2018

Prediction of Novel Anti-Ebola Virus Compounds Utilizing Artificial Neural Network (ANN)

Ronald Bartzatt
University of Nebraska at Omaha, rbartzatt@unomaha.edu

WORLD JOURNAL OF PHARMACEUTICAL RESEARCH
SJIF Impact Factor 8.074
Volume 7, Issue 13, 16-34. Research Article ISSN 2277–7105

PREDICTION OF NOVEL ANTI-EBOLA VIRUS COMPOUNDS UTILIZING ARTIFICIAL NEURAL NETWORK (ANN)

Ronald Bartzatt*

Chemistry Department, Durham Science Center, University of Nebraska at Omaha 6001 Dodge Street, Omaha Nebraska, USA.

Article Received on
06 May 2018,

Revised on 27 May 2018,
Accepted on 18 June 2018

DOI: 10.20959/wjpr201813-12749

***Corresponding Author**
Dr. Ronald Bartzatt
Chemistry Department,
Durham Science Center,
University of Nebraska at
Omaha 6001 Dodge Street,
Omaha Nebraska, USA.

ABSTRACT

Artificial Neural Network (ANN) analysis is shown to predict the molecular properties of new anti-EBOLA compounds following training/learning by use of 60 previously known and studied drugs. Following training/learning by applying properties of 60 known drugs the TIBERIUS ANN system can efficiently predict the molecular properties of comparable new drugs. Molecular weight (MW) is an important and dominant property of perspective drugs considered for clinical trials. TIBERIUS ANN was able to predict comparable values of MW for drugs following training cycles. One-way ANOVA, F and T tests indicate that actual and predicted MW have the same means (*P*=.99). Passing-Bablok regression showed that ANN predicted MW are comparable to actual MW. The coefficient of variation indicated actually less variation in predicted MW as opposed to actual MW. A plot of actual MW values versus ANN predicted MW values, produced a line having no departure from linearity (P=.82), and a 95% ellipses having 55 drugs therein. TIBERIUS ANN allows investigators to input separate property values to predict suitable outcome based on the 60 known drugs. ANN prediction of pharmaceutical properties of new drugs is shown to be efficient and accurate when based on a known set of drugs for training/learning cycle.

KEYWORDS: Ebola, Artificial Neural Network, Viral Hemorrhagic Fever.

INTRODUCTION

The typical organization of neural networks are in layers. These layers are composed of interconnected nodes which contain an activation function. An activation function of a specific node will define the output of that node for the given input (or set of inputs).[1]

Respective patterns are presented to the network through the input layer, followed by communication to one or more of hidden layers, in which the actual processing is accomplished through a system of weighted connections.[1,2] These hidden layers are linked to an output layer from which the answer is ejected.

Artificial neural networks (ANN) are successfully being used for drug design and drug discovery process. For this purpose there are various types of ANN:[2] 1) Multilayered perceptron/backpropagation networks; 2) Kohenen neural networks; 3) Counter propagation networks; 4) Bayesian neural networks; 5) Recurrent neural network. ANN analysis has the capacity to resolve nonlinear relationships, making them advantageous compared to other statistical techniques incorporating pattern analysis.[2] For this study, a particular feature advantageous to ANN is their ability to model complex relationships among various physicochemical properties (i.e. molecular properties), in addition to, mapping correlations of several variables of the particular chemical structures.[2]

Artificial neural networks can form a mathematical model for the way a drug is transported to a target cell or target tissue.[3] This is accomplished by ANN through the solving of nonlinear problems of multivariate and multi-response type systems, which is the scenario for drug transport.[3] ANN analysis can be used to model complex relationships among actual causal variable inputs and resulting outputs.[3] ANN has been used for the prediction of in vitro dissolution profiles of various matrix-controlled release formulations.[4] Attempts to pharmacokinetic modeling pursue the prediction of concentration time profiles of a drug while it is in the body after drug administration. This is the study of the absorption, distribution, metabolism, and elimination (ADME) processes of the drug.[5] The relationship between drug concentration and its positive/negative response can be studied via the simulation and modeling techniques provided by ANN.[5]

This study will utilize the molecular properties of 60 known effective and tested inhibitors of EBOLA virus.[6,7] By applying pattern recognition power of ANN, followed by various numerical analysis methods, the efficiency and usefulness of ANN for prediction of new ever very important anti-EBOLA compounds is accomplished. The power of drug development is expanded and made more efficacious by use of ANN capabilities.

MATERIALS AND METHODS

Molecular Properties and Molecular Modeling

Numerical values of molecular properties (i.e. Log P, polar surface are, molecular weight) for all compounds were determined through heuristic calculation through Molinspiration Chemical Properties Service (Molinspiration Cheminformatics, Nova ulica 61, SK-900 26 Slovensky Grob, Slovak Republic). Elucidation of molecular structural components was accomplished utilizing ACD ChemSketch Modeling v. 12.01 (Advanced Chemistry Development, 110 Yonge Street, Toronto Ontario, M5C 1T4 Canada, http://www.molinspiration.com/services/search.html). Molinspiration Cheminformatics (http://www.molinspiration.com/cgibin/properties) determined molecular properties of Log P, polar surface area (Angstroms^2), molecular weight, number of atoms, molecular volume (Angstroms^3), number of nitrogen, oxygen, amine groups, and hydroxyl groups.

Statistical Analysis and Prediction

Prediction of numerical values for various molecular properties was accomplished by utilizing TIBERIUS v. 7.0.7 (http://www.tiberius.biz/, copyright 2001, Copyright © Tiberius Data Mining, and Copyright © 1999-2007 NeuSolutions). TIBERIUS trial version can be obtained at http://www.tiberius.biz/download.html. Statistical analysis of numerical data to include molecular properties of the compounds in this study was accomplished by Microsoft EXCEL v.14.0.6112.5000 (EXCEL Professional plus 2010). Other statistical tests such as ANOVA analysis, F and T test, coefficient of variation, 95% ellipses, and Kruskal-Wallis test was accomplished utilizing PAST version 2.06 (copyright Oyvind Hammer, D.A.T. Harper, 2011). Numerical outliers were identified, where stated, by using Grubb's analysis for outliers online GraphPad (http://www.graphpad.com/quickcalcs/). Passing-Bablok regression was accomplished utilizing Method Validator (www.multiqc.com). Box plots were accomplished utilizing Smith's Statistical Package v. 2.5 (copyright © 1995-2001, Gary Smith).

