Discovery of RdRp Thumb-1 as a novel broad-spectrum antiviral family of targets and MDL-001 as a potent broad-spectrum inhibitor thereof

Virgil Woods[1,+], Tyler Umansky[1,+], Sean M. Russell[1], David S. 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

Multiple positive sense, single-stranded RNA (+ssRNA) viruses cause human diseases, ranging from mild colds to deadly pandemics. These viruses share a highly conserved RNA-dependent RNA polymerase (RdRp), but developing a non-nucleoside RdRp inhibitor with broad antiviral activity has been elusive. To this end, we first compared RdRp structures across +ssRNA viral families, including homology models created for critical viruses lacking solved crystal structures, and used multiple sequence alignments to observe key loop and pocket residues. Second, we used ChemPrint, a Molecular-Geometric Deep Learning (Mol-GDL) machine learning model, to predict drug inhibition of the Thumb-1 site in RdRp in a panel of +ssRNA viruses of concern. We identified pipendoxifene (MDL-001) as a promising broad-spectrum antiviral that in previous clinical studies has demonstrated safety and tolerability and sufficient evidence to support its oral administration and once a day dosing. These experiments also demonstrated how the Thumb-1 site has been hidden from conventional virtual screening techniques, like docking, and we demonstrated why beclabuvir, the only approved (Japan) Thumb-1 inhibitor, does not work as a broad-spectrum antiviral.

Keywords

+ssRNA Viruses, RNA-dependent RNA polymerase (RdRp), Machine Learning Drug Discovery, Thumb-1 Allosteric Site, Broad-Spectrum Antivirals

Introduction

The rapid adaptability and evolution of viruses present an ongoing risk of new pandemics and create challenges for antiviral drug discovery. 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 antiviral strategy [1]. A drug that has broad antiviral activity could provide treatment for current viruses and possibly provide therapy for new threats.

In the search for broad-spectrum antivirals, the RNA-dependent RNA polymerase (RdRp) has emerged as a prime target due to its central role in the replication processes of positive sense, single-stranded RNA (+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 this active site frequently rely on highly polar phosphate and triphosphate groups in the case of nucleoside and nucleotide analogs. Unfortunately, polar functional groups reduce membrane permeability, which severely limits oral bioavailability. Further, the chemical properties of nucleoside analogs complicate the optimization of potency alongside Absorption, Distribution, Metabolism, Excretion, and Toxicity (ADMET) parameters [3,4]. For example, Remdesivir, initially developed for filoviruses and paramyxoviruses, has been repurposed as an antiviral for SARS-CoV-2; however, its clinical utility is hampered by the need for intravenous administration [3,4]. Another issue with nucleoside RdRp inhibitors is their potential interference with human polymerases. For example, molnupiravir, an oral antiviral for COVID-19, has mutagenicity risk in humans and has poor to no clinical efficacy in post-approval studies [5]. While protease inhibitors, such as those used in HIV [6] and COVID-19 [7] therapies, have been highly successful, they are strictly limited to broad-spectrum activity within a single viral family due to rapid protease evolution [6], whereas our goal is true broad-spectrum inhibition across multiple viral families. There is a great unmet clinical need for broad-spectrum therapies that are safe, effective, and convenient for patients who present with undifferentiated viral infections.

To circumvent these limitations, the focus may need to shift towards exploring allosteric inhibitors of RdRp. Significant advancements have been made in identifying inhibitors of the RdRp of the Hepatitis C Virus (HCV), revealing several allosteric sites. Among these sites is the cryptic Thumb-1 site, whose inhibitors demonstrate cross-genotypic coverage for HCV, hinting at their capability for a broad antiviral range [1]. However, identifying examples of broad-spectrum Thumb-1 inhibitors has proven challenging, exemplified by *in silico* and *in vitro* screening studies of the COVID-19 pandemic. Numerous groups conducted large-scale drug repurposing studies using *in silico* techniques, including those by Dutta et al. [8], to expedite the discovery of antivirals for SARS-CoV-2. These studies employed docking and molecular dynamics (MD) simulations on the apo (closed) structure of RdRp to predict potential antiviral compounds against the RdRp of SARS-CoV-2. Interestingly, the 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, failed to corroborate these predictions [9]. This discrepancy underlines problems in relying solely on physics-based *in silico* methods that do not learn from previous *in vitro* results.

Previous studies on RdRp inhibitors, particularly against HCV, have laid the groundwork for understanding the feasibility of targeting the Thumb-1 allosteric site [10]. 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 1**) [11]. Multiple antivirals have been approved that target HCV NS5B, including beclabuvir (Thumb-1), dasabuvir (Palm-1), and sofosbuvir (active site) [12,13]. In the HCV polymerase NS5B, the Thumb-1 site is cryptic, meaning it is closed without a bound ligand. Typically, cryptic sites occupy a transiently-accessible position on the protein surface that is dynamically revealed. However, in the case of the Thumb-1 site, the cryptic designation is appropriate in the strictest sense. Access to the Thumb-1 site is completely prohibited when the elongating RNA strand is attached, making it more tightly conformationally locked than other cryptic sites and a much more challenging site to drug. The Thumb-1 site 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 [14]. The tip of the NS5B 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. Mechanistic studies show that HCV 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. These studies also show that the pocket becomes unavailable for inhibition once elongation has started [15-16]. The existence of the Thumb-1 site broadly across +ssRNA viruses is so far unsubstantiated, and if it does exist, its druggability—given its strict cryptic nature—outside HCV is unknown. However, if the Thumb-1 site does exist and is indeed druggable across +ssRNA viruses, it could pave the way for a new standard of care in antiviral therapy.

It is becoming increasingly clear that new biology needs to be uncovered to make progress in developing broad-spectrum antiviral therapeutics that can work across +ssRNA viruses; however, modeling the structures of cryptic allosteric sites such as Thumb-1 is inherently difficult. Furthermore, the failure of beclabuvir in SARS-CoV-2 illustrates that discovering a single therapeutic that can successfully drug a site across +ssRNA viral biology, let alone a cryptic allosteric site makes the challenge even greater. In the presented study, we examined if substantial benefits can be gained through ligand-based machine learning methodologies that transcend traditional physics-based simulations by learning in the absence of a protein structure. Specifically, machine learning techniques should be able to learn from *in vitro* data to make predictions on novel bioactivities of compounds, even compounds significantly different in structure from the training data. Thus, we developed an artificial intelligence (AI) drug discovery platform that includes GALILEO with its multiple sequence alignments, homology modeling, and implicit data collection, and ChemPrint, which is a zero-shot capable Molecular-Geometric Deep Learning (Mol-GDL) model [17] (**Figure 2**). We show how this approach identifies new biology that can be readily converted to novel-acting drug compounds with predicted bioactivity across +ssRNA viruses.

