wjpmr, 2025, 11(9), 266-273

WORLD JOURNAL OF PHARMACEUTICAL AND MEDICAL RESEARCH
www.wjpmr.com

SJIF Impact Factor: 6.842

Research Article
ISSN (O): 2455-3301
ISSN (P): 3051-2557

ANALYSIS OF 20 ANTI-EBOLA AGENTS AND USE FOR MULTIPLE REGRESSION ANALYSIS TO PREDICT POTENTIAL ANTI-EBOLA AGENTS

***Ronald Bartzatt**

University of Nebraska at Omaha, 6001 Dodge Street, Chemistry Department, Omaha, Nebraska 68182 USA.

***Corresponding Author: Ronald Bartzatt**

University of Nebraska at Omaha, 6001 Dodge Street, Chemistry Department, Omaha, Nebraska 68182 USA.

Article Received on 25/06/2025 **Article Revised on 15/07/2025** **Article Accepted on 04/08/2025**

ABSTRACT

Algorithmic multiple regression is utilized for analysis of molecular properties and for the predictions of potential anti-EBOLA compounds. In this study, is the analysis of 20 tested anti-EBOLA agents, specifically their molecular properties, with multiple regression analysis for predicting the critical property of molecular weight of potential anti-EBOLA compounds. The molecular properties used are: Log P, molecular volume, polar surface area, and number of atoms. Using these molecular properties, multiple regression analysis successfully generated an equation having R^2 of 0.9318, and therefore accounting for 93.18 % of the variance in the dependent variable of molecular weight. Descriptive statistics of molecular properties for these 20 anti- EBOLA compounds are presented. Application of prediction methods will assist in the development of novel pharmaceuticals, as contemporary techniques are well characterized and have been applied across a broad spectrum of disciplines. Pattern recognition techniques such as cluster analysis, 95% ellipses, and non-metric multidimensional scaling are applied to determine complex relationships found in this multivariate data, such as similarities (or dissimilarities). Descriptive statistics of molecular properties for these 20 anti- EBOLA compounds are presented.

KEYWORDS: EBOLA, multiple regression, pattern recognition.

INTRODUCTION

Multiple regression analysis is utilized in analysis of complex relationships and prediction, as well as understanding how multiple independent variables may influence a single outcome.[1] The single outcome is referred to as the dependent variable that is influenced by multiple independent variables.[1] Multiple regression analysis are successfully being used for drug design and drug discovery process, as well as making forecasts, and testing existing theories.[1]

EBOLA haemorrhagic fever, or EBOLA virus disease (EVD), or simply EBOLA disease, is a severe and often fatal illness in humans that can follow after direct contact with various contaminated body fluids.[2] The illness is characterized by a high fever with various and multiple hemorrhagic manifestations (hemorrhagic conjunctivitis, bleeding ulcerations of mouth and lips, gingival bleeding, hematemesis, ear bleeding, hematuria).[2] The illness can bring about death, many times due to a shock effect resulting from fluid loss, that is often observed from six to 16 days following the first symptoms appearing.[2] High virulence accompanies the rapid progression of the infection, contributing to the high fatality rate.[2] Additional clinical evidence of virus activity is characterized by massive production of pro-inflammatory cytokines, a strong and severe host immunosuppression, rapid viremia, and manifestation of a fulminant hemorrhagic fever.[2]

This study demonstrates that multiple regression analysis of calculated molecular properties of tested anti- EBOLA compounds.[3] Multiple regression analysis can result in an algorithm that could be used to predict molecular properties of similar compounds. Previous studies have shown that a multitude of compounds having anti-EBOLA activity exist, where in production of the virus can be substantially reduced and inhibited.[3]

The molecular structures and determined molecular properties of 20 known anti-EBOLA compounds are presented in this study. These properties are determined in this study through heuristic calculation from available previously tested and proven heuristic platforms. Descriptive statistics of properties are presented. Other studies concerning anti-EBOLA compounds have been completed and have evaluated the drug-likeness of their molecular properties and the potential of the compounds for clinical use as anti-viral pharmaceuticals.[4]

MATERIALS AND METHODS

Molecular Properties and Modeling

The numerical values for molecular properties (i.e. Log P, polar surface are, molecular weight, number of atoms, molecular volume, number of nitrogen, oxygen, amine groups, and hydroxyl groups), for all compounds were determined by Molinspiration Chemical Properties Service, Nova ulica 61, SK-900 26 Slovensky Grob, Slovak Republic (www.molinspiration.com/ cgibin/properties) and/or Mcule (mcule.com, Palo Alto California USA). Elucidation of molecular structures was done by ACD ChemSketch Modeling v. 12.01 (Advanced Chemistry Development, 110 Yonge Street, Toronto Ontario, M5C 1T4 Canada).

