



Article

Learning the Relationship between the Primary Structure of HIV Envelope Glycoproteins and Neutralization Activity of Particular Antibodies by Using Artificial Neural Networks

Cătălin Buiu ^{1,*}, Mihai V. Putz ^{2,3,*} and Speranta Avram ⁴

¹ Department of Automatic Control and Systems Engineering, Faculty of Automatic Control and Computers, Politehnica University of Bucharest, Bucharest 060042, Romania

² Laboratory of Structural and Computational Physical-Chemistry for Nanosciences and QSAR, Biology-Chemistry Department, Faculty of Chemistry-Biology-Geography, West University of Timisoara, Timisoara 300115, Romania

³ Laboratory of Renewable Energies-Photovoltaics, R&D National Institute for Electrochemistry and Condensed Matter, Timisoara 300569, Romania

⁴ Department of Anatomy, Animal Physiology and Biophysics, Faculty of Biology, University of Bucharest, Bucharest 050095, Romania; speranta.avram@gmail.com

* Correspondence: catalin.buiu@acse.pub.ro (C.B.); mv_putz@yahoo.com (M.V.P.); Tel.: +40-214-029-167 (C.B.); +40-256-592-638 (M.V.P.)

Academic Editor: Jesus Vicente De Julián Ortiz

Received: 13 August 2016; Accepted: 3 October 2016; Published: 11 October 2016

Abstract: The dependency between the primary structure of HIV envelope glycoproteins (ENV) and the neutralization data for given antibodies is very complicated and depends on a large number of factors, such as the binding affinity of a given antibody for a given ENV protein, and the intrinsic infection kinetics of the viral strain. This paper presents a first approach to learning these dependencies using an artificial feedforward neural network which is trained to learn from experimental data. The results presented here demonstrate that the trained neural network is able to generalize on new viral strains and to predict reliable values of neutralizing activities of given antibodies against HIV-1.

Keywords: HIV-1; glycoproteins; antibodies; neutralization data; artificial neural network; regression

1. Introduction

HIV-1 entry into target cells is mediated by envelope glycoprotein (ENV) trimers [1]. ENV is a viral protein serving to form the viral envelope, and the glycosylated envelope trimer is synthesized as gp160, a precursor protein which is further cleaved by furin into gp120 and gp41 subunits [2]. A viral spike shows three gp120 glycoproteins which are noncovalent to three gp41 transmembrane molecules [3]. A key step in the viral entry is the binding of this complex to the CD4 receptor on the cell surface. Figure 1 shows the gp120 core (blue) complexed with CD4 (green) and 17b (red and yellow), which is a neutralizing human antibody.

The ENV protein is organized into five conserved regions, namely C1–C5 and five variable regions, namely V1–V5 [4]. It was reported that the exposed surface of the spike is described by the variable regions of gp120 and there is also a variety of carbohydrates that help mask the surface of the protein [4]. The ENV variability, largely reported into variable loop regions V1–V5, and also sequence mutations in gp120 [4], leads to reduced interactions with specific antibodies and represents an attractive target for anti-HIV-1 treatments. From all the V1–V5 loops structures, the V4 and V5 loops are highly disordered [4] while the structure of the V3 loop is well-defined [5].

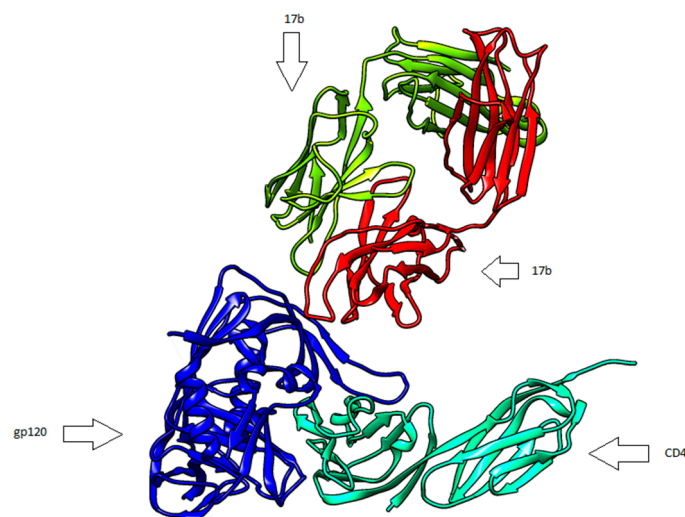


Figure 1. gp120 glycoprotein in complex with CD4 and an antibody (17b) (Protein Data Bank entry 1GC1).

A simplified structure-based model of the V3 loop is used in [6] to model co-receptor tropism in HIV-1. Moreover, the incorporation of ENV determinants outside the V3 loop is demonstrated to be able to improve the reliability of co-receptor usage [7].

The HIV-1 protease is playing a major role in the viral replication and a study of potential protease inhibitors using a QSAR methodology was performed in [8]. A study on the molecular dynamics of the HIV-1 protease was presented in [9]. A structural and docking analysis of HIV-1 integrase and proteins of the nuclear pore complex was investigated in [10].

It is estimated that in ~20% of HIV-1-infected individuals, antibodies that neutralize diverse HIV-1 strains develop in high titers [11]. An important goal for an HIV-1 vaccine development is the identification of broadly neutralizing antibodies (bNAbs) [12–14]. Among the reasons for which this vaccine development is still very challenging are the unusual traits of bNAbs [15].

A number of bNAbs against HIV-1 ENV glycoproteins have been discovered [16–19]. Most of the monoclonal bNAbs target a few major sites on HIV-1 ENV [20]: the CD-4 binding site, two glycan-dependent epitopes involving the V1/V2 and V3 loops, and the membrane-proximal external region (MPER) of the transmembrane gp41 glycoprotein. For example, three bNAbs (2F5, 4E10, and 10E8) are MPER-specific as they target a fusion-intermediate conformation of gp41 [12,21]. In a recent study [22] it is shown that amino acid changes within the MPER epitope can increase the neutralization sensitivity to multiple types of bNAbs.

Partial neutralization by 10E8 was shown to be at least in part influenced by manipulating ENV glycosylation [21]. According to some studies validated by [12], 10E8 is neutralizing HIV-1 with potency and breadth much larger than those of 2F5 and 4E10. Both 2F5 and the m66 antibodies are considered to be the only effective human HIV-1-neutralizing antibodies to recognize the N-terminal region of the MPER of the gp41 subunit of ENV. A crystal structure of m66 in complex with its gp41 epitope is presented in [23]. Antibody accessible sites in the V1–V2 domain of HIV-1 gp120 are the object of several studies, e.g., [24].

