2016

Properties and Drug-likeness of Compounds That Inhibit Ebola Virus Disease (EVD)

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

International Journal of TROPICAL DISEASE & Health
15(2): 1-17, 2016, Article no.IJTDH.25021
ISSN: 2278–1005, NLM ID: 101632866

SCIENCEDOMAIN *international*
www.sciencedomain.org

Properties and Drug-likeness of Compounds That Inhibit Ebola Virus Disease (EVD)

Ronald Bartzatt[1*]

[1]University of Nebraska, Durham Science Center, 6001 Dodge Street, Omaha, Nebraska 68182, USA.

Author's contribution

The sole author designed, analyzed and interpreted and prepared the manuscript.

Article Information

DOI: 10.9734/IJTDH/2016/25021
Editor(s):
(1) Ken-ichiro Inoue, Center for Medical Science, International University of Health and Welfare, Japan.
Reviewers:
(1) Molobe Ikenna Daniel, International Institute of Risk and Safety Management (IIRSM), Nigeria.
(2) Sanjay Mishra, IFTM University, UP, India.
(3) Triveni Krishnan, National Institute of Cholera and Enteric Diseases, Kolkata, India.
Complete Peer review History: http://sciencedomain.org/review-history/13746

Original Research Article

Received 14th February 2016
Accepted 5th March 2016
Published 17th March 2016

ABSTRACT

Aims: To present the molecular structures of compounds that has been shown to inhibit the proliferation of Ebola virus. To elucidate the molecular properties of these virus inhibiting compounds.
Study Design: The molecular properties of virus inhibiting compounds are elucidated and compiled. Pattern recognition methods and statistical analysis are applied to determine optimal properties of this group of compounds.
Place and Duration of Study: Chemistry Department, Durham Science Center, University of Nebraska, Omaha NE. between December 2015 and February 2016.
Methodology: A total of 60 compounds were identified as inhibiting the virus Ebola. The molecular properties such as Log P, molecular weight, and 7 other descriptors were elucidated utilizing heuristic methods. Structures are compared by applying classification methods with statistical tests to determine trends, underlying relationships, and pattern recognition.
Results: For 60 compounds identified the averages determined: for Log P (3.51), polar surface area (89.45 Angstroms^2), molecular weight (432.6), molecular volume (393.96 Angstroms^3), and number of rotatable bonds (7). Molecular weight showed a strong positive correlation to number of oxygen and nitrogen atoms, number of rotatable bonds, and molecular volume. K-means clustering

**Corresponding author: Email: rbartzatt@unomaha.edu;*

indicated seven clusters divided according to highest similarity of members in the cluster. Ranges found: formula weights (157.1 to 822.94), Log P (-2.24 to 8.93), polar surface area (6.48 to 267.04 A^2), and number of atoms (11 to 58). Multiple regression analysis produced an algorithm to predict similar compounds.

Conclusion: The formula weights and Log P values of Ebola virus inhibitors show a broad range in numerical values. Consistency in properties was identified by statistical analysis with grouping for similarity by K-means pattern recognition. Multiple regression analysis enables prediction of similar compounds as drug candidates. Only 29 compounds showed zero violations of rule of 5, an indication of favorable drug-likeness. These compounds are highly varied in structures and properties.

Keywords: Ebola virus; virus; drug-likeness; Ebolavirus; hemorrhagic; EVD.

ABBREVIATIONS

PSA, polar surface area; nAtoms, number of atoms; nON, number of oxygen and nitrogen atoms; nOHNH, number of hydroxyl and amine groups; nRotB, number of rotatable bonds; MV, molecular volume; MW, molecular weight; SMILES, simplified molecular-input line-entry system; EVD, Ebola virus disease.

1. INTRODUCTION

The Ebola virus is one of known viruses within the genus *Ebolavirus* that are generally considered to cause Ebola virus disease (EBV) in humans. The Ebola virus is one five viruses that exist within the genus *Ebolavirus*, order *Mononegavirales*, family *Filoviridae*, and the single member of the species *Zaire ebolavirus* [1]. Four of the five Ebola viruses induce a very severe and very often fatal hemorrhagic fever in mammals, as well as humans [1]. At present, the reservoir for this virus in nature is considered to be bats (i.e. fruit bats) and the primary mode of transference to humans is by body fluids [2]. The Ebola virus is a zoonotic pathogen (a disease that can be passed between animals and humans and zoonotic diseases can be caused by viruses, bacteria, parasites, and fungi) [2].

Some investigators have determined that Ebola virus outbreaks have an increased likelihood to occur when temperatures are lower and humidity is higher [3]. Post recovery, the virus can survive for months in certain organs (i.e. eyes and testes) [3]. A major concern for clinicians is a delayed diagnosis of Ebola resulting in delayed treatment, severe illness, death, and transmission into areas thought cleared of the virus [4]. Deliberate evasion of control intervention actions is very problematic in monitoring possible contacts and identification of potential cases [4].

Direct contact with body fluids is generally considered to be the major risk factor for contracting the virus [5]. Being a viral hemorrhagic fever, it is common to observe blood abnormalities, such as decreased platelets [6]. However, there are other organs involved that include renal, cardiac, lung, gastrointestinal, neurological and hepatic indicators [6]. A high fever and hemorrhagic manifestations (hemorrhagic conjunctivitis, bleeding ulcerations of mouth and lips, gingival bleeding, hematemesis, ear bleeding, hematuria) are generally observed in patients [6].

The Ebola virus is known to be isolable in semen (82 days after illness onset) [7,8], aqueous humor (on day 63 after illness onset) [7], sweat [7], urine (after 26 days after illness onset) [7], vaginal secretions (on day 33 after illness onset), feces [7], and breast milk (15 days after illness onset) [7]. Sexual contact has been shown to be an effective mode of transmitting the virus [8]. Other viral hemorrhagic fevers include Marburg disease and Lassa fever, and all can be easily imported into any part of the industrialized world [9]. For EBV disease identified in Guinea, the symptoms included severe diarrhea, vomiting, and dehydration [10].

The determination and evaluation of pharmacokinetic and pharmacodynamics properties of drugs potentially useful for treatment of Ebola virus disease is a very important consideration for discovery of new pharmaceuticals [11]. Establishing these criteria for the molecular properties of novel potential drug candidates well expedite identification of likely compounds and enhance the

commencement of human trials [11]. Similar approaches in past studies have already been proven to be useful for drug classes such as anti-inflammatories [12], anticancer agents [13], and development of selective COX-2 inhibitors [14]. This study seeks to establish a criteria for molecular properties of a broad range of compounds previously shown to inhibit the proliferation of Ebola virus.

2. METHODOLOGY

2.1 Properties and Molecular Modeling

Numerical values of molecular properties and descriptors of the compound's molecular structures were calculated by utilizing Molinspiration (Molinspiration Cheminformatics, Nova ulica 61, SK-900 26 Slovensky Grob, Slovak Republic). Additional elucidation of structure components was accomplished through the use of ACD ChemSketch Modeling v. 12.01 (Advanced Chemistry Development, 110 Yonge Street, Toronto Ontario, M5C 1T4 Canada, http://www.molinspiration.com/services/search.html).