RESULTS AND DISCUSSION

Ebola virus disease, also known as Ebola haemorrhagic fever, is a severe and often fatal illness in humans. Direct contact with various body fluids is considered a major risk factor for contracting and induction of this illness.[6] Further evidence of contracting the disease is presence of blood abnormalities, such as decreased platelets.[6] Other organs involved include renal, cardiac, lung, gastrointestinal, neurological and hepatic tissue acting as indicators of

the virus presence.[6] A high fever and hemorrhagic manifestations (hemorrhagic conjunctivitis, bleeding ulcerations of mouth and lips, gingival bleeding, hematemesis, ear bleeding, hematuria) are most often observed in patients.[6]

Infection with this virus is characterized by a massive production of pro-inflammatory cytokines, a severe host immunosuppression that accompanies a rapid viremia, with manifestation of a fulminant hemorrhagic fever.[7] The apparent high virulence accompanying rapid progression of infection, contributes to the high fatality rate.[7] For this study, utilizing TIBERIUS ANN based prediction, the molecular properties of 60 compounds that are known to inhibit the proliferation of viral EBOLA have been recognized and compiled.[6,7] These 60 compounds have molecular structures that are presented in Fig 1 and Fig 2. These 60 compounds originate from more than 50 chemical classes of compounds and have been studied because of their antiviral activity.[6] Among these 60 compounds are counted approved drugs, antiviral agents in clinical trials, various lead compounds, various exploratory chemical probes and others from screening hits.[7] Discussed and examined in terms of their molecular properties in previous studies,[6,7] this work will endeavor to demonstrate that ANN is suitable for prediction of new anti-EBOLA compounds when utilizing the molecular properties in Table 1 for “training” the ANN.

All 60 compounds presented in Fig 1, Fig 2, and Table 1 have molecular weight less than 600 daltons, which is considered to be “small molecules” in the fields of pharmacology and molecular biology.[8, 9] Most drugs are small molecules.[8, 9] The molecular weight cut-off of 900 Daltons is a necessary condition for favorable bioavailability in terms of transcellular transport.[8] Small molecule drugs over larger molecules is the ease and efficiency of oral administration.[10] Larger molecules generally requiring injection or other parenteral administration (intramuscular, subcutaneous, and intravenous).[10]

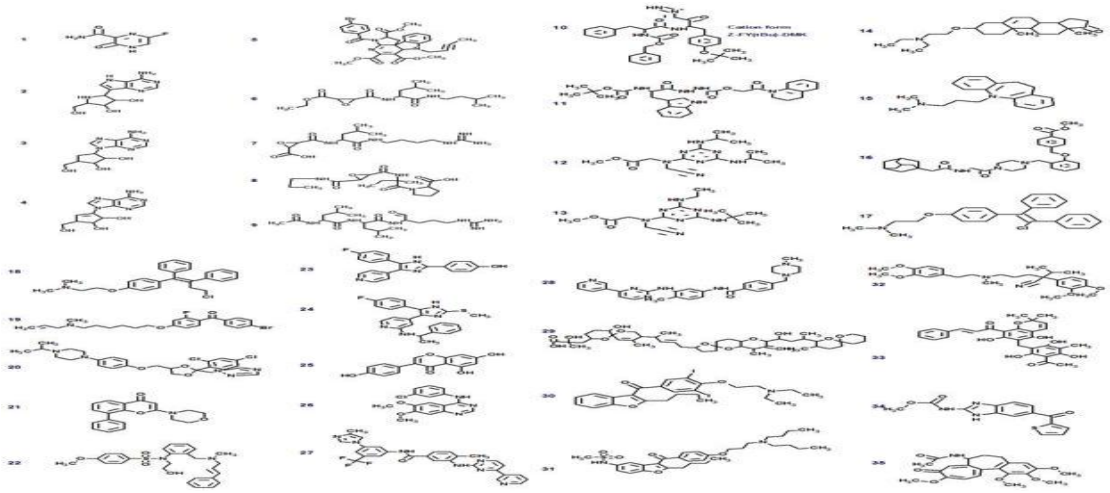

Figure. 1: Compounds 1 to 35 that inhibit growth of EBOLA virus.

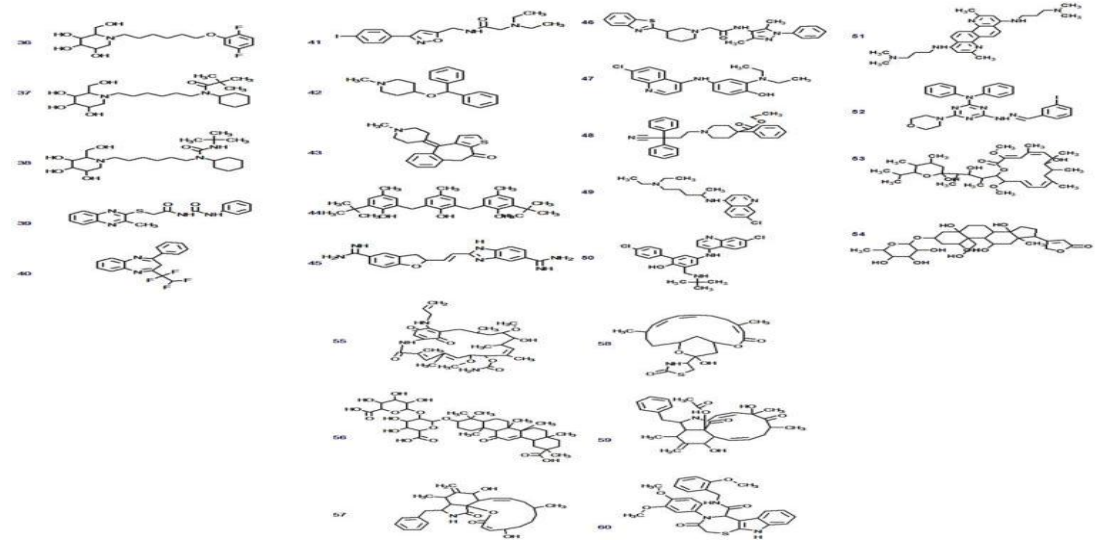

Figure. 2: Compounds 36 to 60 that inhibit EBOLA virus.

