Discovery of RdRp Thumb-1 as a novel broad-spectrum antiviral family of targets and MDL-001 as a potent broad-spectrum inhibitor thereof - Part I: A Bioinformatics and Deep Learning Approach

Virgil Woods[1]*, Tyler Umansky[1]*, Navya Ramesh[1], Sean M. Russell[1], David Garvey[2], Davey Smith[3], Daniel Haders[1]**

[1]Model Medicines, [2]DSG Pharma Consulting LLC, [3]University of California San Diego,

*** These authors contributed equally, **Corresponding author**

ABSTRACT

Positive sense, single-stranded RNA (+ssRNA) viruses consist of 12+ viral families that contain mild pathogens to pandemic-causing viruses like SARS-CoV-2, yet all share a vital and highly conserved RNA-dependent RNA polymerase (RdRp). While RdRp is the target of several viral inhibitors, the active site has several pitfalls when translating in vitro inhibitors to the clinic. The highly polar residues within the active site often necessitate the use of highly polar or charged compounds, especially when designing nucleoside analog inhibitors, posing significant challenges in optimizing drug-likeness and membrane permeability for clinical efficacy. Here, we investigated the broad-spectrum potential of the allosteric Thumb-1 cryptic site of the RdRp, which to date has only been adequately studied in Hepatitis C Virus (HCV). To explore this potential antiviral target, we used a suite of bioinformatics techniques, including homology modeling and multiple sequence alignments, to reveal the conserved landscape of the Thumb-1 site across +ssRNA viruses. We then used ChemPrint, our Mol-GDL (Molecular-Geometric Deep Learning) machine learning model to predict drug inhibition of the Thumb-1 site in RdRp across +ssRNA viruses. Here, we identify MDL-001 as a promising broad-spectrum antiviral candidate with favorable properties that enable oral and once-a-day dosing. We also show how the cryptic nature of the Thumb-1 site masks itself to conventional virtual screening techniques, like docking, where activity prediction is heavily based on solving or predicting an accurate structure of the open pocket. This study demonstrates the utility of this approach in drug discovery for broad-spectrum antivirals that target the Thumb-1 site.

Lay Summary

Background: This paper addresses the challenge of developing broad-spectrum antivirals against +ssRNA viruses, which include pathogens like HCV and SARS-CoV-2. The highly conserved RdRp Thumb-1 site across these viruses is identified as a potential target for broad spectrum antiviral drugs, which would be vital tools in prophylactic and therapeutic responses to future viral threats.

Methodology: Utilizing bioinformatics and deep learning (through our proprietary AI platform called GALILEO), the study maps the conserved landscape of the RdRp Thumb-1 site across a range of human pathogenic +ssRNA viruses and screens for potential inhibitory compounds.

Key Discoveries: MDL-001 emerges as a potential broad-spectrum non-nucleoside antiviral drug candidate and the RdRp Thumb-1 site exists as a family of targets, many of which may be targetable by a single drug.

Impact and Implications: The discovery that the Thumb-1 site of RdRp can be targeted by small molecules for antiviral activity against a range of +ssRNA viruses.

Future Directions: This work paves the way for further development of MDL-001 and research for other broad-spectrum antiviral compounds that target the RdRp Thumb-1 domain.

1. Introduction

The rapid adaptability of viruses presents an ongoing risk of new pandemics and creates challenges for the search for multipurpose antiviral therapies. Pursuing a broad-spectrum antiviral for these viruses presents a significant yet unfulfilled medical need. Historically, antiviral drugs have been developed in response to existing viral threats rather than preemptively. This reactionary approach has constrained the scope of antiviral therapy, limiting it to specific viruses or variants post-outbreak. However, the increasing prevalence and severity of RNA virus outbreaks underscores the necessity for a proactive and versatile antiviral strategy[1]. An orally administered broad spectrum antiviral not only can provide relief for current viruses, but may even enable clinicians to mount effective early responses to future novel threats.

In the search for broad-spectrum antivirals, the RNA-dependent RNA polymerase (RdRp) emerges as a prime target due to its central role in the replication processes of ssRNA viruses[2,3]. During replication, the active site of RdRp is responsible for incorporating free nucleotides into the daughter RNA strand of the replication complex. Broad-spectrum inhibitors of the active site frequently rely on highly polar phosphate and triphosphate groups in the case of nucleoside and nucleotide analogs, respectively. Polar functional groups reduce membrane permeability and severely limit oral bioavailability. This is just one of several limitations that constrain active site inhibitors, as demonstrated by current approved therapies. Remdesivir, initially developed for filoviruses and paramyxoviruses, has been repurposed for SARS-CoV-2, however its clinical utility is hampered by the need for intravenous administration and a high incidence of adverse events in the clinic[3,4]. Also, Molnupiravir, which promised to be an oral antiviral for COVID-19, carries known mutagenicity risk in humans and has poor to no clinical efficacy in post-approval studies[5]. These examples underscore the unmet clinical need for broad-spectrum therapies that are safe, effective, and convenient for patient use. The chemical properties of nucleoside analog complicate the optimization of potency alongside Absorption, Distribution, Metabolism, Excretion, and Toxicity (ADMET) parameters. To circumvent these limitations, the focus may need to shift towards exploring allosteric inhibitors of RdRp. Significant advancements have been made in identifying modulators of the polymerase of Hepatitis C Virus (HCV), revealing a number of allosteric sites. Among these sites is the cryptic Thumb-1 site, whose inhibitors demonstrate superior cross-genotypic coverage for HCV, hinting at their capability for an expansive antiviral range[1,6].

Computer-aided drug design promises to accelerate the discovery of expanded uses of existing drug compounds for new indications. However, identifying broad-spectrum activity of Thumb-1 inhibitors has

proven challenging, exemplified by the *in silico* efforts to fight the COVID-19 pandemic. Numerous groups embarked on large-scale drug repurposing studies using *in silico* techniques, including those by Piplani et al.[7] and Dutta et al.[8], to expedite the discovery of antivirals for SARS-CoV-2. These studies employed docking and molecular dynamics (MD) simulations to predict potential antiviral compounds against the RdRp of SARS-CoV-2. Interestingly, HCV Thumb-1 inhibitor Beclabuvir was nominated in multiple studies as a likely inhibitor of SARS-CoV-2, with one simulation calculating its energy of binding to be consistent with nanomolar activity ($\Delta G_{AutoDock}$ scores of -10 kcal/mol[7] and -9.95 kcal/mol[8]). Despite these computational analyses suggesting significant therapeutic potential, *in vitro* screenings, notably by Janssen[9], failed to corroborate these findings. This discrepancy underlines a critical flaw in relying solely on physics-based *in silico* methods that do not learn from previous *in vitro* results.

It is becoming increasingly clear that previously unknown biology needs to be uncovered to make progress in developing broad-spectrum antiviral therapeutics. However, modeling the structures of cryptic allosteric sites such as Thumb-1 is inherently difficult. Substantial benefits can be gained through ligand-based machine learning (ML) methodologies that transcend traditional physics-based simulations by learning in the absence of a protein structure. These ML techniques can learn from *in vitro* data alone to make predictions on unknown bioactivities of compounds. Given these challenges, our AI drug discovery platform, GALILEO, and its zero-shot capable Molecular-Geometric Deep Learning (Mol-GDL) model[10], ChemPrint, are presented as solutions to discover new biology along with identifying novel inhibitors. Bioinformatics tools within GALILEO, including multiple sequence alignments, homology modeling, and implicit data collection paired with ChemPrint enable the discovery of broad-spectrum antivirals targeting the Thumb-1 therapeutic site that bypass the pitfalls of current RdRp-targeting antivirals. This approach could revolutionize our preparedness for future outbreaks by not only expediting the discovery of effective treatments but also enabling preventative measures against +ssRNA viruses.