Results

Conservation of the Thumb-1 Pocket and Λ1-loop in +ssRNA Viruses

To determine if the Thumb-1 site, deeply characterized in HCV, is conserved across +ssRNA viral families that cause human disease, we began by examining existing Protein Data Bank (PDB) structures of HCV RdRp bound to Thumb-1 inhibitors. Our goal was to identify structural and sequence conservation of this site across viral families and assess its potential as a broad-spectrum druggable target.

HCV Thumb-1 Structure and Key Residues in Ligand Interactions

Ten PDB structures of HCV RdRp bound to Thumb-1 inhibitors (**Supplemental Figure S1A-C**) were analyzed to determine the key residues within 5Å of the inhibitors, highlighting the residues that contribute directly to ligand binding. Fifteen residues in the Thumb-1 pocket were identified (^{392}LA^{393}, ^{395}AA^{396}, T399, ^{424}IL^{425}, ^{428}HF^{429}, ^{492}LGVP^{495}, W500, and R503) across the pdb structures of ligands bound to the Thumb-1 site. Notably, R503 forms a consistent hydrogen bond with inhibitors across all structures.

Structural Comparisons of Thumb-1 and Λ1-loop Interactions

To observe whether the structural interaction between the Thumb-1 pocket and the Λ1-loop observed in HCV was conserved in other +ssRNA viruses, we examined example structures from polymerases of +ssRNA viruses known to cause human disease in the Flaviviridae, Coronaviridae, and Picornaviridae families (**Figure 3A**). The Λ1-loop in HCV adopts a short 9.3 Å α-helix that fits into the Thumb-1 pocket opening. In Coronaviridae, the loop forms a longer 14.1Å α-helix, with conserved asparagine and phenylalanine residues aligning to the hydrophobic pocket. Flaviviridae, such as Japanese Encephalitis Virus (JEV), West Nile Virus, Usutu Virus, and Powassan Virus, exhibit a distinctive horseshoe-shaped structure formed by three α-helices (19 amino acids), with a kink introduced by P333 that helps create the horseshoe's bent conformation when bound to the Thumb domain.

The multiple sequence alignment (MSA) of Λ1-loops across 12 viruses, shown in **Figure 3B**, reveals that the residues projecting deepest into the Thumb-1 site are largely conserved within viral families. These residues are predominantly hydrophobic, typically comprising two conserved hydrophobic side chains, which align well with the hydrophobic nature of the Thumb-1 pocket. This pattern underscores a family-specific conservation of key structural features that may facilitate targeted antiviral strategies.

Picornaviridae, exemplified by Poliovirus and Coxsackievirus, display an 8-amino-acid-long 3_{10} helix within the Λ1-loop that engages the hydrophobic Thumb-1 pocket. In contrast, Norwalk virus (Caliciviridae)

shows a restructured loop with a unique binding interaction. Notably, this virus is the only non-HCV example where some HCV Thumb-1 inhibitors, such as JTK-109, have shown activity, binding to the Thumb domain via a region that serves as an RNA-binding channel. While the loop structures differ significantly in conformation and length across families, they exhibit conserved hydrophobic features compatible with Thumb-1 pocket engagement.

Sequence Conservation Across Viral Families

We then extended our analysis to include multiple sequence alignments (MSA) of both the Λ1-loop and the Thumb-1 pocket across 35 +ssRNA viruses representing eight families. The MSA revealed that some residues proximal to inhibitors in HCV are conserved, particularly among Coronaviridae. Key residues such as A393, A396, L392, and W500 are conserved in coronaviruses SARS-CoV-1, SARS-CoV-2, and MERS-CoV (**Figure 4**). In Flaviviridae, conservation is shared among members of the genera within the family, with residues such as L392 being preserved in members like West Nile Virus and Japanese Encephalitis Virus. For members of Picornaviridae, key residues such as L392 and A396 are conserved.

Implications for Broad-Spectrum Antiviral Development

Despite differences in sequence, the hydrophobic regions of the Thumb-1 pocket demonstrate a high degree of structural consistency. Our findings demonstrate that the Thumb-1 pocket is a structurally conserved site across viral families. This structural consistency, especially within the hydrophobic core of the pocket, highlights its potential as a broad-spectrum antiviral target. The Λ1-loop interaction further supports the feasibility of designing drugs capable of inhibiting RdRp activity across diverse +ssRNA viruses.

Discovery of MDL-001

To discover new chemistry that inhibits the cryptic 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 train on available bioassay and biophysical data. Published structures of the known HCV RdRp inhibitors were used to discover a pharmacophoric model of their shared chemical scaffolds that appear to facilitate Thumb-1-specific binding. Among the 10 aligned PDB structures [18-21], 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 5**). Level 1 data were sourced from papers explicitly claiming Thumb-1 site inhibition in HCV. Level 2 data came from papers that target RdRp of any ssRNA virus without explicit identification of the binding pocket within the protein and match the pharmacophoric profile we identified as Thumb-1. Level 3 compounds were from sources containing assay data from screens performed with infection models of non-HCV viruses. Compounds from levels 2 and 3 containing any of the core Thumb-1 scaffolds were passed through the pharmacophoric filter and presumed to act at the Thumb-1 site, with their bioassay results labeled as ‘implicit' Thumb-1 data. Explicit Thumb-1 data from level 1 and implicit Thumb-1 data from levels 2 and 3 were used to create our broad training dataset. To identify the most potent compounds, we defined those with a bioactivity measurement below 1 μM as active and the rest as inactive. This organized list of active and inactive molecules was then used to train ChemPrint. Notably, the inclusion of implicit data significantly expanded the chemical space that the model trained on (**Figure 6**). The inclusion of the level 3 data, comprising non-HCV Thumb-1 predicted antivirals, allows the model to predict the activity of broad-spectrum Thumb-1 inhibitors. Implicit data from non-HCV viruses allows the machine learning model to predict broad-spectrum Thumb-1 antiviral activity.