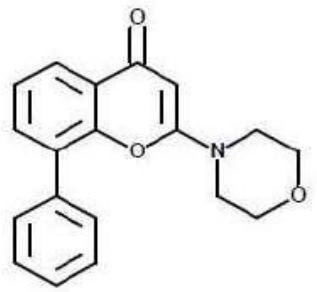
Statistical Analysis and Prediction

Multiple regression analysis as well as numerical statistics (i.e. mean, median, test for normality, standard deviation, minimum and maximum numerical values) was accomplished utilizing Instat Statistical Analysis Package (www. graphpad.com). Pattern recognition analysis for 95% ellipses analysis and non-metric multidimensional scaling was accomplished utilizing PAST version 2.06 (copyright Oyvind Hammer, D.A.T. Harper, 2011). Cluster analysis was accomplished utilizing KyPlot v. 2.0 Beta 15 (copyright Koichi Yoshioka 1997-2001).

RESULTS AND DISCUSSION

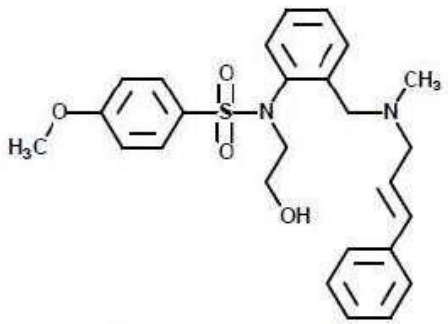
The EBOLA virus inducing severe illness in humans is one of several viruses within the genus *Ebolavirus* that are generally considered to cause EBOLA viral disease in humans.[2,5] Four of the five EBOLA viruses induce a very severe and often fatal hemorrhagic fever in humans and mammals.[2,5] Effective clinical treatment of viral infection is a vital component for surviving a proliferating illness. Studies have been made for identifying compounds that may serve as clnincal pharmaceutical tools in the treatment of this viral illness and multiple compounds have been identified as to having some efficacy for inhibiting the virus.[3] Having identified the structure of the compounds.[3] it is possible to determne the molecular properties utilizing heuristic platforms available, in order to evaluate drug-likeness and make statistical comparisons.[4] The molecular structures (with SMILES notation, Simplified Molecular Input Entry System) of these 20 agents are presented in Fig. 1, 2, and 3. What is immediately observed of the structures presented in Fig. 1, 2, 3, are that the compounds 1 to 20 have a highly diversified molecular framework. This fact is encouraging, for it implies that multiple compounds of various classes could be useful for the clinical treatment of this highly lethal pathogen. Virtually every compound 1 to 20, has a cyclic carbon ring and/or aromatic carbon ring. There is a prolific presence of amine groups (proton donors and acceptors). Relative size, indicated by molecular volume, indicates agent 5 is the smallest (molecular volume is 224.05 Angstroms^3), whereas agent 9 is the largest (molecular volume is 786.22 Angstroms^3), a very considerable variation is size (see Table 1). Visual inspection of the structures suffices to recognize the broad diversity of these compounds.

1



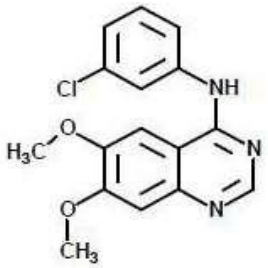
O=C2C=C(Oc1c(cccc12)c3ccccc3)N4CCOCC4

2



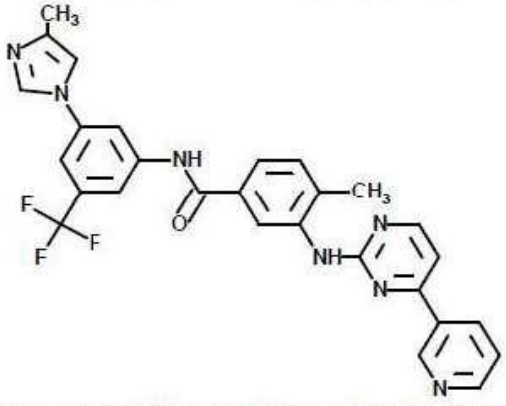
O=S(=O)(N(CCO)c2ccccc2CN(C)C/C=C/c1ccccc1)c3ccc(OC)cc3