A comparison of the neutralization sensitivity for three periods of the epidemic (1987–1991, 1996–2000, 2006–2010) was discussed in [25] which reports that “progressive significantly enhanced resistance to neutralization was observed over calendar time, by both human sera and most of the bNAbs tested (b12, VRC01, VRC03, NIH45-46^{G54W}, PG9, PG16, PGT121, PGT128, PGT145)”. However, a combination of NIH45-46 and PGT128 antibodies was shown to still efficiently neutralize the most contemporary transmitted variants. This analysis is extended to some recently described bNAbs

(PG9-iMab, PG16-iMab, 10E8, 3BNC117, NIH45-46m2, NIH45-46m7, 10-1074, JM4sdAb, 8ANC195, and PG9-16-RSH) in [20].

As pointed out in [26–28], the variation of the neutralization data with respect to various HIV-1 strains is a complicated, unknown function of the ENV primary structure. There are some factors behind this complicated relationship, such as the binding affinity of the antibody to the ENV protein and the intrinsic infection kinetics of the viral strain. An intense effort is carried out in order to identify critical residues of ENV which affect antibody activity. For example, a computational tool to help identify these critical residues is presented in [28] and it is based on the simplifying assumption that the variation of neutralization activities (characterized by IC_{50} values, the concentration at which infectivity is reduced by 50% [29]) is due to amino acid identity or glycosylation state at a small number of sites, each acting independently.

This paper presents the first results of a novel approach, which is based on using machine learning, to extend this analysis of the variation of the neutralization data. Machine learning is a subfield of computer science dedicated to the development and study of algorithms that can learn from and make predictions on data [30]. A more formal definition is provided in [31]: “A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P , if its performance at tasks in T , as measured by P , improves with experience E ”. Artificial neural networks (ANN) are prominent machine learning algorithms which are inspired by the structure and functioning of biological neural networks. There are numerous applications of ANNs in biology and medicine, such as dihedral angles prediction in enzyme loops [32], affinity prediction of protein-ligand complexes [33], cancer prognosis and prediction [34] and computational drug development [35–38].

What we present in this paper is a preliminary work to learn the dependencies between ENV primary structures (amino acids sequences) and neutralization activity of particular antibodies. This is done by training a feedforward ANN with input data (whole ENV primary structures) and output data (neutralization values for particular antibodies) to provide a neural network which is able to generalize on other ENV glycoproteins and to predict neutralization data.

2. Results

A trial-and-error approach has been chosen in order to find the most suitable neural network architecture and parameters for learning the available data, as there are no universally valid guidelines for designing a neural network [39]. For this purpose, a number of experiments have been designed in which some parameters were modified in order to evaluate their impact on the network's performance in learning and generalization.

First, the available data was divided randomly into training, validation and test sets. The MATLAB default values for the sizes of the three datasets were used (75%, 15%, and 10%, respectively). For the data in the training set, a Levenberg–Marquardt backpropagation algorithm [40,41] was used. Monitoring the error on the validation set allows an early stopping of the training, as overfitting is associated with a rise of the validation error. If this error increases for a specified number of iterations, then the training process is stopped. Secondly, each backpropagation training algorithm starts with different initial parameters (weights and biases), so that very different solutions can be obtained with each new training process. Thus, we repeated the above process 100 times and the network with the best generalization was selected. The statistical power of the method was evaluated by the correlation coefficient R . The results for learning the neutralization data for the 2F5 antibody show that suitable correlation coefficients were obtained for the training set ($R = 0.99561$ and dependent variable $Y = 0.99 \times \text{Target} + 0.002$) and for the test set ($R = 0.9674$ and dependent variable $Y = 0.99 \times \text{Target} + 0.016$). These results are shown in Figure 2 which presents the corresponding regression analysis, while Figure 3 shows the error histogram. The mean squared prediction error (MSEP) was 0.015.

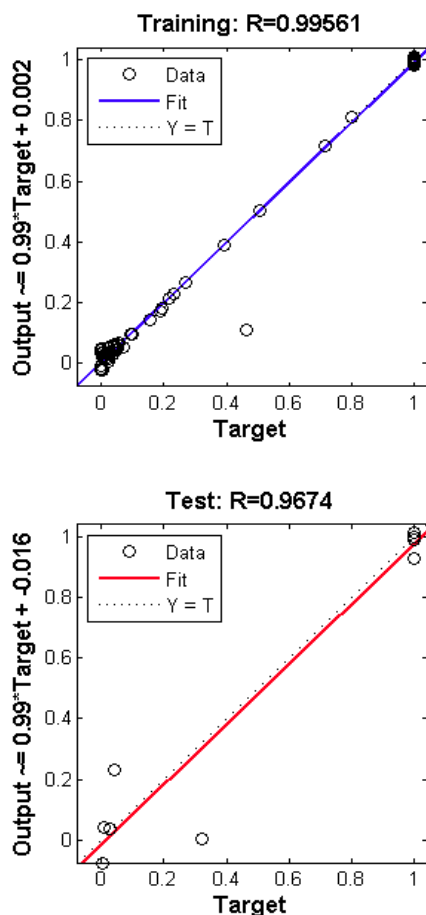


Figure 2. Regression analysis for the training and test data.

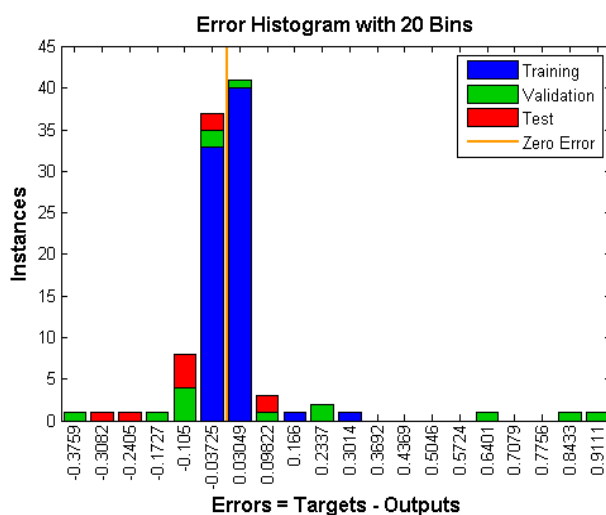


Figure 3. Error histogram.