Table. 1. Molecular Properties of 60 Known Anti-EBOLA Compounds.

Compound	Log P	Polar Surface Area (Angstroms^2)	Number of Atoms	Molecular Weight	Oxygen Nitrogen Atoms	Number of OH and NH_n Groups	Rule of 5 Violations	Rotatable Bonds	Volume (Angstroms^3)
1	-0.98	88.85	11	157.1	5	3	0	1	119.06
2	-2.24	140.31	19	265.27	8	7	1	2	225.72
3	-1.13	130.32	19	265.27	8	5	0	2	226.35
4	-1.14	130.32	19	263.26	8	5	0	2	220.13
5	4.31	114.83	41	620.46	10	0	1	8	492.1
6	1.88	97.03	24	342.44	7	2	0	11	335.67
7	-1.37	169.93	25	357.41	10	7	1	12	330.98
8	-1.93	128.33	26	369.42	9	3	0	8	338.96
9	0.16	166.27	30	426.56	10	7	1	15	422.72
10	1.93	131.69	40	543.64	9	3	1	14	508.68
11	4.95	138.1	39	537.62	11	4	2	10	490.72
12	2.45	116.06	23	321.38	9	2	0	9	304.06
13	2.67	116.06	23	321.38	9	2	0	9	303.7
14	4.64	29.54	28	387.61	3	0	0	6	405.62
15	4.16	6.48	21	280.42	2	0	0	4	287.31
16	5.04	88.18	42	573.73	8	1	2	11	543.75
17	5.78	12.47	27	377.92	2	0	1	7	356.31
18	6.06	12.47	29	405.97	2	0	1	9	389.91
19	6.58	29.54	28	448.38	3	0	1	12	380.93
20	4.55	64.9	36	532.47	8	0	1	8	462.72
21	3.64	42.68	23	307.35	4	0	0	2	278.13
22	4.48	70.08	33	466.6	6	1	0	11	428.61
23	5.07	48.91	25	330.36	3	2	1	3	292.05
24	5.22	53.6	29	404.51	4	2	1	6	360.35
25	2.27	90.89	20	270.24	5	3	0	1	224.05
26	3.93	56.28	22	315.76	5	1	0	4	268.16
27	4.99	97.63	39	529.53	8	2	1	7	446.63
28	3.89	86.28	37	493.62	8	2	0	7	461.44
29	6.11	162.62	58	817.07	12	4	3	11	786.22

30	8.31	42.68	31	645.32	4	0	2	11	437.04
31	7.94	88.85	39	556.77	7	1	2	18	533.46
32	4.55	63.97	33	454.61	6	0	0	13	454.3
33	5.72	144.52	38	516.55	8	5	2	6	457.44
34	2.79	84.09	21	301.33	6	2	0	4	246.9
35	1.1	83.11	29	399.44	7	1	0	5	364.15
36	1.33	93.38	26	375.41	6	4	0	9	338.63
37	2.35	104.46	30	428.61	7	4	0	10	436.32
38	2.72	116.49	31	443.63	8	5	0	10	448.73
39	3.13	83.98	25	352.42	6	2	0	4	305.39
40	4.84	24.73	23	320.29	2	0	0	3	260.99
41	3.58	58.37	22	413.26	5	1	0	7	301.35
42	3.16	12.47	21	281.4	2	0	0	4	283.8
43	3.48	20.31	22	309.43	2	0	0	0	284.23
44	8.93	60.68	34	460.66	3	3	1	6	466.57
45	0.47	137.67	26	346.39	7	7	1	4	306.64
46	3.77	63.05	32	445.59	6	1	0	5	405.9
47	5.38	48.38	24	341.84	4	2	1	5	308.75
48	5.96	53.34	34	452.6	4	0	1	9	440.81
49	5	28.16	22	319.88	3	1	1	8	313.12
50	7.67	57.17	32	466.41	4	3	1	6	409.58
51	4.72	56.31	34	458.65	6	2	0	10	458.26
52	6.6	78.78	35	577.43	8	1	2	7	436.37
53	4.8	114.69	44	620.87	8	3	1	7	630.37
54	-2.18	206.6	41	584.66	12	8	3	4	520.54
55	2.11	166.29	43	599.73	11	5	2	7	570.06
56	1.97	267.04	58	822.94	16	8	3	7	741.93
57	4.11	95.86	35	479.62	6	3	0	2	460.9
58	3.52	84.86	29	421.56	6	2	0	1	391.39
59	2.89	112.93	37	507.63	7	3	1	4	479.11
60	3.84	92.9	37	517.61	8	2	1	7	453.35

The TIBERIUS ANN has been structured and is ready to be trained. Training process includes assigning initial weights for the inputs, which in this study is set to be all equal (setting to 1.00 each). This training is set to a learning pace of 0.70 and to run over a desired number of "EPOCHS", with the error minimized. There are two approaches to training - supervised and unsupervised.[1, 2] Supervised training requires a mechanism of providing the network with a manual grading of network performance or by providing desired outputs along with the inputs. Unsupervised training pushes the network to make sense of the inputs without outside influence. The vast majority of networks utilize supervised training.[1, 2]

Prediction of molecular weights for novel anti-EBOLA compounds due to ease of ANN application and the importance of molecular weight in determining the efficacy of potential drug compounds. Various screening tools for identifying potential drugs will include molecular weight as important criteria for selection. The very popular "Rule of 5" has a criteria of molar mass must be less than 500 Daltons.[11] An extension of the Rule of 5 is the Ghose criteria (CMC Rule) that requires a molecular weight of 180 to 500 Daltons.[12] The QED criteria includes molecular weight in its own parameters.[13] The BDDCS criteria, along with the ECCCS criteria consider a molecular weight of 400 Daltons to be important.[14] This in addition to the finding that more than 80% of all traded drugs have molecular weight below 450 Daltons.[11] Therefore, molecular weight is an excellent choice for utilizing ANN to screen for novel drug properties following training with the properties in Table 1, for 60 proven EBOLA inhibitors.

