Article

anti-Nipah: A QSAR based prediction method to identify the inhibitors against Nipah virus

Akanksha Rajput¹, Archit Kumar² and Manoj Kumar *

Virology Discovery Unit and Bioinformatics Centre, Institute of Microbial Technology, Council of Scientific and Industrial Research (CSIR), Sector 39A, Chandigarh-160036, India.

^{1&2}Equal contributions

¹ akanksha@imtech.res.in

² architkumar@imtech.res.in

* Correspondence: manojk@imtech.res.in; Tel.: +91-172- 6665453

Abstract: Nipah virus (NiV) is responsible to cause various outbreaks in Asian countries, with latest from Kerala state of India. Till date there is no drug available despite its urgent requirement. In the current study, we have provided a computational one-stop solution for NiV inhibitors. We have developed “*anti-Nipah*” web resource, which comprised of a data repository, prediction method, and data visualization modules. The database comprised of 313 (181 unique) inhibitors from different strains and outbreaks of NiV extracted from research articles and patents. However, the quantitative structure–activity relationship (QSAR) based predictors were accomplished using classification approach employing 10-fold cross validation through support vector machine with 120 (68p + 52n) inhibitors. The overall predictor showed the accuracy and Matthew’s correlation coefficient of 88.89% and 0.77 on training/testing dataset respectively. The independent validation dataset also performed equally well. The data visualization modules from chemical clustering and principal component analyses displayed the diversity in the NiV inhibitors. Therefore, our web platform would be of immense help to the researchers working in developing effective inhibitors against NiV. The user-friendly webserver is freely available on URL: <http://bioinfo.imtech.res.in/manojk/antinipah/>

Keywords: Nipah Virus, outbreak, inhibitors, QSAR, database, prediction algorithm

1. Introduction

Nipah virus infection is an emerging zoonotic infectious disease caused by Nipah virus (NiV). It is one of the important public health concerns in the South East Asian Region. The NiV is a negative sense single stranded RNA virus, belongs to genus Henipavirus and is a member of Paramyxoviridae family [1]. The first outbreak of NiV was reported from Malaysia during 1998-1999 and thereafter-yearly outbreaks have been reported from Bangladesh or India (http://www.searo.who.int/entity/emerging_diseases/links/nipah_virus_outbreaks_sear/en/). NiV is known to infect various hosts viz., bats, cat, dog, horse, pig and humans whereas fruit bats (genus Pteropus) remain as the main reservoir. The transmission of the virus can occur through direct contact with the contaminants of the infected bats. However, the human-to-human transmission can also be seen within families and in health care workers [2,3]. The incubation period of the NiV ranges from 5 to 14 days (<https://www.cdc.gov/vhf/nipah/pdf/factsheet.pdf>). Main clinical presentation seen among the NiV infected individuals involves the symptoms like fever and headache followed by encephalitis. Besides neurological manifestations, respiratory involvement was also documented in up to 69% of patients affected in Bangladesh-Indian outbreaks [4]. The NiV infection has been reported to be associated with significant morbidity and mortality. The varied degree of mortality has been reported in various outbreaks. As per WHO, an average case fatality rate of 75% was

observed for NiV infection (http://www.searo.who.int/entity/emerging_diseases/links/nipah_virus_outbreaks_sear/en/). In May 2018, India has witnessed another outbreak of NiV, after 2007, where 19 individuals were affected with a reported mortality of 94% (17/19) (<http://www.who.int/csr/don/07-august-2018-nipah-virus-india/en/>).

Despite being highly pathogenic, no prophylactic or therapeutic intervention is available against NiV. The supportive treatment consists of anticonvulsants is the mainstay for NiV infected patients [4]. The efficacy of Ribavirin has also been evaluated against NiV infection. During the Malaysian outbreak, Ribavirin treatment resulted in a reduction of mortality rate to 32% compared to 54% in the control group [5]. However, the later studies have reported the Ribavirin alone or in combination with Chloroquine were ineffective for the survival of NiV infected hamsters [6,7]. Till now, various studies have been performed on the live virus as well on reporter assay systems. Around 19 compounds have been identified with EC50 or IC50 of less than 1 μ M against live NiV [6,8-13]. A recent study by Dawes et al. have shown the antiviral effect of Favipiravir against live NiV Bangladesh and Malaysian isolates with EC50 of 14.8 and 44.8 respectively [14]. Moreover, their in-vivo studies on hamsters have shown 100% survival rates.