2.2 Pattern Recognition and Multivariate Statistical Analysis

To identify underlying associations and patterns within the numerical properties the use of various pattern recognition techniques was applied. In addition, analysis of non-hierarchical K-means cluster analysis and 95% ellipses were performed by PAST v. 2.06 (copyright Oyvind Hammer, D.A.T. Harper 1999-2008) and KyPlot v. 2.0 Beta 15 (copyright Koichi Yoshioka 1997-2001)

2.3 Various Statistical Analysis Data

Statistical analysis of numerical data, including Pearson r correlation and descriptive statistics were accomplished by Microsoft EXCEL v. 14.0.6112.5000 (EXCEL Professional plus 2010). Multiple regression analysis of molecular property values was accomplished by GraphPad Instat version 3.00 (GraphPad Software, Inc., San Diego, California USA; www.graphpad.com). Determination of any numerical outliers was accomplished by applying Grubb's test (also known as extreme studentized deviate). Standard 2-way plotting and box plots were accomplished for visual representation by PAST v. 2.06.

3. RESULTS AND DISCUSSION

Quick development of efficient drugs for the treatment of Ebola virus infection is a most urgent issue to rescue multitudes of Ebola-infected patients [15]. An antibody therapeutic referred to as ZMapp is under development and testing [15]. ZMapp is comprised of an antibody cocktail that mixes the humanized mAbs with a selected composition of c13C6 from MB-003 (human-mouse chimeric mAbs developed by Mapp Biopharmaceutical Inc., San Diego, CA, USA) and c2G4 and c4G7 from ZMAb (mouse mAbs developed by DeFyrus, Toronto, Canada). Unfortunately, due to the fact that Ebola virus disease did not generate an outbreak frequently, any vaccine research and development was only slowly studied and pursued [15]. A major pursuit of medical professionals encountering infected individuals is to provide Ebola patients with basic supportive care such as sustaining hydration. Clinicians must also ensure that infected patients interaction with others is limited to prevent transmission.

This study will show the structures and molecular properties of 60 compounds that have been shown to inhibit Ebola virus proliferation [16]. This analysis will determine statistical features useful for consideration of similar potential compounds. Pattern recognition did elucidate underlying relationships of the property values. Multiple regression analysis produced a mathematical algorithm useful for determining individual properties of perspective candidate compounds, these being properties important to pharmacodynamics (drugs at the site of action and resulting effect) and pharmacokinetics (time course of drug absorption, distribution, metabolism, and excretion).

The properties of compounds presented here will be seen to vary broadly. The physiological functions of these 60 compounds are also broad and include the following groups of biological activity: viral transcription modulators, protease inhibitors, PCK C1 (NPC-1) small molecules, kinase inhibitors, phosphatase inhibitors, ion channel modulators, microtubule modulators, actin modulators, glycosylation modulators, budding modulators, and various small molecule organic compounds [16].

3.1 Structure and Molecular Properties of EBOLA Inhibiting Compounds

Structures of compounds 1 to 9 are presented in Fig. 1. With compounds 1 to 5 being viral

transcription modulators and compounds 5 to 9 are protease inhibitors [16]. What is quickly apparent from compounds 1 to 9, and descript for others presented here, is the broad range in size, heteroatom rings, functional groups, and functional groups (see Fig. 1). The numerical values of molecular properties for all 60 compounds are shown in Table 1.

Interestingly, upon analysis of numerical values of the descriptors from Table 1 by Grubb's test showed no outliers in values of molecular weight, Log P, the number of atoms, number of rotatable bonds, number of hydroxyl groups (-OH), and number of amine groups ($-NH_n$). This outcome despite the broad variety in molecular structures shown in Figs. 1, 2, 3, 4, 5, 6, and 7. Actually, only in polar surface area is one outlier compound found (56), in addition, only one outlier compound is found in molecular volume (29).

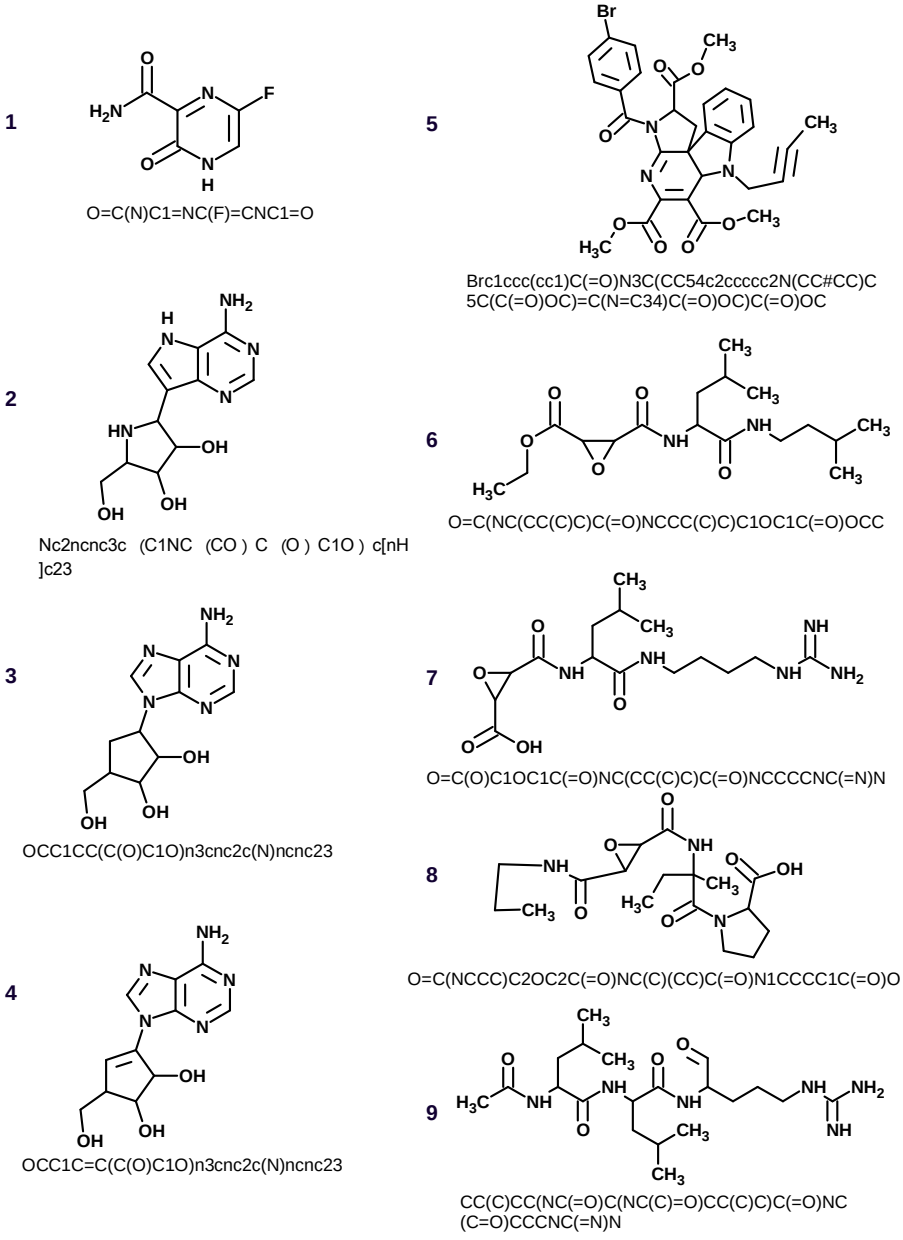
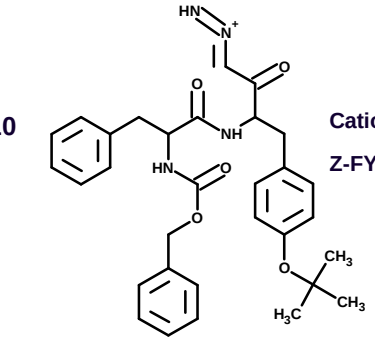