The ChemPrint model [17], fine-tuned on broad-spectrum Thumb-1 data (Level 3), was deployed to make inferences on an extensive dataset of ~17,000 chemical compounds we classify as “known chemical entities” (KCEs). All KCEs in the library have been studied in 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 calculated 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 10 individually trained replicate ChemPrint models. The models identified several promising hits among these compounds. Top-scoring compounds were then filtered with a 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 (pipendoxifene, MDL-001) stood out due to its established clinical safety and tolerability and its ability to be dosed orally, combined with its predicted broad-spectrum Thumb-1 inhibition with a mean inhibition of 0.76 ± 0.02 (95% CI). This indicated a high likelihood of MDL-001 having activity against Thumb-1 sites across multiple viral families.

Using ChemPrint Model Variations to Evaluate MDL-001 Chemical Space Novelty and Broad-Spectrum Potential Relative to Beclabuvir

To further clarify the mechanism of action (MoA) of MDL-001 on the Thumb-1 site, we visualized the chemical space diversity across all of the Level 1-3 data (**Figure 6**). Leveraging our labeled compound dataset (see Supplementary Data), we used Principal Component Analysis (PCA) (**Figure 6C**) and t-SNE (**Figure 6D**) to project molecular embeddings in 2-dimensions (see methods) [30-31]. From these figures, four key results can be observed. First, the PCA and t-SNE projections separate MDL-001 from compounds targeting other sites, indicating a distinct mechanism. Second, MDL-001 is positioned near "explicit" Thumb-1 inhibitors in the PCA, specifically those inhibiting HCV RdRp, and close to "implicit" Thumb-1 inhibitors in non-HCV viruses. Third, in the t-SNE projection, MDL-001 occupies a unique position between HCV-only Thumb-1 inhibitors and non-HCV Thumb-1 inhibitors, suggesting its broad-spectrum potential. Fourth, beclabuvir, the only approved HCV Thumb-1 inhibitor, is chemically dissimilar to MDL-001, as evidenced by its distant positioning in both projections and Tanimoto similarity to MDL-001 of 0.15, further emphasizing the novel nature of MDL-001 as a broad-spectrum Thumb-1 inhibitor.

To understand how the model found MDL-001 and to establish its relation to previous Thumb-1 inhibitors, we compared the scores for MDL-001 across different models trained on different subsets of the compounds in the literature bioassay data, organized into five data frames (**DF1-5, see Figure 6E**). We observed that this compound was predicted to inhibit the Thumb-1 site (as shown in **Figure 7A**). Specifically, **DF2** (Thumb-1 explicit and implicit) and **DF3** (non-Thumb-1) assigned average probability scores of 0.76 and 0.15, respectively, indicating that the drug most likely targets the Thumb-1 site for viral inhibition. To further understand the mechanism of action of MDL-001, we investigated two other models (**Figure 7B**): explicit HCV-only Thumb-1 data (**DF4**) and non-HCV "implicit" Thumb-1 data (**DF5**). We next compared MDL-001 and beclabuvir across the model variants **DF4 and DF5**. Beclabuvir, a known HCV Thumb-1 antiviral, was identified in multiple antiviral studies and represented a structure distinct from that of MDL-001. The average HCV Thumb-1 antiviral activity prediction scores were 0.69 for MDL-001 and 0.98 for beclabuvir in the **DF4** model, using a conservative active/inactive cutoff of 1.0 μM. In the **DF5** model, using a more lenient active/inactive cutoff of 20.0 μM, MDL-001 showed a consistent average prediction score of 0.69, indicating its potential as an inhibitor beyond anti-HCV activity. In contrast, beclabuvir's average prediction score drastically reduced to 0.04 in the **DF5** model, aligning with experimental observations indicating its inactivity against SARS-CoV-2 [9].

Docking Study of MDL-001 and Beclabuvir

In the *in vitro* screen by Janssen, beclabuvir had no significant antiviral effect against SARS-CoV-2 up to a concentration of 20 μM [9]. Since previous docking and MD studies suggested beclabuvir could have activity against SARS-CoV-2 [32-33], we aimed to understand the faulty *in silico* predictions by performing a comparative docking study of the Thumb-1 sites in SARS-CoV-2 and HCV for both drugs (**Figure 8**).

Previous docking campaigns failed to address the Thumb-1 site of SARS-CoV-2 as a cryptically-closed pocket in available apo protein structures. To overcome this limitation, we processed the apo structures by removing the pocket-filling Λ1-loop residues.

While ChemPrint scored MDL-001 with a high probability of successfully inhibiting the Thumb-1 site of RdRp, our comparative docking (Autodock4) indicated that MDL-001 had higher (worse) docking scores in both HCV and SARS-CoV-2 Thumb-1 sites 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**) showed key residue engagements such as Arg503, which is known to form hydrogen bonds with established HCV Thumb-1 inhibitors [34]. 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].

In beclabuvir's top-ranked pose in SARS-CoV-2 RdRp, the cyclohexane at position 3 of its indole scaffold appears to be in a more forced conformation than the top pose in HCV (**Figure 8A-D**). 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. ChemPrint predicted MDL-001 to have activities of < 1 μM for HCV and < 20 μM throughout viruses that comprise the implicit Thumb-1 data set, including SARS-CoV-2. 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]. Since *in vitro* tests determined that beclabuvir was inactive for SARS-CoV-2 (EC50 > 20.0 μM) [9], the 'forced' open structure of the SARS-CoV-2 is likely not a perfect model for *in silico* docking of Thumb-1 inhibitors. However, the AI-driven ChemPrint correctly predicted beclabuvir to be inactive (> 20.0 μM) against SARS-CoV-2, underscoring the potential drawback of relying on rigid or even flexible physics-based docking to protein structures in trying to capture the behavior of ligand binding in cryptic sites like Thumb-1 of RdRp of +ssRNA viruses. Instead, machine learning models like ChemPrint can use existing Thumb-1 chemistry to discover where we might find broad-spectrum inhibition.

Discussion

The impact of epidemics caused by ssRNA viruses, such as Influenza, RSV, and COVID-19, has been profound historically and in recent years [1]. These outbreaks underscore the critical need for versatile treatments capable of addressing a wide array of ssRNA viruses. Since RdRp is highly conserved across these viruses, it is a prime target for direct-acting antivirals. While traditional drug discovery methods have produced remdesivir and other nucleoside analogs targeting the active site of RdRp, these drugs suffer from suboptimal membrane permeability and come with elevated safety and tolerability concerns. Conserved allosteric sites, like Thumb-1 in RdRp, offer a new opportunity, and inhibitors of these sites likely will have better pharmacological properties than nucleoside analogs, such as oral bioavailability. Unfortunately, previous work with Thumb-1 inhibitors has largely only been capable of inhibiting HCV [1].