The flow chart of TIBERIUS ANN compilation of data to "Results and Query Table" is presented in Fig 3. The steps in Fig 3 were applied to this study and should be effective in general application of thisback-propagating ANN methodology.

Presented in Fig 4 is a standard spreadsheet for input of data/information into TIBERIUS ANN, which allows the selection of desired variables (descriptors) for appropriate training and selection of output variables (molecular weight for this study). Note that the first row is appropriate indicators of the molecular properties (see Table 1) (i.e. PSA= polar surface area, natoms = number of atoms, MW = molecular weight, nON = number of oxygen nitrogen, nrotb = number of rotatable bonds, volume = molecular volume). These are selected for input or for output.

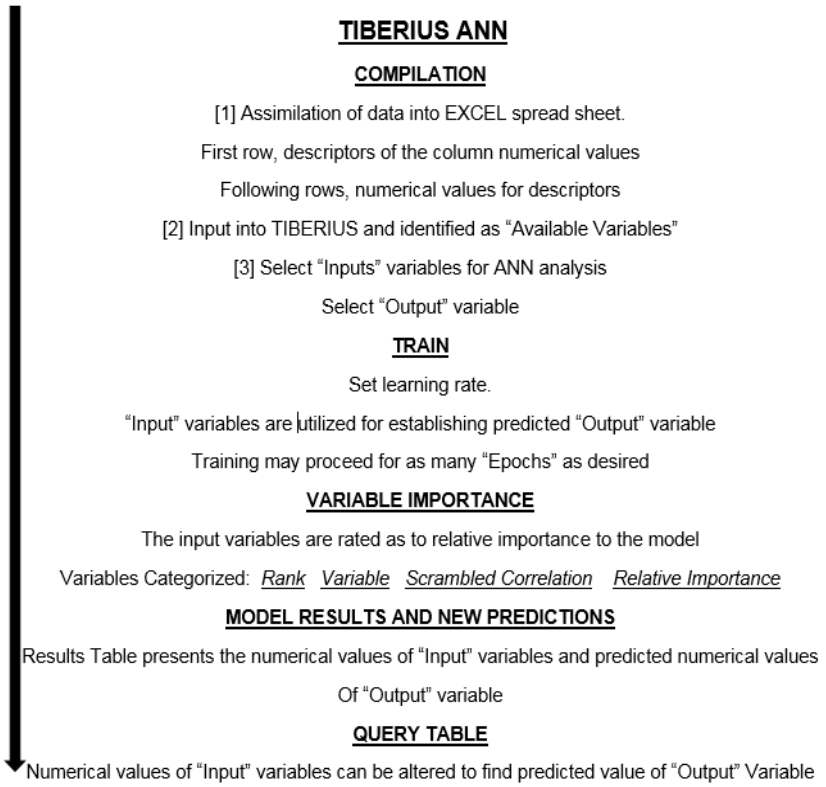

Figure. 3: Flow chart of TIBERIUS ANN application for implementing known molecular properties for training, followed by prediction of new properties of potential drugs.

	A	B	C	D	E	F	G	H	I
1	Log P	PSA	natoms	MW	nON	nOHNH	rule of 5	nrotb	volume
2	-0.98	88.85	11	157.1	5	3	0	1	119.06
3	-2.24	140.31	19	265.27	8	7	1	2	225.72
4	-1.13	130.32	19	265.27	8	5	0	2	226.35
5	-1.14	130.32	19	263.26	8	5	0	2	220.13
6	4.31	114.83	41	620.46	10	0	1	8	492.1
7	1.88	97.03	24	342.44	7	2	0	11	335.67
8	-1.37	169.93	25	357.41	10	7	1	12	330.98
9	-1.93	128.33	26	369.42	9	3	0	8	338.96
10	0.16	166.27	30	426.56	10	7	1	15	422.72
11	1.93	131.69	40	543.64	9	3	1	14	508.68
12	4.95	138.1	39	537.62	11	4	2	10	490.72
13	2.45	116.06	23	321.38	9	2	0	9	304.06
14	2.67	116.06	23	321.38	9	2	0	9	303.7
15	4.64	29.54	28	387.61	3	0	0	6	405.62
16	4.16	6.48	21	280.42	2	0	0	4	287.31
17	5.04	88.18	42	573.73	8	1	2	11	543.75
18	5.78	12.47	27	377.92	2	0	1	7	356.31
19	6.06	12.47	29	405.97	2	0	1	9	389.91
20	6.58	29.54	28	448.38	3	0	1	12	380.93
21	4.55	64.9	36	532.47	8	0	1	8	462.72
22	3.64	42.68	23	307.35	4	0	0	2	278.13

Sheet1

Figure. 4: Input EXCEL spread sheet having all pertinent variables (descriptors) for ANN training and analysis.

Following selection of input variables and out variable (molecular weight in this study), the model is initiated into training utilizing input variables from spreadsheet of molecular properties shown in Fig 4, these derived from the compilation in Table 1. Following training, the model will identify variables having highest importance for generation of results. The highest importance level is 1.000, followed by various levels of numerical values less than 1.00. For this study the outcome of relative importance for variable and relative importance, respectively: Log P (1.000), number of oxygen & nitrogen (0.994), polar surface area (0.459), number of –OH & nHNn (0.050), and number of rotatable bonds (0.002).

