AWARD NUMBER: W81XWH-18-1-0307

TITLE: Hepatitis B Virus Ribonuclease H: Mechanisms of Catalysis and Inhibition

PRINCIPAL INVESTIGATOR: John Tavis, Ph.D.

CONTRACTING ORGANIZATION: Saint Louis University

REPORT DATE: November 2022

TYPE OF REPORT: Final Report

PREPARED FOR: U.S. Army Medical Research and Development Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
NOVEMBER 2022	Final Technical Report	01 Aug 2018-31 July 2022

4. TITLE AND SUBTITLE

Hepatitis B Virus Ribonuclease H: Mechanisms of Catalysis and Inhibition

5a. CONTRACT NUMBER
W81XWH-18-1-0307

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

John Tavis, Ph.D.

E-Mail: john.tavis@health.slu.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Saint Louis University
1 North Grand
Saint Louis, Missouri
63103-2006

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Development Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The proposed project was a biochemical analyses of the hepatitis B virus (HBV) ribonuclease H (RNaseH). RNaseH activity is essential for HBV replication and is a target of drug discovery, but RNaseH enzymology is very poorly understood. RNaseH catalysis requires Mg++ ions. Aim 1 asked "How do Mg++ ions and the heteroduplex substrate interact with the RNaseH?" Aim 2 asked "How do RNaseH inhibitors interact with the RNaseH?" These studies were stymied by unanticipated difficulties producing recombinant HBV RNaseH, so we changed focus to conduct similar studies on human RNase H1 (huRH1). We found that huRH1 inhibitors are unusual noncompetitive inhibitors that bind to the active site. We propose a model in which, by binding to the active site, huRH1 inhibitors stabilize an inactive enzyme–substrate–inhibitor complex to block enzyme function. We also predicted and validated the structure of the HBV polymerase, of which the HBV RNaseH is one domain. This breakthrough structure has strong mechanistic implications and has already launched 3 new drug discovery projects against HBV.

15. SUBJECT TERMS NONE LISTED

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRDC
U	U	U	UU	40	19b. TELEPHONE NUMBER *(include area*

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

TABLE OF CONTENTS

Final Report

Award Number: W81XWH-18-1-0307
Project Title: Hepatitis B Virus Ribonuclease H: Mechanisms of Catalysis and Inhibition
Principal Investigator: John Tavis, PhD
Contracting Organization: Saint Louis University

Introduction

This project was intended to conduct biochemical analyses of the hepatitis B virus (HBV) ribonuclease H (RNaseH). RNaseH activity is essential for the virus to replicate its genome and is a target of ongoing drug discovery projects, but the enzymology of the RNaseH is very poorly understood. The RNaseH needs to bind to both magnesium ions (Mg^{++}) and its heteroduplex substrate to be active, and the inhibitors we discovered appear to work by binding to the Mg^{++} ions to block their function. *Aim 1* asked "*How do Mg^{++} ions and the heteroduplex substrate interact with the RNaseH*?" We planned to conduct basic biochemical binding and enzymatic studies to evaluate how Mg^{++} binds to the enzyme, how Mg^{++} affects the ability of the RNaseH to bind to the substrate, whether the RNaseH's two modes of enzymatic activity (endo- and exo-nucleolytic) have similar Mg^{++} needs, and mutate the RNaseH to identify amino acids in the enzyme that influence substrate binding and RNaseH activity. *Aim 2*, asked "*How do RNaseH inhibitors interact with the RNaseH*?" We planned do extensive molecular binding and enzymatic kinetics studies with wild-type and mutant RNaseH enzymes to evaluate how the inhibitors bind to the enzyme and to identify amino acids in the RNaseH that contribute to action of the inhibitors. Similar techniques will be employed to determine the role of the Mg^{++} ions in the binding, efficacy, and stability of the inhibitors as they interact with the RNaseH.

Keywords

Hepatitis B virus, ribonuclease H, inhibitors, enzymology

Accomplishments

What were the major goals of the project? The major goal of the project was to obtain the first robust biochemical characterization of the HBV RNaseH and its mechanism of inhibition by the RNaseH inhibitors we had identified. These data were sought to guide mechanistic understanding of the enzyme and to provide key foundational data for anti-HBV RNaseH drug discovery.