2. Results

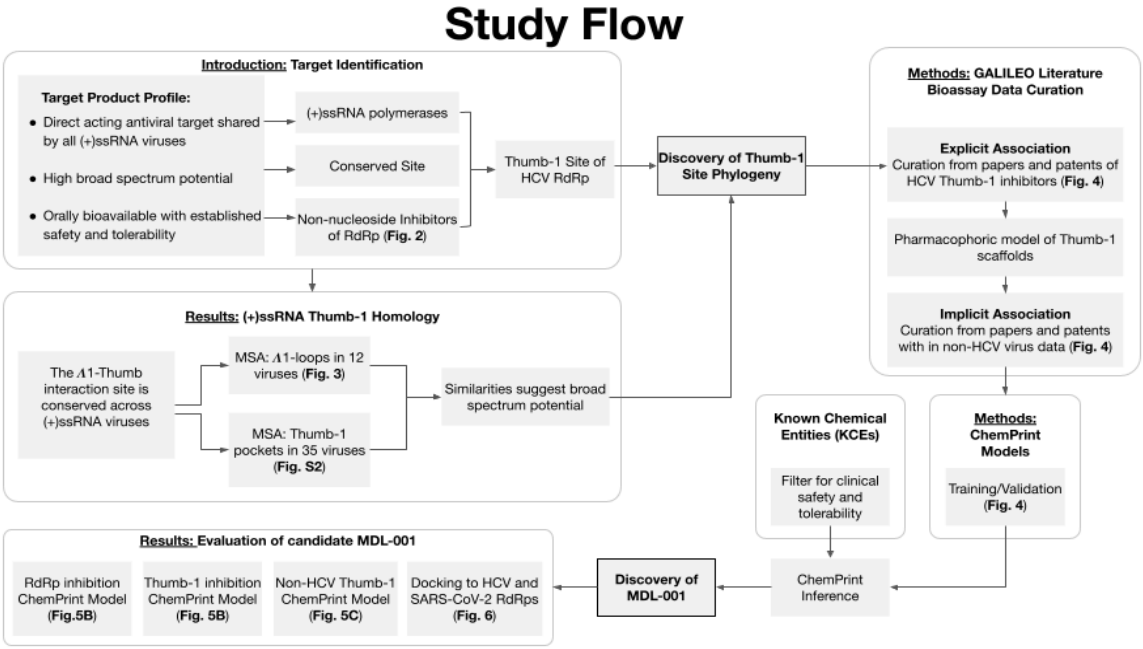

Figure 1: Flow chart of the steps taken in this study to discover and evaluate candidate drug MDL-001 for direct acting, broad spectrum inhibition of positive strand single-stranded RNA +ssRNA viruses. We achieve this via inhibition of viral RNA-dependent RNA polymerase (RdRp) via the allosteric Thumb-1 cryptic site, previously identified and probed only for polymerase Non-Structural Protein 5B (NS5B) of the Hepatitis C Virus (HCV).

2.1 The Thumb-1 Site Mechanism

Previous studies on RdRp inhibitors, particularly in the context of HCV, have laid the groundwork for understanding the feasibility of targeting allosteric sites[11]. As a case study, the HCV RdRp, known as NS5B, features a conserved hand-like structure with three domains—palm, fingers, and thumb—and harbors five identified allosteric sites (Thumb-1, Thumb-2, Palm-1, Palm-2, and Palm-3) for drug discovery efforts (**Figure 2A**)[12]. Multiple antivirals have been approved that target NS5B, including Beclabuvir (Thumb-1), Dasabuvir (Palm-1), and Sofosbuvir (active site)[13,14]. In the HCV polymerase NS5B, the Thumb-1 site is cryptic, meaning it is closed without a bound ligand. It sits at the interface between a loop from the finger domain, known as the Λ1-loop, and a cleft found on the upper thumb domain (**Figure 2B**)[15]. The tip of the Λ1-loop adopts an ***α***-helix conformation in the apoprotein and occupies the upper thumb domain, linking the N-terminal finger and C-terminal thumb domains. Within the HCV Λ1-loop, we observed two key leucine residues that extend deep into the Thumb-1 pocket that overlap well with the structures of bound Thumb-1 inhibitors (**Figure 2B**). This interaction bridges distal domains of the right hand tertiary structure and is required for the RdRp to enclose around replicating RNA strands during initiation of RNA synthesis[16]. Mechanistic studies show that Thumb-1 site inhibitors compete with the Λ1-loop for occupancy of the Thumb-1 pocket, displacing the amino acids in the loop's ***α***-helix[17]. These studies go on to show that the pocket becomes unavailable for inhibition once elongation has started. Amino acids from the Λ1-loop, particularly those in the ***α***-helix, are unresolved across all co-complexed HCV Thumb-1 holoprotein structures from the Protein Data Bank[18,19].

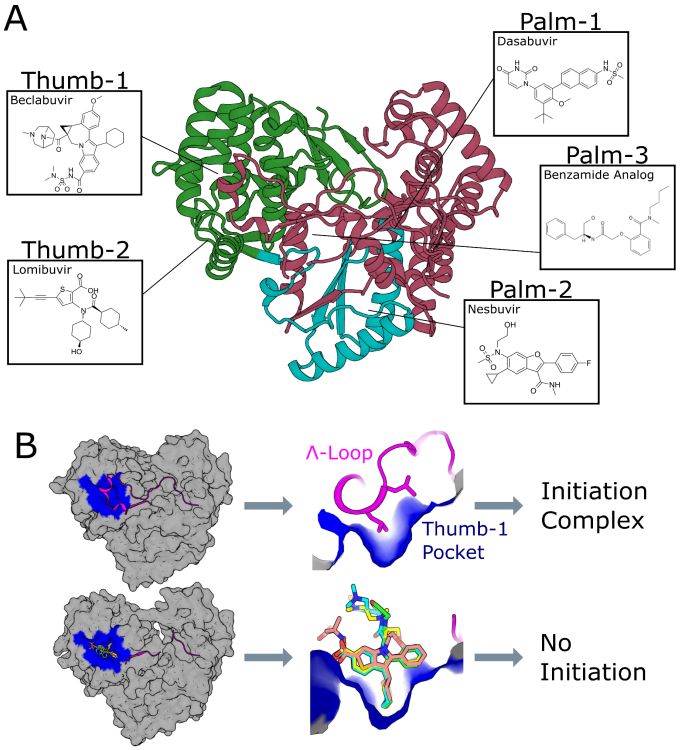

Figure 2 (A) View of HCV RdRp with sites highlighted and one example of an inhibitor for each site (PDB ID: 1C2P)[20]. **(B)** View of the Λ1-thumb interaction and mechanism of thumb-1 inhibition. **Top Left:** HCV RdRp apoprotein structure with a gray surface representation acquired from 1C2P. **Top Right:** A Thumb-1 pocket view of the apoprotein ***α***-helix and its two endogenous leucine side chains occupying the site. **Bottom Left:** Overlay of five Thumb-1 inhibitors with structures available from the PDB (accession codes 2BRK, 2BRL, 2DXS, 2XWY, 2WCX)[21], based on superposition of the holoprotein secondary structure 2XWY, with a gray surface representation. **Bottom Right:** A Thumb-1 pocket view of the holoprotein and its superimposed Thumb-1 inhibitors.

2.2 Conservation of Thumb-1 pocket and Λ1-loop in +ssRNA viruses

Seeking to identify broad spectrum antiviral targets, we investigated the conservation and functional significance of the interaction between the Thumb-1 pocket and the Λ1-loop across +ssRNA viruses. Within the superimposed PDB structures of 10 different HCV NS5B Thumb-1 bound inhibitors, we identified key amino acids (i.e., ^{37}V, ^{392}LA^{393}, ^{395}AA^{396}, ^{399}T, ^{424}IL^{425}, ^{428}HF^{429}, ^{492}LGVP^{495}, ^{500}W, and ^{503}R) interacting between the HCV Thumb-1 pocket and the inhibitors, defined as being located within 5Å of inhibitors) (**Supplemental Figure S1A-D**). Because Thumb-1 site inhibitors act to displace the endogenous interaction of the Λ1-loop with the Thumb-1 site of RdRp, we analyzed how well-conserved the Thumb-1 pocket interaction site is across all +ssRNA viral families depicted in (**Figure 3A**). We performed a multiple sequence alignment (MSA) of the Thumb-1 pocket of 35 +ssRNA viruses (**Figure 3B**). This MSA revealed a remarkable degree of conservation among the Thumb-1 pocket sequences, particularly in residues identified to be within 5Å of the Λ1-loop in the apoprotein of HCV (highlighted green).