Fig. 1. Structures and SMILES notation for antiviral compounds 1 to 9. Compounds 1 to 5 are viral transcription modulators. Compounds 5 to 9 are protease inhibitors

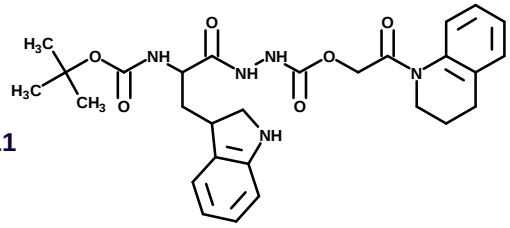
Compounds 10 to 13 are protease inhibitors (see Fig. 2). Compounds 14 to 17 are (NPC1)-dependent small molecule virus inhibitors [16]. Various functional groups, aromatic ring, and alkyl chains are evident on these various scaffoldings.

10 **Cation form Z-FY(tBu)-DMK**



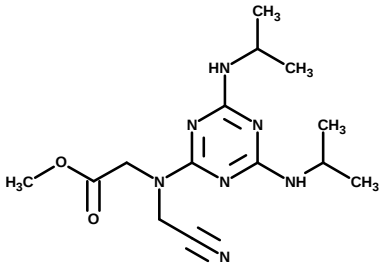
CC(C)(C)Oc1ccc(cc1)CC(NC(=O)C(Cc2ccccc2)NC(=O)OCc3ccccc3)C(=O)\C=[N+]=N

11



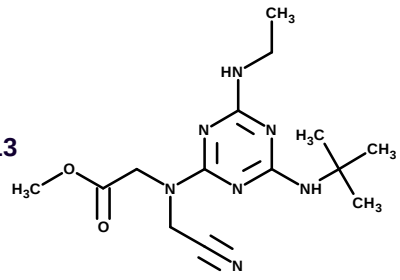
CC(C)(C)OC(=O)NC(CC2CNc1ccccc12)C(=O)NNC(=O)OCC(=O)N4CCCc3ccccc34

12



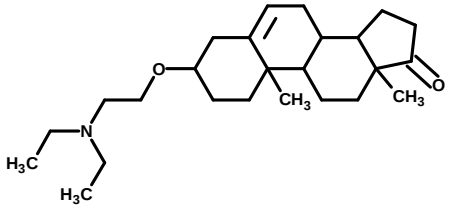
CC(C)Nc1nc(nc(n1)N(CC(=O)OC)CC#N)NC(C)C

13



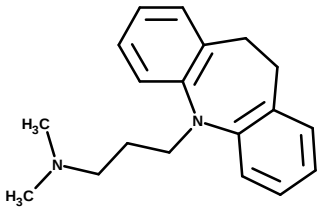
CC(C)(C)Nc1nc(nc(n1)N(CC(=O)OC)CC#N)NCC

14



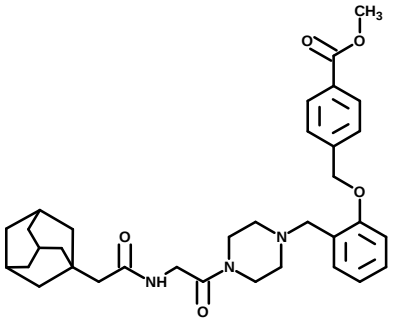
CCN(CC)CCOC2CCC3(C)C4CCC1(C)C(CCC1=O)C4CC=C3C2

15



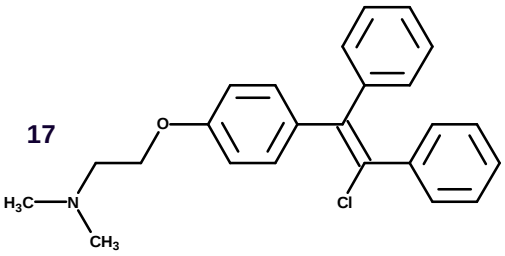
CN(C)CCCN2c3ccccc3CCc1ccccc12

16



O=C(OC)c1ccc(cc1)COc6ccccc6CN2CCN(CC2)C(=O)CNC(=O)CC35CC4CC(C3)CC(C4)C5

17

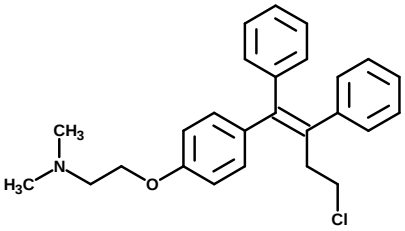


CN(C)CCOc1ccc(cc1)\C(c2ccccc2)=C(/Cl)c3ccccc3

Fig. 2. Molecular structures and SMILES notation for antiviral compounds 10 to 17. Compounds 10 to 13 are protease inhibitors. Compounds 14 to 17 are (NPC1)-dependent virus inhibitors. Note structure 10 is a positively charged species

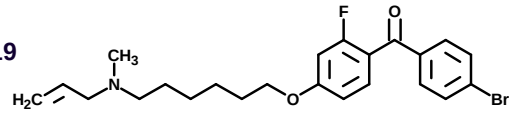
Compounds 18, 19, and 20 are (NPC1)-dependent; with 21 to 27 being phosphatase and kinase inhibitors (see Fig. 3) [16]. Aromatic rings are typical for this group.

18



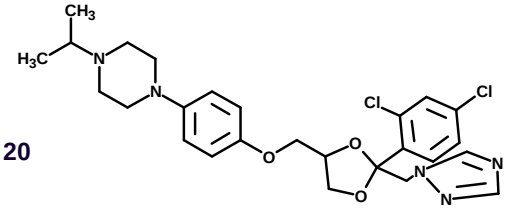
CN(C)CCOc1ccc(cc1)\C(c2ccccc2)=C(/CCCl)c3ccccc3

19



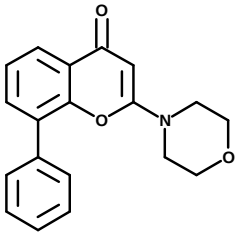
O=C(c1ccc(Br)cc1)c2ccc(OCCCCCCCN(C)CC=C)cc2F

20



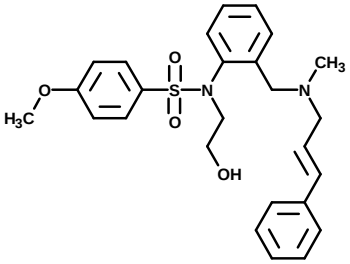
CC(C)N1CCN(CC1)c2ccc(cc2)OCC4COC(Cn3n
cnc3)(O4)c5ccc(Cl)cc5Cl