6



Clc1cccc(c1)Nc3ncnc2cc(OC)c(cc23)OC

7



Cc3ccc(cc3Nc1nc(ccn1)c2cccnc2)C(=O)Nc4cc(cc(c4)n5cc(C)nc5)C(F)(F)F

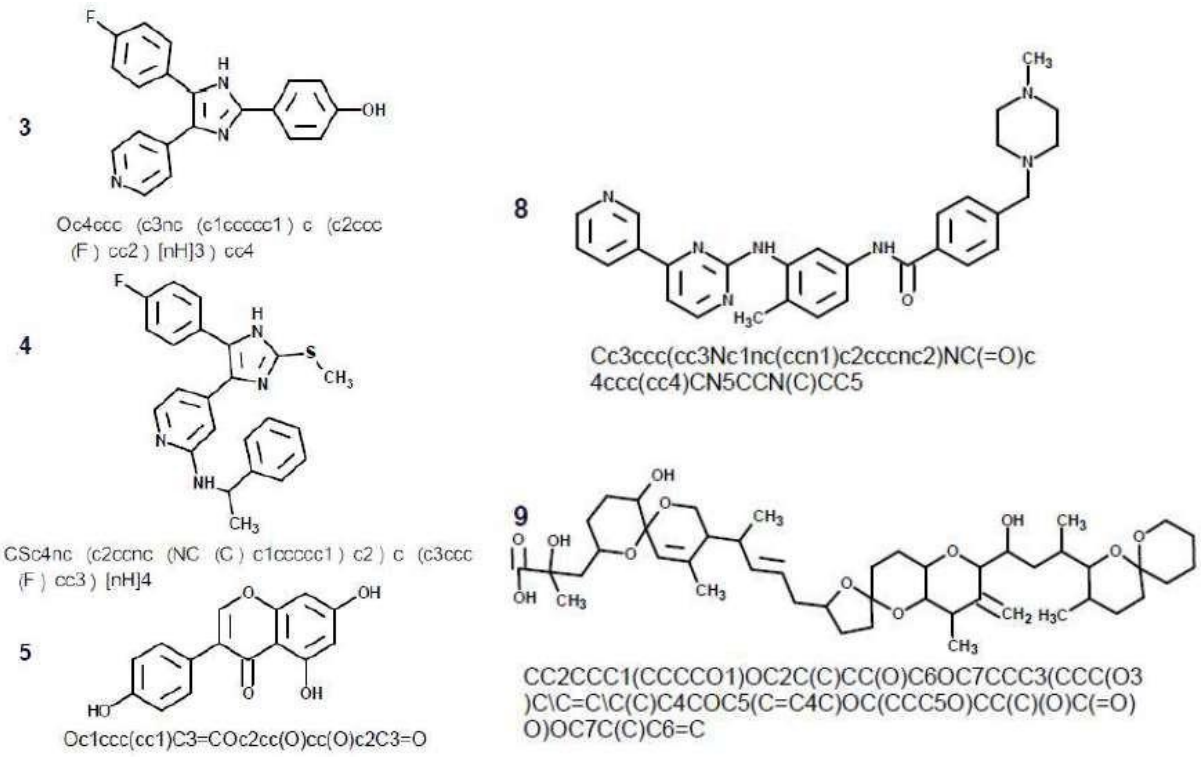
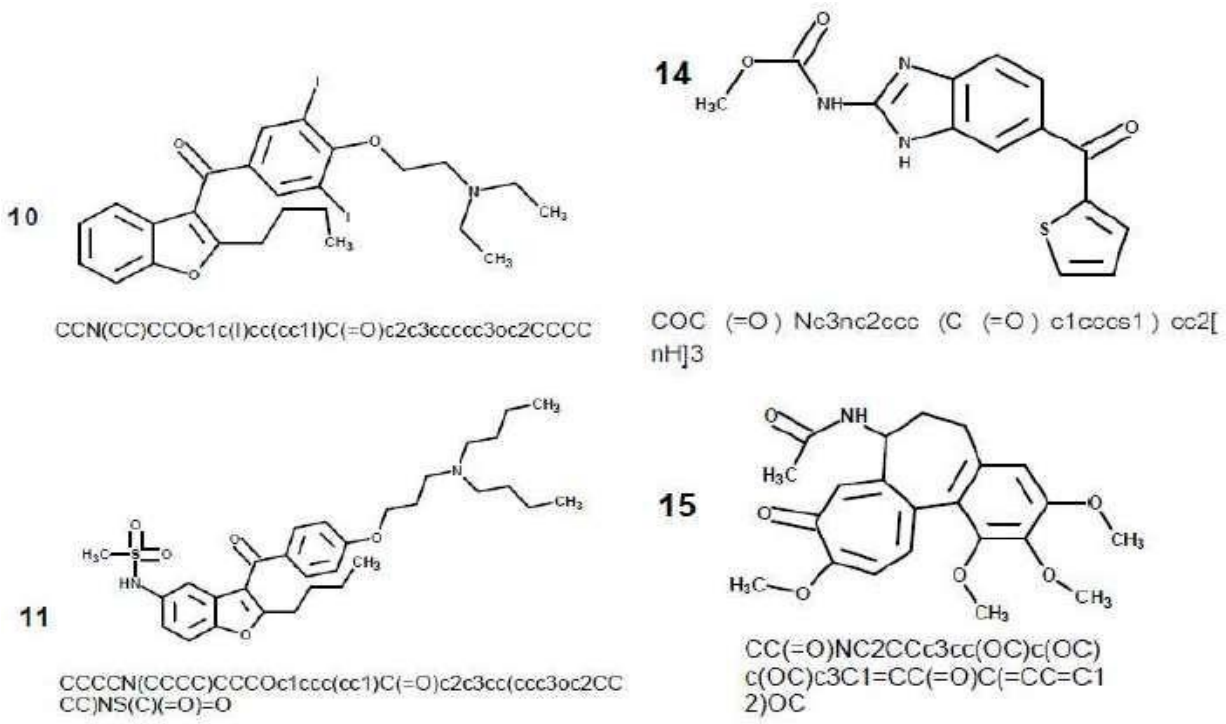