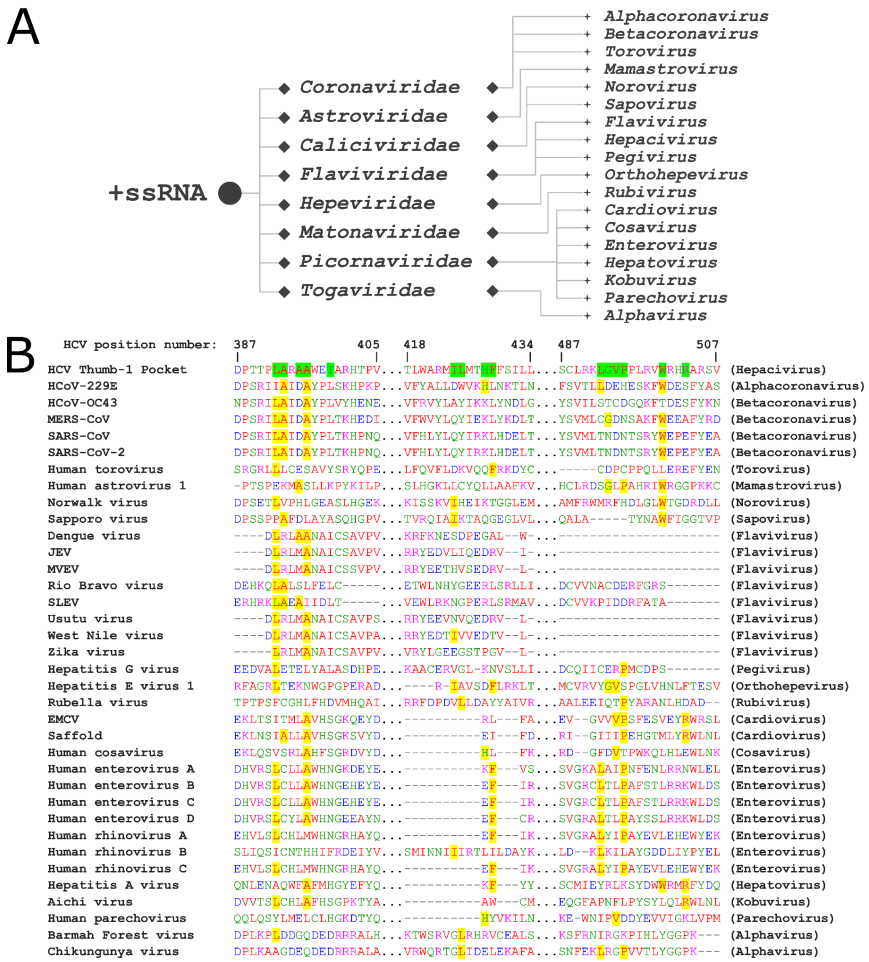

Figure 3: (A) Family tree of 8 human pathogenic +ssRNA families and 18 genera within those families. (**B**) Clustal Omega[22] multiple sequence alignment (MSA) of the Thumb-1 pocket of HCV and representative viruses from each +ssRNA genus (see methods). HCV residues that are found within 5Å of the bound Thumb-1 inhibitors are highlighted green. Identity matches of these residues in other viruses are highlighted yellow.

We next examined available polymerase structures in three families of pathogenic +ssRNA viruses. In addition to published structures, we constructed homology models of polymerases of five viruses that lacked PDB structures using templates obtained from members of the same families. Both the structure and sequence of the Λ1-loop interaction with the Thumb-1 pocket show remarkable conservation within families *Coronaviridae*, *Flaviviridae*, and *Picornaviridae*. As seen in **Figure 4A**, within the *Picornaviridae* family (Poliovirus, Coxsackievirus) the Λ1-loops completely lack structured ***α***-helices in the bound state and have two phenylalanine residues that fit into similar pocket geometries as well as a shared valine residue that appears to also engage the pocket. The Λ1-loops belonging to the *Coronaviridae* (SARS-CoV-1, SARS-CoV-2, MERS-CoV, hCoV-229E, hCoV-OC43) comprise a single ***α***-helix with a significant degree of conservation for residues that face the Thumb-1 site. In particular, an asparagine and two phenylalanine residues are identical across the coronavirus structures. Among non-HCV members of the *Flaviviridae* family (Usutu, Powassan, West Nile, and JEV viruses), conserved tryptophan and proline residues can be found in their Λ1-loop. Next, we performed an MSA on these RdRp Λ1-loop sequences, which revealed a high degree of conservation among the sequences (**Figure 4B**). The preservation of these residues underscores their functional significance in RNA template stabilization and replication fidelity. As a result, the Λ1-loop is suggested to be a well conserved structural motif spanning +ssRNA viruses.

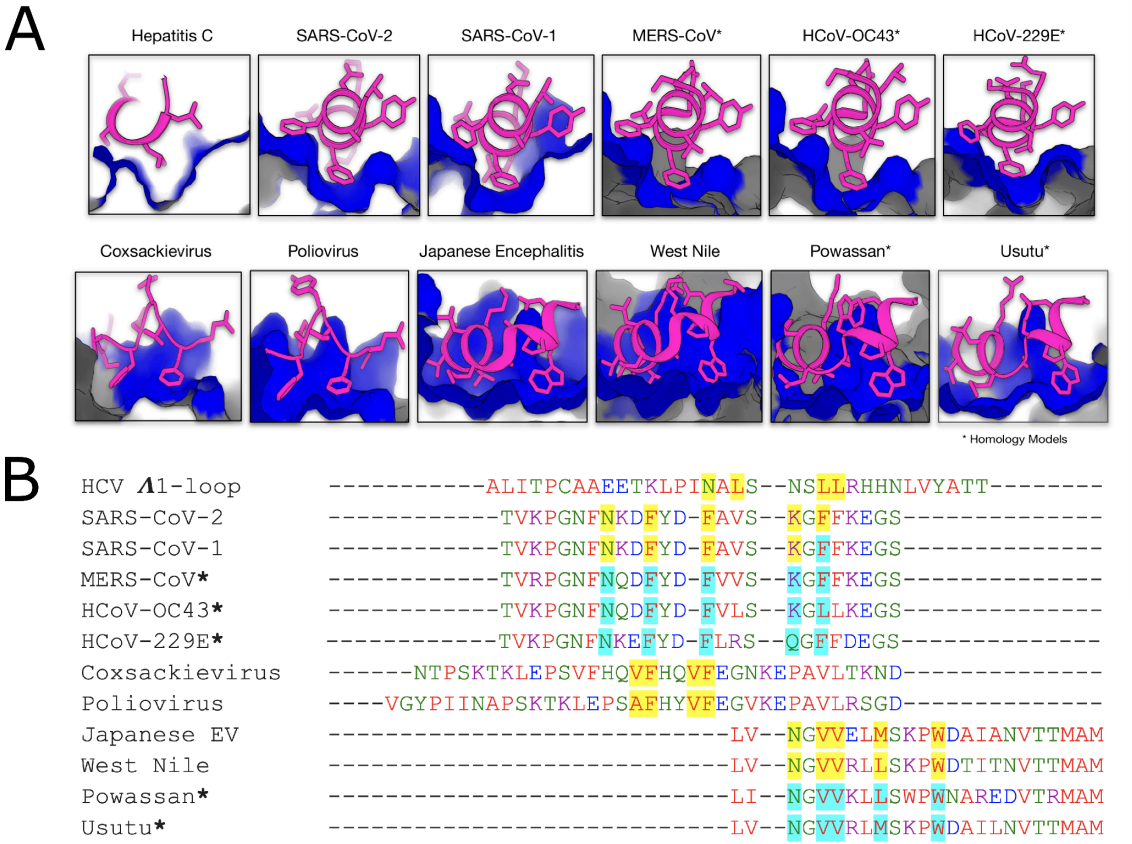

Figure 4: (A) Pymol representations of thumb-Λ1-loop interactions in 12 +ssRNA viruses. *PDB-IDs:* HCV (1C2P)[20], SARS-CoV-2 (6M71)[23], SARS-CoV-1 (6NUS)[24], Coxsackievirus (3CDU)[25], Poliovirus (2IJF)[26], Japanese Encephalitis (4MTP)[27], West Nile (2HFZ)[28]. Structures denoted with * indicate SWISS-MODEL homology models (see methods). **(B)** MSA of Λ1-loops for the same 12 viruses. Residues within 5Å of the Thumb-1 pocket surface in PDB structures (yellow) and homology models (cyan) are highlighted. Note: cyan highlight in SARS-CoV-1 corresponds to an unmodeled phenylalanine in structure 6NUS, but is proximal to the Thumb-1 site in a homology model published on the PDB Model Archive (ma-c1tkv)[29].

2.3 Discovery of MDL-001

To discover new chemistry that inhibits the Thumb-1 pocket across a wide range of viral families, we leveraged our AI drug discovery platform, GALILEO, and our Mol-GDL model, ChemPrint, to curate and train on data grounded in first-principles chemistry (see methods). Published structures of the known HCV RdRp inhibitors were used to determine a pharmacophoric model of their shared chemical scaffolds that appear to facilitate Thumb-1 specific binding. In 25 aligned PDB structures[19], the Thumb-1 pocket is occupied by ten unique ligands, each containing one of three core scaffolds (2-phenylindole, 2-phenylbenzimidazole, 5-Phenyl-4H-thieno[3,2-b]pyrrole).

These scaffolds were then leveraged to classify literature inhibitors of HCV and other viruses into three levels of curated data (**Figure 5A**). *Level 1* data are sourced from papers claiming explicitly to be Thumb-1 site inhibition. *Level 2* data came from papers containing RdRp target assay data without explicit identification of the binding pocket within the protein. *Level 3* compounds are from sources containing assay data from screens performed with infection models of non-HCV viruses. Compounds from *Level 2* and *Level 3* containing any of the core Thumb-1 scaffolds were passed through the pharmacophoric filter and presumed to act at the Thumb-1 site and their bioassay results were labeled as 'implicit' Thumb-1 data. Explicit Thumb-1 data from *Level 1* and implicit Thumb-1 data from *Level 2* and *Level 3* were used to create our broad-spectrum training dataset. To identify the most highly potent compounds, we defined compounds with a bioactivity measurement below 1 μM as active and the rest as inactive. This was then used to train our Mol-GDL model, ChemPrint.