To identify broad-spectrum antiviral agents targeting ssRNA viruses, we analyzed conserved RNA-dependent RNA polymerase (RdRp) elements across viral families from available structures and homology models, followed by multiple sequence Alignments (MSA) of Λ1-loop-Thumb-1 contact residues. We found that the Thumb-1 pocket had structural conservation that could be exploited. Next, we used our AI drug discovery platform, GALILEO, and its Mol-GDL model, ChemPrint [17], to identify a candidate therapeutic that targeted this conserved region, and we found that MDL-001 had a high predicted antiviral efficacy against a variety of +ssRNA viruses (**Figure 5B**).

Upon further investigation, ChemPrint's prediction scores, when trained on Thumb-1 and non-Thumb-1 bioassay results, indicated that MDL-001's mechanism of action is likely targeted at the Thumb-1 site (Figure **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 safety status and incorrectly identified beclabuvir. Notably, MDL-001 and beclabuvir were docked to the crystal structure of the open (ligand-bound) HCV Thumb-1 site, and a modified structure of the SARS-CoV-2 Thumb-1, where the loop residues were removed to open the pocket. While beclabuvir showed favorable docking scores, it lacked inhibitory activity against SARS-CoV-2 at 20 μM from *in vitro* testing [9]. One reason why Thumb-1 inhibition in SARS-CoV-2 might require different chemistry from the optimized HCV NS5B Thumb-1 inhibitors is that in coronaviruses, the RdRp exists as a complex of a catalytic core (NSP12) and cofactors NSP7 and NSP8. These cofactors bind near the Thumb-1 site and the effect of their binding on Λ1-loop-Thumb-1 kinetics. Chemprint may have identified chemistry that can circumvent this hurdle in the discovery of MDL-001. Clinical studies of MDL-001 (pipendoxifene) have demonstrated safety and tolerability in ~100 patients [41-42].

Looking forward, this study paves the way for several critical next steps. First, the assumption that the open Thumb-1 pocket conformation of SARS-CoV-2 is similar to the edited apo structure will be further validated by extending docking studies to other viruses. Second, the non-HCV ChemPrint model, which has so far relied on implicit Thumb-1 data from various whole virus screenings, will be expanded with additional data to enhance its predictive power. Third, the mechanism of action of MDL-001 via Thumb-1 inhibition will undergo further biophysical characterization, animal models, and human studies.

Conclusion

By identifying the broad-spectrum antiviral activity of MDL-001, this study unveils the Thumb-1 site of RdRp as a novel biological target for broad-spectrum drug discovery. This advancement was only possible through integrating AI, notably our GALILEO platform and its ChemPrint prediction model. MDL-001, with its favorable pharmacokinetics and ADMET properties, offers promise as an orally administered therapy that effectively treats a broad range of +ssRNA viral diseases. Beyond its immediate implications for antiviral drug development, our study paves the way for broader applications of allosteric inhibition strategies across a range of biological targets.

Methods

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 3**. To maintain consistency, sequences were retrieved based on their availability and completeness.

Multiple Sequence Alignment (MSA) - Λ1-loop Residues

In the second MSA study (**Figure 3B**), attention was directed towards the Λ1-loop residues of the 12 sequences of viruses in **Figure 3A**, 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].

Multiple Sequence Alignment (MSA) - Thumb-1 Pockets

The 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 4**), with at least one virus from each

genus in **Figure 3**. The Thumb-1 region was identified based on previously annotated structures and functional domains [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.

Sequence Alignment and Parity Score Calculation

Parity scores across 35 viruses were calculated using the ALion web tool, accessible through the Bioinformatics & Evolutionary Genomics group at Ghent University (http://bioinformatics.psb.ugent.be/webtools/alion/). This tool utilizes the Smith-Waterman algorithm for local sequence alignment, employing the Blosum62 substitution matrix to evaluate evolutionary relationships between sequences.

The alignment process incorporated a normalized and weighted gap penalty system, with a gap opening penalty of 5 and a gap extension penalty of 0.5. The normalization was applied to match the scoring matrix, ensuring that penalties were scaled appropriately relative to average match scores. Additionally, gap penalties were weighted according to the amino acid composition of the sequences, to account for potential compositional biases.

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 Λ1-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' Λ1-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.

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 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.

ChemPrint

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 data point. 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. [17].

Train-Test Split

To split our data into training and test sets, we utilized t-distributed Stochastic Neighbor Embedding (t-SNE) to project ECFP4 encoded molecule data into 2-dimensions. This method was performed as previously described in Umansky et al. [17]. The Area Under the Receiver Operating Characteristic curve (AUROC) values of the five alternate models (**DF1**: 0.88, **DF2**: 0.93, **DF3**: 0.86, **DF4**: 0.93, **DF5**: 0.76; **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 even in cases where there were fewer training set compounds.

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.

Dataframes

To provide additional supporting evidence into the mechanism of action of MDL-001, we created several label-specific training datasets (DF1-5) (**Figure 6E**) comprising subsets of the data collected from Levels 1-3. DF1 contains all data from Levels 1-3, DF2 was 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**.

Docking Studies

Protein Preparation

The obtained protein crystal structures were 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.

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. Majerová T, Konvalinka J. Viral proteases as therapeutic targets. Mol Aspects Med. 2022 Dec;88:101159. doi: 10.1016/j.mam.2022.101159. Epub 2022 Nov 29. PMID: 36459838; PMCID: PMC9706241.

7. Yang KS, Leeuwon SZ, Xu S, Liu WR. Evolutionary and Structural Insights about Potential SARS-CoV-2 Evasion of Nirmatrelvir. J Med Chem. 2022 Jul 14;65(13):8686-8698. doi: 10.1021/acs.jmedchem.2c00404. Epub 2022 Jun 22. PMID: 35731933; PMCID: PMC9236210.

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. 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.

11. 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.

12. 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.

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

14. 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.

15. 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.

16. 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.