3. Discussion

The complexity of HIV-1 ENV structural biology asks for complementary information obtained from various techniques such as NMR spectroscopy, X-ray crystallography, cryo-electron microscopy or tomography to understand the virus infectious mechanism, but the limitations of each of these technologies are evident [4]. Given the limitations of each of these approaches, the challenge for the

future HIV-1 ENV studies may be represented by *in silico* methods (e.g., chemical structures-biological activity relationship) for structural biologists in the HIV field to aim higher.

The work presented in this paper is based on our expertise in studying the chemical structures-biological activity relationship HIV-1 protease by using ANNs [42] and also chemical structures-biological activity relationship HIV-1 gp120 in interaction with different antibodies [43]. In [43] we calculated the pharmacological descriptors of the HIV-1 gp 120 binding sites structures for 60 HIV-1 strains. We considered steric molecular descriptors (molecular surfaces, volumes), electronic descriptors (electrostatic energies), counts of atoms and bonds types (number of atoms, number of hydrogen donors or acceptors and number of rigid bonds). We identified: (1) the possible correlation between molecular descriptors of HIV-1 gp 120 and their biological activities; (2) significant fluctuation of descriptors among the strains. Also in [42], we used ANNs to evaluate the biological activity of HIV-1 protease inhibitors for QSAR-like applications and we found that the local mapping of ligand properties, applied to HIV-1 protease, provides accurate results (95%).

This paper presents a novel approach in trying to predict antibody affinities from a primary HIV-1 ENV sequence using a trained feedforward neural network. This has been demonstrated to be an efficient tool to learn dependencies between HIV-1 envelope glycoproteins' primary structure and neutralization activities for particular antibodies. This paper introduced both the idea and the practical realization of a way to model IC_{50} neutralization data variation across a panel of HIV-1 strains.

Results demonstrate that a carefully trained network can learn the nonlinear and complicated dependencies between ENV primary structures and neutralization data for particular antibodies. Partial Least Squares (PLS) regression is widely used in chemometrics [44] for relating two data matrices by a linear multivariate model. We used the Statistics and Machine Learning Toolbox in Matlab in order to relate the input data (aligned ENV sequences) to output data (neutralization data for a particular antibody, 2F5 in our case).

The first step was to fit a PLS regression model with ten PLS components and one response. We generated and analyzed the percent of variance explained in the response variable as a function of the number of components. Figure 4 shows that ten components fully explain the variance.

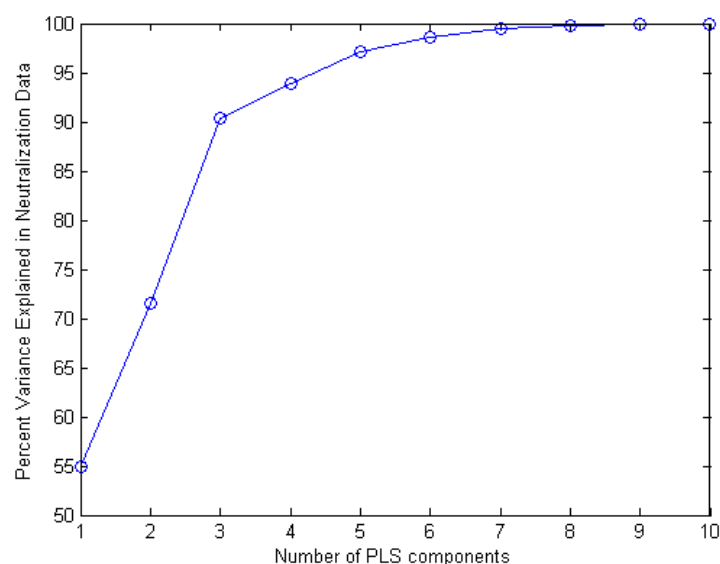


Figure 4. Percent of variance explained in the response variable as a function of the number of Partial Least Squares (PLS) components.

Figure 5 then shows the fitted response vs. the observed response for the PLS regression with ten components with $R = 0.9995$.

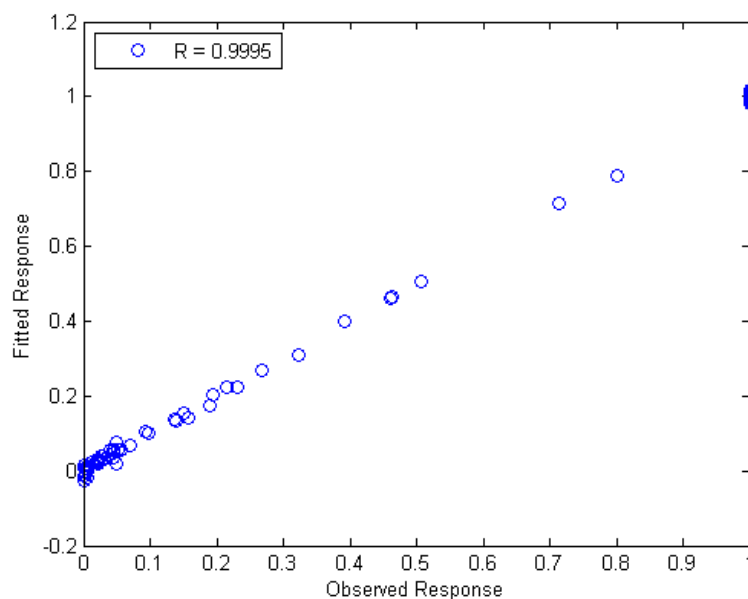


Figure 5. Fitted response vs. observed response for the Partial Least Squares (PLS) regression.

A ten-fold cross-validation technique was then used for estimating the mean squared prediction error (MSEP) which is 0.15 as it can be seen in Figure 6.

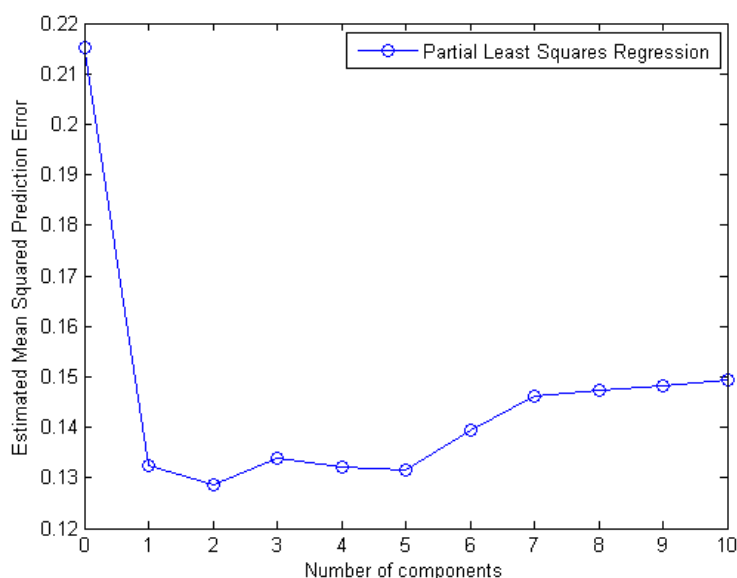


Figure 6. Mean squared prediction error as a function of the number of Partial Least Squares Regression components.