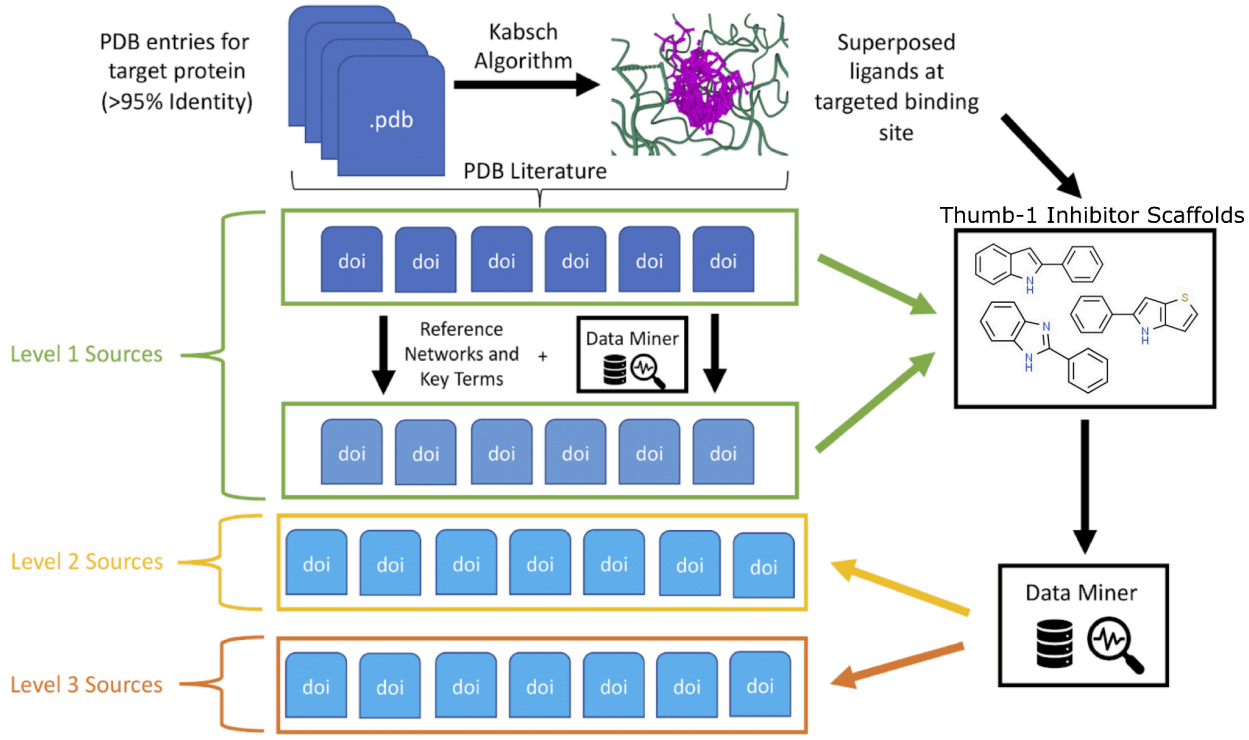

Figure 5: Process overview for bioassay data collection and curation.

ChemPrint exploits the natural connective structure of molecules and directly learns adaptive molecular embeddings for compounds through its graph convolutional network (GCN). Our model leverages the structural information of ligands, allowing for a more complete prediction of their biological activities. ChemPrint was trained and validated using our model benchmarking methodology that leverages t-distributed Stochastic Neighbor Embedding (t-SNE) to split the data (see methods)[10]. The Receiver Operating Characteristic (ROC) curve statistics in **Supplemental Figure S2** illustrate the performance of

ChemPrint in classifying compounds and indicate that ChemPrint is a well-tuned model capable of making accurate predictions.

The ChemPrint model, fine-tuned on broad-spectrum Thumb-1 data, was deployed to make inference on an extensive dataset of ~17,000 chemical compounds we classify as known chemical entities (KCEs). All KCEs in our library have moved beyond Phase I clinical trials, offering an inference library of compounds with established safety and tolerability data that can be used to accelerate the development of a broad spectrum antiviral candidate. For each KCE, ChemPrint calculates a prediction score that represents the probability that a compound, when tested in the same manner as the training set data, will demonstrate potent activity. To evaluate potential Thumb-1 inhibitors, inference compounds were evaluated by individually trained replicate (n=10) ChemPrint models. The models identified several promising hits among these compounds. Top-scoring compounds were then filtered with a freedom to operate search for novelty to the disease indication, and any compounds that had ever been tested in an antiviral assay before were removed from consideration (see methods). One compound (MDL-001) stood out due to its established clinical safety and tolerability combined with its predicted broad-spectrum Thumb-1 inhibition activity. MDL-001 was evaluated to have a probability of potent broad-spectrum Thumb-1 activity with a mean and 95% confidence interval of 0.76 ± 0.02. This indicated a high likelihood of MDL-001 having activity against Thumb-1 sites broadly.

2.4 Using ChemPrint Model Variations to Evaluate MDL-001

To further clarify the mechanism of action (MoA) of MDL-001 on the Thumb-1 site, we visualized the chemical space of all collected data across *Level 1-3* for comparison. Leveraging our labeled compound dataset (see **Supplementary Data**), we used t-SNE to project molecular embeddings to 2-dimensions (see methods)[30]. As shown in **Figure 6A**, the chemical space representation of inhibitors places MDL-001 on top of a cluster of explicit Thumb-1 drugs, cleanly separated from compounds belonging to other target site classes. This placement further supported MDL-001's potential as a Thumb-1 site inhibitor. Additionally, the t-SNE projection showed that MDL-001 fit relatively close to the grouping of implicit Thumb-1 inhibitors that inhibit non-HCV viruses, shown as black circle outlines in the figure, suggesting that MDL-001 may have similar activity outside of HCV.

To provide additional supporting evidence into the mechanism of action of MDL-001, we created several label-specific training datasets (**DF1-5**) (**Figure 6B**)comprising subsets of the data collected from *Levels 1-3*. **DF1** contains all data from *Levels 1-3*, **DF2** is restricted to ligands identified by our pharmacophoric filter, and **DF3** is restricted to all ligands that the pharmacophoric filter did not flag (nucleoside analogs and non-Thumb-1 allosteric inhibitors). **DF4** comprises data only from *Level 1*. **DF5** comprises data only from *Levels 2-3* with ligands identified by our pharmacophoric filter as targeting non-HCV Thumb-1 sites. **DF2** is identical to the dataset used to discover MDL-001 in the previous section. These five training datasets were each used to train ChemPrint models, and the ROC curves are reported in **Supplemental Figure S2.** These curves illustrate the performance of each ChemPrint model using the same benchmarking methodology as before.

