Journal of Cheminformatics

RESEARCH Open Access

Nipah Virus Inhibitor Knowledgebase (NVIK): a combined evidence approach to prioritise small molecule inhibitors

Bhupender Singh[1†], Nishi Kumari[2†], Ayush Upadhyay[3†], Bhavini Pahuja[4], Eugenia Covernton[5], Kishan Kalia[6], Kanika Tuteja[7], Priyanka Rani Paul[8,9], Rakesh Kumar[1,9,10], Mayur Sudhakar Zarkar[1] and Anshu Bhardwaj[1,9*]

Abstract

Nipah Virus (NiV) came into limelight due to an outbreak in Kerala, India. NiV infection can cause severe respiratory and neurological problems with fatality rate of 40–70%. It is a public health concern and has the potential to become a global pandemic. Lack of treatment has forced the containment methods to be restricted to isolation and surveillance. WHO's 'R&D Blueprint list of priority diseases' (2018) indicates that there is an urgent need for accelerated research & development for addressing NiV. In the quest for druglike NiV inhibitors (NVIs) a thorough literature search followed by systematic data curation was conducted. Rigorous data analysis was done with curated NVIs for prioritising curated compounds. Our efforts led to the creation of Nipah Virus Inhibitor Knowledgebase (NVIK), a well-curated structured knowledgebase of 220 NVIs with 142 unique small molecule inhibitors. The reported IC50/EC50 values for some of these inhibitors are in the nanomolar range—as low as 0.47 nM. Of 142 unique small-molecule inhibitors, 124 (87.32%) compounds cleared the PAINS filter. The clustering analysis identified more than 90% of the NVIs as singletons signifying their diverse structural features. This diverse chemical space can be utilized in numerous ways to develop druglike anti-nipah molecules. Further, we prioritised top 10 NVIs, based on robustness of assays, physicochemical properties and their toxicity profiles. All the NVIs related information including their structures, physicochemical properties, similarity analysis with FDA approved drugs and other chemical libraries along with predicted ADMET profiles are freely accessible at https://datascience.imtech.res.in/anshu/nipah/. The NVIK has the provision to submit new inhibitors as and when reported by the community for further enhancement of the NVIs landscape.

Scientific contribution

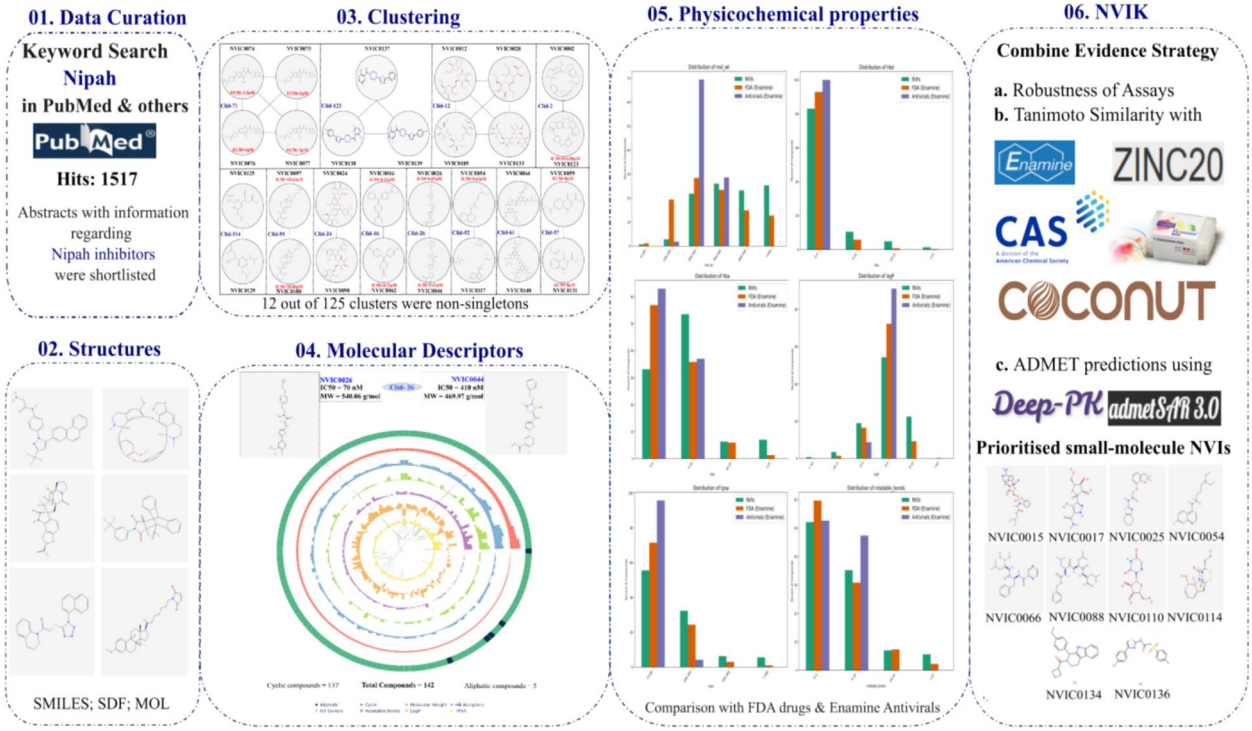
The NVIK is a dedicated resource for NiV drug discovery containing manually curated NVIs. The NVIs are structurally mapped with known chemical space to identify their structural diversity and recommend strategies for chemical library expansion. Also, in NVIK a combined evidence-based strategy is used to prioritise these inhibitors.

Keywords Nipah Virus, Inhibitors, Cheminformatics, Antivirals, Crowdsourcing, ADMET

[†]Bhupender Singh, Nishi Kumari, and Ayush Upadhyay have equal contribution.

*Correspondence:
Anshu Bhardwaj
anshu@imtech.res.in
Full list of author information is available at the end of the article

Graphical Abstract



Introduction

Nipah Virus (NiV) is a deadly human and animal pathogen from the paramyxoviridae family of genus henipavirus. Its genome consists of a negative-sense single-stranded RNA which encodes for six structural proteins including nucleocapsid protein, phosphoprotein, large protein/RNA polymerase, glycoprotein, fusion protein and matrix protein [1]. The genome size of NiV is 18.2 kb and the viral particles vary in their morphology (spherical and filamentous) and size (40–1900 nm). Two genetic lineages of NiV from Malaysia and Bangladesh are known to infect humans [2]. Pteropid bats are the known reservoir of NiV and they are also able to spread to humans and some other species. This virus can cause lethal respiratory and neurological conditions across their hosts [3].

First human outbreak of NiV, was reported in 1998–1999 in Malaysia due to zoonotic transfer in humans which were in direct contact with infected pigs [4] [5]. Following this, similar outbreaks with encephalitis were reported in 2001 and 2003 in Bangladesh and serologically confirmed antibodies were found in local *Pteropus* bats possibly acting as NiV reservoir [6, 7]. Since then, NiV outbreaks have been reported in India, Singapore and Philippines with 40–70% mortality [7]. The major transmission pathways of NiV in humans include transmission from bats (consumption of bat infected date palm sap by humans in Bangladesh), livestock (consumption of food infected with bat's secretions following transfer to humans) and human to human transmission [8, 9]. The virus had garnered attention due to its resurgence in the Kozhikode district of south Indian state of Kerala in 2021, where it caused the death of a 12-year-old child. Notably, the district was the epicentre of the 2018 nipah outbreak in the state, first ever NiV outbreak in southern India [10, 11]. In September 2023, sixth outbreak of NiV via zoonotic mode of transmission was reported in India, resulting in total six cases with two deaths. The genome sequence analysis of this outbreak of NiV clustered it into the Indian clade of 2018 and 2019 outbreaks [12].

NiV causes cell-to-cell fusion in the host forming multinucleated cells called syncytia. This allows the virus to spread despite the absence of viral budding and greatly influences its infectivity [13, 14]. The incubation period of the virus ranges from 4 to 14 days [7]. Primarily the virus infects the central nervous system but the NiV strain from Bangladesh has been shown to have significant respiratory involvement [2]. The capacity of NiV to acquire frequent mutations, high mortality rate and non-availability of standard treatments makes it a potential bioterrorism agent [15, 16].