21



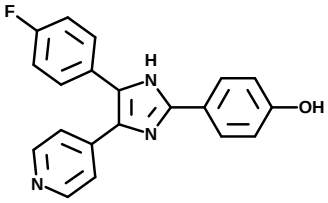
O=C2C=C(Oc1c(cccc12)c3ccccc3)N4CCOCC4

22



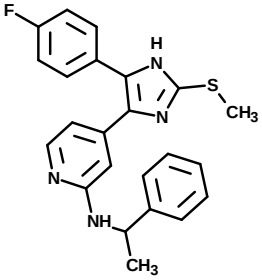
O=S(=O)(N(CCO)c2ccccc2CN(C)C/
C=C/c1ccccc1)c3ccc(OC)cc3

23



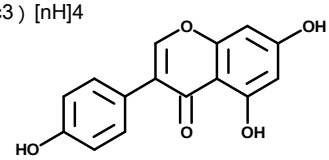
Oc4ccc (c3nc (c1ccccc1) c (c2ccc
(F) cc2) [nH]3) cc4

24



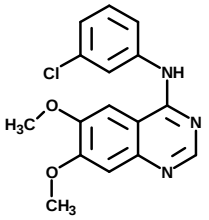
CSc4nc (c2ccnc (NC (C) c1ccccc1) c2) c (c3ccc
(F) cc3) [nH]4

25



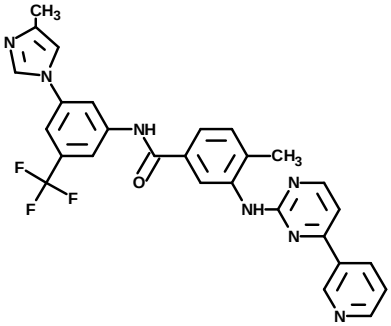
Oc1ccc(cc1)C3=COc2cc(O)cc(O)c2C3=O

26



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

27

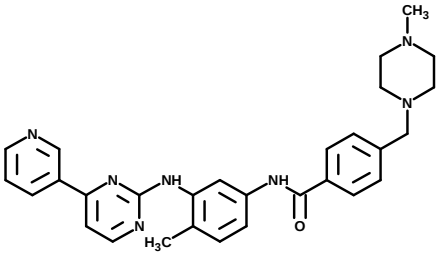


Cc3ccc(cc3Nc1nc(ccn1)c2cccnc2)C(=O)Nc
4cc(cc(c4)n5cc(C)nc5)C(F)(F)F

Fig. 3. Molecular structures and SMILES notation for antiviral compounds 18 to 27. Compounds 18 to 20 are (NPC1)-dependent small virus inhibitors. Compounds 21 to 27 are kinase and phosphatase inhibitors

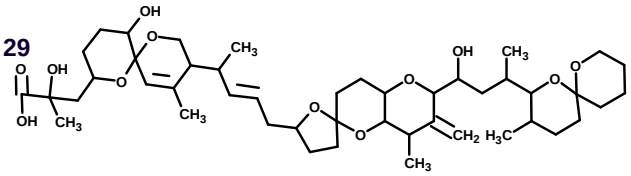
Aromatic rings are typical for the kinase and phosphatase inhibitors of 28 to 29; ion channel compounds 30 to 33; and actin/microtubule modulators 34, 35 (see Fig. 4) [16].

28



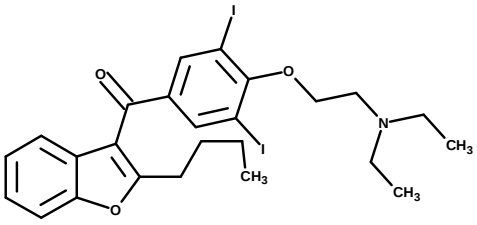
Cc3ccc(cc3Nc1nc(ccn1)c2cccnc2)NC(=O)c4ccc(cc4)CN5CCN(C)CC5

29



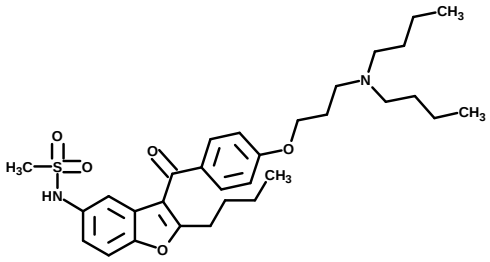
CC2CCC1(CCCCO1)OC2C(C)CC(O)C6OC7CCC3(CCC(O3)C\C=C\C(C)C4COC5(C=C4C)OC(CCC5O)CC(C)(O)C(=O)O)OC7C(C)C6=C

30



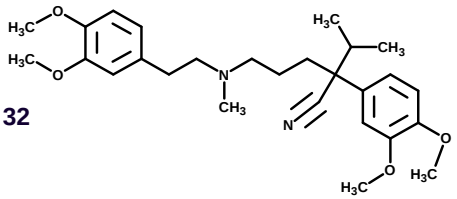
CCN(CC)CCOc1c(I)cc(cc1I)C(=O)c2c3ccccc3oc2CCCC

31



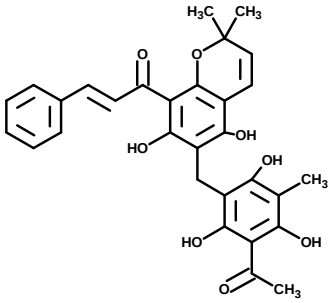
CCCCN(CCCC)CCCOc1ccc(cc1)C(=O)c2c3cc(ccc3oc2CCCC)NS(C)(=O)=O

32



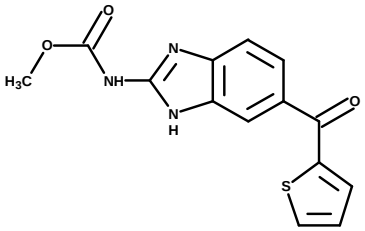
COc1cc(ccc1OC)C(C#N)(CCCN(C)CCc2ccc(OC)c(OC)c2)C(C)C

33



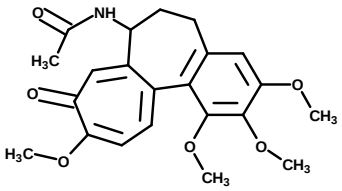
CC(=O)c4c(O)c(Cc2c(O)c(c1OC(C)(C)C=Cc1c2O)C(=O)/C=C/c3ccccc3)c(O)c(C)c4O

34



COC(=O)Nc3nc2ccc(C(=O)c1cccs1)cc2[nH]3

35

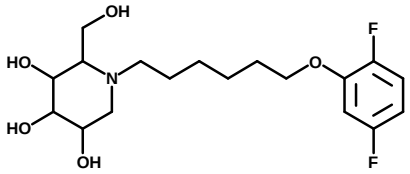


CC(=O)NC2CCc3cc(OC)c(OC)c(OC)c3C1=CC(=O)C(=CC=C12)OC

Fig. 4. Molecular structures and SMILES notation of antiviral compounds 28 to 35. Compounds 28 to 29 are kinase and phosphatase inhibitors. Compounds 30 to 33 are ion channel modulators. Compounds 34 and 35 are microtubule and actin modulators

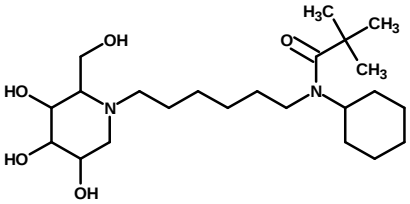
In Fig. 5, there are glycosylation (36, 37, 38) and budding (i.e. budding from membrane) modulators. Structures 41 to 45 are small molecules expressing Ebola virus inhibitors [16].