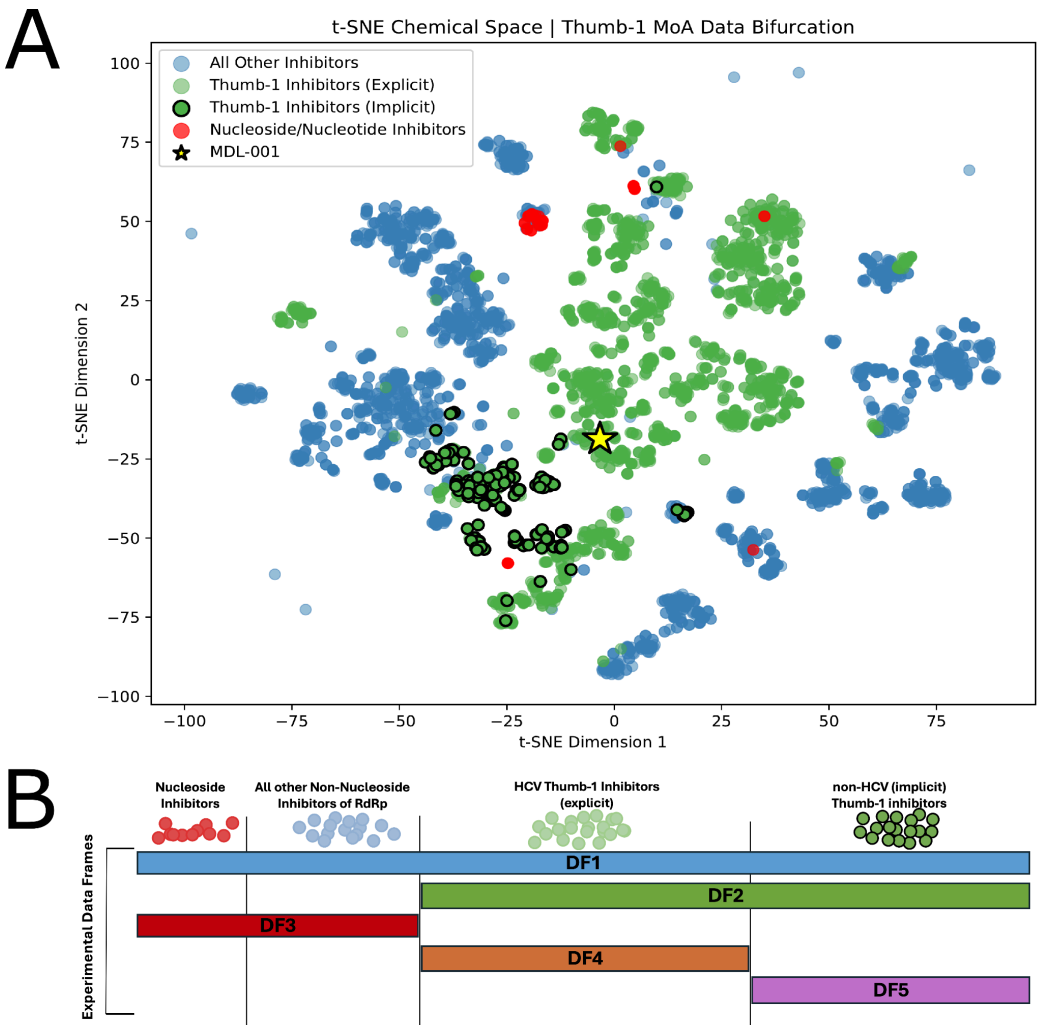

Figure 6: (A) TSNE of allosteric compounds that are known to target RdRp, showing the different classifications of nucleoside, non-nucleoside, and thumb-1 inhibitors (explicit and implicit data). MDL-001 is depicted as a yellow star. **(B)** Schematic of the compositions of Datasets: (**DF1**) All Data (**DF2**) Thumb-1 (Explicit and Implicit) (**DF3**) All non-Thumb-1 Inhibitors (**DF4**) Thumb-1 (Explicit) (**DF5**) Thumb-1 (Implicit)

By comparing the scores for MDL-001 in the five different training datasets, we found that MDL-001 was predicted to inhibit the Thumb-1 site (**Figure 7A**), with **DF2** (Thumb-1 explicit and implicit) and **DF3** (non-Thumb-1) assigning it average(n=10) probability scores of 0.76 to 0.15, showing reliance on Thumb-1 training data to predict MDL-001 as active. We further probed MDL-001's activity via two other models (**Figure 7B**): “explicit” HCV-only Thumb-1 data (**DF4**) and non-HCV “implicit” Thumb-1 data (**DF5**). Alongside MDL-001 we also screened Beclabuvir because it had been nominated in multiple studies as likely having antiviral activity beyond HCV[7,8]. With an active/inactive cutoff at 1.0 μM, the average HCV Thumb-1 antiviral activity prediction scores were 0.69 and 0.98 for MDL-001 and Beclabuvir, respectively, in the **DF4** model. In the **DF5** model with a more lenient active/inactive cutoff of 20 μM, we found that MDL-001 maintained a similar average prediction score of 0.69, indicating its potential as an inhibitor beyond anti-HCV activity, while Beclabuvir's average prediction score drastically reduced to 0.04, consistent with experimental observations that Beclabuvir is inactive against SARS-CoV-2[9].

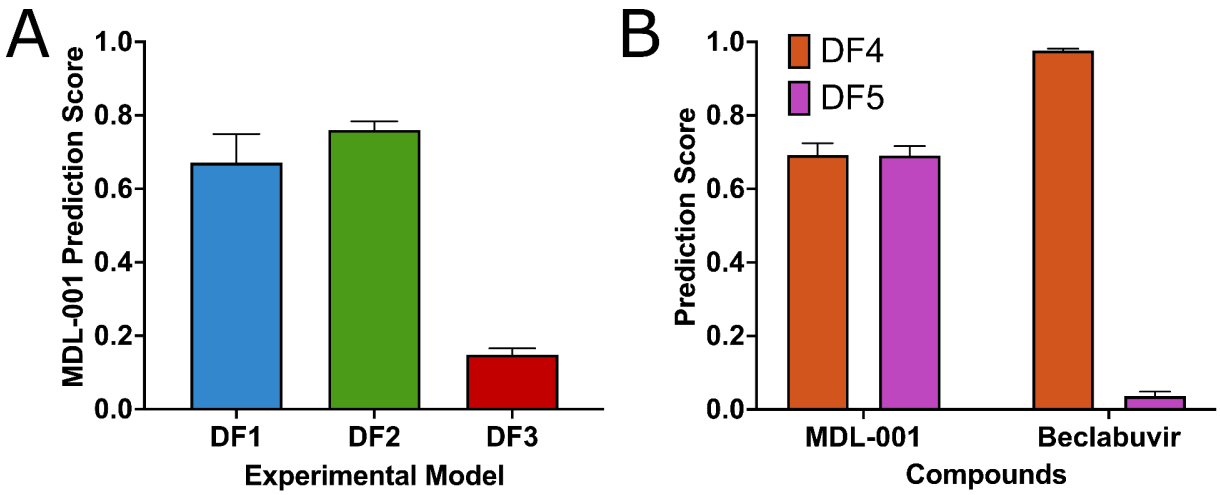

Figure 7: (A) Prediction scores for MDL-001 across the three experimental models **DF1** (All training data), **DF2** (Thumb-1 only data), and **DF3** (only non-Thumb-1 data). **(B)** Bar graph of the prediction scores of MDL-001 and Beclabuvir across models **DF4** (HCV-only Thumb-1 data) and **DF5** (non-HCV implicit Thumb-1 data).

2.5 Retroactive Docking Study of MDL-001 and Beclabuvir

In the *in vitro* screen by Janssen, Beclabuvir had no significant antiviral effect up to concentrations of 20 μM against SARS-CoV-2[9] Previous docking and MD studies suggest Beclabuvir, which targets the HCV Thumb-1 pocket, could also have activity against SARS-CoV-2[31,32]. We aimed to understand these faulty *in silico* predictions by performing a docking study of our own. We evaluated the broad spectrum antiviral potential of Beclabuvir and MDL-001 by conducting a comparative docking study of the Thumb-1 sites in SARS-CoV-2 and HCV (**Figure 8**).

While ChemPrint scored MDL-001 with a high probability of successfully inhibiting the Thumb-1 site of RdRp, Autodock4 indicates that MDL-001 had higher (worse) docking scores to the Thumb-1 site in both HCV and SARS-CoV-2 than did Beclabuvir (**Figure 8A-D**). Specifically, MDL-001 showed consistent docking scores of -7.82 kcal/mol and -7.11 kcal/mol for HCV and SARS-CoV-2 respectively. MDL-001 docked to the HCV NS5B Thumb-1 pocket (**Figure 8A**) displayed key residue engagements such as Arg503, which is known to form hydrogen bonds with established HCV Thumb-1 inhibitors[33]. Comparatively, Beclabuvir had docking scores of -10.18 kcal/mol and -8.29 kcal/mol, which are in-line with previous work that predicted Beclabuvir's potential to bind and inhibit SARS-CoV-2 RdRp in docking and molecular dynamics studies[7,8].

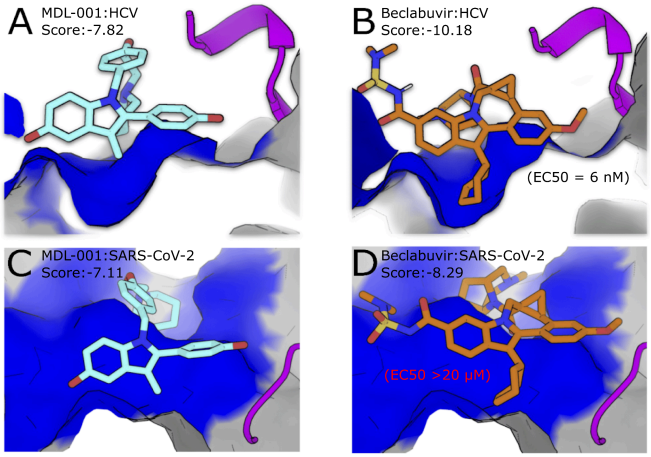

Figure 8: (A-B) Poses of MDL-001 (**A**) and Beclabuvir (**B**) docked into the Thumb-1 pocket of HCV RdRp acquired from PDB ID: 4DRU[34]. **(C-D)** MDL-001 (**C**) and Beclabuvir (**D**) docked into the Thumb-1 pocket of SARS-CoV-2 RdRp acquired from PDB ID: 6M71[23].