Currently, there is no known treatment for the NiV. Broad-spectrum antiviral ribavirin has shown contradictory results with even the most optimistic ones significantly below 50% in causing reduction of the mortality rate [1718, 19]. Other small-molecule inhibitors, namely, favipiravir and remdesivir have shown positive results but their efficacy is still under investigation [20, 21]. Also, few vaccines namely- mRNA-1215 (status: completed, no results posted), PHV02 (status- active) and HeV-sG-V (status: completed, results not available publically) are under various phases of clinical trials [22]. Despite being a lethal pathogen there is not a single drug under clinical trial against this virus. Due to several challenges in discovery and development of NVIs, along with lack of therapeutic intervention and the pathogenic propensity of the virus, nipah infections are included in the WHO's list of priority diseases, and they have proposed strategies, key activities and milestones to prevent and contain nipah infection in the South-East Asia region [23] [24].

Various methods including viral titre reduction assays, syncytium formation and screening of large chemical libraries have been performed and led to identification of NVIs such as sulfonamide antimicrobials with low EC50, broad-spectrum antiviral ribavirin, favipiravir- a purine analogue antiviral showing protection to Syrian hamster model challenged with a lethal dose of NiV and monoclonal antibodies to prevent the NiV fusion and entry [17, 20, 25–27]. Despite reports of NVIs, no systematic evaluation of the reported compounds is performed. A few antiviral inhibitor databases and studies exists, namely, The Influenza Research Database [28, 29], The Virus Pathogen Database and Analysis Resource [30] and Enamine antiviral-library [31] etc. They contain very limited or no information with the focus on NiV. The ones with focus on NVIs are either scattered datasets or curated versions with limited or no evaluation of the anti-Nipah compounds as starting points for drug discovery and development of anti-Nipah inhibitor(s) [32].

Towards this, we adopted a crowdsourcing-based model to systematically curate, evaluate and prioritise NVIs following combined evidence strategy and developed a dedicated resource for NVIs-NVIK. The NVIs were prioritised based on assay robustness, their toxicity profiles and structural similarity with FDA approved drugs, Enamine antivirals and other chemical libraries. In addition, the ADMET predictions were carried out using deep learning-based methods and the NVIs were also screened for the presence of PAINS substructures. Based on a combined evidence approach, a few candidate NVIs are proposed as starting points for drug discovery of NiV. Further, to expand the chemical space, drug-like compounds similar to these prioritised candidates are reported from FDA approved drugs, Enamine antivirals, ZINC lead-like compounds and COCONUT libraries. All these analyses and curated data is publically available on a web-based platform which also allows the submission and assessment of new molecules reported as NVIs. The NVIK is freely available at https://datascience.imtech.res.in/anshu/nipah/.

Methods

NVIs data curation and structured representation

The biomedical and life sciences literature resource- PubMed (https://www.ncbi.nlm.nih.gov/pubmed) & other web-resources were utilised for retrieval of publications related to NVIs. Studies related to NVIs were retrieved using the keyword 'Nipah' and at the time of curation (21 May 2024) this keyword listed 1517 papers (Supplementary File 1). The papers with confirmed information regarding NVIs were shortlisted to perform manual curation and create a NVIK-Open resource sheet to organise data under following fields (Table 1).

Table 1 NVIK data structure showing various fields/parameters of the NVIs

S. no	Feature	Description
1	Compound ID & Name	NVIC_ID (Internal ID for NVIK), Inhibitor Name (as reported in literature)
2	Compound representation and external IDs	SMILES format of compounds, InChIKey, External IDs include PubChemID/DrugBank ID/ChemSpider ID
3	Physicochemical properties	Molecular weight (MW), Number of H-bond acceptor (HBA), Number of H-bond donor (HBD), Number of rotatable bonds (RB), LogP, Topological polar surface area (TPSA)
4	Target	Target reported as mentioned in literature, along with mechanism wherever reported
5	Assay Information	Assay type, Assay description, Assay source, Assay cell type
6	Outcome	Outcome of the assay, IC50 / EC50 in nM unit
7	Data Source	Reference, preferably Pubmed ID, URL, Structure citation
8	Curator information	Curator_email_id

The structure of the NVIs were either drawn using MarvinSketch 5.11.4 or retrieved from their corresponding research paper, Pubchem, Chemspider and DrugBank identifiers. In order to ensure consistency, the NVIs were designated with a unique 'NVIC_ID' identifier and their corresponding IC50/EC50 were mentioned in nM concentration. In this way the NVIK-Open resource sheet is created and the email address of the curator (Curator_email_id) is also provided in this sheet corresponding to their curated NVIs.

NVIK platform architecture

The NVIK is built using the LAMP stack: Linux (Ubuntu 20.04.2 LTS), Apache/2.4.41 (Ubuntu), MySQL 8.0.26, and PHP 7.4.3.0. PHP handles the server-side operations, connecting with MySQL to manage the database of NVIs. The platform offers search features, including text, structure, and complex query-based searches, to quickly find the required information. This setup ensures that NVIK is secure, scalable, and user-friendly, making it a valuable tool for research and drug discovery. The NVIK homepage is shown in Fig. 1 which provides an overview of the various data fields and features of the platform.

Utility

NVIK user interface

NVIK allows the user to Browse, Search and Submit NVIs. It also provides the physicochemical properties of NVIs, their ADMET predictions and database statistics. The user can also access structural similarity data of NVIs with FDA approved drugs, Enamine antivirals and other chemical libraries.

NVIK search options

The user can search the database to retrieve the desired results using 'Simple Search', 'Query Builder' and 'Structure Search' options. The 'simple search' option enables the user to perform text based search among all or selected fields of NVIK whereas in 'query builder' the user can customize the query using a combination of different data fields and logical operators to retrieve specific hits. Finally, the 'structure search option' enables the user to retrieve the structurally similar NVIs corresponding to their query molecule submitted in SMILES/SDF/MOL format or drawn using JSME browser.

Browsing interface

The 'Browse' option in the NVIK homepage directs the user to access the various fields of NVIK as shown in Fig. 2. The user can access the NVIK information field-wise such as by inhibitor name, targets/mechanisms, assay type, IC50 and EC50 values. Apart from this the 'Browse' option also provides a link to access inhibitors other than small molecules.

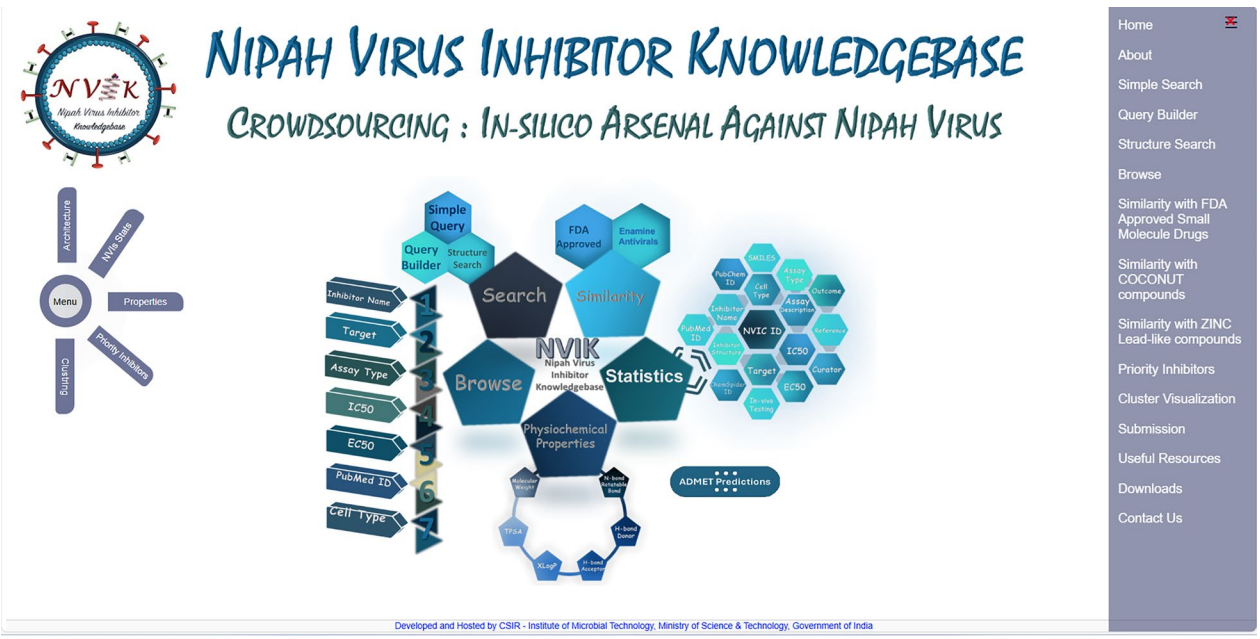


Fig. 1 NVIK homepage architecture illustrating various fields including 'Browse', 'Search', 'Similarity', 'Statistics', 'Physicochemical properties' and 'ADMET' options