However, there is still not any therapeutic modality available for NiV infection and not a single drug is under clinical trial against NiV. So there is a need for the identification of other putative compounds or drugs against NiV. But owing to the BSL4 pathogen, limited studies could be conducted on the live NiV. Therefore, a wide range of compounds remains unexplored. Thus, there is a need for a computational tool that can identify the unexplored putative inhibitor against NiV. We have previously developed the antiviral prediction servers mainly for Zika virus, Human immunodeficiency virus (HIV), Hepatitis B virus (HBV) and Hepatitis C virus (HCV) [15-17]. However, in the present study, we have collected the overall anti-nipah inhibitors available in the literature and developed first quantitative structure-activity relation (QSAR) based prediction algorithm using support vector machine learning for the identification of anti-NiV compounds along the data visualization modules.

2. Materials and Methods

2.1. Data collection

The experimentally validated compounds with anti-nipah activity were collected from research articles and patents. We used Pubmed (178 articles) and Orbit Intelligence (76 patents) using the search terms "Nipah", "antiviral" OR "inhibit*". The chemical information was fetched from PubChem or Chemspider or drawn using MarvinSketch. The data, representing inhibitory concentration 50 (IC50), effective concentration 50 (EC50), percentage inhibition and viral titers against NiV was obtained from 17 PMIDs and 01 patent. From the overall 181 unique NiV inhibitors we proceeded to develop prediction algorithm. The positive dataset contains 68 non-redundant compounds with percentage inhibition of more than 80% and IC50/EC50 less than 10 μ M concentration, while negative dataset consists of 52 non-redundant compounds with percentage inhibition of less than 30% and IC50/EC50 more than 50 μ M concentration. Hence, a total of 120 non-redundant compounds were used in the successive steps of descriptor calculation and model development.

2.2. Prediction algorithm data preparation

Descriptors calculation: The chemical descriptors were calculated using PaDel [18]. Initially, a total of 16383 descriptors were extracted, which includes 1D, 2D, 3D, and fingerprints.

2.3. Feature selection

The feature selection was performed to remove non-desirable descriptors using mRMR [19]. For classification approach the mRMR resulted in 20 most relevant and minimum redundant features from the 16383. The detail of the relevant features is given in Supplementary Table S1.

2.4. Machine learning

The QSAR model was developed for anti-Nipah compounds using selected descriptors through support vector machine (SVM)[20,21]. The architecture for the development of prediction model is given in Figure 1. We employed the ten-fold cross-validation approach for model development. The overall (68p+52n) dataset was divided into training/testing (61p+47n) and independent validation (7p+5n) data set using randomization method. The robustness of the model was evaluated by performing internal as well as external cross-validation.

Figure 1. Overall architecture for the development of anti-Nipah prediction server.

2.5. Model evaluation

The performance of the QSAR model was evaluated by the sensitivity, specificity, accuracy, Matthew's correlation coefficient (MCC), by using below stated formulas:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \times 100$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \times 100$$

$$\text{MCC} = \frac{(\text{TP} \times \text{TN}) - (\text{FP} \times \text{FN})}{\sqrt{(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TP} + \text{FP})(\text{TN} + \text{FN})}}$$

where, TP = true positive; TN = true negative; FP = false positive; FN = false negative.

2.6. Chemical clustering

The clustering of 120 compounds or drugs with their activity (positive or negative) were made using 'compounds specific bioactivity dendrogram' (C-SPADE) an online web based server [22]. Further for the visualization of positive and negative data in 3 dimensional chemical space principal component analysis (PCA) was performed on the SMILES of the same data set using WebMolCS [23].