In these analyses, we see that the conformation of MDL-001 docked to the Thumb-1 pocket of SARS-CoV-2 assumes a more relaxed state than that of Beclabuvir. In Beclabuvir's top ranked pose in SARS-CoV-2, the cyclohexane at position 3 of its indole scaffold appears to be in a more strained and forced conformation than the top pose in HCV. MDL-001's docking scores, interpreted as proxies for the free energy of protein-ligand binding in the Gibbs free energy equation, equate to K_D values of 1.83 μM and 6.06 μM at room temperature (298K) for HCV and SARS-CoV-2, respectively. In comparison, ChemPrint predicted MDL-001 to have activities of < 1 μM for HCV and < 20 μM for non-HCV viruses. By contrast, the docking scores of Beclabuvir with SARS-CoV-2 RdRp are consistent with previously published docking and MD studies that predicted nanomolar activities for Beclabuvir in SARS-CoV-2 RdRp[7,8]. Again, *in vitro* tests determined that Beclabuvir was found to be inactive (EC50 > 20.0 μM)[9] indicating this as a *in silico* false positive. It is possible that the open structure of the SARS-CoV-2 Thumb-1 site generated by deletion of the apo structure's Λ1-loop is not suitable for accurate *in silico* docking of Thumb-1 inhibitors. In contrast to the docking results, ChemPrint correctly predicted Beclabuvir to be inactive (> 20.0 μM) against non-HCV. This suggests that docking using modified apo structures alone is not sufficient to accurately predict Thumb-1 inhibition. This underscores the potential drawback of relying on physics-based docking of rigid, or even moderately flexible, protein structures to capture the behavior of ligand binding in cryptic sites like Thumb-1.

3. Discussion

The impact of ssRNA virus epidemics and pandemics has been profound historically and in recent years. Diseases like Influenza, RSV, and COVID-19, caused by ssRNA viruses, have had devastating global health, social, and economic impacts[1]. These outbreaks underscore the critical need for versatile treatments capable of addressing a wide array of ssRNA viruses, underscoring the urgency and importance of our research. RdRp is a validated and highly conserved protein target for direct acting antivirals. However, while traditional drug discovery methods have produced Remdesivir and other nucleoside analogs targeting the active site, these drugs suffer sub-optimal membrane permeability and come with elevated safety and tolerability concerns. RdRp allosteric sites, like Thumb-1, and their inhibitors offer a compromise between site conservation and preferable pharmacological properties like oral bioavailability. Previously, Thumb-1 inhibitors have largely only been capable of inhibiting HCV, out of all human pathogenic viruses[1].

Comparative studies of the Thumb-1 pocket across various viral families revealed structural similarities previously uncharacterized, supported by MSA and homology modeling. We used our AI drug discovery platform, GALILEO, and its Mol-GDL model, ChemPrint[10], to identify a candidate therapeutic for this

purpose. We discovered MDL-001 with a ChemPrint model trained with comprehensive explicit and implicit Thumb-1 data collected from the literature (**Figure 5B**). The selection of MDL-001 was primarily based on its established clinical safety and ADME profile, underscoring the advantage of integrating AI-driven platforms in the early stages of drug discovery for infectious diseases.

Upon further investigation, prediction scores determined by ChemPrint, when trained on Thumb-1 and non-Thumb-1 bioassay results, indicated MDL-001's MoA is likely targeted at the Thumb-1 site, differentiating it from active site inhibitors and allosteric inhibitors of other sites (**Figures 7A**). The model consistently predicted MDL-001's activity, even after excluding explicit HCV Thumb-1 data, suggesting potential broad-spectrum efficacy (**Figure 7B**). These findings challenged conventional *in silico* screening methods used during the early pandemic, which failed to identify MDL-001 as a candidate despite its known clinical status and incorrectly identified Beclabuvir. Notably, MDL-001 and Beclabuvir were docked to the HCV Thumb-1 site and a contrived SARS-CoV-2 Thumb-1 site, illustrating limitations in traditional docking approaches. While Beclabuvir showed favorable docking scores, it lacked inhibitory activity against SARS-CoV-2 at 20 μM.

Conversely, MDL-001 demonstrated the ability to inhibit SARS-CoV-2 Variants of Concern (VoCs) with nanomolar efficacy in a cell based assay of disease[35]. MDL-001 has also demonstrated activity against an *in vivo* SARS-CoV-2 infection model. Administering MDL-001 orally significantly lowered the amount of SARS-CoV-2 virus in the lungs of a mouse model specifically adapted to study SARS-CoV-2 infection[36]. In another animal model, MDL-001 has demonstrated potential best-in-class lung penetration (AUC Lung/AUC plasma ratio) that is 1-2 orders of magnitude greater than nirmatrelvir/ritonavir (Paxlovid™)[37]. IND-enabling testing of MDL-001's is ongoing, and further *in silico* work is enabling predictions of MDL-001 activity in additional viruses. These promising results highlight the novel insights provided by our GALILEO platform and its Mol-GDL ChemPrint model. In particular, MDL-001's proposed Thumb-1 MoA needs to be confirmed in future studies.

The SARS-CoV-2 structure prepared for our docking study relied on an assumption that the open pocket conformation is close enough to the available apo structure. While we did not perform docking on other viruses, it is possible that some contain closer-matching apo and ligand-bound Thumb-1 site conformations. In addition, the non-HCV ChemPrint model relied on implicit Thumb-1 data taken from many whole virus screening sources, resulting in a relatively small dataset size compared to the other model variants in this study. Extensive testing of the known Thumb-1 inhibitors in non-HCV viruses would need to be performed to create a higher-performance ChemPrint model of broad spectrum Thumb-1 inhibition. The non-HCV ChemPrint model relied on implicit Thumb-1 data taken from many whole virus screening sources, resulting in a relatively small dataset size compared to the other model variants in this study.

4. Conclusion

Through the discovery of MDL-001 and in proposing its Thumb-1 MoA, this study has unveiled the Thumb-1 site of RdRp as a novel biological target for broad-spectrum drug discovery. This advancement was propelled by the integration of AI, notably through our GALILEO platform and its ChemPrint prediction model. The discovery of MDL-001, with its favorable pharmacokinetics and ADMET properties and potential as an orally administered therapy, offers hope as an orally administered therapy that effectively treats a broad swath of ssRNA viral diseases.

5. Methods

5.1 Multiple sequence alignments

Sequence Retrieval and Compilation

35 RdRp viral sequences were sourced from publicly available databases, encompassing various viral families. These sequences were selected to ensure representation across each viral genus identified in **Figure 3A**. To maintain consistency, sequences were retrieved based on their availability and completeness.

Multiple Sequence Alignment (MSA) - Thumb-1 Pockets

The first MSA study aligned the Thumb-1 pockets of the selected RdRp viral sequences using the Clustal Omega algorithm via the online tool from EMBL-EBI[22] (**Figure 3B**), at least one virus from each genus in **Figure 3A**. The Thumb-1 region was identified based on previously annotated structures and functional domains[12].

Multiple Sequence Alignment (MSA) - Λ1-loop Residues

In the second MSA study (**Figure 4B**), attention was directed towards the lambda loop residues of the 12 sequences of viruses in **Figure 4A** sequences, utilizing the Clustal Omega algorithm. The Λ1-loop, was identified in viral sequences through structural analyses and previous identifications of the Thumb domain of non-HCV RdRps in the literature[12].

Validation and Quality Control

Stringent validation and quality control measures were implemented to ensure the reliability of the Clustal Omega alignments. Sequences with ambiguous annotations or incomplete regions were carefully curated or excluded from the analysis. Additionally, the Clustal Omega-generated alignments underwent manual inspection to verify the correct alignment of functionally relevant residues.

5.2 Homology Modeling with SWISS-MODEL

SWISS-MODEL, a widely used homology modeling tool[38], was employed in this study to generate three-dimensional structures for the interaction site of the lambda loop and Thumb-1 pockets of select viruses. Specifically, structures for MERS-CoV, HCoV-OC43, HCoV-229E, Powassan, and Usutu viruses were modeled to elucidate the molecular details of these critical regions.

Sequence Retrieval

Sequences for the target viruses (MERS-CoV, HCoV-OC43, HCoV-229E, Powassan, and Usutu) were obtained from curated UniProtKB entries. These sequences were utilized in the Multiple Sequence Alignment (MSA) studies, ensuring consistency across analyses.

Homology Modeling Parameters

SWISS-MODEL was configured to perform homology modeling based on the crystal structures of homologous proteins with known three-dimensional structures. The aim was to model the interaction site of the selected viruses' lambda loop and Thumb-1 pockets. The choice of template structures was guided by their relevance to the viral RdRp domains and the availability of high-quality crystal structures.

Model Generation and Validation

Five models corresponding to MERS-CoV, HCoV-OC43, HCoV-229E, Powassan, and Usutu viruses were generated using SWISS-MODEL. These models were subsequently validated for structural integrity and

accuracy, focusing on the interaction site near the finger loop and Thumb domains. Metrics such as Ramachandran plots and stereochemical quality assessments were used to ensure the reliability of the generated models.