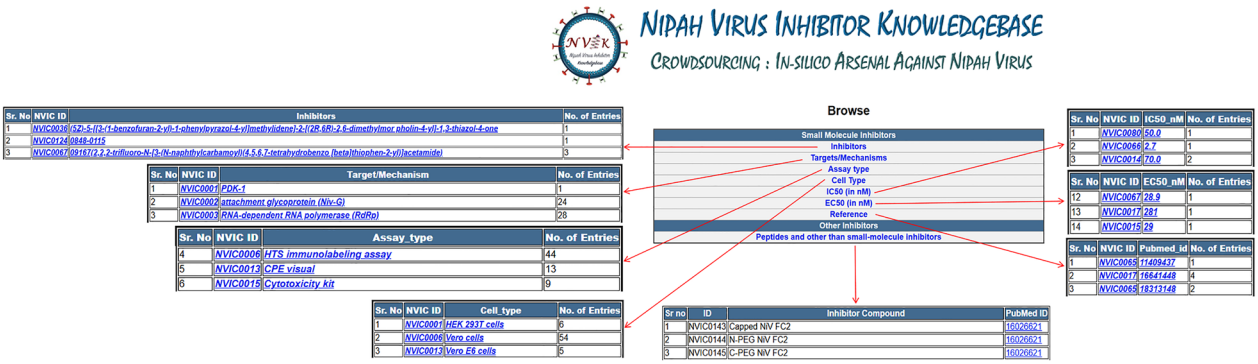


Fig. 2 NVIK browsing option to access the various fields namely- Inhibitors, Targets/Mechanisms, Assay type, Cell type, IC50 (in nM), EC50 (nM), Reference and Other Inhibitors of the curated small-molecule NVIs. This tab also provides the option to access other types of inhibitors including peptide-based inhibitors

Clustering, Similarity analysis and visualisation of the small-molecule NVIs

Clustering

The GenerateMD (https://docs.chemaxon.com/display/docs/fingerprints_fingerprint-and-descriptor-generation-generatemd.md) application program from ChemAxon was used to generate the chemical fingerprints (CF) of the NVIs and later were used for clustering using Jarp (version 6.0.2) (https://docs.chemaxon.com/display/docs/Jarvis-Patrick+clustering), to get variable-length clusters at 0.15 dissimilarity threshold (1-Tanimoto coefficient (Tc)).

Physicochemical property comparison and structural similarity of NVIs

The properties namely MW, HBD, HBA, LogP, TPSA and RB were calculated using RDKit (version 2024.03.5). Physicochemical properties-based comparison of the NVIs was done with FDA approved drugs (1123) (https://enamine.net/compound-libraries/bioactive-libraries/fda-approved-drugs-collection) and Enamine antiviral library (3200) (https://enamine.net/compound-libraries/targeted-libraries/antiviral-library). Additionally, similar comparison was also done with the COCONUT (COlleCtion of Open NatUral producTs), which is a comprehensive resource of natural compounds (https://coconut.naturalproducts.net/), Pathogen box compounds, which contains structurally diverse library of compounds for infectious and neglected diseases (https://www.mmv.org/mmv-open/pathogen-box/about-pathogen-box) and CAS (Chemical Abstracts Service) COVID-19 library which provides unique and unambiguous identifier to the compounds (https://www.cas.org/press-releases/open-access-covid-19-dataset).

Further, structural similarity of the NVIs was carried out with FDA approved drugs, known antivirals (Enamine), COCONUT and CAS libraries using Morgan fingerprints (radius = 2, nBits = 2048) at $Tc \geq 0.8$. The SwissSimilarity (http://www.swisssimilarity.ch) [33] web server was used to find the number of Zinc lead-like compounds overlapping with NVIs using electroshape and Extended-Connectivity Fingerprints (ECFPs). Finally, Pan Assay Interference Compounds (PAINS) in NVIs were identified using RDKit (version 2024.03.5).

Visualisation

The circular plot of the NVIs common physicochemical properties were represented graphically using plotly (version 5.22.0) (https://plotly.com/python/), which is an open-source, data-visualisation library in python. Distinct colours were used to represent different properties. The outermost circle/ring in the plot represents the aliphatic/cyclic nature of the compound and rest of the properties were represented as bar plots using barpolar graph object in plotly. The bar plots were sorted based on the molecular weights.

Utilising Deep Learning (DL) platforms for ADMET prediction of NVIs

Recently developed, web-based DL platforms namely admetSAR 3.0 [34] and DeepPK [35] were used to predict the ADMET properties of the NVIs. Both admetSAR 3.0 and DeepPK are graph neural network based ADMET prediction models. The admetSAR 3.0 is a contrast learning based multitask graph neural network framework, which utilises transformation rules and scaffold hopping optimization strategies to provide predictions on 119 ADMET endpoints. On the other hand, DeepPK uses 449 graph-level features for DMPNN (Directed Message Passing Neural Network) based learning and Bayesian hyperparameter optimization to provide predictions on 64 ADMET endpoints. The SMILES format of the NVIs

were used as input to get the ADMET predictions from both servers.

Combined evidence strategy for NVIs prioritisation

The curated NVIs were prioritized based on the robustness of assays, physicochemical properties, PAINS filter and their toxicity profiles. The physicochemical properties based comparison of NVIs was done with FDA approved drugs and Enamine antivirals considering Lipinski [36] and Veber [37] rules. The toxicity profiles were analysed based on a set of toxicity endpoints. Finally a prioritised list of NVIs is reported using a combined evidence approach.

Results and discussion

NVIK statistics

The current version of the NVIK contains a total of 220 NVIs (Supplementary File 2). There are a total of 142 unique small-molecule NVIs with 178 entries as some compounds are reported in multiple studies. The IC50/EC50 value is available for 66 (47.48%) inhibitors mostly ranging from 0.47 nM to 1030.79 μM and a few ranging from 1848.2 μM to 7562.71 μM and structures are provided for all the 142 NVIs.

NVIK physicochemical properties and structural diversity

An overall distribution of physicochemical properties of NVIK is presented in Fig. 3 in which the NVIs are arranged in the ascending order of their molecular weights. The NVIs distribution of MW, HBA, HBD, RB, LogP and TPSA ranges between 55.84–1626.23 (MW), 0–21 (HBA), 0–16 (HBD), 0–27 (RB), -10.36–7.79 (LogP) and 0–463.93 (TPSA), respectively (Supplementary File 2). Majority (137 out of 142) of the small molecule NVIs are cyclic in nature. The innermost connection lines in this figure represent the unique cluster connections obtained through clustering analysis. Interestingly, we can see compounds with different molecular weights are part of the same cluster (represented by those connections which are passing through the centre of the Fig. 3). These connections belong to following clusters- Clid52 (NVIC0054 (MW = 319.87 g/mol; IC50 = 620 nM) and NVIC0117 (MW = 515.9 g/mol)), Clid12 (NVIC0012 (MW = 504.43 g/mol) and NVIC0028 (MW = 344.31 g/mol), Clid12 (NVIC0028 (MW = 344.31 g/mol) and NVIC0105 (MW = 504.43 g/mol)) and Clid12 (NVIC0028 (MW = 344.31 g/mol) and NVIC0133 (MW = 666.57 g/mol)) with difference in their molecular weights. As seen in Fig. 3, Clid26 comprises of NVIC0026 (MW = 540.06 g/mol; IC50 = 70 nM) and NVIC0044 (MW = 469.97 g/mol; IC50 = 410 nM) which have similar structure and difference in their activity.