Figure 1: Molecular structures of anti-EBOLA agents 1 to 9.[3] Agent 9 has the largest molecular weight and largest molecular volume of this set of 20 anti-EBOLA compounds. Note that there is also a substantial number of amine groups (-NHn). Virtually every compound also has a cyclic carbon ring and/or aromatic carbon ring.



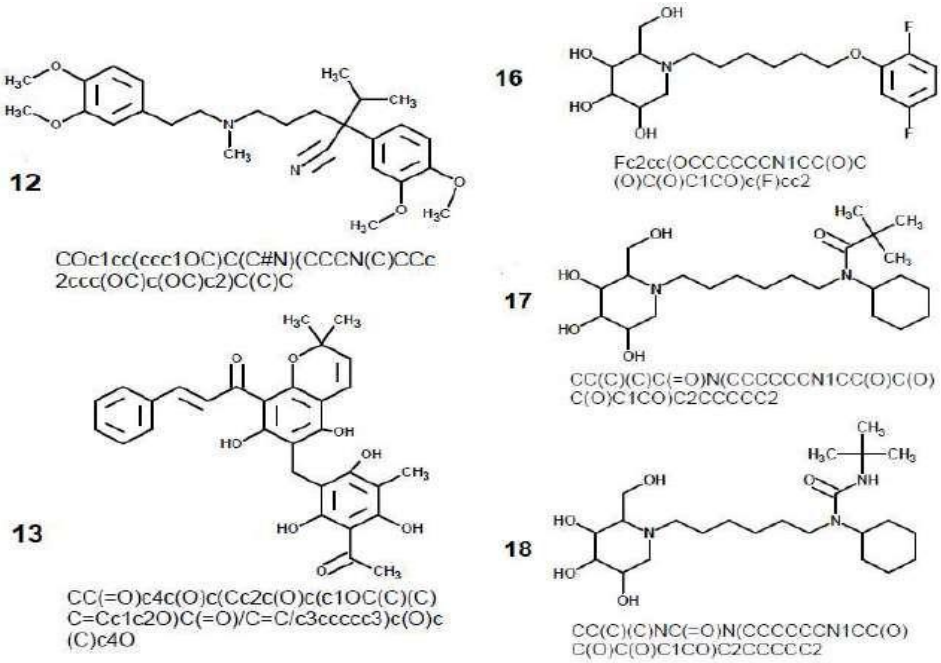


Figure 2: Molecular structures of anti-EBOLA agents 10 to 18.[3] Virtually every compound also has a cyclic carbon ring and/or aromatic carbon ring. There are a considerable number of -OH functional groups within this set of compounds.

The molecular properties of this diverse set of 20 anti-EBOLA compuonds have been determined and presented in Table 1. From these numercal values it is possible to evaluate their relative drug-likeness and therefore their potential as tools for clinical application. Various criteria for evaluating drug-likeness have been studied, with useful results, one study showed that molecules having 10 or fewer rotatable bonds and a polar surface area (PSA) of 140 Angstroms^2 or less will show favorable oral absorption (Veber et al.).[6]