So, the neural network based approach has generated an MSEP ten times smaller than the Partial Least Squares regression.

In this preliminary study, our results improve the knowledge about the HIV-1 ENV protein, its molecular and possible neutralization properties. This ANN-based method can be applied on a large number of HIV-1 ENV structures with large variability. The trained neural network is able to generalize and to predict neutralization data for particular antibodies across HIV-1 strains which were not included in the training set.

Future work will include the acquisition of more neutralization data, and more aligned ENV sequences. Particular attention will be paid to the study of the influence of the glycosylation sites and amount of glycosylation. A sensitivity analysis will be implemented for the trained network, in order to determine which inputs (which residues of ENV glycoproteins) affect the output (neutralization activity) most. This sensitivity analysis can be implemented based on two methods: a backward stepwise method in which one variable (ENV residue) is blocked (rejected), and the effect on the output is quantified; and a second weight method which is based on the weights magnitude. This sensitivity analysis will suggest critical residues as candidates for mutagenesis studies.

A better understanding of the biological activity of HIV-1 ENV structures can be achieved by performing both experimental and *in silico* studies and we will focus our next studies in this direction. We are sure that, in the near future, our study can be extended by experimental techniques which are able to explore more precisely the molecular features of HIV-1 ENV structures. Even though the biological processes in HIV-1 ENV structures involved are very complex and difficult to replicate *in vivo*, the extension of our study by *in vivo* analyses is crucial.

4. Materials and Methods

The goal of the research reported here was to find an implicit model for the relationship between the primary structure of HIV-1 ENV proteins and neutralization data (IC₅₀ values (μg/mL)). All the programs were written in MatlabTM (R2012A) and the Matlab's Neural Networks Toolbox (version 7.0.3, MathWorks[®], Natick, MA, USA) was used. Microsoft Excel was used for archiving neutralization data. The experiments were run on a computer with an Intel(R) Core(TM) i7-3160QM CPU @ 2.30 GHz, 16 GB installed memory and a 64-bit operating system.

The critical importance of ENV regions variability for the HIV-1 infective process and also for the virus escape from antibody interactions was already mentioned. Our aim in these preliminary experiments was to predict, in the most accurate way possible, the interactions of a large number of HIV-1 ENV strains.

The input sequences (primary structures for various HIV-1 strains) can vary in length. As the input of the neural network is fixed in length, our approach was to use aligned sequences. Other possible approaches were to use sparse-encoding [45] or interpolation [46].

Aligned ENV sequences (input data for our network) were collected from the HIV Sequence Database (<http://www.hiv.lanl.gov/>) [47] which we used for downloading ENV data in a FASTA file which contains 4907 aligned ENV sequences. These ENV alignments are based on the complete genome nucleotide alignment. So, the input data for our approach is represented by global alignments of ENV proteins from a large number of HIV-1 strains. The length of the global alignment is 1369.

The output (neutralization) data was collected from literature [16,17,19,29,48–52] and stored in Microsoft Excel files where each row corresponds to a different viral strain and each column corresponds to a different antibody, e.g., 2F5, VRC01, NIH45-46, 3BNC117, PG9, and PG16. Data in these files is represented by IC₅₀ (the half maximal inhibitory concentration) values and where the IC₅₀ for a particular case is known only to be greater than or less than some value (e.g., 50 μg/mL), then that specific value was selected. The FASTA file and a sample Excel file with neutralization data (178 HIV-1 strains and neutralization data for six antibodies) are publicly available for download [53].

The sample neutralization data looks like in Table 1, while Figure 7 shows the distribution of the IC₅₀ values for two antibodies (PG16 and 2F5) against the 178 HIV-1 strains: the distribution of the majority of values around 50 μg/mL and below 1 μg/mL for PG16, while for 2F5 the values are more scattered between 0 and 50 μg/mL.

These data files are read into MATLAB and further used for training the ANN. A coverage curve can be generated for these antibodies using a Matlab function we designed, in order to compare neutralization across a panel of HIV-1 strains. In Figure 8, coverage curves were generated for the data in the sample file (a coverage curve shows the cumulative frequencies of IC₅₀ values up to the concentration which is shown on the *x* axis [13]).

Table 1. Sample data used for training the network (HIV-1 strains in the first column, neutralization data for six selected antibodies in the other columns).

HIV-1 Strain	2F5	VRC01	NIH45-46	3BNC117	PG9	PG16
0260.v5.c36	50	0.529	0.397	0.2	2.18	2.1
0330.v4.c3	14.6	0.064	0.049	0.013	0.018	0.006
0439.v5.c1	4.43	0.052	0.185	0.215	50	50
3415.v1.c1	43.9	0.092	0.082	0.094	0.149	0.036
3718.v3.c11	3.88	0.218	0.871	50	0.05	0.019
398-F1_F6_20	0.28	0.058	0.157	0.071	50	50
BB201.B42	2.92	0.343	0.303	3.35	0.014	0.003
BB539.2B13	0.136	0.094	0.022	0.033	0.106	0.012
BI369.9A	0.249	0.149	0.043	0.02	0.029	0.007
BS208.B1	1.1	0.029	0.006	0.002	0.031	0.004
KER2008.12	6.98	0.563	0.567	0.248	0.017	0.006
KER2018.11	2.01	0.07	0.828	0.417	0.001	0.001
KNH1209.18	2.24	0.087	0.246	0.04	0.367	0.678
MB201.A1	0.436	0.237	0.165	0.464	0.024	0.001
MB539.2B7	2.49	0.544	0.402	0.087	0.058	0.025
MI369.A5	1.44	0.162	0.074	0.033	0.058	0.011
MS208.A1	1.1	0.147	0.09	0.019	0.071	0.047
Q168.a2	7.83	0.14	0.138	0.05	0.106	0.031
Q23.17	10.8	0.086	0.106	0.017	0.007	0.002
Q259.17	16.1	0.051	0.046	0.017	0.045	0.028
Q461.e2	13.4	0.41	0.212	0.069	3.01	4.11
Q769.d22	0.609	0.015	0.013	0.007	0.007	0.01
Q769.h5	50	0.014	0.019	0.006	0.002	0.002
Q842.d12	50	0.006	0.015	0.002	0.005	0.001
QH209.14M.A2	50	0.024	0.011	0.008	50	50
RW020.2	7.55	0.303	0.144	0.02	0.103	0.07
UG037.8	0.202	0.035	0.056	0.02	0.021	0.001
3301.V1.C24	50	0.084	0.055	0.046	0.281	0.023
6540.v4.c1	40	50	50	50	0.035	0.017
6545.V4.C1	26	50	50	50	0.095	0.068
0815.V3.C3	7.37	0.036	0.055	0.018	50	50
6095.V1.C10	0.147	0.464	0.601	0.096	0.242	0.023
3468.V1.C12	3.51	0.04	0.104	0.073	2.09	2.38
620345.c1	0.455	50	50	50	0.393	50
C1080.c3	0.056	1.5	0.539	0.096	0.004	0.001
C2101.c1	0.344	0.097	2.38	0.064	0.026	0.009

