immunization Can Confer Protection against Lethal Challenge with African Swine Fever Virus. *J. Virol.* **2014**, *88*, 13322–13332. [[CrossRef](#)]
11. Lokhandwala, S.; Waghela, S.D.; Bray, J.; Sangewar, N.; Charendoff, C.; Martin, C.L.; Hassan, W.S.; Koynarski, T.; Gabbert, L.; Burrage, T.G.; et al. Adenovirus-vectored novel African Swine Fever Virus antigens elicit robust immune responses in swine. *PLoS ONE* **2017**, *12*, e0177007. [[CrossRef](#)]
12. O'Donnell, V.; Risatti, G.R.; Holinka, L.G.; Krug, P.W.; Carlson, J.; Velazquez-Salinas, L.; Azzinaro, P.A.; Gladue, D.; Borca, M.V. Simultaneous Deletion of the 9GL and UK Genes from the African Swine Fever Virus Georgia 2007 Isolate Offers Increased Safety and Protection against Homologous Challenge. *J. Virol.* **2016**, *91*. [[CrossRef](#)] [[PubMed](#)]
13. Reis, A.L.; Abrams, C.C.; Goatley, L.C.; Netherton, C.; Chapman, D.G.; Sanchez-Cordon, P.; Dixon, L.K. Deletion of African swine fever virus interferon inhibitors from the genome of a virulent isolate reduces virulence in domestic pigs and induces a protective response. *Vaccine* **2016**, *34*, 4698–4705. [[CrossRef](#)]
14. Zhu, Z.; Fan, Y.; Liu, Y.; Jiang, T.; Cao, Y.; Peng, Y. Prediction of antiviral drugs against African swine fever viruses based on protein-protein interaction analysis. *PeerJ* **2020**, *8*, e8855. [[CrossRef](#)] [[PubMed](#)]
15. Arabyan, E.; Hakobyan, A.; Kotsinyan, A.; Karalyan, Z.; Arakelov, V.; Arakelov, G.; Nazaryan, K.; Simonyan, A.; Aroutiounian, R.; Ferreira, F.; et al. Genistein inhibits African swine fever virus replication in vitro by disrupting viral DNA synthesis. *Antivir. Res.* **2018**, *156*, 128–137. [[CrossRef](#)] [[PubMed](#)]
16. Freitas, F.B.; Frouco, G.; Martins, C.; Leitão, A.; Ferreira, F. In vitro inhibition of African swine fever virus-topoisomerase II disrupts viral replication. *Antivir. Res.* **2016**, *134*, 34–41. [[CrossRef](#)]
17. Galindo, I.; Hernaez, B.; Berná, J.; Fenoll, J.; Cenis, J.; Escribano, J.; Alonso, C. Comparative inhibitory activity of the stilbenes resveratrol and oxyresveratrol on African swine fever virus replication. *Antivir. Res.* **2011**, *91*, 57–63. [[CrossRef](#)] [[PubMed](#)]
18. Gil-Fernández, C.; Páez, E.; Vilas, P.; Gancedo, G. Effect of Disodium Phosphonoacetate and Iododeoxyuridine on the Multiplication of African Swine Fever Virus in vitro. *Chemotherapy* **1979**, *25*, 162–169. [[CrossRef](#)]
19. Hakobyan, A.; Arabyan, E.; Avetisyan, A.; Abroyan, L.; Hakobyan, L.; Zakaryan, H. Apigenin inhibits African swine fever virus infection in vitro. *Arch. Virol.* **2016**, *161*, 3445–3453. [[CrossRef](#)]

20. Hakobyan, A.; Galindo, I.; Nañez, A.; Arabyan, E.; Karalyan, Z.; Chistov, A.A.; Streshnev, P.P.; Korshun, V.A.; Alonso, C.; Zakaryan, H. Rigid amphipathic fusion inhibitors demonstrate antiviral activity against African swine fever virus. *J. Gen. Virol.* **2018**, *99*, 148–156. [[CrossRef](#)] [[PubMed](#)]
21. Hernández, B.; Tarragó, T.; Giralt, E.; Escribano, J.M.; Alonso, C. Small Peptide Inhibitors Disrupt a High-Affinity Interaction between Cytoplasmic Dynein and a Viral Cargo Protein. *J. Virol.* **2010**, *84*, 10792–10801. [[CrossRef](#)] [[PubMed](#)]