17. Umansky, T., Ramesh, N., Woods, V., Russell, S. M., Smith, D. M., & Haders, D. (2024). ChemPrint: An AI-Driven Framework for Enhanced Drug Discovery. bioRxiv. https://doi.org/10.1101/2024.03.22.586314

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*. Science 368,779-782(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 cofactors*. 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. Need PCA reference: J. Medina-Franco, K. Martinez-Mayorga, M. Giulianotti, R. Houghten, and C. Pinilla, "Visualization of the Chemical Space in Drug Discovery," Current Computer Aided-Drug Design, vol. 4, no. 4. Bentham Science Publishers Ltd., pp. 322–333, Dec. 01, 2008. doi: 10.2174/157340908786786010.

31. 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.

32. 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.

33. 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.

34. 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.

35. 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

36. 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.

37. What's new in the Guidelines (February 29, 2024) National Institutes of Health. Available at: https://www.covid19treatmentguidelines.nih.gov/about-the-guidelines/whats-new/.

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 modeling 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. Cotreau, M. M. et al. *Multiple-Dose, Safety, Pharmacokinetics, and Pharmacodynamics of a New Selective Estrogen Receptor Modulator, ERA-923, in Healthy Postmenopausal Women*. The Journal of Clinical Pharmacology 42, 157–165 (2002).

42. Sorbera, L. A., Castaner, J. & Silvestre, J. S. *Pipendoxifene*. Drugs of the Future 27, 0942 (2002).

43. Craig G. Nevill-Manning, Cecil N. Huang & Douglas L. Brutlag, (1997) "Pairwise protein sequence alignment using Needleman-Wunsch and Smith-Waterman algorithms," personal communication.

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.

Author Contributions

D.H. conceived the study design that led to the discovery of MDL-001 as an inhibitor of RdRp. The evaluation of MDL-001 through machine learning and RdRp homology studies, which led to the identification of RdRp Thumb-1 as a family of inhibitors that can be targeted with broad-spectrum chemistry, was conceived by V.W., T.U., and D.H.. V.W. performed the Multiple Sequence Alignment analysis. V.W. and T.U. conducted the machine learning experiments, and T.U. built the machine learning models used and performed docking experiments. V.W., T.U., S.R., D.G., D.S., and D.H. analyzed the results. All authors reviewed the manuscript.

Data Availability

The source sequences used in the multiple sequence alignments are included in a supplementary data file along with the list of papers that comprise the training dataset.

Competing Interests

The authors affiliated with Model Medicines declare the existence of a financial competing interest. Davey Smith reports the following competing interests: Consulting fees from Bayer, Lucira, Pharma Holdings, Evidera, Vx Biosciences, Gilead, and Red Queen Biosciences. Stock options from Fluxergy, Linear Therapies, and Model Medicines. Payments to his institution from the NIH. These entities have provided financial compensation, stock options, or institutional support within the past 36 months, potentially influencing, or that could give the appearance of potentially influencing the submitted work.

Figures

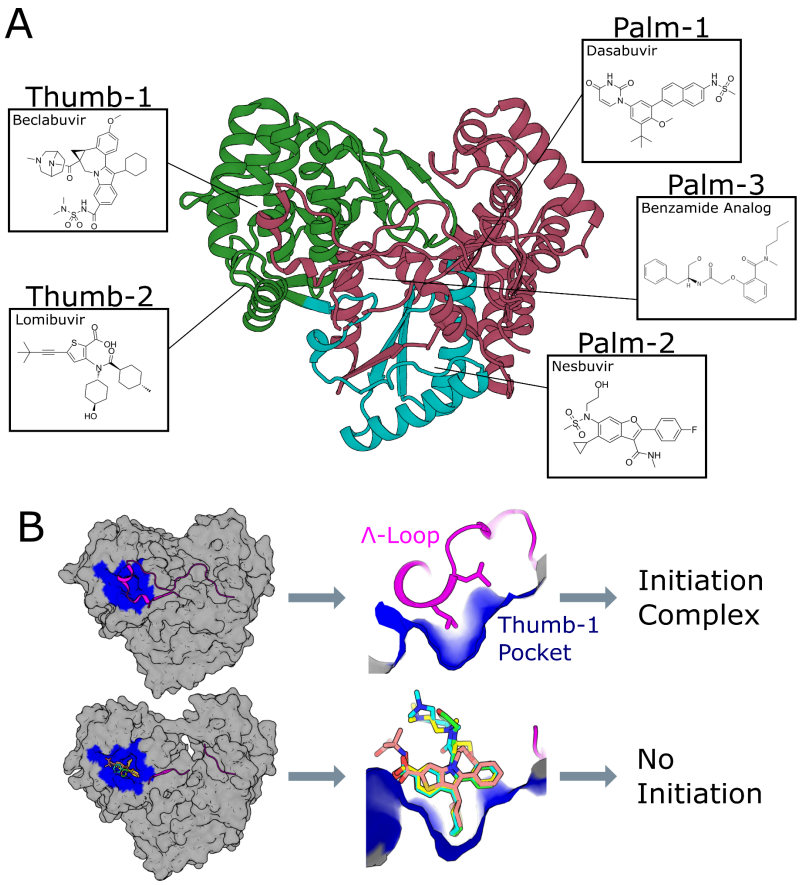

Figure 1 (A) View of HCV RdRp with sites highlighted and one example of an inhibitor for each site (PDB ID: 1C2P) [20]. **(B)** View of the *Λ*1-thumb interaction and mechanism of Thumb-1 inhibition. **Top Left:** HCV RdRp apoprotein structure with a gray surface representation acquired from 1C2P. **Top 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.