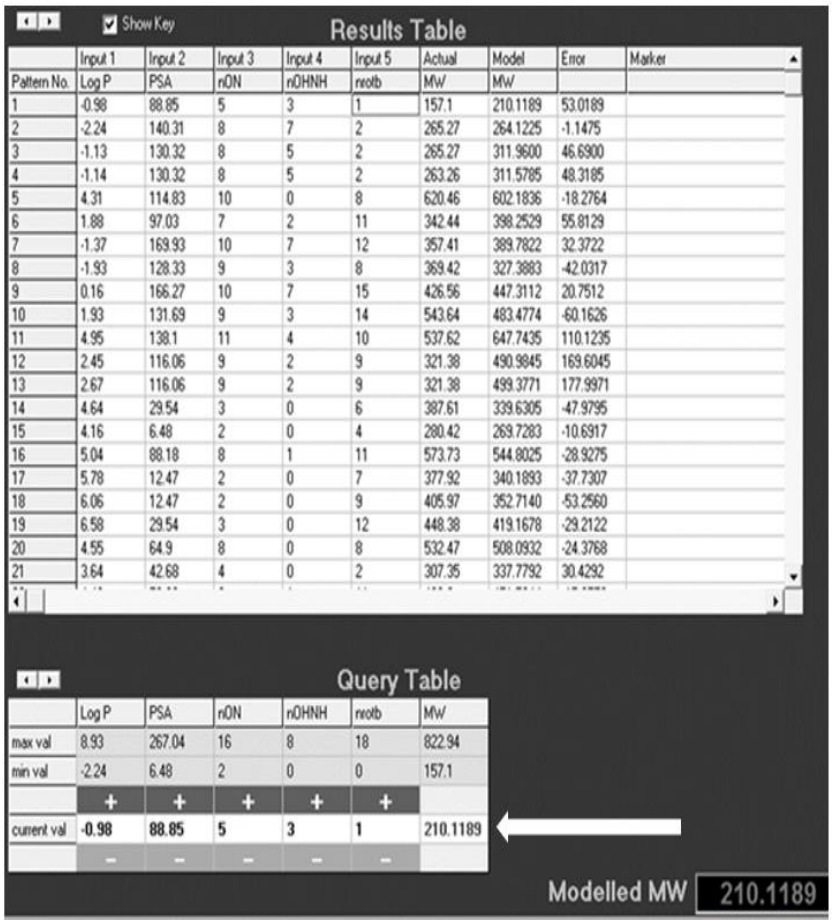

Figure. 5: Results Table and Query Table where new molecular property values can be entered into "current val" row to determine best fit molecular weight (MW) based on ANN analysis of 60 proven anti-EBOLA virus compounds (see inset arrow).

In the "Results and Query" window the final predicted molecular weight values are compared to the actual molecular weight values of the training set, and with the numerical values of error calculated (Fig 5). The actual molecular weight values and predicted molecular weight values are compiled in Table 2 for comparison and statistical analysis. The actual molecular weights of compounds 1 to 60, having structures presented in Fig 1 and Fig 2, and nine

molecular properties in Table 1, with ANN predicted molecular weights are shown in Table 2. The following statistical evaluation of the predicted MW to actual MW will show that ANN is effective, efficient, and possessing high level of accuracy.

Table. 2. Comparison of Actual values of MW to ANN Predicted Values of MW.

Compound	Actual MW	Predicted MW	Compound	Actual MW	Predicted MW	Compound	Actual MW	Predicted MW
1	157.1	210.11	21	307.35	337.77	41	413.26	374.92
2	265.27	264.12	22	466.6	451.52	42	281.4	237.47
3	265.27	311.95	23	330.36	357.00	43	309.43	253.71
4	263.26	311.57	24	404.51	397.16	44	460.66	510.94
5	620.46	602.18	25	270.24	336.10	45	346.39	339.69
6	342.44	398.25	26	315.76	383.45	46	445.59	411.98
7	357.41	389.78	27	529.53	540.84	47	341.84	397.20
8	369.42	327.38	28	493.62	487.71	48	452.6	443.22
9	426.56	447.31	29	817.07	744.10	49	319.88	346.18
10	543.64	483.47	30	645.32	524.22	50	466.41	486.47
11	537.62	647.74	31	556.77	635.49	51	458.65	438.54
12	321.38	490.98	32	454.61	457.68	52	577.43	591.37
13	321.38	499.37	33	516.55	590.93	53	620.87	542.72
14	387.61	339.63	34	301.33	386.72	54	584.66	434.03
15	280.42	269.72	35	399.44	356.92	55	599.73	556.72
16	573.73	544.80	36	375.41	329.45	56	822.94	762.80
17	377.92	340.18	37	428.61	407.24	57	479.62	439.16
18	405.97	352.71	38	443.63	452.59	58	421.56	412.57
19	448.38	419.16	39	352.42	399.59	59	507.63	438.31
20	532.47	508.09	40	320.29	312.71	60	517.61	492.32

Carrying out the One-way ANOVA evaluation of actual and predicted MW, indicated that the overall means for the two groups are equal (*P*=.99).[15] Applying the F and T test, the outcome shows that actual and predicted MW values have the same mean (*P*=.99) and the two population variances are equal (*P*=.36).[15] The Pearson r correlation between the two sets of values, actual and predicted, shows very strong positive relationship (r = 0.8879). The paired t-test and Wilcoxon indicated that the sets of MW values have the same median (P=.69) and same median (P=.13).[15] The Kolmogorov-Smirnov test indicates the two sets of MW values are taken from populations having equal distributions (*P*=.72). In addition, the Kruskal-Wallis test indicates the two sets have equal medians (*P*=.99).

Both sets of MW values were examined for numerical outliers within the training MW (actual MW) and predicted MW. The Grubbs' test, also called the ESD method (extreme studentized deviate), determines whether one of the values in either list of MW values is a signficant

outlier from the rest.[15] The Grubb's test did not find any outliers within the 60 values of actual MW or the predicted MW values.

The coefficient of variation (CV) is a measure of relative variability. It is the ratio of the standard deviation to the mean (average).[15] CV is useful for comparing results from two different tests that have differing values. The higher the numerical values of CV, the higher the variation. The actual MW of 60 anti-EBOLA compounds showed a CV of 30.15%, whereas the predicted MW values showed CV of 26.78%. Therefore, the variation in MW of the ANN predicted values is less than the variation in actual MW for these 60 compounds.