Other screening mechanisms for identifying compounds having favorable drug-likenss is the Rule of 5 (Lipinski et al.).[7] This guideline holds to the following parameters, for favorable drug-likeness if there are no more than one violation of the Rule of 5. The Rule of 5 is as follows: 1) Molecular mass should be less than 500 Daltons; 2) No more than 5 hydrogen bond donors and 10 nitrogen or oxygen atoms total in structure; 3) Calculated value of Log P less than 5.[7] Additionally, another criteria (Ghose et al.) focusing on oral bioavailability, requires: 1) Molecular weight between 160 to 480; 2) Log P between -0.4 to 5.6; 3) Molar refractivity (relationship of the polarizability of a molecule) from 40 to 130 cm^3/mole; and 4) Total atoms in the molecule of 20 to 70.[8]

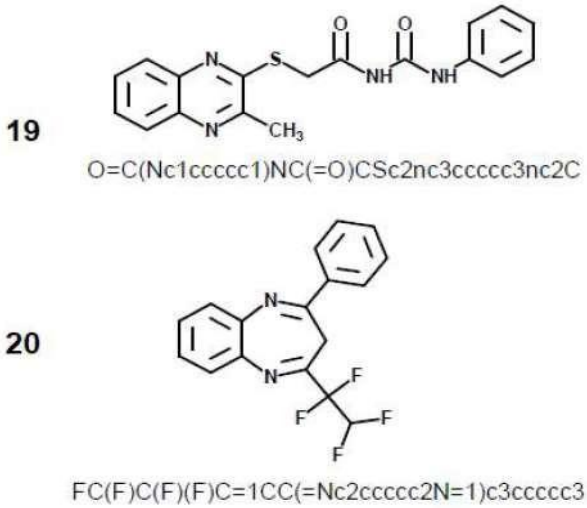


Figure 3: Molecular structures of anti-EBOLA agents 19 and 20.[3] In these two compounds, each compound also has a cyclic carbon ring and aromatic carbon ring. Interestingly, compound 20 has four fluorine atoms adjacent to each other.

The molecular properties for all 20 anti-EBOLA agents are presented in Table 1. These properties include those important for drug-likeness consideration, meaning suitability for pharmaceutical application. The favorable criteria established by Veber, et al.[6] indicates that agents 1, 3, 4, 5, 6, 7, 8, 14, 15 16, 17, 18, 19, and 20 have favorable drug-likeness properties. However, utilizing the Rule of 5 (Lipinski, et. al., see Table 1), all agents will have favorable oral activity except agents 9, 10,11, and 13. However, applying the criteria for favorable oral bioavailability for molecular weight, Log P, and number of atoms, determined by Ghose, et al.,[8] indicates all agents except for agents 7, 8, 9, 10, 11, and 13 have favorable properties.

Table 1: Molecular Properties of anti-EBOLA Agents 1 to 20.

Agent	Log P	Polar Surface Area (A^2)	Number of Atoms	Molecular Weight	Number of Oxygen & Nitrogen	Number of –OH & -NHn	Rule of 5	Number of Rotatable Bonds	Molecular Volume (A^3)
1	3.64	42.68	23	307.36	4	0	0	2	278.13
2	4.48	70.08	33	466.60	6	1	0	11	428.61
3	5.07	48.91	25	330.36	3	2	1	3	292.05
4	5.22	53.60	29	404.51	4	2	1	6	360.35
5	2.27	90.89	20	270.24	5	3	0	1	224.05
6	3.98	56.28	22	315.76	5	1	0	4	268.16
7	4.99	97.63	39	529.53	8	2	1	7	446.63
8	3.89	86.28	37	493.62	8	2	0	7	461.44
9	6.11	162.62	58	817.07	12	4	3	11	786.22
10	8.31	42.68	31	645.32	4	0	2	11	437.04
11	7.94	88.85	39	556.77	7	1	2	18	533.46
12	4.55	63.97	33	454.61	6	0	0	13	454.30
13	5.72	144.52	38	516.55	8	5	2	6	457.44
14	2.79	84.09	21	301.33	6	2	0	4	246.90
15	1.10	83.11	29	399.44	7	1	0	5	364.15
16	1.33	93.38	26	375.41	6	4	0	9	338.63
17	2.35	104.46	30	428.61	7	4	0	10	436.32
18	2.72	116.49	31	443.63	8	5	0	10	448.73
19	3.13	83.98	25	352.42	6	2	0	4	305.39
20	4.84	24.73	23	320.29	2	0	0	3	260.99

Descriptive statistical analysis of the numerical values of the molecular properties, are shown in Table 2, and provides useful information for evaluating and determining the characteristics of these and potential compounds to express anti-viral activity. Studies of perspective compounds, whether natural compounds or synthesized, that have the molecular properties analogous to previous established anti-EBOLA agents would likely be successful and worthy of further investigation.

Table 2: Statistics Of Molecular Properties For Anti-Ebola Agents 1 To 20.