Variations of these neutralization values against different strains are complicated functions of ENV sequences, as noted in the introduction. Our idea was to model these complicated dependencies using the powerful learning capabilities of a feedforward neural network trained to minimize the error between the target and actual neutralization data and, further on, to use the generalization abilities of these networks to predict IC_{50} values for different strains.

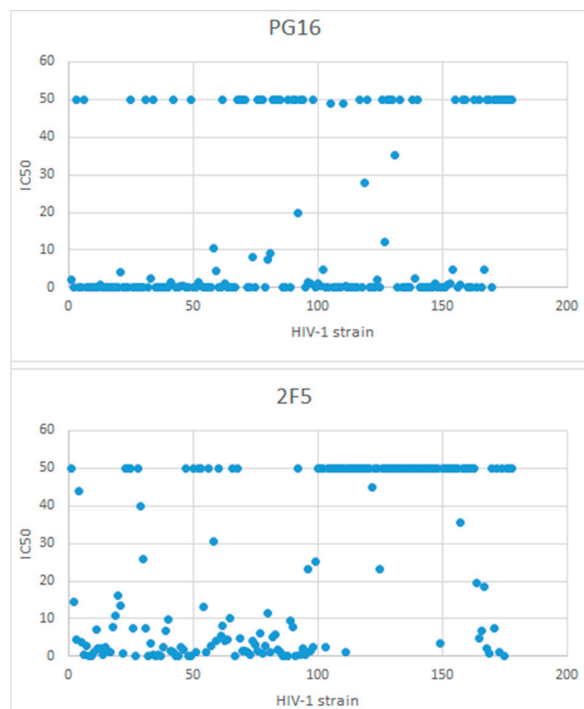


Figure 7. Distribution of IC₅₀ values for selected antibodies (PG16 above, and 2F5 below) against all the 178 strains in the sample file.

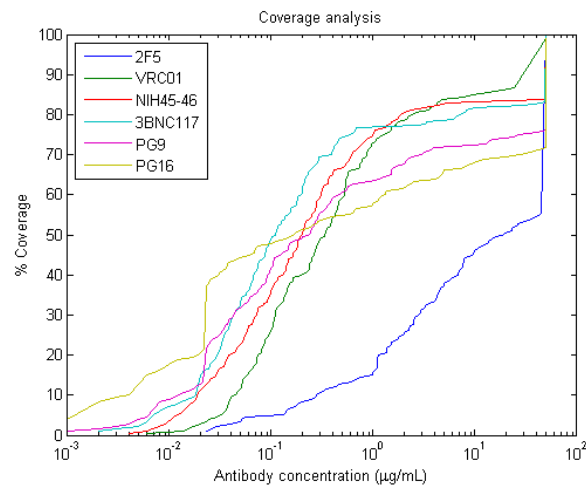


Figure 8. Coverage curves for given antibodies.

In Figure 9, we present the generic structure of our neural network as a function approximator between inputs (ENV primary structures) and outputs (IC₅₀ values).

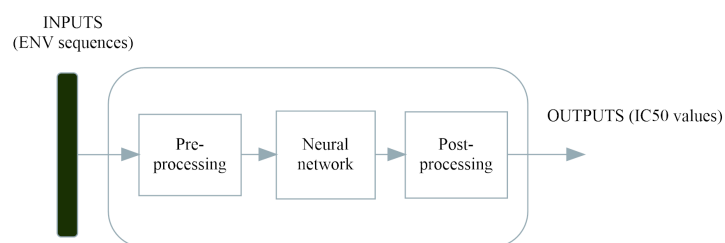


Figure 9. Modeling (learning) neutralization data using a feedforward neural network.

It is known that such a neural network may learn, in appropriate conditions, any nonlinear relationship between input and output data. The overall process has a number of steps which are detailed in the following for our specific application.

4.1. Collecting Data

Input (ENV sequences)/output (neutralization values) data is collected as indicated above.

4.2. Creating the Network

This step is about creating a Matlab neural network object using the predefined *fitnet* function which produces a feedforward neural network whose parameters will be specified during the next steps.

4.3. Configure the Network

At this step, we specify the number of inputs, number of hidden layer neurons, and number of outputs. As the input vectors are aligned ENV sequences of a length of 1369 characters, these 1369 positions are provided to the input neurons. The network can be fed with other aligned ENV sequences, and if the length of the alignment is different, then the number of input nodes will change accordingly. The input data has to be converted to a numeric format. This is done in a simple way by using the correspondence table in Figure 10, where B is D or N (aspartic), Z is E or Q (glutamic), X represents any amino acid, * represents an end terminator, - is a gap, and ? is an unknown amino acid. This mapping is the amino acid letter codes to integers coding used in the Bioinformatics Toolbox from Matlab.

A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V	B	Z	X	*	-	?
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	0

Figure 10. Correspondences between amino acids and integers in our codification scheme.

The network will have one output for to the IC₅₀ value corresponding to a particular antibody, e.g., 2F5. The number of neurons in the hidden layer is adjustable (the implicit value is 10). The output data is normalized to (0,1).

4.4. Initializing the Network

This is accomplished by generating random values for the network's weights and biases.