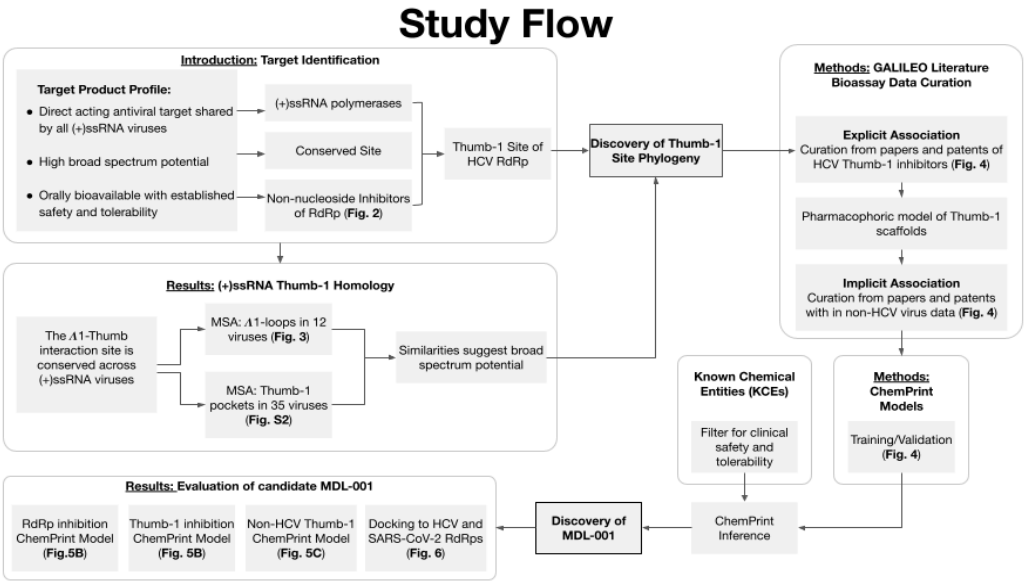

Figure 2: Flow chart of the steps taken in this study to discover and evaluate candidate drug MDL-001 for direct-acting, broad-spectrum inhibition of positive-strand single-stranded RNA (+ssRNA) viruses. We achieve this via predicted inhibition of viral RNA-dependent RNA polymerase (RdRp) via the allosteric Thumb-1 cryptic site, previously identified and probed only for polymerase Non-Structural Protein 5B (NS5B) of the Hepatitis C Virus (HCV).

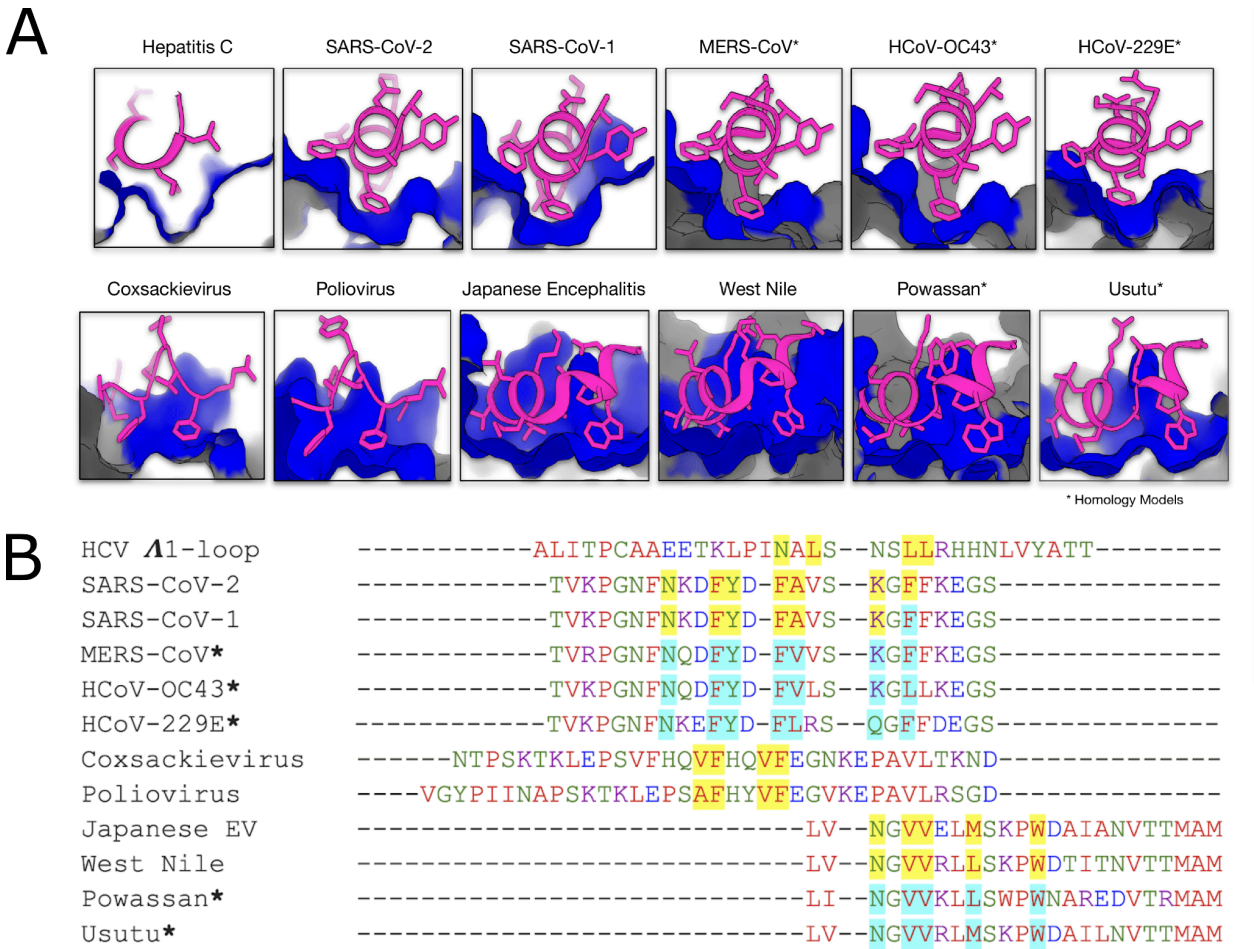

Figure 3: (A) Pymol representations of thumb-Λ1-loop interactions in 12 +ssRNA viruses. *PDB-IDs:* HCV (1C2P) [20], SARS-CoV-2 (6M71) [23], SARS-CoV-1 (6NUS) [24], Coxsackievirus (3CDU) [25], Poliovirus (2IJF) [26], Japanese Encephalitis (4MTP) [27], West Nile (2HFZ) [28]. Structures denoted with * indicate 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].