3. Results

3.1. Database

The anti-Nipah database contains total 313 entries with 181 unique against Nipah virus with EC50, IC50, and percentage inhibition values. The database contains the fields like inhibitor name, NiV strain, experimental approaches, time and duration on inhibitor delivery, mode of delivery, type of inhibition, survival and cytotoxicity activity, etc. The anti-Nipah compounds experimentally tested found majorly are Flavipiravir in 11 cases followed by AAHL 13, AAHL 16, AAHL 18, AAHL 22, AAHL 23, AAHL 33, AAHL 7 in 10 experiments each. The top-most acting drugs are shown in Figure 2. Various NiV strains were targeted in developing the anti-Nipah compounds e.g. Malaysia-1999, NiV-pVSV, Malaysia, fusion expression vector in 160, 37, 19, and 14 experiments respectively (Figure 3).

Figure 2. Frequency distribution of anti-nipah drugs.

Figure 3. Frequency distribution of Nipah virus strains targeted in the experiments.

Moreover, several assays are being used to check the inhibition of compounds against NiV. Maximally used assay is Chemiluminescent in 89 experiments followed by CatL-Peptide cleavage, Plaque assay, immunolabelling assay, Cytopathic effect assay in 50, 42, 41, and 22 entries respectively (Supplementary Figure 1).

3.2. Prediction

3.2.1. Performance *the* QSAR model

For the identification of the significant features of NiV the classification correlation test was performed between the chemical features and the EC₅₀/IC₅₀/percentage inhibition of NiV compounds using dataset of 120 (68p+52n) compounds. The descriptors viz, FP502, FP169, KRFP4293, FP265, FP644, GATS5e, GraphFP87, FP731, nsssN, ExtFP308, FP490, ExtFP857, FP626, GraphFP101, FP541, PubchemFP824, FP551, GraphFP980, ExtFP666, and FP172 were found most relevant for classification based model development. During the 10-fold cross validation approach the training/testing dataset displayed an accuracy and MCC of 88.89% and 0.77 respectively. However, the independent validation dataset showed accuracy and MCC of 75.00% and 0.51 correspondingly. The values for all the statistical parameters are given in Table 1.

Table 1. Performance of Support Vector Machine models on training/testing and independent validation data sets using 10-fold cross validation

Dataset	Sensitivity	Specificity	Accuracy	MCC	ROC
Training/Testing data set (108=61p+47n)	91.80	85.11	88.89	0.77	0.92
Validation data set (12=7p+5n)	71.43	80.00	75.00	0.51	0.86

MCC, Matthews' correlation coefficient; ROC, Receiver operating characteristics;

3.2.2. Web server

The QSAR based prediction model of anti-Nipah compounds or drugs was amalgamated into an openly available and easy to execute web server, '**anti-Nipah**'. The users can use anti-Nipah for the prediction of antiviral activity of the compounds or drugs against NiV. The "*anti-Nipah*" is freely available on url: <http://bioinfo.imtech.res.in/manojk/antinipah/> and consisted of following features:

3.2.3. Database browse and search:

User can get the information of available NiV inhibitors available in the literature as well as patents. The anti-Nipah contains browse can be performed on four important fields like NiV inhibitors, Strains, Assays, and mode of inhibitor delivery. Moreover, the user can check the information of respective compounds in the database through the search tool option of the server like viral strains, experimentation, assays performed, etc.

3.2.4. Predictor (Input and Output):

The user can submit the query compound to the server. The server in return will predict the potency of the query compound or drug against NiV. It will suggest the user for the potential antiviral activity of the query compound or drug against NiV. Moreover, the user can also check the other drug-likeness properties such as, molecular formula, formal charges, H-bond acceptor and donor, Lipinski acceptor and donor, rotatable bonds, etc. in the server.

3.2.5. Draw:

The Marvin sketch tool is incorporated into the "anti-Nipah" server so that user can also draw their query compounds of drugs. After submission of the structure, the user will know about the activity and chemical properties of the query compound or drug.