4.5. Training the Network

The goal is to learn, in the best possible way, the input–output relationship which is implicit in the ENV primary structure—neutralization data dependency. Training an artificial neural network generally means finding the optimal values for the network's weights and biases in order to minimize a performance index F , which is usually the mean square error (*mse*) as indicated below:

$$F = mse = \frac{1}{N} \sum_{i=1}^N (e_i)^2 = \frac{1}{N} \sum_{i=1}^N (t_i - a_i)^2 \quad (1)$$

where N is the number of input–output pairs, t (for target) is the desired output (experimental IC₅₀) of the network and a (for actual) is the actual output of the network.

Regularization is usually used for improving the generalization abilities of the neural network. In this case, the performance function (1) above is modified to:

$$msereg = \gamma \times msw + (1 - \gamma) \times mse \quad (2)$$

where:

$$msw = \frac{1}{n} \sum_{i=1}^n w_j^2 \quad (3)$$

and w are the network's weights, and γ is the performance ratio (usually 0.5).

The standard feedforward network training is based on the Levenberg–Marquardt algorithm [40], and other widely used alternatives are Bayesian regularization and BFGS quasi-Newton methods.

4.6. Validating the Network

After training has been finished, one has to check the network's performance. This can be done by checking the training record (an output of the training process together with the trained network) in order to see if changes with regards to training procedure, networks architecture and parameters, are needed. Dynamic values for the performance index and the gradient are also available from the training record. The next step is the generation of a regression plot (network response vs. corresponding targets), which provides a view of the dependency between the desired output and the actual output of the network. It is also possible to plot the error histogram plot which shows the distribution of the network errors.

4.7. Utilizing the Network

As previously indicated, the utility of having such a network available is the possibility to provide any new ENV sequence at its input and thus to predict the neutralization data for the modeled antibody (one of the six antibodies in our case: 2F5, VRC01, NIH45-46, 3BNC117, PG9 and PG16).

Acknowledgments: Cătălin Buiu acknowledges discussions with Anthony P. West from the Division of Biology of the California Institute of Technology, who kindly clarified some aspects of his approach to computational analysis of anti-HIV-1 antibody neutralization panel data and provided his computational tool together with neutralization data. Speranța Avram acknowledges the financial support from the ERA-NET 4-004/2013 grant.

Author Contributions: Cătălin Buiu and Mihai V. Putz proposed and designed the approach; Cătălin Buiu and Speranța Avram analyzed the results and refined the approach; Cătălin Buiu and Mihai V. Putz wrote the paper.

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

References

1. Julien, J.-P.; Cupo, A.; Sok, D.; Stanfield, R.L.; Lyumkis, D.; Deller, M.C.; Klasse, P.-J.; Burton, D.R.; Sanders, R.W.; Moore, J.P.; et al. Crystal structure of a soluble cleaved HIV-1 envelope trimer. *Science* **2013**, *342*, 1477–1483. [[CrossRef](#)] [[PubMed](#)]
2. Foley, B.; Apetrei, C.; Mizrachi, I.; Rambaut, A.; Korber, B.; Kuiken, C.; Leitner, T.; Hahn, B.; Mullins, J.; Wolinsky, S.; et al. *HIV Sequence Compendium 2012*; Los Alamos National Laboratory, Theoretical Biology and Biophysics: Los Alamos, NM, USA, 2012.
3. Pancera, M.; Majeed, S.; Ban, Y.-E.A.; Chen, L.; Huang, C.; Kong, L.; Do Kwon, Y.; Stuckey, J.; Zhou, T.; Robinson, J.E.; et al. Structure of HIV-1 gp120 with gp41-interactive region reveals layered envelope architecture and basis of conformational mobility. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 1166–1171. [[CrossRef](#)] [[PubMed](#)]
4. Merk, A.; Subramaniam, S. HIV-1 envelope glycoprotein structure. *Curr. Opin. Struct. Biol.* **2013**, *23*, 268–276. [[CrossRef](#)] [[PubMed](#)]
5. Huang, C.; Tang, M.; Zhang, M.-Y.; Majeed, S.; Montabana, E.; Stanfield, R.L.; Dimitrov, D.S.; Korber, B.; Sodroski, J.; Wilson, I.A.; et al. Structure of a V3-containing HIV-1 gp120 core. *Science* **2005**, *310*, 1025–1028. [[CrossRef](#)] [[PubMed](#)]
6. Heider, D.; Dybowski, J.N.; Wilms, C.; Hoffmann, D. A simple structure-based model for the prediction of HIV-1 co-receptor tropism. *BioData Min.* **2014**, *7*, 14–25. [[CrossRef](#)] [[PubMed](#)]
7. Cashin, K.; Sterjovski, J.; Harvey, K.L.; Ramsland, P.A.; Churchill, M.J.; Gorry, P.R. Covariance of charged amino acids at positions 322 and 440 of HIV-1 Env contributes to coreceptor specificity of subtype B viruses, and can be used to improve the performance of V3 sequence-based coreceptor usage prediction algorithms. *PLoS ONE* **2014**, *9*, e109771. [[CrossRef](#)] [[PubMed](#)]