36



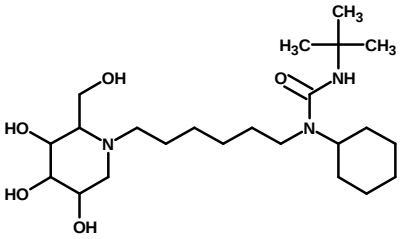
Fc2cc(OCCCCCCN1CC(O)C(O)C(O)C1CO)c(F)cc2

37



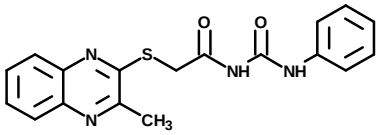
CC(C)(C)C(=O)N(CCCCCCN1CC(O)C(O)C(O)C1CO)C2CCCCC2

38



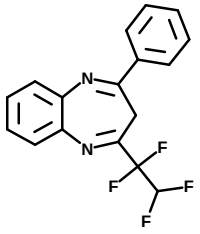
CC(C)(C)NC(=O)N(CCCCCCN1CC(O)C(O)C(O)C1CO)C2CCCCC2

39



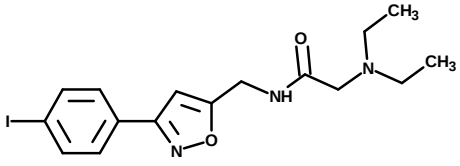
O=C(Nc1ccccc1)NC(=O)CSc2nc3ccccc3nc2C

40



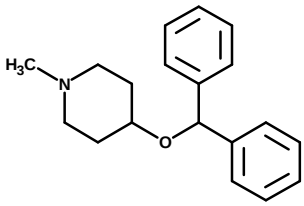
FC(F)C(F)(F)C=1CC(=Nc2ccccc2N=1)c3ccccc3

41



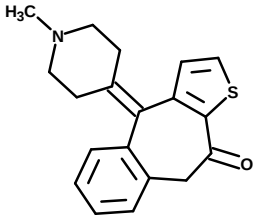
CCN(CC)CC(=O)NCc1cc(no1)c2ccc(I)cc2

42



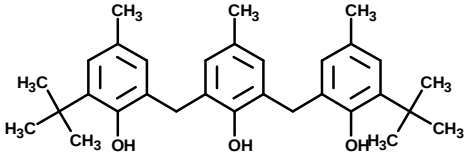
CN3CCC(OC(c1ccccc1)c2ccccc2)CC3

43



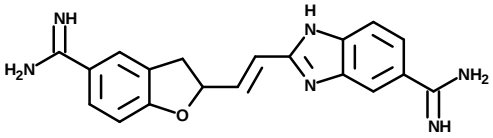
CN1CC/C(CC1)=C3\c4ccccc4CC(=O)c2sccc23

44



CC(C)(C)c3cc(C)cc(Cc2cc(C)cc(Cc1cc(C)cc(c1O)C(C)(C)C)c2O)c3O

45

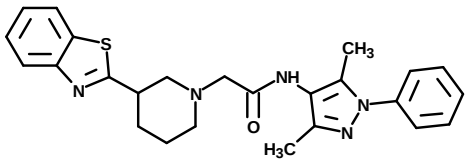


NC(=N)c4ccc3OC(C=Cc2nc1cc(C(N)=N)ccc1[nH]2)Cc3c4

Fig. 5. Molecular structures and SMILES notation of antiviral compounds 36 to 45. Compounds 36 to 38 are glycosylation modulators. Compounds 39 and 40 are budding modulators. Compounds 41 to 45 are various small molecule Ebola virus inhibitors [16]

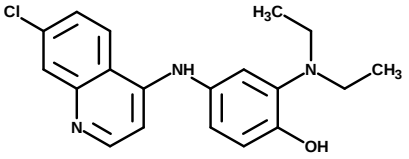
Compounds shown in Fig. 6 are various small molecule viral inhibitors (save for 46 being a budding modulator) [16]. Compounds in Fig. 7 are among the largest of the 60 members and incorporate large internal ring structures and also diverse heteroatom groups.

46



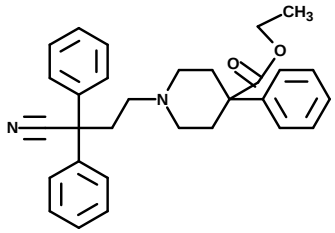
Cc4n(nc(C)c4NC(=O)CN1CCCC(C1)c2nc3ccccc3s2)c5ccccc5

47



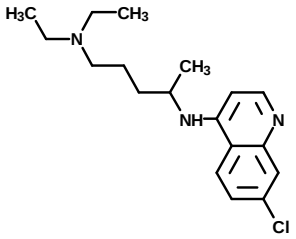
CCN(CC)c1cc(ccc1O)Nc2ccnc3cc(Cl)ccc23

48



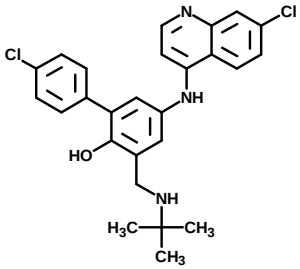
O=C(OCC)C3(CCN(CCC(C#N)(c1ccccc1)c2ccccc2)CC3)c4ccccc4

49



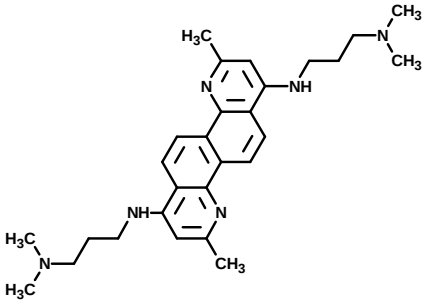
CCN(CC)CCCC(C)Nc1ccnc2cc(Cl)ccc12

50



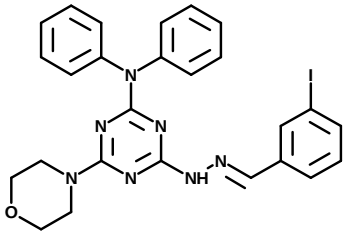
Clc1ccc(cc1)c2cc(cc(CNC(C)(C)C)c2O)Nc3ccnc4cc(Cl)ccc34

51



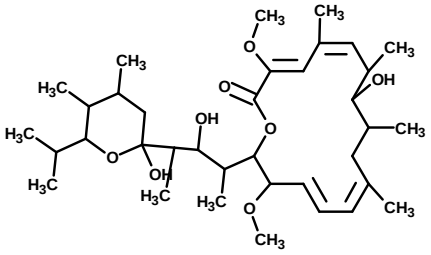
CN(C)CCCNc3cc(C)nc2c4ccc1c(cc(C)nc1c4ccc23)NCCCN(C)C

52



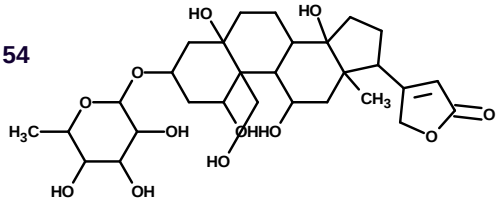
Ic1cccc(c1)/C=N/Nc2nc(nc(n2)N(c3ccccc3)c4ccccc4)N5CCOCC5