What was accomplished under these goals? We did not achieve our primary goals with the HBV RNaseH due to unanticipated, intractable technical issues that appeared with producing recombinant HBV RNaseH of suitable quality and quantity.

Approved Statement of Work:

Specific Aim 1: ***Evaluate RNaseH binding activity to Mg^{++} and heteroduplex substrate***	**Timeline** Months	**Site 1**	**Site 2**	**Status**
Major Task 1: Investigate affinity of Mg^{++} for the RNaseH	1-4	Drs. Tavis & Pozzi		90% completed
			NA	
Major Task 2: Measure binding of the heteroduplex substrate to the HBV RNaseH	3-10	Drs. Tavis & Pozzi	NA	90% complete
Major Task 3: Evaluate the effects of varying Mg^{++} concentration on substrate binding affinity.	8-14	Drs. Tavis & Pozzi	NA	90% completed
Major Task 4: Measure exonucleolytic RNA degradation at varying Mg^{++} concentrations	12-16	Dr. Tavis		15% completed

Major Task 5: Determine impact of key mutations that inhibit substrate binding and/or catalysis on viral replication.	12-18	Dr. Tavis		5% completed
Milestone(s) achieved: Determine how RNaseH binds Mg^{++} and the substrate and identify RNaseH residues key to these processes.		Drs. Tavis & Pozzi		Not completed
Specific Aim 2: ***Evaluate binding of inhibitors to the RNaseH***				
Major Task 6: Assess how Mg^{++} affects binding of RNaseH inhibitors	16-22	Drs. Tavis & Pozzi		90% completed
Major Task 7: Determine how substrate affects inhibitor binding	20-24	Drs. Tavis & Pozzi		25% completed
Major Task 8: Assess the affinity and kinetics of inhibitor binding to the RNaseH	20-30	Drs. Tavis & Pozzi		Not started
Major Task 9: Evaluate the mode of RNaseH inhibition	28-32	Dr. Tavis		70% completed
Major Task 10: Determine whether the endo- and exonuclease activities differ in sensitivity to inhibitors	32-36	Dr. Tavis		Not started
Major Task 11: Identify RNaseH residues that affect inhibition efficacy	24-36	Dr. Tavis		10% complete
Milestone(s) Achieved: Determine how Mg++ and the substrate affect inhibitor binding affinity, kinetics, and function; Determine if RNaseH inhibition is competitive; Compare inhibition of the exo- and endo-nuclease reactions of the RNaseH, and reveal residues that affect inhibition.	36	Drs. Tavis & Pozzi	NA	Not completed

Task #1. The K_d for Mg^{++} binding to the RNaseH was measured as 0.26 mM, and binding was found to be cooperative based on the Hill slope value of 1.6 (Fig. 1). As cooperativity requires two or more binding ligands, this provides strong experimental support for binding of two Mg^{++} ions to the RNaseH active site, as predicted by the enzyme's possession of a DEDD two Mg^{++} ion binding motif. These studies were not finalized because we lost confidence in the quality of the enzyme we produced, consequently they need to be confirmed with a more robust enzyme.

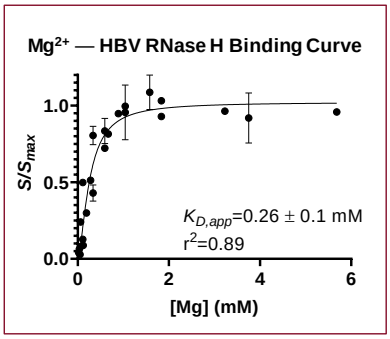

Fig. 1. Binding curve for Mg^{++} to the HBV RNaseH.