Variable	Log P	Polar Surface Area (A^2)	Number of Atoms	Molecular Weight (Daltons)	Number of Oxygen and Nitrogen	Number of –OH and -NHn	Rule of 5	Number of Rotatable Bonds	Molecular Volume (A^3)
Mean	4.21	81.9	30.6	436.47	6.1	2.05	0.6	7.25	391.45
Standard deviation	1.93	34.1	8.83	133	2.22	1.63	0.94	4.30	129.1
Error of Mean	0.43	7.61	1.98	29.8	0.497	0.366	0.21	0.96	28.87
Minimum	1.10	24.7	20.0	270.24	2.00	0.00	0.00	1.00	224.05
Median	4.21	84.0	29.5	416.56	6.00	2.00	0.00	6.50	396.38
Maximum	8.31	162.6	58.0	817.07	12.0	5.00	3.00	18.00	786.22
Pass the Normality Test	Yes	Yes	Yes	Yes	Yes	No	No	Yes	No

Not only are screening methods such as those of Veber, et al.,, Lipinski, et al, and Ghose, et al. useful and successful for initial screening for perspective drug candidates, there are pattern recognition methods for studying underlying relationships within the multivariate data. Such methods such as cluster analysis, 95% ellipses, and non-metric multidimensional scaling will be applied to the molecular properties.[1] A very high resolution for differentiation of the 20 agents can be achieved by applying cluster analysis of molecular property values of Table 1.[1] The outcome is shown in Fig. 4. The cluster analysis, utilizing Eucledian distance (or the distance between two points as measured by the line segment from one point to the other point), will show that the compounds being most similar to one another (by properties) are adjacent and found under the

same node.

Initially all agents are under node A (see Fig.4), but are divided to major node B and node C (agents 10 and 11, are most similar), and agent 9 (dissimilar from all remaining agents). Further resolution breaks down to node D (agents 2, 12, 7, 8, most similar), node E (agents 16, 17, 18, most similar), node F (agents 3, 4, most similar) and node G (agents 1, 6, 20, 5, 14, 19, most similar).

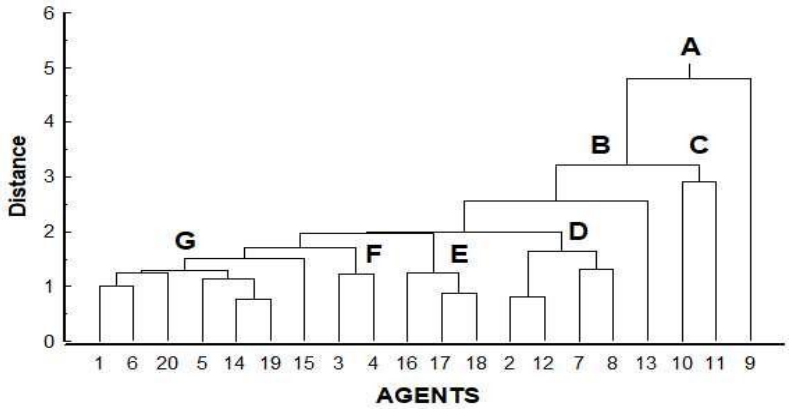


Figure 4: Cluster analysis of the molecular properties shown in Table 1, for all 20 anti-EBOLA agents, utilizing Euclidean distance and single linkage analysis parameters. Under node A there is separation of agent 9 (having the highest molecular weight and molecular volume) from all other agents. Node C shows agents 10 and 11, from all remaining agents (under node B). Node D shows agents 2, 12, 7, and 8. Node E shows agents 16, 17, and 18. Node F shows agents 3 and 4. Finally, node G shows agents 1, 6, 20, 5, 14, and 19. Anti-EBOLA agents sequestered under the same node are most similar to each other based on molecular properties.

Cluster analysis distinguishes the 20 agents from each other with high resolution and based on their molecular properties. Another analysis, 98% ellipses, indicates where the actual population mean as well as 95% of the joint population distributions would occur within the indicated region.[1] Also considered a confidence interval, 95% ellipses indicate the region where the true mean of variables involved will repeatedly exist.[1] All 20 agents fall within the 95% ellipsis, which at least in this analysis the actual population mean is found within the 95% ellipses (see Fig. 5).