3.3. Chemical clustering

The dendrogram of the positive dataset (68 highly effective compounds) and a negative dataset (53 less effective/inactive compounds) along with their activity was constructed (**Figure 4**). The smiles information was used for clustering using extended-connectivity fingerprint 4 (ECFP 4) module of the C-SPADE. Large shape spheres represent the active compounds while the less effective/inactive compounds can be seen as small spheres.

Further, the PCA was performed using ECFP4 module of WebMolCS. The coordinates obtained for x-, y-, and z-axis were used for generating scatter plots for highly effective (red) and less effective/inactive (green) dataset (**Supplementary Figure S2**). After clustering analysis, the demarcation between the highly effective and less effective/inactive can be seen. The active compounds or drugs are mainly seen in clusters while the non-active compounds are dispersed through the 3D space.

Figure 4. Dendrogram of highly effective and less effective/inactive compounds against Nipah virus (NiV): The blue nodes represent the respective compound tested against NiV. The yellow spheres represent the activity of the compounds. The large spheres represent the compounds having the anti-Nipah activity while small spheres represent the less effective/inactive compounds.

4. Discussion

The presence of antiviral prediction web servers is important for combating the emerging viral infections speedily. It become more important, especially for the viruses against whom no treatment modality is available. Although, the limited antiviral prediction algorithms are available for the prediction of compounds against viral infections, which includes the AVCpred and HIVprotI [15,16]. The AVCpred is an antiviral compound prediction server, especially for viruses HIV, HCV, HBV, HHV and also, include a general prediction tool for 26 viruses. While HIVprotI is a dedicated prediction server for compounds specifically targeting the integrase, protease and reverse transcriptase of HIV. However, in the present study, we have developed first QSAR based 'anti-Nipah' prediction web server for the screening of compounds having antiviral activity against NiV.

We have developed a database (313 entries) with all the inhibitors of Nipah Virus, which include 120 chemical inhibitors which are highly diverse and varies in molecular weight from 182 to 1000 g/mol. The maximally validated anti-nipah inhibitor is Flavipiravir, which is a purine analog and one amongst the effective broad-spectrum antiviral effective against RNA viruses [14]. Followed by, several novel antivirals designed to target the replication stage of NiV [11]. Likewise, the user can get the details of all the NiV inhibitors available in research articles as well as patents at a platform using our web resource. Further, using the experimentally validated data, we developed a robust QSAR based prediction method. This is the first QSAR based web server dedicated to Nipah virus. The user can predict the anti-NiV activity of any unknown compound with the accuracy of ~89.00%. Therefore, our predictor would be very helpful to experimental biologists for speeding up their research towards developing novel and effective anti-NiV scaffolds.

The analyses showed that diversity of the anti-NiV compounds. PCA analyses showed that highly effective (Percentage inhibition >80% & IC₅₀ <10 μ M) compounds against NiV are clustered together, while the less effective/inactive (Percentage inhibition <30% & IC₅₀ >50 μ M) compounds are scattered throughout the 3D space. Thus, the presence of similar scaffold or chemical modifications might have resulted in the clustering of the active compounds. Though most of the less effective/inactive compounds are dispersed, a few of them were found in proximity to the active compounds. The chemical modifications in functional groups might have increased their activity against NiV. Most of LJ series compounds, synthesised using side chain modification, were found to be clustered together. Interestingly, while performing the chemical clustering analyses, we found that in a few cases, highly efficient and less efficient/inactive compounds were also clustered together. A modification in either carbon chain, carbon ring or side group may have caused the transient shift in the activity of these compound. Therefore, the construction of libraries using side group modifications can be useful for screening the effective antiviral compounds against NiV. Moreover, our anti-Nipah prediction tool will be useful in the screening of such compounds against NiV.