22. Keita, D.; Heath, L.; Albina, E. Control of African swine fever virus replication by small interfering RNA targeting the A151R and VP72 genes. *Antivir. Ther.* **2010**, *15*, 727–736. [[CrossRef](#)] [[PubMed](#)]
23. Paez, E.; Garcia, F.; Gil Fernandez, C. Interferon cures cells lytically and persistently infected with African swine fever virus in vitro. *Arch. Virol.* **1990**, *112*, 115–127. [[CrossRef](#)] [[PubMed](#)]
24. Alcamí, A.; Carrascosa, A.L.; Viñuela, E. Interaction of African swine fever virus with macrophages. *Virus Res.* **1990**, *17*, 93–104. [[CrossRef](#)]
25. Akaike, T. Role of free radicals in viral pathogenesis and mutation. *Rev. Med Virol.* **2001**, *11*, 87–101. [[CrossRef](#)] [[PubMed](#)]
26. Forman, J.H.; Torres, M. Redox signaling in macrophages. *Mol. Asp. Med.* **2001**, *22*, 189–216. [[CrossRef](#)]
27. Chen, Y.; Chen, X.; Huang, Q.; Shao, Z.; Gao, Y.; Li, Y.; Yang, C.; Liu, H.; Li, J.; Wang, Q.; et al. A unique DNA-binding mode of African swine fever virus AP endonuclease. *Cell Discov.* **2020**, *6*, 1–13. [[CrossRef](#)]
28. Lamarche, B.J.; Showalter, A.K.; Tsai, M.-D. An Error-Prone Viral DNA Ligase. *Biochemistry* **2005**, *44*, 8408–8417. [[CrossRef](#)]
29. Oliveros, M.; Yáñez, R.J.; Salas, M.L.; Salas, J.; Viñuela, E.; Blanco, L. Characterization of an African Swine Fever Virus 20-kDa DNA Polymerase Involved in DNA Repair. *J. Biol. Chem.* **1997**, *272*, 30899–30910. [[CrossRef](#)]
30. Lamarche, B.J.; Tsai, M.-D. Contributions of an Endonuclease IV Homologue to DNA Repair in the African Swine Fever Virus. *Biochemistry* **2006**, *45*, 2790–2803. [[CrossRef](#)]
31. Achenbach, J.E.; Gallardo, C.; Nieto-Pelegrín, E.; Rivera-Arroyo, B.; Degefa-Negi, T.; Arias, M.; Jenberie, S.; Mulisa, D.D.; Gizaw, D.; Gelaye, E.; et al. Identification of a New Genotype of African Swine Fever Virus in Domestic Pigs from Ethiopia. *Transbound. Emerg. Dis.* **2017**, *64*, 1393–1404. [[CrossRef](#)]
32. Boshoff, C.I.; Bastos, A.D.; Gerber, L.; Vosloo, W. Genetic characterisation of African swine fever viruses from outbreaks in southern Africa (1973–1999). *Vet. Microbiol.* **2007**, *121*, 45–55. [[CrossRef](#)] [[PubMed](#)]
33. Showalter, A.K.; Tsai, M.-D. A DNA polymerase with specificity for five base pairs. *J. Am. Chem. Soc.* **2001**, *123*, 1776–1777. [[CrossRef](#)]
34. García-Escudero, R.; García-Díaz, M.; Salas, M.L.; Blanco, L.; Salas, J. DNA Polymerase X of African Swine Fever Virus: Insertion Fidelity on Gapped DNA substrates and AP lyase Activity Support a Role in Base Excision Repair of Viral DNA. *J. Mol. Biol.* **2003**, *326*, 1403–1412. [[CrossRef](#)]
35. Redrejo-Rodríguez, M.; García-Escudero, R.; Yáñez-Muñoz, R.J.; Salas, M.L.; Salas, J. African Swine Fever Virus Protein pE296R Is a DNA Repair Apurinic/Apyrimidinic Endonuclease Required for Virus Growth in Swine Macrophages. *J. Virol.* **2006**, *80*, 4847–4857. [[CrossRef](#)] [[PubMed](#)]
36. Chen, Y.; Zhang, J.; Liu, H.; Gao, Y.; Li, X.; Zheng, L.; Cui, R.; Yao, Q.; Rong, L.; Li, J.; et al. Unique 5'-P recognition and basis for dG: dGTP misincorporation of ASFV DNA polymerase X. *PLoS Biol.* **2017**, *15*, e1002599. [[CrossRef](#)]