8. Shityakov, S.; Dandekar, T. Lead expansion and virtual screening of Indinavir derivate HIV-1 protease inhibitors using pharmacophoric—Shape similarity scoring function. *Bioinformatics* **2010**, *4*, 295–299. [[CrossRef](#)] [[PubMed](#)]
9. Harte, W.E.; Swaminathan, S.; Beveridge, D.L. Molecular dynamics of HIV-1 protease. *Proteins* **1992**, *13*, 175–194. [[CrossRef](#)] [[PubMed](#)]
10. Shityakov, S.; Rethwilm, A.; Dandekar, T. Structural and docking analysis of HIV-1 integrase and Transportin-SR2 interaction: Is this a more general and specific route for retroviral nuclear import and its regulation? *Online J. Bioinform.* **2010**, *11*, 19–33.
11. Kwong, P.D.; Mascola, J.R. Human antibodies that neutralize HIV-1: Identification, structures, and B cell ontogenies. *Immunity* **2012**, *37*, 412–425. [[CrossRef](#)] [[PubMed](#)]
12. Chen, J.; Frey, G.; Peng, H.; Rits-Volloch, S.; Garrity, J.; Seaman, M.S.; Chen, B. Mechanism of HIV-1 neutralization by antibodies targeting a membrane-proximal region of gp41. *J. Virol.* **2014**, *88*, 1249–1258. [[CrossRef](#)] [[PubMed](#)]
13. Hessel, A.J.; Rakasz, E.G.; Pognard, P.; Hangartner, L.; Landucci, G.; Forthal, D.N.; Koff, W.C.; Watkins, D.I.; Burton, D.R. Broadly neutralizing human anti-HIV antibody 2G12 is effective in protection against mucosal SHIV challenge even at low serum neutralizing titers. *PLoS Pathog.* **2009**, *5*, e1000433. [[CrossRef](#)] [[PubMed](#)]
14. Yang, L.; Wang, P. Passive immunization against HIV/AIDS by antibody gene transfer. *Viruses* **2014**, *6*, 428–447. [[CrossRef](#)] [[PubMed](#)]
15. Kepler, T.B.; Liao, H.-X.; Alam, S.M.; Bhaskarabhatla, R.; Zhang, R.; Yandava, C.; Stewart, S.; Anasti, K.; Kelsoe, G.; Parks, R.; et al. Immunoglobulin gene insertions and deletions in the affinity maturation of HIV-1 broadly reactive neutralizing antibodies. *Cell Host Microbe* **2014**, *16*, 304–313. [[CrossRef](#)] [[PubMed](#)]
16. Wu, X.; Yang, Z.-Y.; Li, Y.; Hogerkorp, C.-M.; Schief, W.R.; Seaman, M.S.; Zhou, T.; Schmidt, S.D.; Wu, L.; Xu, L.; et al. Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. *Science* **2010**, *329*, 856–861. [[CrossRef](#)] [[PubMed](#)]
17. Scheid, J.F.; Mouquet, H.; Ueberheide, B.; Diskin, R.; Klein, F.; Oliveira, T.Y.K.; Pietzsch, J.; Fenyo, D.; Abadir, A.; Velinzon, K.; et al. Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. *Science* **2011**, *333*, 1633–1637. [[CrossRef](#)] [[PubMed](#)]
18. Wu, X.; Zhou, T.; Zhu, J.; Zhang, B.; Georgiev, I.; Wang, C.; Chen, X.; Longo, N.S.; Louder, M.; McKee, K.; et al. Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing. *Science* **2011**, *333*, 1593–1602. [[CrossRef](#)] [[PubMed](#)]
19. Walker, L.M.; Huber, M.; Doores, K.J.; Falkowska, E.; Pejchal, R.; Julien, J.-P.; Wang, S.-K.; Ramos, A.; Chan-Hui, P.-Y.; Moyle, M.; et al. Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nature* **2011**, *477*, 466–470. [[CrossRef](#)] [[PubMed](#)]
20. Bouvin-Pley, M.; Morgand, M.; Meyer, L.; Goujard, C.; Moreau, A.; Mouquet, H.; Nussenzweig, M.; Pace, C.; Ho, D.; Bjorkman, P.J.; et al. Drift of the HIV-1 envelope glycoprotein gp120 toward increased neutralization resistance over the course of the epidemic: A comprehensive study using the most potent and broadly neutralizing monoclonal antibodies. *J. Virol.* **2014**, *88*, 13910–13917. [[CrossRef](#)] [[PubMed](#)]
21. Kim, A.S.; Leaman, D.P.; Zwick, M.B. Antibody to gp41 MPER Alters Functional Properties of HIV-1 Env without Complete Neutralization. *PLoS Pathog.* **2014**, *10*, e1004271. [[CrossRef](#)] [[PubMed](#)]
22. Bradley, T.; Trama, A.; Tumba, N.; Gray, E.; Lu, X.; Madani, N.; Jahanbakhsh, F.; Eaton, A.; Xia, S.-M.; Parks, R.; et al. Amino acid changes in the HIV-1 gp41 membrane proximal region control virus neutralization sensitivity. *EBioMedicine* **2016**, *16*, 30402–30409. [[CrossRef](#)] [[PubMed](#)]
23. Ofek, G.; Zirkle, B.; Yang, Y.; Zhu, Z.; McKee, K.; Zhang, B.; Chuang, G.-Y.; Georgiev, I.S.; O'Dell, S.; Doria-Rose, N.; et al. Structural basis for HIV-1 neutralization by 2F5-like antibodies m66 and m66.6. *J. Virol.* **2014**, *88*, 2426–2441. [[CrossRef](#)] [[PubMed](#)]
24. Shmelkov, E.; Grigoryan, A.; Krachmarov, C.; Abagyan, R.; Cardozo, T.J. Sequence conserved and antibody accessible sites in the V1V2 domain of HIV-1 gp120 envelope protein. *AIDS Res. Hum. Retrovir.* **2014**, *30*, 927–931. [[CrossRef](#)] [[PubMed](#)]
25. Bouvin-Pley, M.; Morgand, M.; Moreau, A.; Jestin, P.; Simonnet, C.; Tran, L.; Goujard, C.; Meyer, L.; Barin, F.; Braibant, M. Evidence for a continuous drift of the HIV-1 species towards higher resistance to neutralizing antibodies over the course of the epidemic. *PLoS Pathog.* **2013**, *9*, e1003477. [[CrossRef](#)] [[PubMed](#)]