5. Conclusions

Nipah Virus is known for several outbreaks, the recent one is from Kerala, India from May-June 2018. NiV is responsible for causing significant mortality, varies from 0%-100%, among affected individuals. Therefore, in the current study we tried to investigate the anti-Nipah agents. Our first approach is to compile all the anti-Nipah agents available in pubmed and patents. Followed by developing the QSAR based predictor with all the inhibitors, and provide them in form of user-friendly web interface named "anti-Nipah". We hope that our web resource would prove beneficial for researchers to predict effective anti-nipah agents.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, Figure S1: Frequency distribution of assays used to check the inhibition of anti-Nipah compounds, Figure S2: The principal component of the highly effective and less effective/inactive compounds against NiV. The red spheres (left) represent active compounds and green spheres (right) represent less effective/inactive compounds against NiV, Table S1: Details

of the most relevant descriptors extracted from mRMR feature selection algorithm and used for prediction model development.

Author Contributions: Idea was conceived by MK and also helped in interpretation, analysis and overall supervision. Data collection by AK and curation was performed by AK, AR. Predictive models and web server development by AR and analysis was done by AR, AK, MK. Manuscript writing AK, AR, MK.

Funding: Please add: This research received no external funding.

Acknowledgments: We acknowledge the support of Council of Scientific and Industrial Research (CSIR) (OLP-501) and Department of Biotechnology, and infrastructure support Government of India (GAP0001).