53



OC1(CC(C)C(C)C(O1)C(C)C)C(C)C(O)C(C)C2OC(=O)C(OC)=CC(C)=CC(C)C(O)C(C)CC(C)=CC=CC2OC

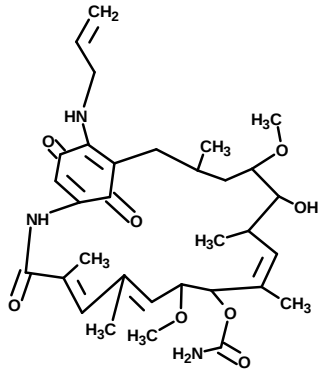
54



O=C1C=C(CO1)C3CCC4(O)C2CCC5(O)CC(CC(O)C5(CO)C2C(O)CC34C)OC6OC(C)C(O)C(O)C6O

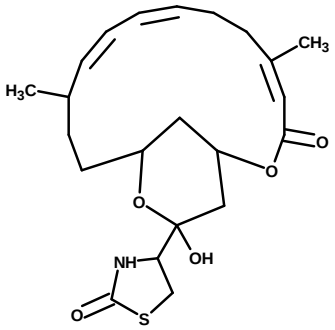
Fig. 6. Molecular structures of antiviral compounds 46 to 54. Compound 46 is a budding modulator. Compounds 47 to 54 are small molecule Ebola virus inhibitors

55



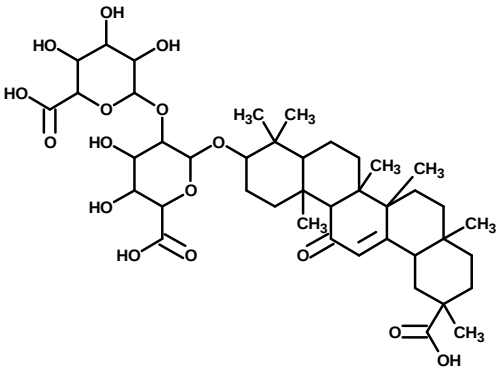
NC(=O)OC1C(C)=CC(C)C(O)C(OC)CC(C)C
C2=C(NCC=C)C(=O)C=C(NC(=O)C(C)=CC(
C)=CC1OC)C2=O

58



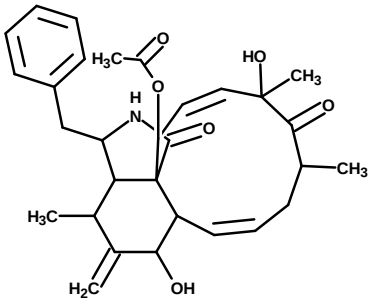
CC3CCC1CC(CC(O)(O1)C2NC(=O)SC2)OC(=O)
C=C(C)CCC=CC=C3

56



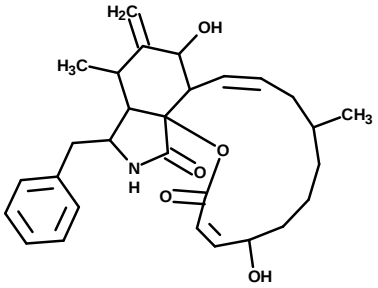
O=C(O)C7OC(OC6C(O)C(O)C(OC6OC2CCC3(C)C4C
(=O)C=C1C5CC(C)(CCC5(C)CCC1(C)C4(C)CCC3C2(
C)C)C(=O)O)C(=O)O)C(O)C(O)C7O

59



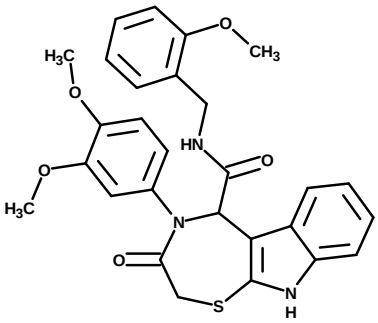
CC(=O)OC2C=CC(C)(O)C(=O)C(C)CC=CC4
C(O)C(=C)C(C)C3C(Cc1ccccc1)NC(=O)C234

57



CC1CC=CC3C(O)C(=C)C(C)C2C(NC(=O)C23OC(=O
)C=CC(O)CCC1)Cc4ccccc4

60



COc1ccccc1CNC (=O) C4N (c2ccc (OC) c
(OC) c2) C (=O) CSc5[nH]c3ccccc3c45

Fig. 7. Molecular structures and SMILES notation of antiviral compounds 55 to 60. Compounds 55 and 56 are budding modulators. Compounds 57 to 59 are microtubule and actin modulators. Compound 60 is a budding modulator

Statistical analysis and pattern recognition analysis of properties will reveal underlying relationships of these numerical values determined for compounds 1 to 60 (see Table 1).

Table 1. Molecular properties of 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

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)
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 descriptor polar surface area [17] have been demonstrated to be good a predictor of drug absorption and membrane permeation [17]. Previous studies have shown that the values of polar surface area of less than 80 Angstroms^2 can result in greater than 50% of an orally administered drug to be absorbed from the gastro-intestines [17]. This encompasses 26, or about half the total of 60 compounds. Comparison of molecular weight to Log P has results seen in Fig. 8. The Log P values range from -2.24 to 8.93, within a molecular weight range of 157.1 to 822.94 (see Plot A in Fig. 8).

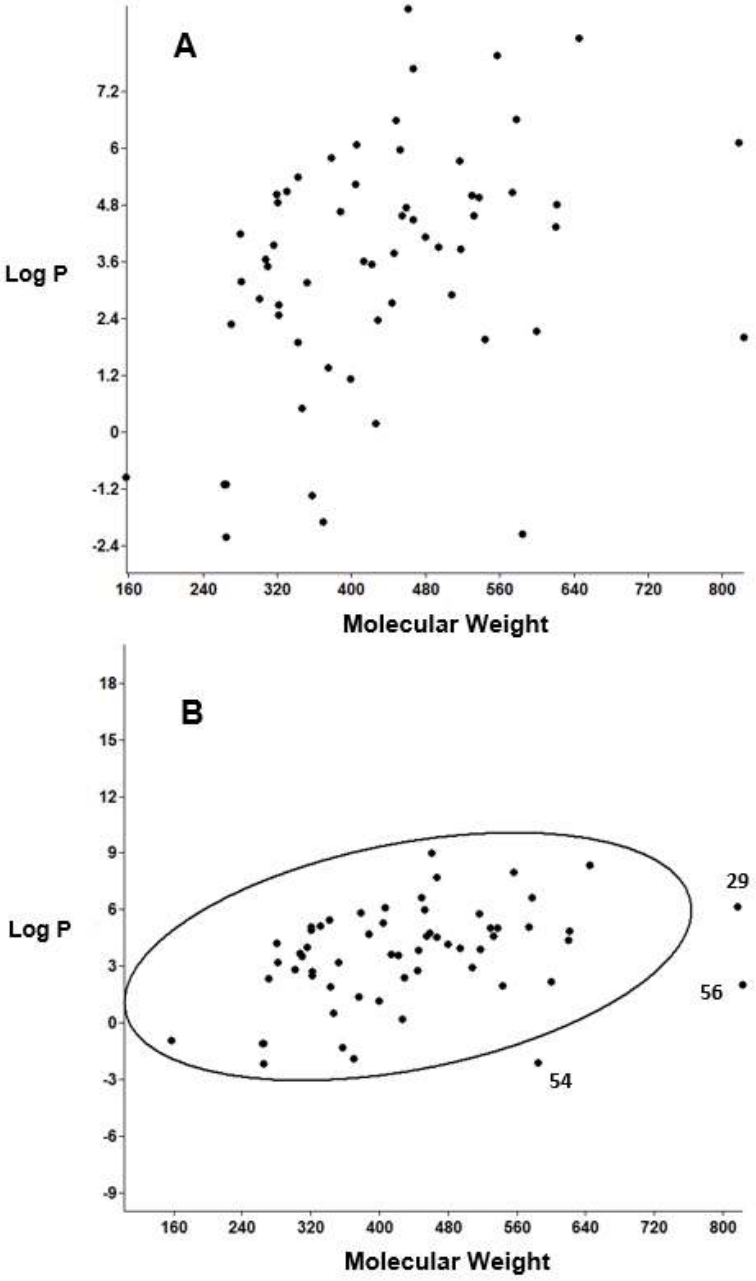


Fig. 8. Comparison of molecular weight to Log P for all compounds (Plot A). Results of 95% ellipses analysis (Plot B), with only outliers 29, 56, and 54