37. Kinyanyi, D.; Amwayi, P.; Wamalwa, M.; Obiero, G. Comparative in silico study of congocidine congeners as potential inhibitors of African swine fever virus. *PLoS ONE* **2019**, *14*, e0221175. [[CrossRef](#)]
38. Vamathevan, J.; Clark, D.; Czodrowski, P.; Dunham, I.; Ferran, E.; Lee, G.; Li, B.; Madabhushi, A.; Shah, P.; Spitzer, M.; et al. Applications of machine learning in drug discovery and development. *Nat. Rev. Drug Discov.* **2019**, *18*, 463–477. [[CrossRef](#)]
39. Gupta, R.; Srivastava, D.; Sahu, M.; Tiwari, S.; Ambasta, R.K.; Kumar, P. Artificial intelligence to deep learning: Machine intelligence approach for drug discovery. *Mol. Divers.* **2021**, 1–46. [[CrossRef](#)]
40. Lam, L.H.T.; Le, N.H.; Van Tuan, L.; Ban, H.T.; Hung, T.N.K.; Nguyen, N.T.K.; Dang, L.H.; Le, N.Q.K. Machine Learning Model for Identifying Antioxidant Proteins Using Features Calculated from Primary Sequences. *Biology* **2020**, *9*, 325. [[CrossRef](#)]
41. Le, N.Q.K.; Do, D.T.; Hung, T.N.K.; Lam, L.H.T.; Huynh, T.-T.; Nguyen, N.T.K. A Computational Framework Based on Ensemble Deep Neural Networks for Essential Genes Identification. *Int. J. Mol. Sci.* **2020**, *21*, 9070. [[CrossRef](#)] [[PubMed](#)]
42. Ros-Lucas, A.; Correa-Fiz, F.; Bosch-Camós, L.; Rodriguez, F.; Alonso-Padilla, J. Computational Analysis of African Swine Fever Virus Protein Space for the Design of an Epitope-Based Vaccine Ensemble. *Pathogens* **2020**, *9*, 1078. [[CrossRef](#)] [[PubMed](#)]
43. Halgren, T.A. Identifying and characterizing binding sites and assessing druggability. *J. Chem. Inf. Modeling* **2009**, *49*, 377–389. [[CrossRef](#)]
44. Halgren, T.A.; Murphy, R.B.; Friesner, R.A.; Beard, H.S.; Frye, L.L.; Pollard, W.T.; Banks, J.L. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.* **2004**, *47*, 1750–1759. [[CrossRef](#)]
45. Friesner, R.A.; Banks, J.L.; Murphy, R.B.; Halgren, T.A.; Klicic, J.J.; Mainz, D.T.; Repasky, M.P.; Knoll, E.H.; Shelley, M.; Perry, J.K.; et al. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47*, 1739–1749. [[CrossRef](#)]
46. Medina-Franco, J.L.; Martínez-Mayorga, K.; Bender, A.; Marín, R.M.; Giulianotti, M.A.; Pinilla, C.; Houghten, R.A. Characterization of Activity Landscapes Using 2D and 3D Similarity Methods: Consensus Activity Cliffs. *J. Chem. Inf. Model.* **2009**, *49*, 477–491. [[CrossRef](#)] [[PubMed](#)]
47. Medina-Franco, J.L.; Martínez-Mayorga, K.; Giulianotti, M.A.; Houghten, R.A.; Pinilla, C. Visualization of the chemical space in drug discovery. *Curr. Comput.-Aided Drug Des.* **2008**, *4*, 322–333. [[CrossRef](#)]

-
48. Geppert, H.; Vogt, M.; Bajorath, J. Current Trends in Ligand-Based Virtual Screening: Molecular Representations, Data Mining Methods, New Application Areas, and Performance Evaluation. *J. Chem. Inf. Model.* **2010**, *50*, 205–216. [[CrossRef](#)] [[PubMed](#)]
 49. Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37–52. [[CrossRef](#)]
 50. Lê, S.; Josse, J.; Husson, F. FactoMineR: An R package for multivariate analysis. *J. Stat. Softw.* **2008**, *25*, 1–18. [[CrossRef](#)]