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

References

1. Wang, L.; Harcourt, B.H.; Yu, M.; Tamin, A.; Rota, P.A.; Bellini, W.J.; Eaton, B.T. Molecular biology of Hendra and Nipah viruses. *Microbes Infect* 2001, 3, 279-287.
2. Luby, S.P.; Gurley, E.S.; Hossain, M.J. Transmission of human infection with Nipah virus. *Clin Infect Dis* 2009, 49, 1743-1748, doi:10.1086/647951.
3. Chadha, M.S.; Comer, J.A.; Lowe, L.; Rota, P.A.; Rollin, P.E.; Bellini, W.J.; Ksiazek, T.G.; Mishra, A. Nipah virus-associated encephalitis outbreak, Siliguri, India. *Emerg Infect Dis* 2006, 12, 235-240, doi:10.3201/eid1202.051247.
4. Ang, B.S.P.; Lim, T.C.C.; Wang, L. Nipah Virus Infection. *J Clin Microbiol* 2018, 56, doi:10.1128/jcm.01875-17.
5. Chong, H.T.; Kamarulzaman, A.; Tan, C.T.; Goh, K.J.; Thayaparan, T.; Kunjapan, S.R.; Chew, N.K.; Chua, K.B.; Lam, S.K. Treatment of acute Nipah encephalitis with ribavirin. *Ann Neurol* 2001, 49, 810-813.
6. Freiberg, A.N.; Worthy, M.N.; Lee, B.; Holbrook, M.R. Combined chloroquine and ribavirin treatment does not prevent death in a hamster model of Nipah and Hendra virus infection. *J Gen Virol* 2010, 91, 765-772, doi:10.1099/vir.0.017269-0.
7. Georges-Courbot, M.C.; Contamin, H.; Faure, C.; Loth, P.; Baize, S.; Leyssen, P.; Neyts, J.; Deubel, V. Poly(I)-poly(C)12U but not ribavirin prevents death in a hamster model of Nipah virus infection. *Antimicrob Agents Chemother* 2006, 50, 1768-1772, doi:10.1128/aac.50.5.1768-1772.2006.
8. Lo, M.K.; Nichol, S.T.; Spiropoulou, C.F. Evaluation of luciferase and GFP-expressing Nipah viruses for rapid quantitative antiviral screening. *Antiviral Res* 2014, 106, 53-60, doi:10.1016/j.antiviral.2014.03.011.
9. Lo, M.K.; Jordan, P.C.; Stevens, S.; Tam, Y.; Deval, J.; Nichol, S.T.; Spiropoulou, C.F. Susceptibility of paramyxoviruses and filoviruses to inhibition by 2'-monofluoro- and 2'-difluoro-4'-azidocytidine analogs. *Antiviral Res* 2018, 153, 101-113, doi:10.1016/j.antiviral.2018.03.009.
10. Aljofan, M.; Sganga, M.L.; Lo, M.K.; Rootes, C.L.; Porotto, M.; Meyer, A.G.; Saubern, S.; Moscona, A.; Mungall, B.A. Antiviral activity of gliotoxin, gentian violet and brilliant green against Nipah and Hendra virus in vitro. *Virol J* 2009, 6, 187, doi:10.1186/1743-422x-6-187.
11. Aljofan, M.; Saubern, S.; Meyer, A.G.; Marsh, G.; Meers, J.; Mungall, B.A. Characteristics of Nipah virus and Hendra virus replication in different cell lines and their suitability for antiviral screening. *Virus Res* 2009, 142, 92-99, doi:10.1016/j.virusres.2009.01.014.
12. Lo, M.K.; Jordan, R.; Arvey, A.; Sudhamsu, J.; Shrivastava-Ranjan, P.; Hotard, A.L.; Flint, M.; McMullan, L.K.; Siegel, D.; Clarke, M.O., et al. GS-5734 and its parent nucleoside analog inhibit Filo-, Pneumo-, and Paramyxoviruses. *Sci Rep* 2017, 7, 43395, doi:10.1038/srep43395.
13. Aljofan, M.; Lo, M.K.; Rota, P.A.; Michalski, W.P.; Mungall, B.A. Off Label Antiviral Therapeutics for Henipaviruses: New Light Through Old Windows. *J Antivir Antiretrovir* 2010, 2, 1-10, doi:10.4172/jaa.1000014.
14. Dawes, B.E.; Kalveram, B.; Ikegami, T.; Juelich, T.; Smith, J.K.; Zhang, L.; Park, A.; Lee, B.; Komono, T.; Furuta, Y., et al. Favipiravir (T-705) protects against Nipah virus infection in the hamster model. *Sci Rep* 2018, 8, 7604, doi:10.1038/s41598-018-25780-3.
15. Qureshi, A.; Rajput, A.; Kaur, G.; Kumar, M. HIVprofl: an integrated web based platform for prediction and design of HIV proteins inhibitors. *J Cheminform* 2018, 10, 12, doi:10.1186/s13321-018-0266-y.
16. Qureshi, A.; Kaur, G.; Kumar, M. AVCpred: an integrated web server for prediction and design of antiviral compounds. *Chem Biol Drug Des* 2017, 89, 74-83, doi:10.1111/cbdd.12834.
17. Gupta, A.K.; Kaur, K.; Rajput, A.; Dhanda, S.K.; Sehgal, M.; Khan, M.S.; Monga, I.; Dar, S.A.; Singh, S.; Nagpal, G., et al. ZikaVR: An Integrated Zika Virus Resource for Genomics, Proteomics, Phylogenetic and Therapeutic Analysis. *Sci Rep* 2016, 6, 32713, doi:10.1038/srep32713.
18. Yap, C.W. PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. *J Comput Chem* 2011, 32, 1466-1474, doi:10.1002/jcc.21707.
19. Peng, H.; Long, F.; Ding, C. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans Pattern Anal Mach Intell* 2005, 27, 1226-1238, doi:10.1109/tpami.2005.159.
20. Rajput, A.; Gupta, A.K.; Kumar, M. Prediction and analysis of quorum sensing peptides based on sequence features. *PLoS One* 2015, 10, e0120066, doi:10.1371/journal.pone.0120066.

21. Rajput, A.; Thakur, A.; Sharma, S.; Kumar, M. aBiofilm: a resource of anti-biofilm agents and their potential implications in targeting antibiotic drug resistance. *Nucleic Acids Res* 2018, 46, D894-d900, doi:10.1093/nar/gkx1157.
22. Ravikumar, B.; Alam, Z.; Peddinti, G.; Aittokallio, T. C-SPADE: a web-tool for interactive analysis and visualization of drug screening experiments through compound-specific bioactivity dendrograms. *Nucleic Acids Res* 2017, 45, W495-w500, doi:10.1093/nar/gkx384.
23. Awale, M.; Probst, D.; Reymond, J.L. WebMolCS: A Web-Based Interface for Visualizing Molecules in Three-Dimensional Chemical Spaces. *J Chem Inf Model* 2017, 57, 643-649, doi:10.1021/acs.jcim.6b00690.