Clustering of 142 small molecule NVIs resulted in 12 non-singleton clusters (9.6% of NVIs) (maximum number of compounds per cluster is four) and the remaining 113 clusters (90.4% of NVIs) were singletons (one compound per cluster). Also, the average dissimilarity among NVIs was 0.80 which clearly shows the structurally diverse nature of the NVIs. Interestingly, a difference in IC50/EC50 values were observed between the compounds of a particular cluster as shown in Fig. 4. For example, the Clid71 in Fig. 4 has 4 compounds in which NVIC0074 has EC50 of 1.5 μM while other three compounds in the cluster namely NVIC0075 (EC50 = 3 μM), NVIC0076 (EC50 = 4 μM) and NVIC0077 (EC50 = 3 μM) have relatively higher EC50 values. Similarly, in Clid26, NVIC0026 has an IC50 value of 0.07 μM while NVIC0044 has IC50 of 0.41 μM. Same differences in IC50 values were also observed for Clid-16 and Clid-91 as shown in Fig. 4. These analyses suggest that small-molecule NVIs are not only diverse in their chemical structure but also minor structural variation within members of a cluster can lead to difference in their potency. In cases where activity values for one of the cluster members is reported, the activities of the other members can be explored utilizing the common and different structural features among these compounds.

Comparative distribution of physicochemical properties among NVIs, FDA approved drugs, enamine-antivirals, and other chemical libraries

Physicochemical properties based comparative distribution of NVIs was done by plotting the percentage distribution of the six properties (MW, HBD, HBA, LogP, TPSA and RB) among compounds from NVIs, FDA drugs and enamine antivirals as shown in Fig. 5. Based on physicochemical properties as suggested by Lipinski [36] and Veber [37], it is observed that the majority of NVIs are well-within their prescribed range for oral bioavailability.

It is observed that the distribution of these six properties for NVIs is similar to FDA drugs and also in some cases with enamine antivirals. Similar comparison was made with other chemical libraries, namely, COCONUT, Pathogen box and CAS compounds as shown in Fig. 6. Most of the NVIs also had distribution similar to the COCONUT and CAS libraries followed by the Pathogen box compounds.

Structural similarity of NVIs with FDA approved drugs and targeted libraries

Structural similarity analysis identified 27 NVIs ($Tc \geq 0.8$) similar to the FDA approved drugs out of which 14 NVIs ($Tc = 1$) are structurally identical (Supplementary File 3). The NVIs, including NVIC0015,

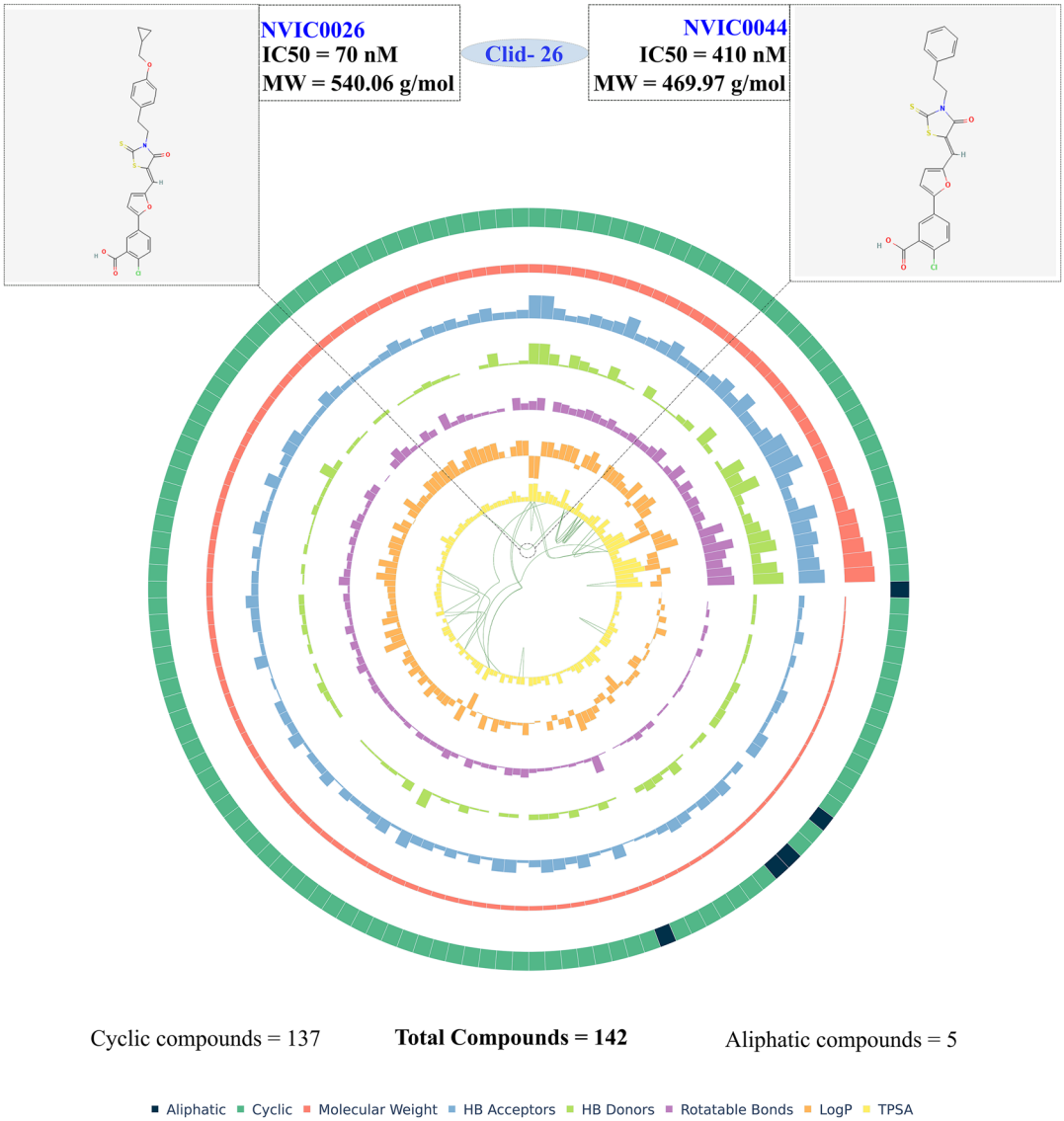


Fig. 3 Circular plot of 142 unique small molecule NVIs showing distribution of their physicochemical properties. The outermost ring represents the aliphatic (blue) and cyclic compounds (green) among NVIs. The NVIs are sorted according to the increasing order of their MW and the subsequent rings represent the MW, HBA, HBD, RB, LogP and TPSA values respectively. The innermost lines represent the connections among NVIs determined through clustering analysis in which the pointed cluster- Clid26 with two members as NVIC0026 and NVIC0044 are in the same cluster despite differences in their MW and IC50 values

NVIC0065 and NVIC0117 are Remdesivir, Ribavirin and Chloroquine which are approved antivirals; NVIC0096 (Cyclosporine) is an immunosuppressive agent; NVIC0102 is Bosutinib which is a kinase inhibitor; NVIC0066 (Bortezomib) is antineoplastic agent and NVIC0061 is Gabapentin, an anticonvulsant according to the FDA Pharm Classes (https://pubchem.ncbi.nlm.nih.gov/source/FDA%20Pharm%20Classes). This shows that a substantial number of NVIs include FDA approved drugs with reported anti-Nipah activity. Such information can be vital towards developing the drug-repurposing strategies against NiV.