Task #2. Directly detecting binding of the heteroduplex substrate to the HBV RNaseH requires physical methods. The substrate:enzyme complex is not stable enough to be measured by co-precipitation methods, and the enzyme quality was inadequate for us to use the planned plasmon resonance approach. However, Mg^{++} is an essential cofactor in RNaseH reactions and free Mg^{++} is found at

0.3-0.7 mM in cells, so the studies addressing how Mg^{++} affects substrate binding by assessing substrate K_M in activity assays in Task 3 are pertinent. Fig. 2A reveals that the K_M for substrate is ~50 nM at ~0.5 mM Mg^{++}, which is a proxy for the K_d for enzyme:substrate binding. These studies were not finalized because we lost confidence in the quality of the enzyme we produced.

Task #3. The effects of Mg^{++} binding on heteroduplex substrate binding by the HBV RNaseH were measured by observing changes in the K_M for substrate over a wide range of Mg^{++} concentrations. We observed a biphasic response in which a dose-dependent reduction in K_M occurred until 1-2 mM Mg^{++}, above which K_M increased in a dose-dependent manner (Fig. 2A). Given the free intracellular Mg^{++} concentration is ~0.3 – 0.7 mM, these data suggest Mg^{++} contributes to affinity of the enzyme-substrate complex in physiological conditions. V_{max} was also dependent on Mg^{++}, and the magnitude of this effect was much greater than the effect on K_M (Fig. 2B). Non-linear curve fitting shows that the 50% effective concentration (EC_{50}) of Mg^{++}'s effect on V_{max} is 1.1 ± 0.3 mM. Unlike Mg^{++}'s effect on the RH's K_M, V_{max} was not inhibited at high Mg^{++} concentration. Mg^{++}'s biphasic effects on K_M and its stimulatory effect on catalysis, each with distinct EC_{50}s indicate that Mg^{++} impacts RNA hydrolysis by HBV RH via multiple mechanisms. These studies were not finalized because we lost confidence in the quality of the enzyme we produced, consequently they need to be confirmed with a more robust enzyme.

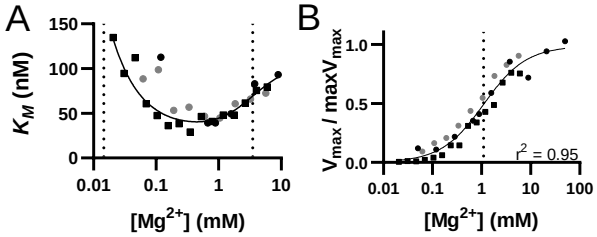

Fig. 2. Effects of Mg++ on substrate binding by HBV RNaseH. A. Bi-phasic effects of Mg^{++} on substrate binding revealed by alterations in substrate K_M. B. Mg^{++} increases V_{max} for HBV RNaseH hydrolysis.

Task #4. Preliminary studies were conducted to determine conditions under which to quantitatively measure the effects of Mg^{++} on the exonuclease activity of the HBV RNaseH. The assay was successfully converted from one requiring [^{32}P]-labeled RNA to one that could be detected by staining the RNA with the fluorescent stain SyberGold, and we confirmed that rates could be determined by quantifying degradation of the full-length RNA. We stopped this project at this point because we lost confidence in the enzyme.

Task #5. We constructed 15 mutant HBV RNaseHs in preparation for structure-function assays with the HBV RNaseH. Work was largely stopped at that point because we lost confidence in the enzyme.

Task #6. Inhibition of the HBV RNase H was found to be strongly dependent upon Mg^{++} concentration, with the half-maximal effect of Mg^{++} on inhibition efficiency being similar to the K_d of Mg^{++}. This provides very strong support for the hypothesis that inhibition involves direct interaction between the inhibitors and the Mg^{++} ions in the active site. Fig. 3 shows the Mg^{++} dependence on 50% inhibitory concentration (IC_{50}) for compounds #110, 1, and 208. For reference, the K_d of Mg^{++} is 0.26 mM. Furthermore, given that the inhibition constant K_i reflects Kd and that IC_{50} changes proportionally to K_i when substrate is held constant, the dependence of IC_{50} on Mg^{++} concentration directly implies that inhibitor affinity for the enzyme is strongly dependent upon Mg^{++} concentration. These results have not been published because we lost confidence in the quality of the enzyme we produced and they need to be confirmed with a more robust enzyme. However, they are consistent with what was measured with human ribonuclease H1 (below), so we are confident they reflect the overall situation with HBV RNaseH.