5.3 GALILEO Data Collection and Curation

Leveraging our data collection pipeline within GALILEO, we created the training dataset with a highly specialized and well-defined Thumb-1 mechanism of action. To catalog the structures of ligands that bind to the Thumb-1 cryptic site, we mined structural information from the available published literature in the Protein Data Bank (PDB). We used a modified version of the Kabsch algorithm[39] to superpose every NS5B structure found in the PDB onto a single protein structure and extracted only the organic ligands from each (**Figure 5**). The algorithm first calculated the center of mass, the ligand's principal axes, and the corresponding binding site on the protein structure. Then, it aligns the principal axes of the ligand to those of the binding site and rotates the ligand accordingly to minimize the root-mean-square deviation (RMSD) between the ligand and the binding site. The resulting superposed model allowed us to locate the Thumb-1 site, identify the ligands bound, and compare their experimentally determined binding poses.

We created our Thumb-1 training set by first curating existing small-molecule bioassay data from all relevant virus families, such as, but not limited to, *Coronaviridae*, *Picornaviridae*, *Flaviviridae*, and *Togaviridae*. More specifically, data was extracted from RdRp-specific and whole virus assays (**DOI's in Supplementary Data**). Protein inhibition and viral reduction assays are just as informative for our machine learning model as target-site-specific data if the protein pocket can be surmised from the structural features of the compound tested. In cross-referencing these data against known Thumb-1 chemistries, many were included in the training set due to their high likelihood of targeting the Thumb-1 site of RdRp. The dataset was assembled using our GALILEO pipeline, which focused on collecting compounds with potential interactions at the Thumb-1 site of RdRp from various virus families. The data curation was crucial for ensuring the quality and relevance of the training data for the ChemPrint model.

From the identified primary source literature, GALILEO extracted protein-specific QSAR data that was validated by our experts and filtered for exclusion factors such as Pan Assay Interference Compounds (PAINS), molecular weight, and duplicate chemistry. Protein interaction measurements were optimized uniquely for each ligand using proprietary techniques. Results of appropriate bioactivity assays split compounds into active and inactive classes based on a selected cutoff defined according to the needs of the experiments. The cutoff was 1.0 μM for **DF1**, **DF2**, **DF3**, and **DF4**, while **DF5** used a cutoff of 20.0 μM.

5.4 ChemPrint

Train-Test Split

To split our data into training and test sets, we utilized t-distributed Stochastic Neighbor Embedding (t-SNE) to project ECFP-4 encoded molecule data into 2-dimensions. This method was performed as previously described in Umansky et al.[10].

ChemPrint Architecture

ChemPrint leverages mol-GDL to learn the adaptive embeddings derived from its training data to make predictions when given novel inference sets. The input data takes the form of a geometric graph, a molecular representation that encapsulates the structural information of each datapoint, to which one-hot encoded atoms types are passed in as features. Our model architecture encompasses an end-to-end Graph Convolutional Network (GCN) with a Multilayer Perceptron (MLP) module to facilitate positive and negative classification. Chemprint architecture and validation are described in Umansky et al.[10].

Inference

After the model was trained and deployed, we evaluated compounds in an inference set of KCEs known to exhibit favorable properties according to the ideal Target-Product Profile of a once-a-day oral therapeutic with known safety in human trials. Compounds were given a score taken to be the probability of being active against the target in an in vitro bioassay. Top scoring compounds were then investigated in a freedom to operate (FTO) search that flagged compounds that had already been tested in antiviral screening and removed them from consideration.

5.5 Docking Studies

Protein Preparation

The obtained protein crystal structure was prepared and refined for molecular docking study. The water molecules were removed while polar hydrogens were added to the PDB structure. Kollam's united atom force field was used to add charges to the protein structure by the AutoDock 4.2 tool[40]. Then, the structure was saved in a PDBQT format for further molecular docking calculations.

Receptor Grid Preparation

The grid box for HCV and SARS-CoV-2 RdRp was generated through the AutoDock tool[40] around the active site defined by an already co-crystalized ligand present in the PDB structure with the binding pocket. In the case of 'apo' or cryptically-locked pockets, apo structures of RdRp were processed by removing the pocket-filling Λ1-loop residues.

6. Acknowledgments

The authors extend their gratitude for the invaluable support and insightful discussions provided by a multitude of consultants, advisors, partners, and past employees. Particular appreciation is expressed towards Dr. Tushar Menon for his expert consultation and detailed scrutiny of this manuscript. The contributions of ChatGPT 4 in enhancing the manuscript's clarity and readability through language editing and refinement are also gratefully acknowledged[41].

7. References

1. L. Tian et al., *RNA-dependent RNA polymerase (RdRp) inhibitors: The current landscape and repurposing for the COVID-19 pandemic*, European Journal of Medicinal Chemistry, vol. 213. Elsevier BV, p. 113201, Mar. 2021. doi: 10.1016/j.ejmech.2021.113201.

2. Pathania S, Rawal RK, Singh PK. *RdRp (RNA-dependent RNA polymerase): A key target providing anti-virals for the management of various viral diseases*. J Mol Struct. 2022 Feb 15;1250:131756. doi: 10.1016/j.molstruc.2021.131756. Epub 2021 Oct 17. PMID: 34690363; PMCID: PMC8520695.

3. Warren T.K., Jordan R., Lo M.K., Ray A.S., Mackman R.L., Soloveva V., Siegel D., Perron M., Bannister R., Hui H.C., et al. *Therapeutic Efficacy of the Small Molecule GS-5734 against Ebola Virus in Rhesus Monkeys*. Nature. 2016;531:381–385. doi: 10.1038/nature17180.

4. Gandham R, Eerike M, Raj GM, Bisoi D, Priyadarshini R, Agarwal N. *Adverse events following remdesivir administration in moderately ill COVID-19 patients - A retrospective analysis*. J Family Med Prim Care. 2022 Jul;11(7):3693-3698. doi: 10.4103/jfmpc.jfmpc_2468_21. Epub 2022 Jul 22. PMID: 36387660; PMCID: PMC9648220.

5. Focosi D. *Molnupiravir: From Hope to Epic Fail?.* Viruses. 2022 Nov 19;14(11):2560. doi: 10.3390/v14112560. PMID: 36423169; PMCID: PMC9694535.

6. Liu, Y., Hill, M. G., & Klose, T. (2020). *Identification of a Broad-Spectrum Viral Inhibitor Targeting a Novel Allosteric Site in the RNA-Dependent RNA Polymerases of Dengue Virus and Norovirus*. Frontiers in Microbiology, 11, Article 1440.

7. Piplani S, Singh PK, Winkler DA, Petrovsky N. *Computationally repurposed drugs and natural products against RNA dependent RNA polymerase as potential COVID-19 therapies*. Mol Biomed. 2021;2(1):28. doi: 10.1186/s43556-021-00050-3. Epub 2021 Sep 20. PMID: 34766004; PMCID: PMC8450033.

8. Dutta, K., Shityakov, S., Morozova, O., Khalifa, I., Zhang, J., Panda, A., & Ghosh, C. (2020). *Beclabuvir can Inhibit the RNA-dependent RNA Polymerase of Newly Emerged Novel Coronavirus (SARS-CoV-2)*. Preprint. https://doi.org/10.20944/preprints202003.0395.v1

9. E. Van Damme, *Screening of ~5500 FDA-approved drugs and clinical candidates for anti-SARS-CoV-2 activity*. Zenodo, May 20, 2021. doi: 10.5281/ZENODO.4774709.

10. Umansky, T., Ramesh, N., Woods, V., Russell, S. M., Smith, D. M., & Haders, D. (2024). ChemPrint: *An AI-Driven Framework for Enhanced Drug Discovery. Cold Spring Harbor Laboratory*. Preprint. https://doi.org/10.1101/2024.03.22.586314

11. Eltahla AA, Luciani F, White PA, Lloyd AR, Bull RA. *Inhibitors of the Hepatitis C Virus Polymerase; Mode of Action and Resistance*. Viruses. 2015 Sep 29;7(10):5206-24. doi: 10.3390/v7102868. PMID: 26426038; PMCID: PMC4632376.

12. S. Harper et al., *Development and Preliminary Optimization of Indole-N-Acetamide Inhibitors of Hepatitis C Virus NS5B Polymerase*, Journal of Medicinal Chemistry, vol. 48, no. 5. American Chemical Society (ACS), pp. 1314–1317, Feb. 08, 2005. doi: 10.1021/jm049122i.

13. Chou R, Hartung D, Rahman B, Wasson N, Cottrell EB, Fu R. *Comparative effectiveness of antiviral treatment for hepatitis C virus infection in adults: a systematic review*. Ann Intern Med. 2013 Jan 15;158(2):114-23. doi: 10.7326/0003-4819-158-2-201301150-00576. PMID: 23437439.