Further, we have identified 1314 high similarity hits ($Tc \geq 0.8$) of NVIs in COCONUT natural compounds library (Supplementary File 4). We noticed that a single NVI entry with $Tc = 1$ was matching with multiple COCONUT identifiers (for example: NVIC0012, NVIC0024, NVIC0056, NVIC0133 and others showed hits to multiple COCONUT compounds). We recommend not to consider $Tc = 1$ as an identical hit in this

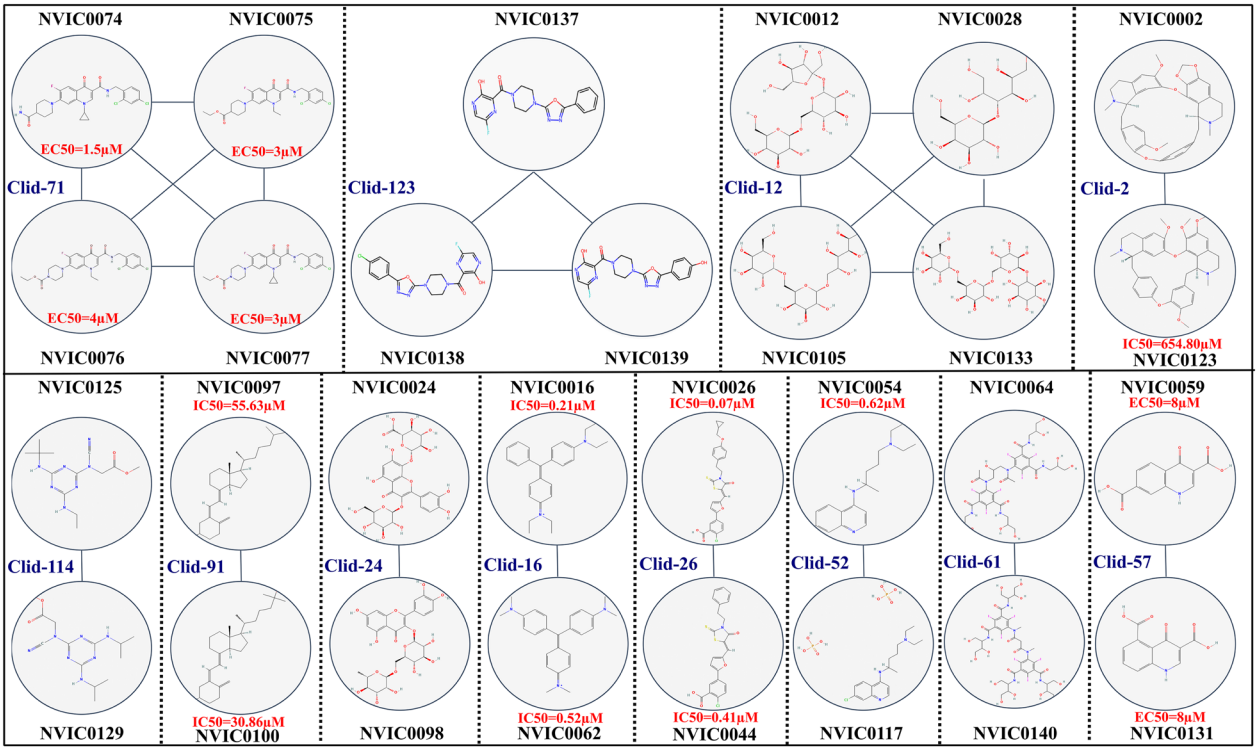


Fig. 4 Structural representation of the clusters observed in the small-molecule NVIs. The IC50/EC50 values of NVIs are mentioned as reported in literature. Differences in IC50/EC50 values within the compounds of clusters having Clid, 16, 26, 71 and 91 is reported here. For NVIC0054 minimum IC50 value is reported

analysis as the Tc was computed using Morgan fingerprints which performs substructure-based search and thus the resulting index is representative of substructure identity rather than the entire structural identity. The above explanation can be better understood by considering an example of NVIC0133 which has 5 COCONUT hits (CNP0264570.1, CNP0414235.2, CNP0264570.3, CNP0414235.1 and CNP0264570.2) with Tc = 1 in which CNP0264570.1, CNP0264570.2 and CNP0264570.3 has different molecular weights and number of aromatic rings as compared to NVIC0133 but still has Tc = 1 signifying only the substructure identity. Similar comparisons with targeted libraries, namely, Enamine-antivirals found one hit with maximum Tc of 0.58 (maximum among all the antivirals) and among pathogen box compounds got one hit with maximum Tc of 0.85 (Supplementary File 5).

Further, PAINS substructure identification in small-molecule NVIs led to screening of frequent hitters. PAINS substructures were identified in 18 out of the 142 compounds. This shows that the majority (87.32%) of the curated NVIs were free from PAINS substructures.

ADMET predictions of the NVIs

DL based ADMET predictions of NVIs can be accessed at NVIK homepage as shown in Fig. 1. Many drugs fail to reach the market due to an imbalance between the pharmacokinetic and pharmacodynamic profile, which includes absorption, distribution, metabolism, excretion and toxicity (ADMET) characteristics [38]. Therefore, two DL methods namely admetSAR 3.0 and DeepPK were applied to NVIK for prioritisation of drug-like inhibitors. The analysis was focused on the toxicity-based predictions from these servers owing to its importance in the drug discovery process to screen and prioritise compounds based on their safety profile [41]. The admetSAR 3.0 has 43 human health toxicity endpoints and DeepPK has 35 toxicity endpoints. Out of these, 10 common toxicity endpoints, namely, hepatotoxicity [(DILI (Drug Induced Liver Injury) in admetSAR 3.0; Liver Injury II in Deep-PK (as it considers only human hepatotoxicity)], hERG blockers [(hERG (< 1 μM > 10 μM) for admetSAR 3.0 (as it is the best developed-model for the used dataset) and hERG in Deep-PK], mutagenicity, respiratory toxicity, Glucocorticoid Receptor (GR), Thyroid Receptor (TR), Mitochondrial Membrane Potential (MMP), micronucleus, Estrogen Receptor (ER) and p53 were selected based on their relevance to the current study. The admetSAR 3.0 predictions of the above-mentioned toxicity endpoints identified 7 non-toxic NVIs (NVIC0012, NVIC0017, NVIC0028, NVIC0059, NVIC0105,