The 95% confidence ellipses shown in Plot B, Fig. 8, are derived from the confidence ellipse that is the smallest ellipse that will cover 95% of the all points. Studies have shown that Log P is quite useful for predicting membrane permeation [18]. Note that only compounds 29, 56, and 54 are the only non-assimilated compounds of the entire group of 60 compounds, within the ellipses. This is an interesting outcome for 60 diverse compounds and indicates a consistent feature within this broad group of Ebola viral inhibiting compounds.

The Rule of 5 states that a compound is more likely to be membrane permeable and easily absorbed by the body if it matches the following criteria [18]: 1) Molecular weight is less than 500; 2) The compound's lipophilicity, known as Log P (the logarithm of the partition coefficient between water and 1-octanol), is less than 5; 3) The sum of hydroxyl and amine groups in a drug molecule is less than 5; and 4) The sum of oxygen and nitrogen atoms is less than 10.

A total of 29 of the 60 compounds show zero violations of the Rule of 5 (see Table 1). These 29 compounds would be considered to be orally active drugs, membrane permeable, and easily absorbed by the body due to matching the criteria of Rule of 5 [18].

The box plot is a standardized way of displaying the distribution of data based on a five number summary: minimum, first quartile, median, third quartile, and maximum (see Fig. 9) [19].

Although the range of molecular weight is quite broad (157.1 to 822.94), this is not the case for the number of oxygen atoms & nitrogen atoms (nON), hydroxyl groups & amine groups($nOHNH_n$), and rotatable bonds (nRotB) (see box plots, Fig. 9). Ranges in values for nON, $nOHNH_n$, and nRotB are as follows, respectively: 2 to 16, 0 to 8, and 1 to 18. These criteria could be effective in identifying potential drug candidates by restricting these descriptors to match these parameters. The number of atoms ranged from 11 to 58, with compound 56 as an outlier.

Correlation (Pearson's r) among molecular properties (see Table 1) varied widely. A moderate positive correlation ($0.3000 < r < 0.3900$) exists between Log P to nAtoms, and Log P to molecular volume. A strong positive correlation ($0.7000 < r < 1.000$) exists between polar surface area to $nOHNH_n$, nAtoms to the Rule of 5, and molecular weight to volume. A very negative correlation (-0.4000 to -0.6900) exists between Log P to nON, and Log P to $nOHNH_n$. Again, correlation relationships among properties of candidate drugs may be used for selection of optimal candidates.

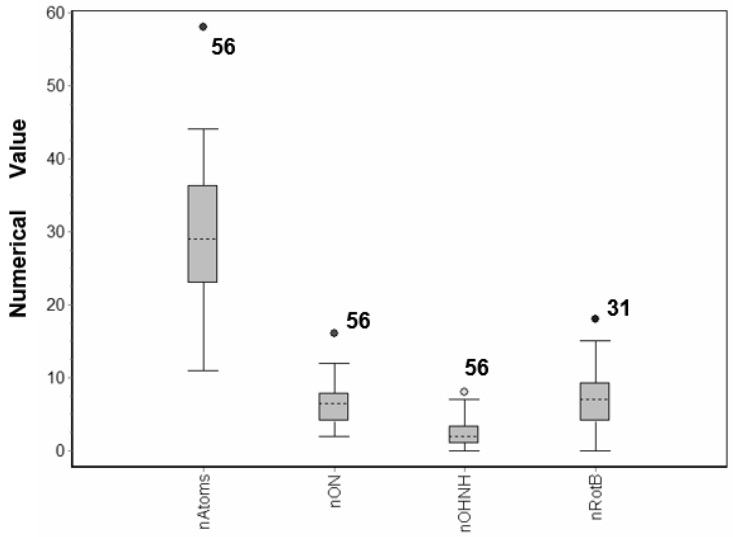


Fig. 9. Box plots of all numerical values of number of atoms (nAtoms), number of oxygen & nitrogen atoms (nON), number of hydroxyl (-OH) & amine ($-NH_n$) groups (nOHNH), and number of rotatable bonds (nRotB). Note that compound 56 is an outlier for nAtoms, nON, and nOHNH. Compound 31 is an outlier for nRotB

3.2 Prediction by Multiple Regression Analysis

There are two major purposes for multiple regression [19]: 1) First, one use of multiple regression is a prediction or estimation of an unknown, dependent value corresponding to a set of independent values; 2) A second use of multiple regression is to understand the functional relationships between the dependent and independent variables, in order to determine the cause of the variation in the dependent variable.

Multiple regression analysis is a powerful technique used for predicting the unknown value of a variable from the known value of two or more variables- also called the predictors. Applying the multivariate properties in Table 1 to multiple regression analysis, with the purpose of predicting molecular weight (dependent variable) based on proposed values for seven properties (independent variables). Prediction of molecular weight MW-molecular weight utilizing seven properties including PSA- polar surface area and MV-molecular volume. The resulting mathematical model (shown below) accounts for 94.25% of the variance in the molecular weight:

MW = 6.018 + 8.750 (Log P) + 0.3072(PSA) + 6.597(nAtoms) + 5.667(nON) – 5.342(nOHNH) – 0.2900(nRotB) + 0.3706(MV)

The independent variables Log P and number of atoms (nAtoms), provide the greatest contribution within the model. Only the property nOHNH (number of hydroxyl and amine groups), did not pass the normality test (numerical values that are sample from a Gaussian distribution [19]) within these descriptors utilized here. Note that only the independent variable total number of hydroxyl and amine groups ($nOHNH_n$) and rotatable bonds (nRotB) have negative coefficients (indicating a decrease in molecular weight when these variables increases by one). The remaining descriptors Log P, polar surface area, number of atoms, number of oxygen-nitrogen atoms, and molecular volume have positive coefficients (indicating increase in molecular weight when these variables increase by one).

3.3 Pattern Recognition Utilizing K-means Cluster Analysis

K-means is one of the simplest unsupervised learning algorithms that solve the clustering problem [19]. The method follows a simple way to classify a given data set into a certain number of clusters that is assigned and fixed prior to analysis [19]. This cluster method is non-hierarchical. Members of each cluster are determined to be the most similar to each other based on numerical values of the multivariate table of properties.