The simple two-way plot of actual MW versus predicted MW is shown in Fig 6. Here is seen the regression line through the origin intercept (solid line) with equation (y = 0.9828x) with Pearson r = 0.8581. Interpretation of this model means that for every one point increase in actual MW values there is a 0.99 (nearly 1.00) for predicted MW, or in other words a linear slope of 1.00. This is accomplished by ANN prediction of MW, noting the slope of the origin intercepting regression line is 0.9828 (see Fig 6). The non-origin intercept regression model (dotted line) with equation (y = 0.7885x + 91.52) with Pearson r =0.8879. This line is also shown to have 34 runs by the runs test (test to determined test whether the line fit by linear regression deviates systematically your data), indicating no significant departure from linearity (P=.82).[15] In both cases, the Pearson r-values indicate a very strong positive relationship.

Further interpretation of the effectiveness and accuracy of ANN prediction of MW for anti-EBOLA compounds can be achieved by applying Passing-Bablok regression analysis.[16] Passing–Bablok regression is a statistical method for non-parametric regression analysis suitable for method comparison studies.[16] The outcome are interpreted so that if 0 is in the confidence interval for the y-axis intercept, and 1 is in the confidence interval of the slope, the two methods are comparable within the investigated range. If 0 is not in the confidence interval of the y-axis intercept then there is a systematic difference, however, if 1 is not in the confidence interval of.

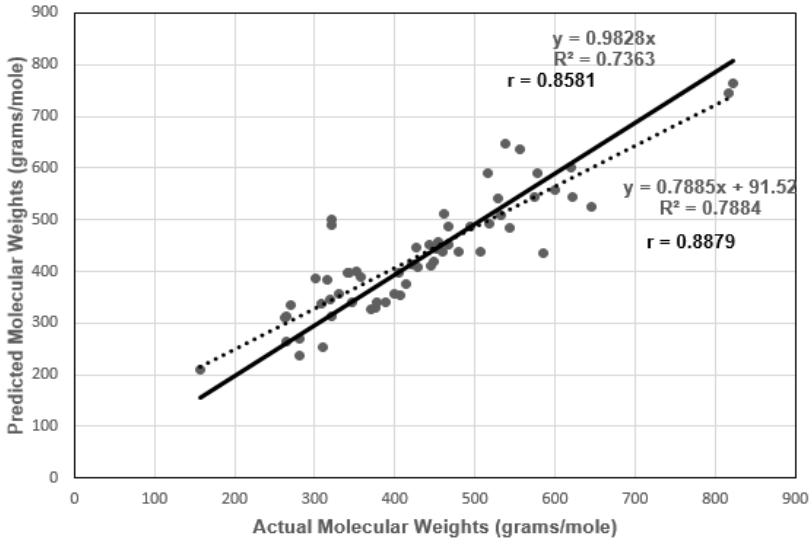

Figure. 6: Linearity of actual values molecular weights (x-axis) to predicted values molecular weights (y-axis). Pearson r correlation of line with regression through origin is 0.8561 (y = 0.9828x), slope equal to 0.9828 which is > 98% of 1.000. Pearson r correlation for line with non-zero y-axis intercept line is 0.8879 (y=0.7885x + 91.52).

The slope then there is a proportional difference between the two methods.[16] Passing-Bablok plot of actual MW versus predicted MW (see Fig 7) shows that 1 is in the confidence interval of the slope and 0 is in the confidence interval of the y-axis intercept. Therefore, the actual and predicted MW values are systematically the same and there is no proportional difference between the two sets of MW values.

Box plots are drawn for groups of data to enable the study of the distributional characteristics of more than one group. They are excellent for detecting and illustrating location and variation changes between different groups of data.[15] Any outliers are identified as individual points and the line in the box represents the median. The actual MW values and predicted MW are presented in the form of Box plots in Fig 8. Clearly, the median lines indicated are essentially coincident with predicted values of MW contained within the box for actual values of MW. This finding corroborates the outcome of coefficient of variation as well as Passing-Bablok regression analysis.

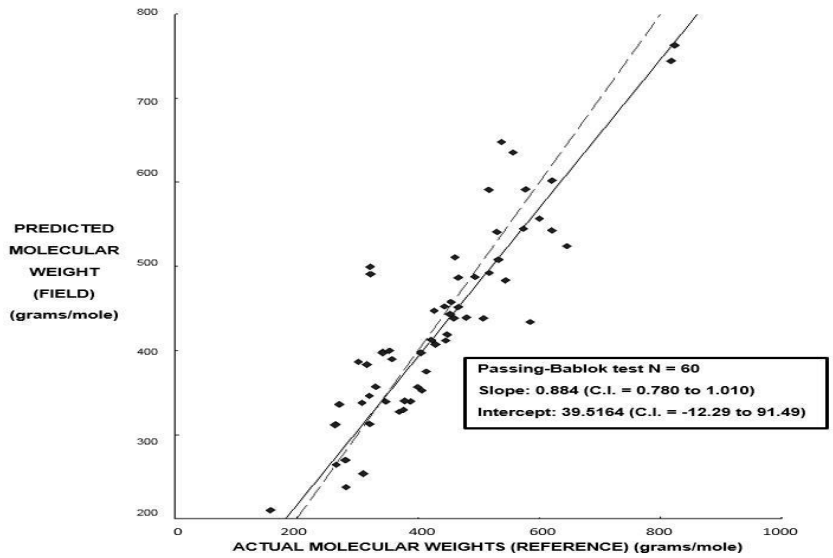

Figure. 7: Passing-Bablok plot comparing actual molecular weights to predicted molecular weight obtained from TIBERIUS ANN analysis.

An additional graphical approach for visualizing the similarity of actual MW and ANN predicted MW is through 95% confidence ellipses, presented in Fig 9. Essentially a 95% ellipses can be applied if you were to replicate your sampling from the underlying distribution many times and each time calculate a confidence ellipse, then 95% of the ellipses so constructed would contain the underlying mean and demonstrate consistency of numerical values.[15]

The two-way plot of actual MW values (independent variable) versus predicted MW values by TIBERIUS ANN (dependent variable) is shown in Fig 9. The position of each compound within the 95% ellipses is indicated by using the numbering system in Table 1. Note that the plotted points for compounds 13, 12, 29, 11, and 56 are the only drugs that fall outside the 95% ellipses. This finding further corroborates the overall effectiveness of ANN prediction of MW for anti-EBOLA compounds by modeling with TIBERIUS system.