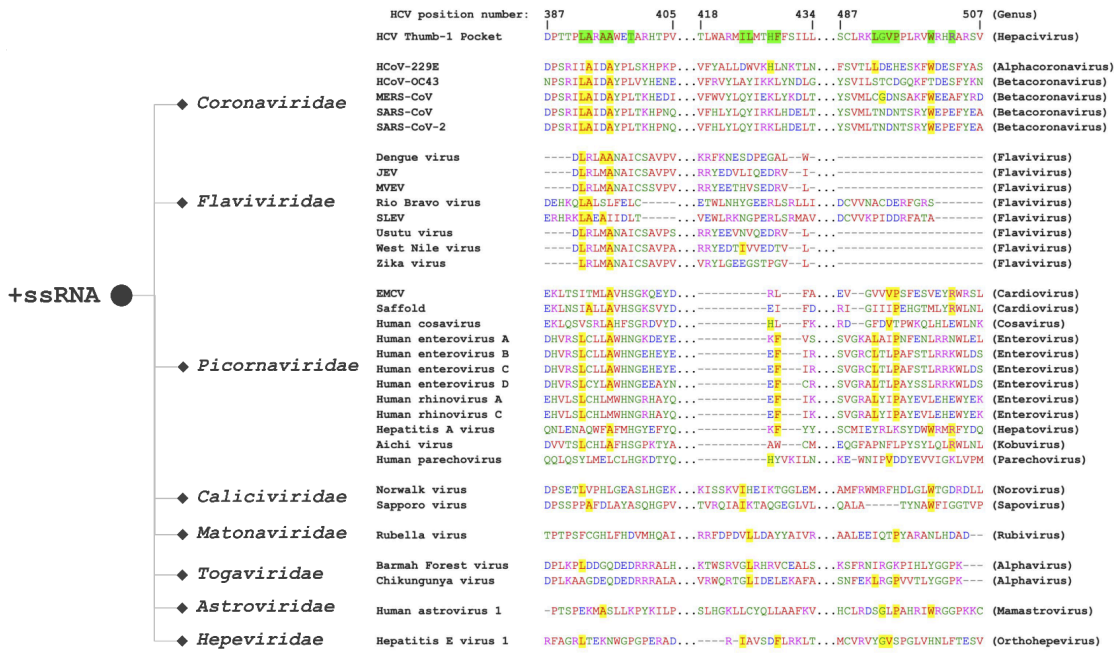

Figure 4: Family tree of 8 human pathogenic +ssRNA families and genera within those families. Clustal Omega [22] multiple sequence alignment (MSA) of the Thumb-1 pocket of HCV and representative viruses from each +ssRNA genus (see methods). HCV residues that are found within 5Å of the bound Thumb-1 inhibitors are highlighted green. Identity matches of these residues in other viruses are highlighted yellow.

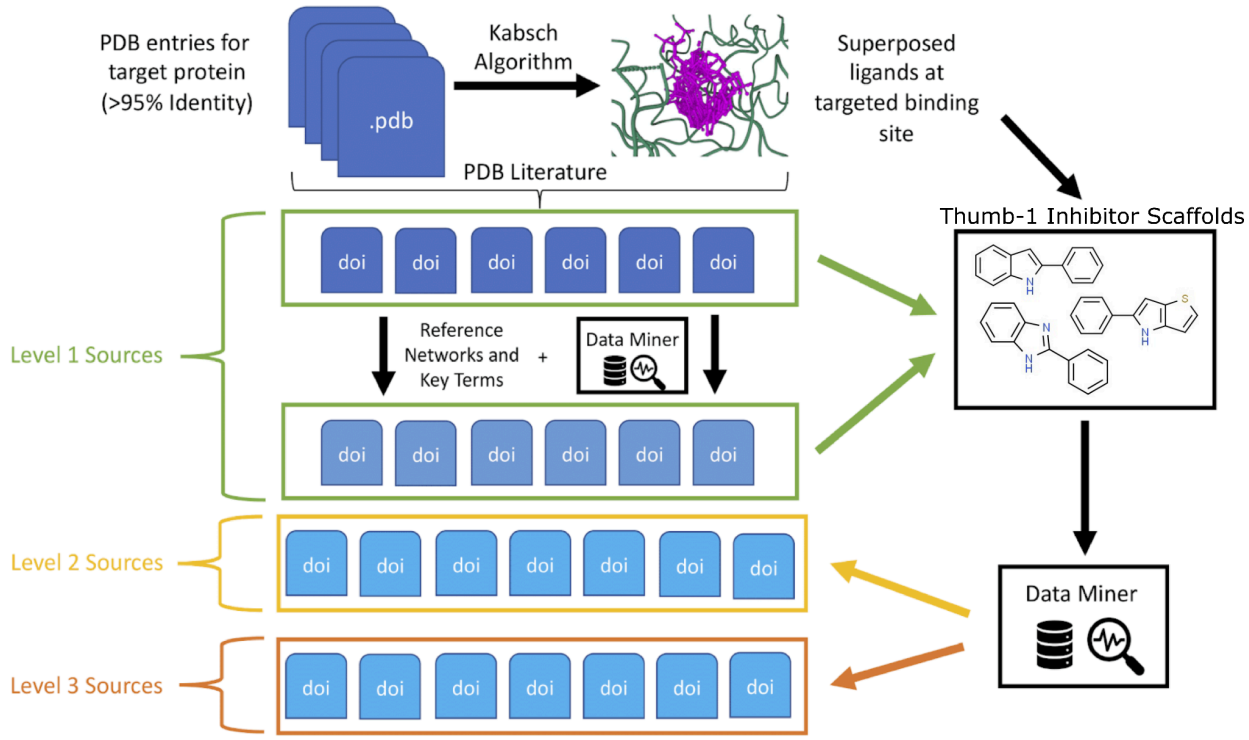

Figure 5: Process overview for bioassay data collection and curation.