K-means cluster analysis of Table 1 into seven clusters provides a separation and discernment of these 60 diverse compounds into groups of compounds having highest similarity to each of the other members of the cluster [19]. Results of K-means cluster analysis into seven cluster outcome, are shown (members, common structure features):

Cluster 1: Compounds 1, 2, 3, 4, 25 (heterocyclic ring structures, aromatic rings, carbonyl carbons, hydroxyl groups).

Cluster 2: Compounds 5, 10, 11 16, 30, 31, 52, 53, 54, 55 (aromatic rings, carbonyl carbon, amide groups, tertiary amines).

Cluster 3: Compounds 15, 21, 23, 26, 34, 40, 42, 43, 47, 49 (aromatic ring, heterocyclic ring, tertiary amine, carboxylate ester group).

Cluster 4: Compounds 14, 17, 18, 19, 24, 35, 44, 46, 50, 58 (aromatic rings, tertiary amine, methoxy substituents, halogen atoms).

Cluster 5: Compounds 6, 7, 8, 12, 13, 36, 39, 41, 45 aromatic rings, tertiary amine, hetero atom ring, ester group, branched alkyl chain).

Cluster 6: Compounds 9, 20, 22, 27, 28, 32, 33, 37, 38, 48, 51, 57, 59, 60 (large rings, aromatic rings, phenolic hydroxyl group).

Cluster 7: Compounds 29, 56 (hydroxyl groups, heterocyclic rings, carboxyl groups, cycloalkane ring).

Members of each cluster are determined to have the highest similarity after analysis of the multivariate data Table 1. After identification into clusters, members within clusters may express similar pharmacodynamics and/or pharmacokinetics. Inferences of potential clinical activity may be determined by examination of cluster members. Cluster analysis provides investigators some means to understand

underlying relationships among this diverse population of 60 compounds.

Ebola virus remains a pathogen inducing high morbidity and mortality. The identification of drugs to treat EVD is an urgent necessity, along with new methods for screening for such drugs [20]. The investigation of compounds previously shown to inhibit viral proliferation could enhance the effectiveness of recognizing new candidate drugs for clinical treatment of infection. Although further study is required and crucial, this study and similar studies present the molecular structures of known virus inhibitors and the molecular properties are useful for elucidating pharmacokinetics and pharmacodynamics. The search for effective virus inhibitors is ongoing.

4. CONCLUSION

The molecular structures of 60 compounds shown in previous studies to inhibit viral proliferation of Ebola virus are presented with numerical values of eight important descriptors. Their molecular scaffolding is found to be highly diverse, having various heteroatom rings, alkyl structures, functional groups, and various aromatic substituents. The numerical values of eight molecular properties are determined with statistical analysis to find ranges and correlations. Pattern recognition K-means cluster analysis divided the 60 compounds into seven clusters having members with the greatest similarity based on molecular properties. Multiple regression analysis determined a mathematical model explaining 94.25% of the variance in the molecular weight. This model will enable the prediction of expected properties for potential drug candidates. The Rule of 5 identified 29 compounds out of the total of 60, considered to be orally active, membrane permeable, and easily absorbed. Box plots identified a narrower criteria in the number of oxygen atoms, nitrogen atoms, hydroxyl groups, amine groups, and rotatable bonds for potential drug candidates. Further studies are vital for the identification of new drugs to treat patients infected with Ebola virus disease.

CONSENT

It is not applicable.

ETHICAL APPROVAL

It is not applicable.

COMPETING INTERESTS

Author has declared that no competing interests exist.

REFERENCES

1. Kuhn JH, Becker SE, Ebihara H, Geisbert T, Johnson KM, Kawaoka Y, et al. Proposal for a revised taxonomy of the family Filoviridae: Classification, names of taxa and viruses, and virus abbreviations. Archives of Virology. 2010;155(12):2083-103.
2. Feldmann H. Ebola- A growing threat. N Engl J Med. 2014;371(15):1375-8.
3. Ng S, Cowling B. Association between temperature, humidity and ebola virus disease outbreaks in Africa. Euroserveillance. 2014;19(35):20892.
4. Hersey S, Martel L, Jambai A, Keita S, Yoti Z, Meyer E, et al. Ebola virus disease-Sierra leone and guinea. Morbidity and Mortality Weekly Report. 2015;64(35):981-984.
5. Bausch DG, Towner JS, Dowell SF, Kaducu F, Lukwiya M, Sanchez A, et al. Assessment of the risk of ebola virus transmission from bodily fluids and formites. J Inf Dis. 2007;196(Suppl 2): S142-S147.
6. Wiwanitkit V. Ebola virus infection: what should be known? North American Journal of Medical Sciences. 2014;6(11):549-552.
7. Dowell SF, Mukunu R, Ksiazek T, Khan A, Rollin P, Peters C, et al. Transmission of Ebola hemorrhagic fever: A study of risk factors in family members, Kikwit, Democratic Repurblic of the Congo. J Infect Dis. 1995;179(Suppl 1):s87-91.
8. Fischer R, Judson S, Miazgowicz K, Bushmaker T, Munster V. Ebola virus persistence in semen ex vivo. Emerg Infect Dis. 2016;22(2):289-91.
9. Hotta H. Viral hemorrhagic fever-Ebola hemorrhagic fever, Marburg disease and Lassa fever. Rinsho Byori. 1998;46(7):651-5.
10. Bah EI, Lamah MC, Fletcher T, Jacob ST, Brett-Major DM, Sall AA, et al. Clinical presentation of patients with *Ebola virus* disease in Conakry, Guinea. N Engl J Med. 2015;372(1):40.
11. Madelain V, Nguyen T, Olivo A, Lamballerie X, Guedj J, Taburet AM, et al. Ebola virus infection: Review of the

pharmacokinetic and pharmacodynamic properties of drugs considered for testing in human efficacy trials. Clin Pharmacokinet. 2016;10(1007):1-17.
12. Bartzatt R. Anti-iinflammatory drugs and prediction of new structures by comparative anlaysis. Antiinflamm Antiallergy Agents Med Chem. 2012;11(2): 151-60.
13. Bartzatt R. Lomustine analogous drug structures for intervention of brain and spinal cord tumors: The benefit of in silico substructure search and analysis. Chemother Res Pract. 2013; 2013(360624):1-7.
14. Bartzatt R. Drug analogs of COX-2 selective inhibitors lumiracoxib and valdecoxib derived from in silico search and optimization. Antiinlamm Antiallergy Agents Med Chem. 2014;13(1):17-28.
15. Choi WY, Hong KJ, Hong JE, Lee WJ. Progress of vaccine and drug development for Ebola preparedness. 2015;4(1):11-16.
16. Picazo E, Giordanetto F. Small molecule inhibitors of *Ebola virus* infection. Drug Discovery Today. 2015;20(2):277-286.
17. Palm K, Stenberg P, Luthman K, Artursson P. Polar molecular surface properties predict the intestinal absorption of drugs in humans. Pharm Res. 1997;14:568-571.
18. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches in estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev. 2001;46(1-3):3-26.
19. Davis JC. Statistics and data analysis in geology. New York: John Wiley & Sons; 1986.
20. Kouznetsova J, Sun W, Martinez-Romero C, Tawa G, Shinn P, Chen CZ, et al. Identification of 53 compounds that block Ebola virus-like particle entry via a repurposing screen of approved drugs. Emerging Microbes & Infections. 2014;3: e84-e89.

Peer-review history:
The peer review history for this paper can be accessed here:
http://sciencedomain.org/review-history/13746