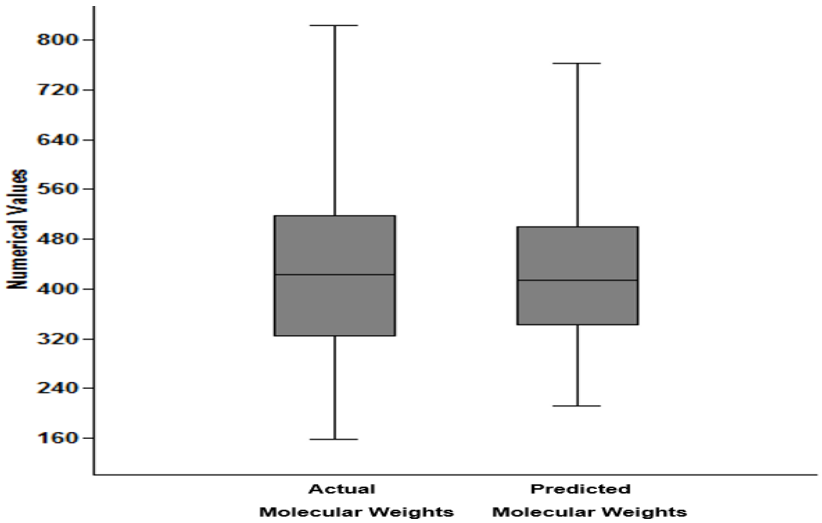

Figure. 8: Box plots for comparing actual molecular weights to predicted molecular weights obtained from ANN analysis of 60 known anti-EBOLA compounds.

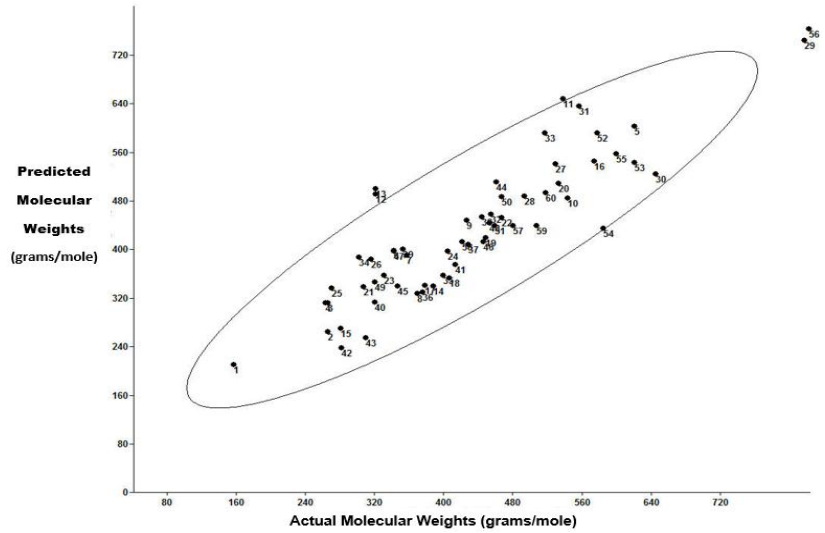

Figure. 9: 95% ellipses comparing actual molecular weights to predicted molecular weights.

The Query Table (see inset arrow for Fig 5) allows the investigator to input differing values of critical properties, and then determine the corresponding MW. This is a powerful capability of TIBERIUS ANN that allows direct prediction of the expected structural molecular properties of potential drugs from the 60 training known anti-EBOLA inhibitors (see Fig 1, Fig 2, and Table 1). Examples shown in Table 3, show that the change of critical descriptors Log P, polar surface area, number of oxygen & nitrogen, number of –OH and –NHn, and rotatable bonds, will result in the predicted molecular weight values. This capability allows investigators to build potential drug candidates based on 60 known anti-EBOLA compounds and with rational predicted MW, PSA, etc.

Table. 3. Example of Property Prediction from Query Table.

Log P	Polar Surface Area (Angstroms^2)	Number of oxygen & Nitrogen Atoms	Number of -OH and –NH_n	Number of Rotatable Bonds	Predicted Molecular Weight (grams/mole)
-0.98	88.85	5	3	1	210.1189
-0.98	88.85	10	3	1	345.3661
-0.98	88.85	10	5	1	330.0396
3.5	180	10	5	1	590.6643
3.5	180	10	5	12	600.8023
3.5	30	10	5	12	512.2134
3.5	90	5	5	12	376.9663
3.5	90	5	2	12	399.9559
2.5	88.85	4	3	1	315.8251

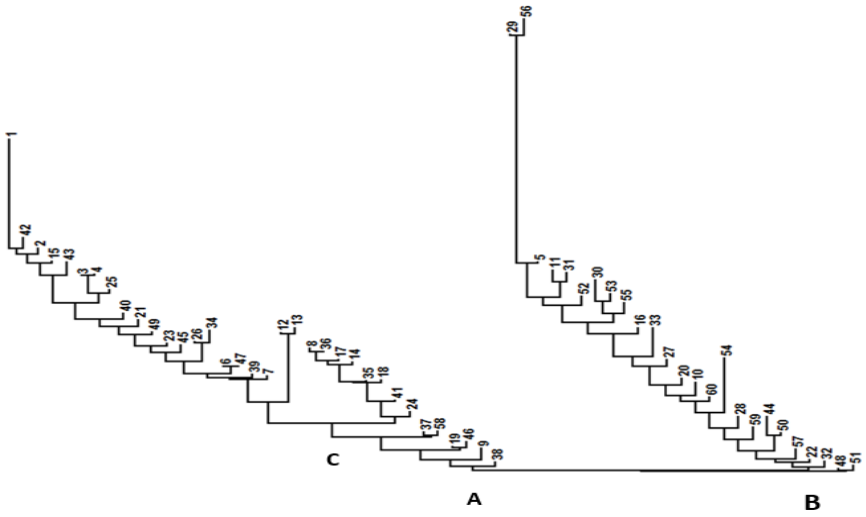

Figure. 10: Neighbor joining cluster analysis. Drugs have closest proximity are most similar.