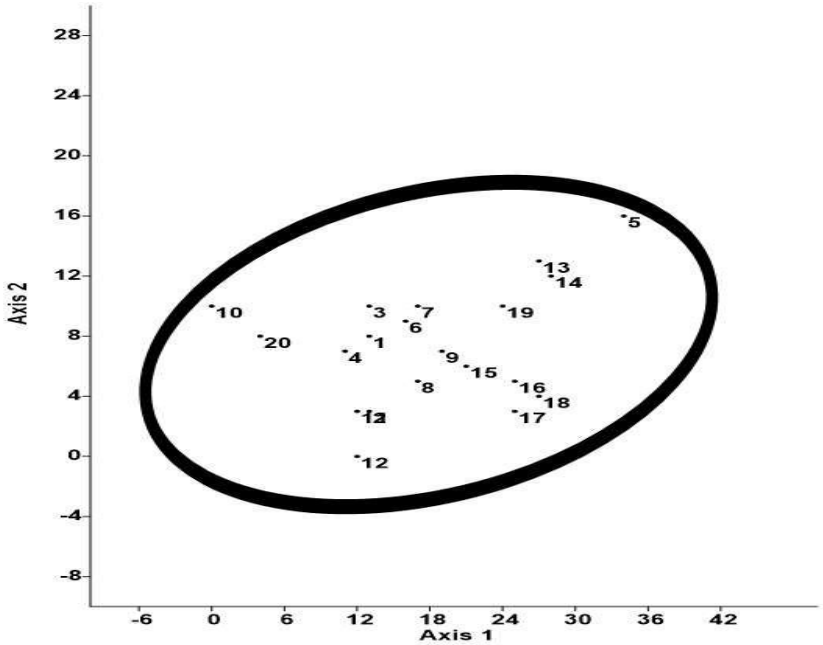


Figure 5: A 95% ellipses analysis of all molecular properties presented in Table 1, for all 20 anti- EBOLA agents. All 20 agents are contained within the 95% ellipse. Indicating that all agents fall within the region where the actual population mean, as well as 95% of the joint population distributions will occur in the indicated region. Also considered a confidence interval, 95% ellipses indicate the region where the true mean of variables involved will repeatedly exist.[1]

Non-metric multidimensional scaling is an analysis methodology for visualizing complex relationships within multivariate data through reduction of dimensionality, but preserving the dissimilarities.[1] Some investigators consider the resolution of objects analyzed by non-metric multidimensional scaling to be less than that provided by cluster analysis. In that, the outcome of non-metric multidimensional scaling of properties in Table 1, is presented in Fig. 6. Clearly the outcome indicates that agents 9 and 10 are substantially distinct from each other (see arrows) and from all other agents (see inset square, Fig. 6). These types of analysis assist in the determination of the efficacy of tested compounds and which among the population are best for actual phamaceutical application. Pattern recognition analysis of the properties from a group of compounds can resolve which of the compounds are most alike to one another, and also which are most alike to known successful drugs. In this manner, the agents that are among the larger population of compounds being evaluated, and express the desired bioligical activity, will be identified.[4]