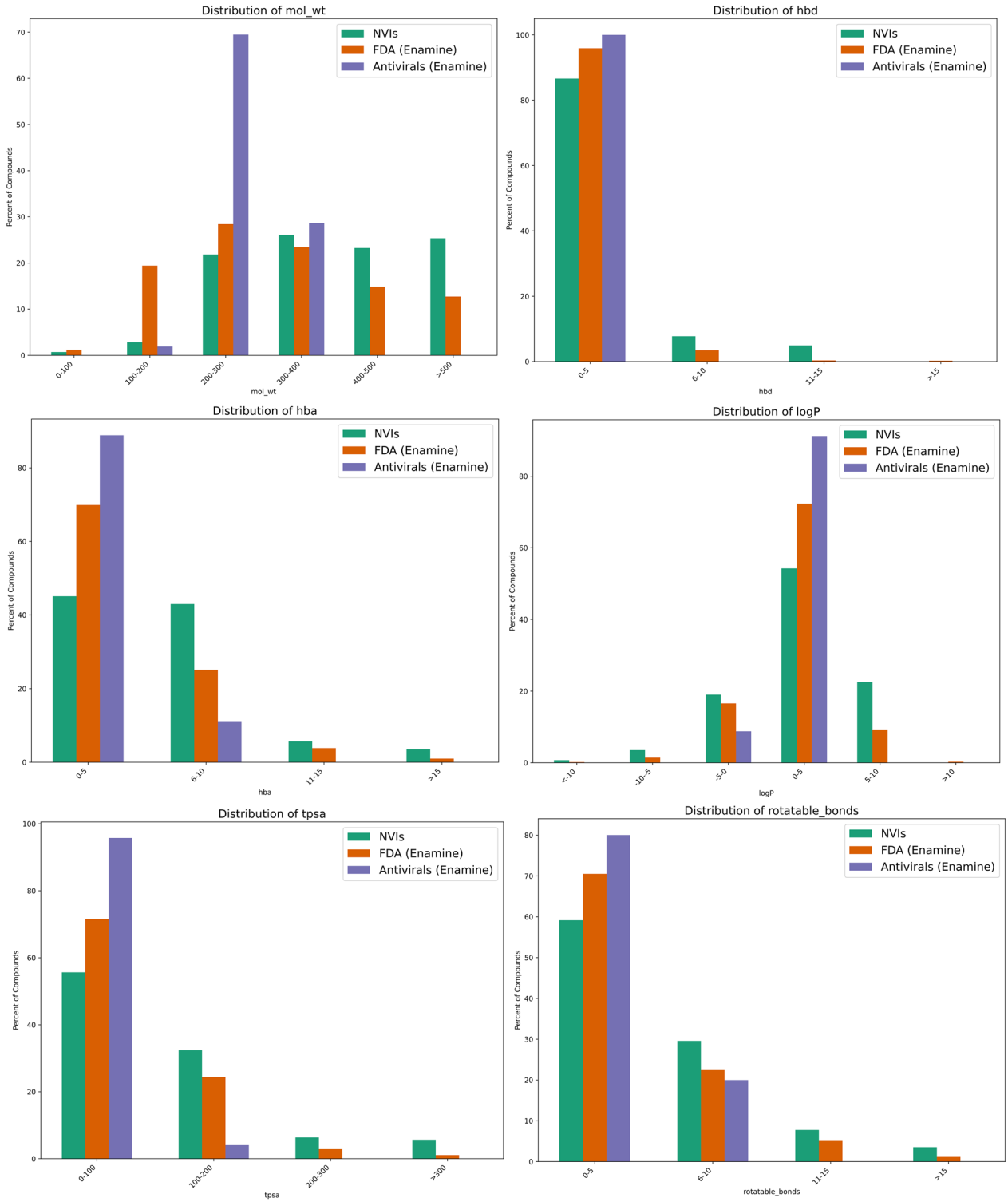


Fig. 5 Comparative distribution of six physicochemical properties among NVIs (green), FDA approved drugs (orange) and Enamine antivirals (purple). The x-axis represents a range of physicochemical properties, and the y-axis represents the percentage of compounds in that range for all three datasets

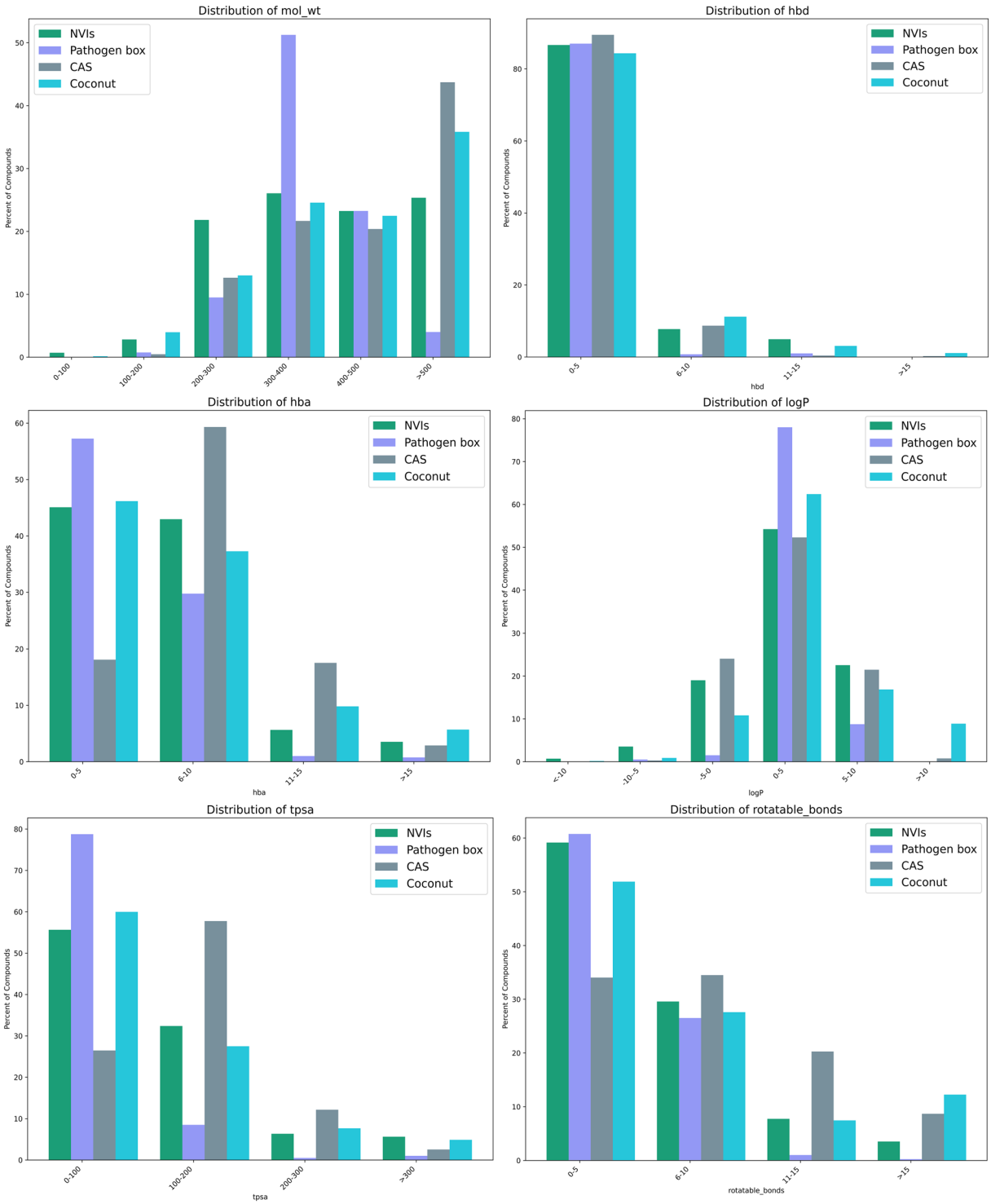


Fig. 6 Comparative distribution of six physicochemical properties among NVIs (green), Pathogen box (sage), CAS (grey) and COCONUT (teal). The x-axis represents a range of various physicochemical properties, and the y-axis represents the percentage of compounds in that range for all four datasets

Table 2 Details of the top 10 prioritised small-molecule NVIs along with their biological activity, presence of PAINS flag, structural similarity with FDA drugs, other chemical libraries and number of toxicity endpoints cleared based on predictions from Deep-PK and admetSAR3.0

NVIK identifier (Name)	IC50/EC50 (μM)	PAINS flag	Tanimoto similarity (Tc)			Swiss similarity	Toxicity endpoints cleared (out of 10)	
			No. of FDA drugs(Tc ≥ 0.8)	No. of COCONUT molecules (Tc ≥ 0.8)	No. of CAS molecules (Tc ≥ 0.8)	No. of ZINC lead like molecules (Electroshape) (Score ≥ 0.85)	DeepPK	admetSAR 3.0
NVIC0015 (Remdesivir)	0.029*	No	DB14761	–	22	97	6	7
NVIC0054 (Chloroquine)	0.62*	No	DB00608	4	32	400	5	7
NVIC0066 (Bortezomib)	0.0027	No	DB00188	1	-	400	6	6
NVIC0017 (Pyrazofurin)	0.281*	No	–	2	–	–	8	10
NVIC0025 (ZHAWOC21026)	0.08	No	–	–	–	400	7	6
NVIC0088 (MG132)	0.00047	No	–	2	–	400	7	5
NVIC0110 (6-Azauridine)	0.25**	No	–	1	1	16	6	6
NVIC0114 (Gliotoxin)	0.149*	No	–	10	-	400	7	7
NVIC0134 (C1)	1.83	Yes	–	–	–	400	5	5
NVIC0136 (F1)	4.15	No	–	–	–	400	8	7

The prioritised NVIs identified as approved drugs are given with their DrugBank accession number. None of the compounds mentioned in this table showed similarity (Tc ≥ 0.8) with enamine-antivirals and pathogen box compounds

* Minimum value has been reported

** Minimum effective concentration in μg/ml

NVIC0119 and NVIC0131) whereas Deep-PK predictions identified two safe (NVIC0028 and NVIC0105) compounds which cleared all 10 toxicity endpoints, respectively (Supplementary File 6).