14. Johnson, M. et al., 2016, *Dasabuvir: A review of its use in Hepatitis C treatment*, Hepatology International

15. S. Bressanelli, L. Tomei, F. A. Rey, and R. De Francesco, *Structural Analysis of the Hepatitis C Virus RNA Polymerase in Complex with Ribonucleotides*, Journal of Virology, vol. 76, no. 7. American Society for Microbiology, pp. 3482–3492, Apr. 2002. doi: 10.1128/jvi.76.7.3482-3492.2002.

16. S. Di Marco et al., *Interdomain Communication in Hepatitis C Virus Polymerase Abolished by Small Molecule Inhibitors Bound to a Novel Allosteric Site*, Journal of Biological Chemistry, vol. 280, no. 33. Elsevier BV, pp. 29765–29770, Aug. 2005. doi: 10.1074/jbc.m505423200.

17. K. L. Rigat et al., *Mechanism of Inhibition for BMS-791325, a Novel Non-nucleoside Inhibitor of Hepatitis C Virus NS5B Polymerase*, Journal of Biological Chemistry, vol. 289, no. 48. Elsevier BV, pp. 33456–33468, Nov. 2014. doi: 10.1074/jbc.m114.613653.

18. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *The Protein Data Bank*. Nucleic Acids Res*.* 2000, 28, 235.

19. RCSB. PDB codes 1C2P, 1NB4, 1NB7, 2BRK, 2BRL, 2DXS, 2XWY, 2WCX, 3MWW, 4DRU, 4GMC, 4NLD, 5TRK, 3Q0Z: Protein Data Bank; 2013 Available from: http://www.rcsb.org/pdb/home/home.do.

20. Lesburg, C., Cable, M., Ferrari, E. et al. *Crystal structure of the RNA-dependent RNA polymerase from hepatitis C virus reveals a fully encircled active site*. Nat Struct Mol Biol 6, 937–943 (1999). https://doi.org/10.1038/13305

21. Martin Hernando, J., Ontoria, J., Malancona, S., Attenni, B., Fiore, F., Bonelli, F., Koch, U., Di Marco, S., Colarusso, S., Ponzi, S., Gennari, N., Vignetti, S., del Rosario Rico Ferreira, M., Habermann, J., Rowley, M. and Narjes, F. (2009), *Optimization of Thienopyrrole-Based Finger-Loop Inhibitors of the Hepatitis C Virus NS5B Polymerase*. ChemMedChem, 4: 1695-1713. https://doi.org/10.1002/cmdc.200900184

22. Madeira F, Pearce M, Tivey ARN, et al. *Search and sequence analysis tools services from EMBL-EBI in 2022*. Nucleic Acids Research. 2022 Jul;50(W1):W276-W279. DOI: 10.1093/nar/gkac240. PMID: 35412617; PMCID: PMC9252731.

23. Yan Gao et al. ,*Structure of the RNA-dependent RNA polymerase from COVID-19 virus*. Science368,779-782(2020).DOI:10.1126/science.abb7498

24. Kirchdoerfer, R.N., Ward, A.B. *Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors*. Nat Commun 10, 2342 (2019). https://doi.org/10.1038/s41467-019-10280-3

25. Gruez ASelisko BRoberts M, Bricogne G, Bussetta C, Jabafi I, Coutard B, De Palma AM, Neyts J, Canard B 2008. *The Crystal Structure of Coxsackievirus B3 RNA-Dependent RNA Polymerase in Complex with Its Protein Primer VPg Confirms the Existence of a Second VPg Binding Site on Picornaviridae Polymerases*. J Virol 82:. https://doi.org/10.1128/jvi.00631-08

26. Marcotte LL, Wass AB, Gohara DW, Pathak HB, Arnold JJ, Filman DJ, Cameron CE, Hogle JM 2007. *Crystal Structure of Poliovirus 3CD Protein: Virally Encoded Protease and Precursor to the RNA-Dependent RNA Polymerase*. J Virol 81:. https://doi.org/10.1128/jvi.02306-06

27. Parag Surana, Vijaya Satchidanandam, Deepak T. Nair, *RNA-dependent RNA polymerase of Japanese encephalitis virus binds the initiator nucleotide GTP to form a mechanistically important pre-initiation state*, Nucleic Acids Research, Volume 42, Issue 4, 1 February 2014, Pages 2758–2773, https://doi.org/10.1093/nar/gkt1106

28. Malet H, Egloff MP, Selisko B, Butcher RE, Wright PJ, Roberts M, Gruez A, Sulzenbacher G, Vonrhein C, Bricogne G, Mackenzie JM, Khromykh AA, Davidson AD, Canard B. *Crystal structure of the RNA polymerase domain of the West Nile virus non-structural protein 5*. J Biol Chem. 2007 Apr 6;282(14):10678-89. doi: 10.1074/jbc.M607273200. Epub 2007 Feb 7. PMID: 17287213.

29. Xu X, Liu Y, Weiss S, Arnold E, Sarafianos SG, Ding J. *Molecular model of SARS coronavirus polymerase: implications for biochemical functions and drug design*. Nucleic Acids Res. 2003 Dec 15;31(24):7117-30. doi: 10.1093/nar/gkg916. PMID: 14654687; PMCID: PMC291860.

30. L. van der Maaten and G. Hinton, *Visualizing Data using t-SNE*, in Journal of Machine Learning Research, vol. 9, no. Nov, pp. 2579-2605, 2008.

31. Faisal S, Badshah SL, Kubra B, Sharaf M, Emwas AH, Jaremko M, Abdalla M. *Computational Study of SARS-CoV-2 RNA Dependent RNA Polymerase Allosteric Site Inhibition*. Molecules. 2021 Dec 30;27(1):223. doi: 10.3390/molecules27010223. PMID: 35011458; PMCID: PMC8746673.

32. Jukič M, Janežič D, Bren U. *Potential Novel Thioether-Amide or Guanidine-Linker Class of SARS-CoV-2 Virus RNA-Dependent RNA Polymerase Inhibitors Identified by High-Throughput Virtual Screening Coupled to Free-Energy Calculations*. Int J Mol Sci. 2021 Oct 15;22(20):11143. doi: 10.3390/ijms222011143. PMID: 34681802; PMCID: PMC8540652.

33. K. L. Rigat et al., *Mechanism of Inhibition for BMS-791325, a Novel Non-nucleoside Inhibitor of Hepatitis C Virus NS5B Polymerase*, Journal of Biological Chemistry, vol. 289, no. 48. Elsevier BV, pp. 33456–33468, Nov. 2014. doi: 10.1074/jbc.m114.613653.

34. Cummings, M.D., Lin, T.-I., Hu, L., Tahri, A., McGowan, D., Amssoms, K., Last, S., Devogelaere, B., Rouan, M.-C., Vijgen, L., Berke, J.M., Dehertogh, P., Fransen, E., Cleiren, E., van der Helm, L., Fanning, G., Van Emelen, K., Nyanguile, O., Simmen, K., Raboisson, P. and Vendeville, S. (2012), *Structure-Based Macrocyclization Yields Hepatitis C Virus NS5B Inhibitors with Improved Binding Affinities and Pharmacokinetic Properties*. Angew. Chem. Int. Ed., 51: 4637-4640. https://doi.org/10.1002/anie.201200110

35. Nguyen, L. (2023, December 20). *MDL-001 demonstrates potential best-in-class combination of viral load reduction, symptom reduction, activity against SARS-COV-2 variants, and Pan antiviral activity*. Model Medicines. www.modelmedicines.com/newsroom/mdl-001-demonstrates-potential-best-in-class-combination

36. Nguyen, L. (2023, December 20). *Model medicines' oral anti-covid-19 drug candidate MDL-001 found to significantly reduce viral load in lungs; accepted into NIH's Antiviral Program for Pandemics*. Model Medicines. https://www.modelmedicines.com/newsroom/mdl-001-viral-load-covid19

37. White, K.M. (2022, September). *Identification of MDL-001 as an oral, potent broad-spectrum antiviral agent*. [Poster presentation]. Model Medicines, FluOMICS 7th Annual Host-Pathogen Interactions: Circuits to Systems Symposium.

38. Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R., Schwede, T. *SWISS-MODEL: homology modelling of protein structures and complexes*. Nucleic Acids Res. 46(W1), W296-W303 (2018)

39. Sapundzhi, F., Popstoilov, M., Lazarova, M. (2023). *RMSD Calculations for Comparing Protein Three-Dimensional Structures*. Numerical Methods and Applications. NMA 2022. Lecture Notes in Computer Science, vol 13858. Springer, Cham. https://doi.org/10.1007/978-3-031-32412-3_25

40. Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S. and Olson, A. J. (2009) *Autodock4 and AutoDockTools4: automated docking with selective receptor flexibility*. J. Computational Chemistry 2009, 16: 2785-91

41. OpenAI. "ChatGPT-4." OpenAI, 2023. [Online]. Available: https://openai.com/chatgpt.