The interrelationships of all 60 anti-EBOLA compounds presented in Fig 1, Fig 2, and Table 1 can be visualized by neighbor-joining cluster analysis, which is presented in Fig 10. Neighbor joining cluster analysis is a bottom-up (agglomerative) clustering method for the creation of trees to find the members of the various branches of the tree that have the highest similarity to each other.[17] In the neighbor-joining method, a modified distance matrix is constructed in which the separation between each pair of nodes is adjusted based on their average divergence from all other nodes.[17] In this manner, the distance of members of the tree can be optimized and be based on their numerical similarities in the various molecular properties utilized in this study. Neighbor cluster analysis in Fig 10, produced super nodes A and B, and whereas, node A is further divided into two sub-branches at node C. Compounds found in each of these major branches are determined to be most similar to each other based

on molecular properties (see Table 1), with further discrimination into sub-nodes where compounds closest to each other are most similar. Presumably, this will allow expectations for pharmacokinetic activity and which compound would have activity.

The application of TIBERIUS ANN for training with 60 known anti-EBOLA compounds (see Fig 1 and Fig 2) allowed the prediction of molecular weights that were statistically the same as the original training set (see Table 1). Various statistical analysis proved the numerical efficiency of ANN prediction of molecular property. In addition, the TIBERIUS ANN system permits the investigator to input various numerical values for critical properties, followed by accurate prediction of the desired "output" variable (molecular weight for this study). Although applied in this study for the prediction of new properties of new anti-EBOLA compounds, this same approach can be applied to other drugs that are clinically proven for the treatment of many other threatening diseases (i.e. malaria). The prediction, identification, and study of novel drugs for clinical application is an ever important endeavor within the field of medicinal chemistry. ANN is now shown to be a powerful tool for that endeavor.

CONCLUSION

Results of this study showed that TIBERIUS ANN can effectively predict important molecular properties of novel anti-EBOLA compounds following training utilizing properties of 60 known drugs that inhibit proliferation of EBOLA virus. Specifically demonstrated for values of molecular weight, TIBERIUS ANN can be utilized to achieve the same objective for any desired descriptor of new drugs. TIBERIUS allows the input of varying numerical values of any property placed into the "input" nodes, for consequential effect on "output" node (molecular weight for this study). Predicted MW values were shown to be statistically the same as actual MW of the training set by one-way ANOVA, F and T test. Comparison of actual and predicted molecular weight values were shown to be highly comparable with equal means by Passing-Bablok regression and one-way ANOVA, respectively. Box plots and coefficient of variation showed predicted and actual molecular weights to be comparable and effective. The use of ANN is a powerful tool for predicting properties of novel drug designs when effective training/learning data is administered to "input" nodes.

ACKNOWLEDGEMENTS

The College of Arts & Sciences, Durham Science Center, University of Nebraska at Omaha, Omaha, Nebraska 68182 USA, supported this project.

REFERENCES

1. Adya M, Collopy F. How effective are neural networks at forecasting and prediction? A review and evaluation. Journal of Forecasting, 1998; 17: 481-495.
2. Mandlik V, Bejugam PR, Singh S. Application of artificial neural networks in modern drug discovery. Artificial neural network for drug design, New York; Elsevier: 2015, pp. 123-130.
3. Chatterjee S, Pandya AS. Artificial neural networks in drug transport modeling and simulation-II. Artificial neural network for drug design, New York; Elsevier: 2015, pp. 243-250.
4. Das M, Chakraborty T. ANN in pharmaceutical product and process development. Artificial neural network for drug design, New York; Elsevier: 2015, pp. 277-282.
5. Kareem SS, Pathak Y. Clinical applications of artificial neural networks in pharmacokinetic modeling. Artificial neural network for drug design, New York; Elsevier: 2015, pp. 393-398.
6. Bartzatt R. Properties and drug-likeness of compounds that inhibit ebola virus disease (EVD. International Journal of Tropical Disease & Health, 2016; 15(2): 1-17.
7. Picazo D, Giordanetto F. Small molecule inhibitors of ebola virus infection. Drug Discovery Today, 2015; 20(2): 277-286.
8. Macie lag M. Chemical properties of antibacterials and their uniqueness. in T. Dougherty, M Pucci (ed), Antibiotic Discovery and Development, New York; Springer: 2012, pp. 801-802.
9. Veber DF, Johnson SR, Cheng HY, Smith BR, Ward KW, Kopple K. Molecular properties that influence the oral bioavailability of drug candicates. Journal of Medicinal Chemistry, 2002; 45(12): 2615-2623.
10. Samanen J. Chapter 5.2 How do SMDs differ from biomolecular drugs. in C. Ganelin, R. Fefferis, S. Roberts (ed), Introduction to biological and small molecule drug research and development, New York; Academic Press: 2013, pp. 5-35.
11. Lipinski C, Lombardo F, Dominy B, Feeney P. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Advanced Drug Delivery Reviews, 2001; 46(1-3): 3-26.
12. Ghose A, Viswanadhan V, Wendoloski J. A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A quantitative characterization of known drug databases. Journal Comb Chemistry, 1999; 1(1): 55-68.

13. Bickerton G, Paolin G, Bisnard J, Muresan S, Hopkins A. Quantifying the chemical beauty of drugs. Nature Chemistry, 2012; 4: 90-98.
14. Benet L, Hosey C, Ursu O, Oprea T. BDDCS, the rule of 5 and drug-ability. Advanced Drug Delivery Review, 2016; 101: 89-98.
15. Zar J. Biostatistical analysis, New York; NY: Prentice Hall: 1996.
16. Passing H, Bablok W. A new biometrical procedure for testing the equality of measurement from two different analytical Methods. Application of linear regression procedures for method comparison studies in clinical chemistry, part I. Journal of Clinical Chemistry & Clinical Biochemistry, 1983; 21: 709-720.
17. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution, 1987; (4): 406-425.