A favourable PK-PD profile requires that the drug is well absorbed, typically through the gastrointestinal tract if administered orally, achieving sufficient systemic circulation. The NVIK consists of small molecule NVIs, most of which can be administered orally and hence F50 is an important criteria for classification. The admetSAR 3.0 predicted F50 value of 65 out of 139 (46.76%) NVIs to be greater than 0.5 whereas, for Deep-PK predictions, 75 out of 139 NVIs (53.95%) were predicted as orally bioavailable. Once in the bloodstream, the drug should be efficiently distributed to its target tissues. Steady state volume of distribution (VDss) is reported to be optimal (in between 0.71 L/Kg to 2.81 L/Kg) for 86 out of 139 (61.87%) NVIs as per Deep-PK predictions whereas admetSAR 3.0 predicted distribution of 26 NVIs (18.7%) in the medium (13 out of 139 NVIs) and high (13 out of 139 NVIs) range for VDss. This parameter is considered relevant because it is used to determine the effective dose [39]. The drug should undergo metabolism in the liver, in a way that minimises toxic metabolite production, ensuring maintenance of therapeutic efficacy. admetSAR 3.0 classified 15 out 139 (10.79%) NVIs as CYP2D6 inhibitors while Deep-PK predicted 18 out of 139 (12.94%) NVIs as CYP2D6 inhibitors. Individuals with a deficiency of this enzyme are classified as poor metabolizers [40]. Finally, the drug must be eliminated efficiently via renal and hepatic pathways to prevent accumulation in the body cells. For half life, Deep-PK predicted 30 out of 139 (21.58%) NVIs with half life ≥ 3 h while no conclusive half life estimates can be made from admetSAR 3.0 (Supplementary File 6). These observations can facilitate prioritization of drug-like NVIs and offer starting points for expanding the chemical space for furthering drug discovery efforts for this pathogen.

Using combined evidence-based strategy for NVIs prioritisation

The small-molecule NVIs were prioritised based on robustness of assays, physicochemical properties and

number of toxicity endpoints cleared. Based on robustness of the assays, we prioritised 16 NVIs (NVIC0014, NVIC0015, NVIC0016, NVIC0017, NVIC0025, NVIC0026, NVIC0044, NVIC0054, NVIC0062, NVIC0066, NVIC0067, NVIC0088, NVIC0110, NVIC0114, NVIC0134 and NVIC0136) with IC50 value ranging from as low as 0.47 nM (NVIC0088) to 4150 nM (NVIC0136). Three of these 16 NVIs are NVIC0015 (GS-5734/Remdesivir), NVIC0054 (Chloroquine) and NVIC0066 (Bortezomib) and are approved drugs. In the prioritisation criteria, compounds which cleared at least five toxicity endpoints from both toxicity prediction servers were selected. From the 16 NVIs, 10 compounds (NVIC0015, NVIC0054, NVIC0066, NVIC0017, NVIC0025, NVIC0088, NVIC0110, NVIC0114, NVIC0134 and NVIC0136) qualified in this criteria. The final priority list consisted of 10 small-molecule NVIs with reported IC50/EC50 values, cleared toxicity and PAINS filters as shown in Table 2. Structural similarity ($Tc \geq 0.8$) based mapping of the NVIs identified one hit each for three NVIs in FDA drugs and multiple hits for six and three NVIs in COCONUT and CAS libraries (Supplementary File 7), respectively as shown in Table 2. Based on structural similarity with ZINC lead-like library, no hits were observed for ECFP4 similarity search with set threshold score ≥ 0.85 and several hits were observed for electroshape based similarity search (Table 2) (Supplementary File 8).

We have developed a strategy to curate NVIs, perform their evaluation based on robustness of assays, physicochemical properties, structural diversity and ADMET predictions to prioritise a set of evidence based high confidence inhibitors for NiV. Further, to keep the NVIK updated and advance the prioritisation of potential NVIs, a web server is made available at https://datascience.imtech.res.in/anshu/nipah/.

Conclusion

In this study, we have developed NVIK, which is a dedicated resource for small molecule NVIs reported through experimental and computational approaches. We adopted a combined evidence strategy including assay robustness and toxicity endpoints to prioritise the NVIs. Our efforts led to the creation of a well-curated structured knowledgebase of 220 NVIs with 142 small molecule inhibitors. We also prioritised top 10 NVIs based on combined evidence and reported structurally similar compounds to these in FDA approved drugs and other important chemical libraries to develop strategies for NVIs library expansion using known chemical space. The WHO has clearly indicated a need to accelerate research & development for addressing NiV by including it in the list of priority diseases. We believe that this platform will act as a centralized resource for the scientific community to access, submit and improve NVIs drug discovery landscape.

Abbreviations

NiV	Nipah Virus
WHO	World Health Organization
NVIs	Nipah Virus Inhibitors
NVIK	Nipah Virus Inhibitor Knowledgebase
MW	Molecular weight
HBA	Number of H-bond acceptor
HBD	Number of H-bond donor
RB	Number of rotatable bonds
TPSA	Topological polar surface area
ADMET	Absorption, Distribution, Metabolism, Excretion, Toxicity
FDA	Food & Drug Administration
CF	Chemical Fingerprints
COCONUT	COlleCtion of Open NatUral product
CAS	Chemical Abstracts Service
ECFPs	Extended-Connectivity Fingerprints
PAINS	Pan Assay Interference Compounds
DMPNN	Directed Message Passing Neural Network
CPE	Cytopathic effect
DILI	Drug Induced Liver Injury
GR	Glucocorticoid Receptor
TR	Thyroid Receptor
MMP	Mitochondrial Membrane Potential
ER	Estrogen ReceptorSupplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13321-025-01049-6.

Additional file

Supplementary file 1 (CSV 491 KB)

Supplementary file 2 (XLSX 56 KB)

Supplementary file 3 (XLSX 2916 KB)

Supplementary file 4 (CSV 4099709 KB)

Supplementary file 5 (XLSX 10350 KB)

Supplementary file 6 (XLSX 478 KB)

Supplementary file 7 (CSV 266516 KB)

Supplementary file 8 (ZIP 2167 KB)

Acknowledgements

BS acknowledges support from HCP047, RK acknowledges fellowship from UGC, PRP acknowledges support from IAS summer training programme, MZ acknowledges support from GAP240.

Author contributions

Conceptualisation: AB, Methodology: AB, BS, NK, AU, BP and EC. Software: RK, KK, MSZ, BS and BP. Formal analysis: BS, NK, AU, BP, KT, PRP, EC, AB. Data Curation: BS, NK, AU, BP, KT, PRP, EC and AB. Writing—Original Draft: BS, NK, AU, BP, AB. Writing—Review & Editing: BS, NK, AU, BP, KT, PRP, EC, RK, KK, MSZ, AB. Visualisation: BS, NK, KK, AB. Funding acquisition: AB.

Funding

OLP0172.

Availability of data and materials
All the Supplementary files and datasets supporting the conclusions of this article are available in the following Zenodo repository, https://doi.org/10.5281/zenodo.15129022. The source code and analysis repository of this article is available in the following Github repository, https://github.com/AB-DataScienceLab/NVIK.

Declarations

Competing interests
The authors declare no competing interests.

Author details
[1]Bioinformatics Centre, CSIR-Institute of Microbial Technology, Chandigarh, India. [2]Department of Biotechnology, University Institute of Engineering and Technology, Panjab University, Chandigarh, India. [3]Department of Microbiology, Bhaskaracharya College of Applied Sciences, Delhi University, New Delhi, India. [4]Department of Pharmacy, Birla Institute of Technology and Sciences, Pilani, Rajasthan, India. [5]Learning Planet Institute, Paris, France. [6]Department of Biotechnology and Bioinformatics, D.A.V. College, Sector-10, Chandigarh, India. [7]Centre of Systems Biology and Bioinformatics, Panjab University, Chandigarh, India. [8]Genomics and Molecular Medicine Unit, Institute of Genomics and Integrative Biology (IGIB), Council of Scientific and Industrial Research (CSIR), Mall Road, Delhi 110007, India. [9]Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201002, India. [10]Translational Bioinformatics Group, International Centre for Genetic Engineering and Biotechnology (ICGEB), New Delhi, Delhi 110067, India.