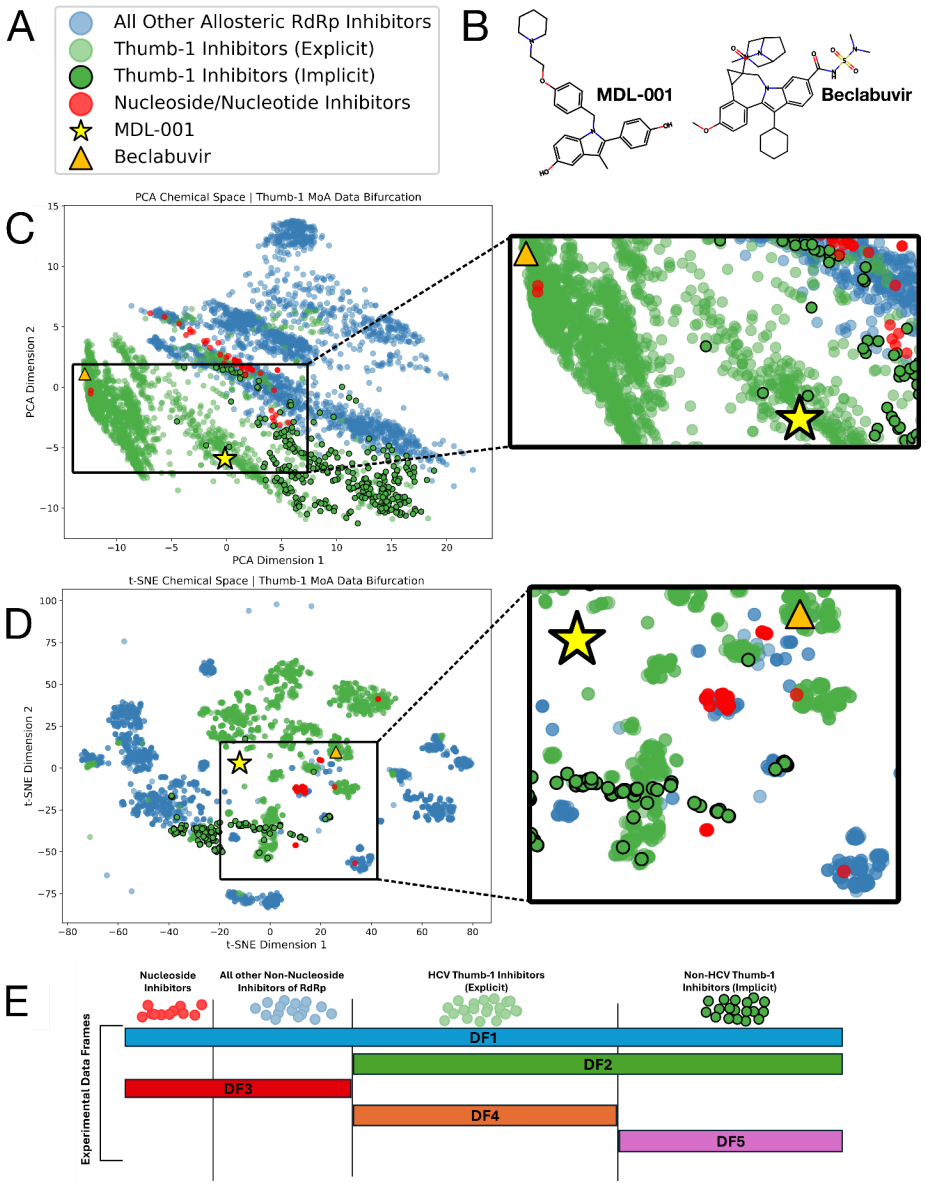

Figure 6: (A) The legend for PCA and t-SNE visualizations. **(B)** Drawn structures of MDL-001 and beclabuvir. **(C)** PCA and **(D)** t-SNE chemical space projections of antiviral compounds that are known to target RdRp, showing the different classifications of nucleoside, non-nucleoside, and Thumb-1 inhibitors (explicit and implicit data). MDL-001 is depicted as a yellow star and Beclabuvir is depicted as an orange triangle. The first two principal components in the PCA explain 24% of the total variance. **(E)** 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)

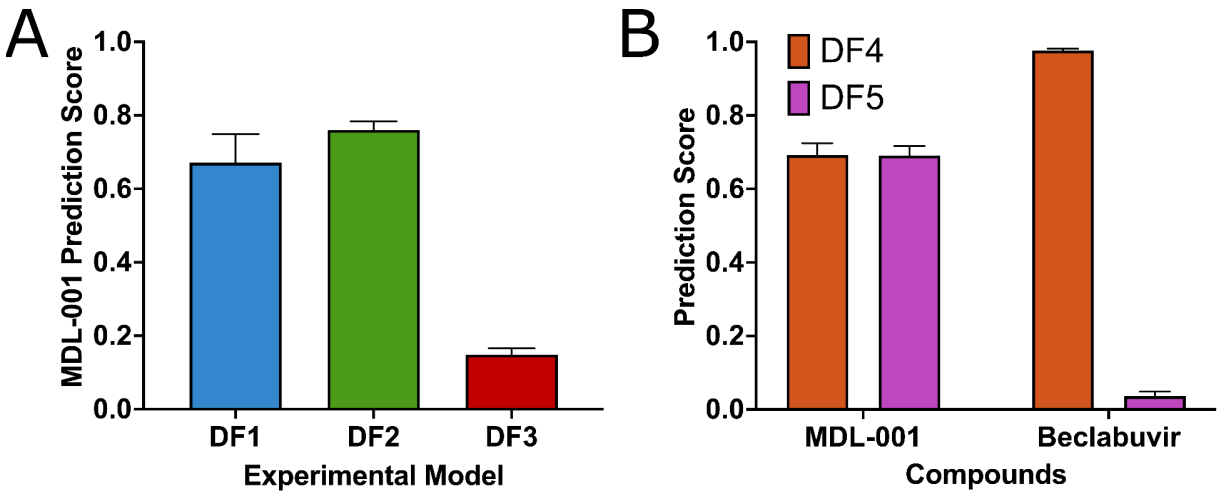

Figure 7: (A) Prediction scores for MDL-001 across the three experimental models **DF1** (All training data), **DF2** (Thumb-1 only data), and **DF3** (only non-Thumb-1 data). **(B)** Bar graph of the prediction scores of MDL-001 and beclabuvir across models **DF4** (HCV-only Thumb-1 data) and **DF5**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

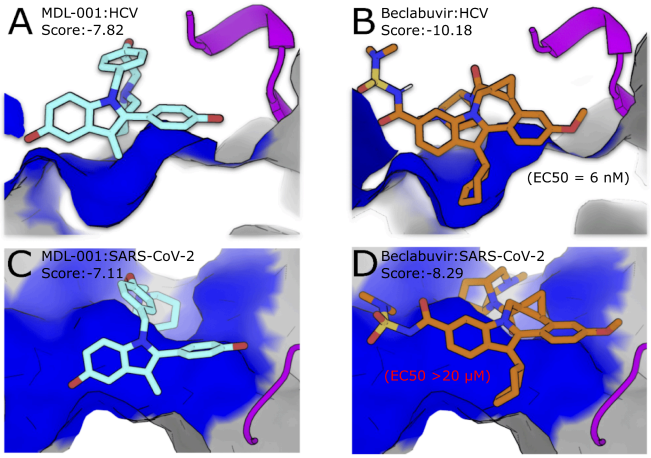

Figure 8: (A-B) Poses of MDL-001 (**A**) and beclabuvir (**B**) docked into the Thumb-1 pocket of HCV RdRp acquired from PDB ID: 4DRU [35]. **(C-D)** MDL-001 (**C**) and beclabuvir (**D**) docked into the Thumb-1 pocket of SARS-CoV-2 RdRp acquired from PDB ID: 6M71 [23].