26. Frey, G.; Peng, H.; Rits-Volloch, S.; Morelli, M.; Cheng, Y.; Chen, B. A fusion-intermediate state of HIV-1 gp41 targeted by broadly neutralizing antibodies. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 3739–3744. [[CrossRef](#)] [[PubMed](#)]
27. Chakrabarti, B.K.; Walker, L.M.; Guenaga, J.F.; Ghobbeh, A.; Poignard, P.; Burton, D.R.; Wyatt, R.T. Direct antibody access to the HIV-1 membrane-proximal external region positively correlates with neutralization sensitivity. *J. Virol.* **2011**, *85*, 8217–8226. [[CrossRef](#)] [[PubMed](#)]
28. West, A.P.; Scharf, L.; Horwitz, J.; Klein, F.; Nussenzweig, M.C.; Bjorkman, P.J. Computational analysis of anti-HIV-1 antibody neutralization panel data to identify potential functional epitope residues. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 10598–10603. [[CrossRef](#)] [[PubMed](#)]
29. Montefiori, D.C. Evaluating neutralizing antibodies against HIV, SIV, and SHIV in luciferase reporter gene assays. *Curr. Protoc. Immunol.* **2005**, *12*. [[CrossRef](#)]
30. Kohavi, R.; Provost, F. Glossary of Terms. *Mach. Learn.* **1998**, *30*, 271–274.
31. Mitchell, T.M. *Machine Learning*; McGraw-Hill, Inc.: New York, NY, USA, 1997.
32. Al-Gharabli, S.I.; Al-Agtash, S.; Rawashdeh, N.A.; Barqawi, K.R. Artificial neural networks for dihedral angles prediction in enzyme loops: A novel approach. *Int. J. Bioinform. Res. Appl.* **2015**, *11*, 153–161. [[CrossRef](#)] [[PubMed](#)]
33. Ashtawy, H.M.; Mahapatra, N.R. BgN-Score and BsN-Score: Bagging and boosting based ensemble neural networks scoring functions for accurate binding affinity prediction of protein-ligand complexes. *BMC Bioinform.* **2015**, *16*, 8–20. [[CrossRef](#)] [[PubMed](#)]
34. Kourou, K.; Exarchos, T.P.; Exarchos, K.P.; Karamouzis, M.V.; Fotiadis, D.I. Machine learning applications in cancer prognosis and prediction. *Comput. Struct. Biotechnol. J.* **2015**, *13*, 8–17. [[CrossRef](#)] [[PubMed](#)]
35. Schneider, G.; Wrede, P. Artificial neural networks for computer-based molecular design. *Prog. Biophys. Mol. Biol.* **1998**, *70*, 175–222. [[CrossRef](#)]
36. Douali, L.; Villemin, D.; Ziad, A.; Cherqaoui, D. Artificial neural networks: Non-linear QSAR studies of HEPT derivatives as HIV-1 reverse transcriptase inhibitors. *Mol. Divers.* **2004**, *8*, 1–8. [[CrossRef](#)] [[PubMed](#)]
37. Winkler, D.A. Neural networks as robust tools in drug lead discovery and development. *Mol. Biotechnol.* **2004**, *27*, 139–168. [[CrossRef](#)]
38. Durrant, J.D.; McCammon, J.A. NNscore: A neural-network-based scoring function for the characterization of protein-ligand complexes. *J. Chem. Inf. Model.* **2010**, *50*, 1865–1871. [[CrossRef](#)] [[PubMed](#)]
39. Shah, J.V. Chi-Sang Poon Linear independence of internal representations in multilayer perceptrons. *IEEE Trans. Neural Networks* **1999**, *10*, 10–18. [[CrossRef](#)] [[PubMed](#)]
40. Marquardt, D.W. An algorithm for least-squares estimation of nonlinear parameters. *J. Soc. Ind. Appl. Math.* **1963**, *11*, 431–441. [[CrossRef](#)]
41. Hagan, M.T.; Menhaj, M.B. Training feedforward networks with the Marquardt algorithm. *IEEE Trans. Neural Networks* **1994**, *5*, 989–993. [[CrossRef](#)] [[PubMed](#)]
42. Milac, A.-L.; Avram, S.; Petrescu, A.-J. Evaluation of a neural networks QSAR method based on ligand representation using substituent descriptors: Application to HIV-1 protease inhibitors. *J. Mol. Graph. Model.* **2006**, *25*, 37–45. [[CrossRef](#)] [[PubMed](#)]
43. Calborean, O.; Mernea, M.; Avram, S.; Mihailescu, D.F. Pharmacological descriptors related to the binding of Gp120 to CD4 corresponding to 60 representative HIV-1 strains. *J. Enzym. Inhib. Med. Chem.* **2013**, *28*, 1015–1025. [[CrossRef](#)] [[PubMed](#)]
44. Wold, S.; Sjostrom, M.; Eriksson, L. PLS-regression: A basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109–130.
45. Qian, N.; Sejnowski, T.J. Predicting the secondary structure of globular proteins using neural network models. *J. Mol. Biol.* **1988**, *202*, 865–884. [[CrossRef](#)]
46. Heider, D.; Hoffmann, D. Interpol: An R package for preprocessing of protein sequences. *BioData Min.* **2011**, *4*, 16–22. [[CrossRef](#)] [[PubMed](#)]
47. Kuiken, C.; Korber, B.; Shafer, R.W. HIV sequence databases. *AIDS Rev.* **2003**, *5*, 52–61. [[PubMed](#)]
48. Diskin, R.; Scheid, J.F.; Marcovecchio, P.M.; West, A.P.; Klein, F.; Gao, H.; Gnanapragasam, P.N.P.; Abadir, A.; Seaman, M.S.; Nussenzweig, M.C.; et al. Increasing the potency and breadth of an HIV antibody by using structure-based rational design. *Science* **2011**, *334*, 1289–1293. [[CrossRef](#)] [[PubMed](#)]

49. Diskin, R.; Klein, F.; Horwitz, J.A.; Halper-Stromberg, A.; Sather, D.N.; Marcovecchio, P.M.; Lee, T.; West, A.P.; Gao, H.; Seaman, M.S.; et al. Restricting HIV-1 pathways for escape using rationally designed anti-HIV-1 antibodies. *J. Exp. Med.* **2013**, *210*, 1235–1249. [[CrossRef](#)] [[PubMed](#)]
50. Mouquet, H.; Scharf, L.; Euler, Z.; Liu, Y.; Eden, C.; Scheid, J.F.; Halper-Stromberg, A.; Gnanapragasam, P.N.P.; Spencer, D.I.R.; Seaman, M.S.; et al. Complex-type N-glycan recognition by potent broadly neutralizing HIV antibodies. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 3268–3277. [[CrossRef](#)] [[PubMed](#)]
51. Huang, J.; Ofek, G.; Laub, L.; Louder, M.K.; Doria-Rose, N.A.; Longo, N.S.; Imamichi, H.; Bailer, R.T.; Chakrabarti, B.; Sharma, S.K.; et al. Broad and potent neutralization of HIV-1 by a gp41-specific human antibody. *Nature* **2012**, *491*, 406–412. [[CrossRef](#)] [[PubMed](#)]
52. Doria-Rose, N.A.; Louder, M.K.; Yang, Z.; O'Dell, S.; Nason, M.; Schmidt, S.D.; McKee, K.; Seaman, M.S.; Bailer, R.T.; Mascola, J.R. HIV-1 neutralization coverage is improved by combining monoclonal antibodies that target independent epitopes. *J. Virol.* **2012**, *86*, 3393–3397. [[CrossRef](#)] [[PubMed](#)]
53. Buiu, C. Neutralization Data and Aligned ENV Sequences for Predicting Antibody Affinities Using Artificial Neural Networks. Mendeley Data. Available online: <http://dx.doi.org/10.17632/bhcjwjtwh4.1> (accessed on 8 October 2016).



© 2016 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 (<http://creativecommons.org/licenses/by/4.0/>).