Received: 3 April 2025 Accepted: 29 June 2025
Published online: 24 November 2025

References

1. Wang LF et al (2001) Molecular biology of Hendra and Nipah viruses. Microbes Infect 3:279–287. https://doi.org/10.1016/S1286-4579(01)01381-8
2. Ang BSP, Lim TCC, Wang L (2018) Nipah virus infection. J Clin Microbiol. https://doi.org/10.1128/JCM.01875-17
3. Halpin K et al (2011) Pteropid bats are confirmed as the reservoir hosts of henipaviruses: a comprehensive experimental study of virus transmission. Am J Trop Med Hyg 85:946–951. https://doi.org/10.4269/AJTMH.2011.10-0567
4. Outbreak of Hendra-Like Virus— Malaysia and Singapore (1998–1999) CDC. https://www.cdc.gov/mmwr/preview/mmwrhtml/00056866.html. Accessed 09 Dec 2024.
5. Parashar UD et al (2000) Case-control study of risk factors for human infection with a new zoonotic paramyxovirus, Nipah virus, during a 1998–1999 outbreak of severe encephalitis in Malaysia. J Infect Dis 181:1755–1759. https://doi.org/10.1086/315457
6. Hsu VP et al (2004) Nipah virus encephalitis reemergence, Bangladesh. Emerg Infect Dis 10:2082–2087. https://doi.org/10.3201/EID1012.040701
7. About Nipah Virus | Nipah Virus | CDC. https://www.cdc.gov/nipah-virus/about/index.html?CDC_AAref_Val=https://www.cdc.gov/vhf/nipah/outbreaks/distribution-map.html. Accessed 09 Dec 2024.
8. Chadha MS et al (2006) Nipah virus-associated encephalitis outbreak, Siliguri, India. Emerg Infect Dis 12:235–240. https://doi.org/10.3201/EID1202.051247
9. Luby SP, Gurley ES, Hossain MJ (2009) Transmission of human infection with Nipah virus. Clin Infect Dis 49:1743–1748. https://doi.org/10.1086/647951
10. Nipah Virus Outbreak in Kerala. https://www.who.int/southeastasia/outbreaks-and-emergencies/health-emergency-information-risk-assessment/surveillance-and-risk-assessment/nipah-virus-outbreak-in-kerala. Accessed 09 Dec 2024.
11. Nipah Virus in Kerala: How did Kerala handle the earlier outbreaks of the disease? https://indianexpress.com/article/explained/nipah-outbreak-kerala-symptoms-treatment-coronavirus-7489944/. Accessed 09 Dec 2024.
12. Anoop Kumar AS et al (2024) Clinico-epidemiological presentations and management of Nipah virus infection during the outbreak in Kozhikode district, Kerala state, India 2023. J Med Virol 96:3. https://doi.org/10.1002/JMV.29559
13. Maisner A, Neufeld J, Weingartl H (2009) Organ- and endotheliotropism of Nipah virus infections in vivo and in vitro. Thromb Haemost 102:1014–1023. https://doi.org/10.1160/TH09-05-0310
14. Diederich S, Thiel L, Maisner A (2008) Role of endocytosis and cathepsin-mediated activation in Nipah virus entry. Virology 375:391–400. https://doi.org/10.1016/J.VIROL.2008.02.019
15. Kulkarni S, Volchkova V, Basler CF, Palese P, Volchkov VE, Shaw ML (2009) Nipah virus edits its P gene at high frequency to express the V and W proteins. J Virol 83:3982–3987. https://doi.org/10.1128/JVI.02599-08
16. Lam SK (2003) Nipah virus—a potential agent of bioterrorism? Antiviral Res 57:113–119. https://doi.org/10.1016/S0166-3542(02)00204-8
17. Snell NJC (2004) Ribavirin therapy for Nipah virus infection. J Virol 78:10211–10211. https://doi.org/10.1128/JVI.78.18.10211.2004
18. Chong HT et al (2001) Treatment of acute Nipah encephalitis with ribavirin. Ann Neurol 49:810–813. https://doi.org/10.1002/ANA.1062
19. Georges-Courbot MC et al (2006) Poly(I)-poly(C12U) but not ribavirin prevents death in a hamster model of Nipah virus infection. Antimicrob Agents Chemother 50:1768–1772. https://doi.org/10.1128/AAC.50.5.1768-1772.2006
20. Dawes BE et al (2018) Favipiravir (T-705) protects against Nipah virus infection in the hamster model. Sci Rep. https://doi.org/10.1038/S41598-018-25780-3
21. Lo MK et al (2019) Remdesivir (GS-5734) protects African green monkeys from Nipah virus challenge. Sci Transl Med 11:494. https://doi.org/10.1126/SCITRANSLMED.AAU9242
22. Search for: Nipah Virus Infection | Card Results | ClinicalTrials.gov. https://clinicaltrials.gov/search?cond=Nipah%20Virus%20Infection. Accessed: 22 Mar 2025.
23. Prioritizing diseases for research and development in emergency contexts, WHO. https://www.who.int/activities/prioritizing-diseases-for-research-and-development-in-emergency-contexts. Accessed 22 Mar 2025.
24. WHO South-East Asia Regional Strategy for the prevention and control of Nipah virus infection 2023–2030, WHO. https://www.who.int/publications/i/item/9789290210849. Accessed 22 Mar 2025.
25. Tigabu B et al (2014) A BSL-4 high-throughput screen identifies sulfonamide inhibitors of Nipah virus. Assay Drug Dev Technol 12:155. https://doi.org/10.1089/ADT.2013.567
26. García-Murria MJ, Expósito-Domínguez N, Duart G, Mingarro I, Martinez-Gil L (2019) A bimolecular multicellular complementation system for the detection of syncytium formation: a new methodology for the identification of Nipah virus entry inhibitors. Viruses. https://doi.org/10.3390/V11030229
27. Dang HV et al (2019) An antibody against the F glycoprotein inhibits Nipah and Hendra virus infections. Nat Struct Mol Biol 26:980–987. https://doi.org/10.1038/S41594-019-0308-9
28. Zhang Y et al (2017) Influenza research database: an integrated bioinformatics resource for influenza virus research. Nucleic Acids Res 45:D466–D474. https://doi.org/10.1093/NAR/GKW857
29. De Clercq E, Li G (2016) Approved antiviral drugs over the past 50 years. Clin Microbiol Rev 29:695–747. https://doi.org/10.1128/CMR.00102-15
30. Pickett BE et al (2012) Virus pathogen database and analysis resource (ViPR): a comprehensive bioinformatics database and analysis resource for the coronavirus research community. Viruses 4:3209–3226. https://doi.org/10.3390/V4113209
31. Antiviral Library—Enamine. https://enamine.net/compound-libraries/targeted-libraries/antiviral-library. Accessed 09 Dec 2024.
32. Rajput A, Kumar A, Kumar M (2019) Computational identification of inhibitors using QSAR approach against Nipah virus. Front Pharmacol 10:430000. https://doi.org/10.3389/FPHAR.2019.00071/BIBTEX
33. Zoete V, Daina A, Bovigny C, Michielin O (2016) SwissSimilarity: a web tool for low to ultra high throughput ligand-based virtual screening. J Chem Inf Model 56:1399–1404. https://doi.org/10.1021/ACS.JCIM.6B00174
34. Gu Y et al (2024) admetSAR3.0: a comprehensive platform for exploration, prediction and optimization of chemical ADMET properties. Nucleic Acids Res 52:W432–W438. https://doi.org/10.1093/NAR/GKAE298

35. Myung Y, De Sá AGC, Ascher DB (2024) Deep-PK: deep learning for small molecule pharmacokinetic and toxicity prediction. Nucleic Acids Res 52:W469–W475. https://doi.org/10.1093/NAR/GKAE254
36. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 46:3–26. https://doi.org/10.1016/S0169-409X(00)00129-0
37. Veber DF, Johnson SR, Cheng HY, Smith BR, Ward KW, Kopple KD (2002) Molecular properties that influence the oral bioavailability of drug candidates. J Med Chem 45:2615–2623. https://doi.org/10.1021/JM020017N
38. Darvas F, Keseru G, Papp A, Dorman G, Urge L, Krajcsi P (2005) in silico and ex silico ADME approaches for drug discovery. Curr Top Med Chem 2:1287–1304. https://doi.org/10.2174/1568026023392841
39. Holford N, Yim DS (2023) Volume of distribution. Transl Clin Pharmacol 24:74–77. https://doi.org/10.12793/tcp.2016.24.2.74.
40. Bertilsson L, Dahl ML, Dalén P, Al-Shurbaji A (2002) Molecular genetics of CYP2D6: clinical relevance with focus on psychotropic drugs. Br J Clin Pharmacol 53:111. https://doi.org/10.1046/J.0306-5251.2001.01548.X
41. Van Tran TT, Surya Wibowo A, Tayara H, Chong KT (2023) Artificial intelligence in drug toxicity prediction: recent advances, challenges, and future perspectives. J Chem Inf Model 63:2628–2643. https://doi.org/10.1021/ACS.JCIM.3C00200

Publisher's Note