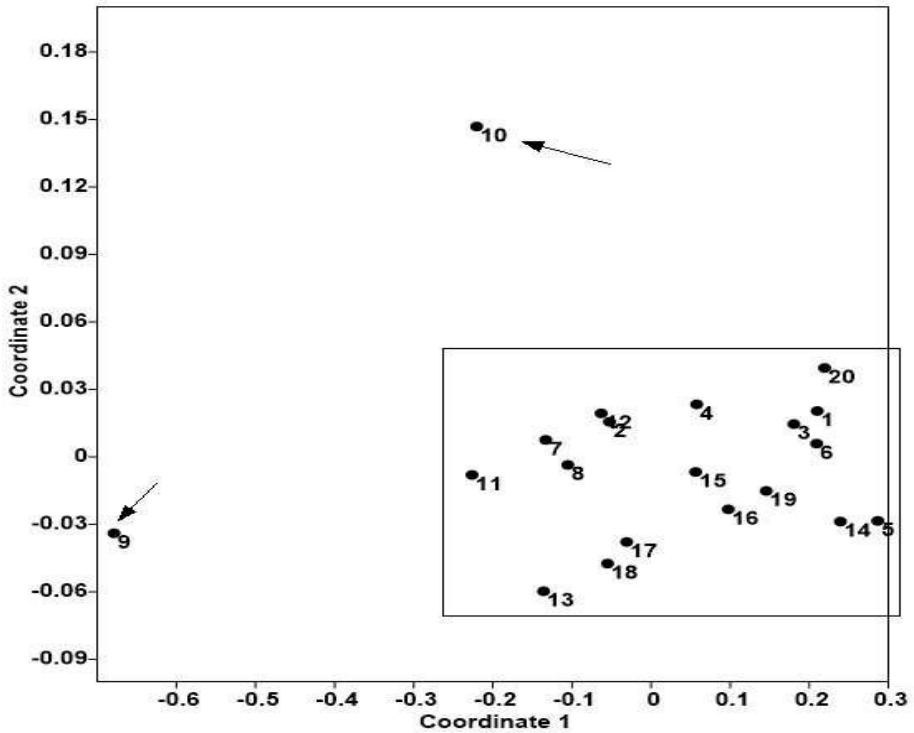


Figure 6: Non-metricmultidimensional scaling analysis of molecular properties presented in Table 1. Utilizing Euclidean distance between numerical values of molecular properties, this analysis also shows similarity amoung agents based on proximity in the 2-dimensional plot. Agent 9 and agent 10 (see arrows inset in plot) are substantially distinct from all the remaining agents (see inset square). In terms of pharmaceutical potential, these pattern recognition analysis aids in identifying compounds having significant potential to act with similar bioactivity.

The various screening modalities, such as Rule of 5, allow the judicious selection of drug candidates from a broader group of compounds. Such approach saves expense, time, and an improved likelihood of success as opposed to "trial and error" methodology.[1,4] The application of the Rule of 5 indicated all agents will have favorable oral activity except agents 9, 10,11, and 13. The pattern recognition methods cluster analysis, 95% ellipses, and non-metric multidimensional scaling provided high resolution in distinguishing the similarity (or dissimilarity), of these 20 agents to one another. This being highly useful, which after following screening for general drug-likenss, will enable the determination of the most favorable drug structures from among the initial group. This is in addition to identifying the agents having optimal oral availability and/or bioactivity.

Multiple Regression Prediction of MW Outcome

From the molecular properties shown in Table 1, four propertes can be utilized for multiple regresson calculation, and these properties reaching across the three screening methodologies discussed above. These will be Log P, molecular volume, polar surface area, and number of atoms. The multiple regression analysis was succussful and showed an outcome of R^2 of 0.9318, indicating that 93.18% of the variance in molecular weight (MW) is explained by the multiple regression model.[1] This very high value indicates the model is a extremely good fit for the data (molecular properties). The very high R^2 also indictes a very high level of the model's explanatory potential and very high level of predictive ability.[1]

The dependent variable as molecular weight (MW), is

described by independent variables Log P (Log P), molecular volume (Volume), polar surface area (PSA), and number of atoms (nAtoms).

MW = 26.129 + 16.055(Log P) + 0.7895(Volume) + 0.04276(PSA) + 0.9821(nAtoms)

In application, a potential anti-EBOLA compound structure will have its appropriate properties determined utilizing the available heuristic platforms available for this purpose, and determined prior to insertion into the model. Those values inserted in the model and molecular weight calculated to determine if the results are alike to the training set of 20 agents presented here. Of the four independent variables, the greatest contribution to the model is from Log P (P=.03) and molecular volume (P=.02), and lesser extent from polar surface area (P=.92), number of atoms (P=.84), and the constant 26.129 (P=.48). Alternatively, investigators may insert desired values for the independent variables and calculate the molecular weight (dependent variable) of the resulting model, to ascertain if the independent variables are suitable (or feasible) for a potential pharmaceutical.

CONCLUSION

EBOLA disease, is a severe and often fatal illness in humans. Efficacious clinical treatment of this viral infection, applied swiftly after diagnosis, is a must for patient survival. Studies for identifying additional compounds for treatment are a highly sought-after necessity. In this study, 20 compounds shown in previous studies to express anti-EBOLA activity have their structures and molecular properties disclosed and examined by pattern recognition algorithms to reveal similarities (or dissimilarities) of members of the group. The evaluation for drug-likeness also indicated favorable oral availability for many of these 20 agents. Agents 9 and 10 were shown to be substantially dissimilar, by properties, from the remaining 18. The outcome of multiple regression analysis produced a model explaining 93.18% of the variance in molecular weight. The properties of Log P and molecular volume contributed the greatest to the regression model. Further studies are highly desired to identify additional anti-EBOLA compounds for clinical treatment of this viral illness.

ACKNOWLEDGEMENTS

This study was completed at the Chemistry Department of the University of Nebraska at Omaha, 6001 Dodge Street, Omaha, Nebraska 68182 USA.

REFERENCES

1. Zar J. Biostatistical analysis. New York; Prentice Hall, 1996.
2. Singh SK, Ruzek D. Viral Hemorrhagic Fevers. BocaRaton; CRC Press Taylor & Francis Group, 2014.
3. Picazo D, Giordanetto F. (Small molecule inhibitors of ebola virus infection). Drug Discovery Today, 2015; 20(2): 277-286.
4. 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.
5. Feldmann H. (Ebola-A growing threat). N Engl J Med, 2014; 371(15): 1375-8.
6. 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.
7. 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.
8. 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.