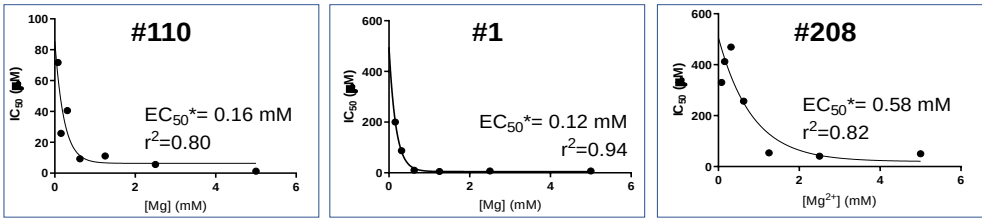

Fig. 3. Dependence of IC_{50} on Mg^{++} concentration for representative RNaseH inhibitors.

Task #7. The effect of inhibitors on substrate binding to the HBV RNaseH was investigated as part of the inhibition studies described under Task #9. The competitive nature of inhibition indicates that the inhibitors compete for binding to the enzyme active site with the substrate. However, we were not able to directly

quantify the effects of the compounds on substrate binding. That will require a more robust source of HBV RNaseH.

Task #8. We did not pursue studies to define the kinetics and affinity of the inhibitors for the HBV RNaseH active site because these studies require highly sensitive surface plasmon resonance measurements, and the enzyme we could produce was of insufficient quality.

Task #9. The mechanisms of inhibition for the three major chemotypes of RNaseH inhibitors were evaluated against HHBV RNaseH. The α-Hydroxytropolone #110 was mixed-mode with a competitive-dominant character. This means that most inhibition occurs by impeding binding of the substrate to the enzyme, but that catalysis was also directly disrupted to some degree. Inhibition by the HIDs (compound #1) was also mixed mode with somewhat weaker competitive-dominant character. The inhibition mechanism by the HPDs (compound #208) was more difficult to define because the data did not fit the classical models as well did as the data for #1 and #110. The best interpretation of the data for #208 is that inhibition is probably non-competitive, indicating that catalysis is inhibited with little effect on substrate binding. Together, these data indicate that the three leading chemotypes of HBV RNaseH inhibitors work by somewhat different mechanisms. Example Lineweaver-Burke plots illustrating the inhibition patterns are in Fig. 4; intersection of the lines in the left half of the plots is indicative of mixed-mode inhibition, whereas an intersection on the Y axis is indicative of non-competitive inhibition. Substrate binding was found to be independent of Mg^{++} because varying Mg^{++} concentration had no effect on K_M of the reactions. These studies were not finalized because we lost confidence in the quality of the enzyme we produced, consequently they need to be confirmed with a more robust enzyme.

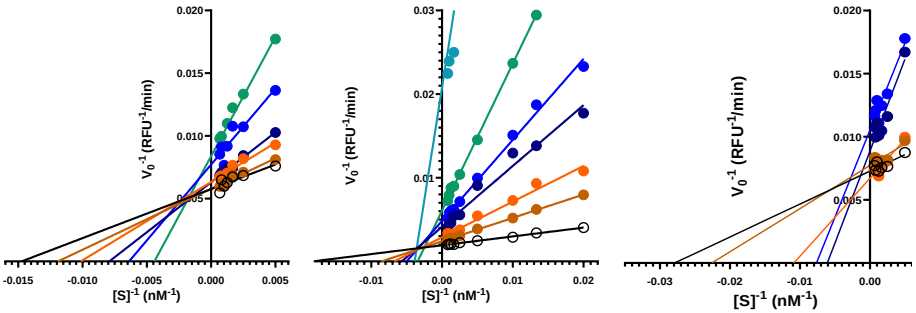

Fig. 4. Lineweaver-Burke plots of inhibition of the HBV RNaseH by representative αHT, HID, and HPD compounds

Task #10. We did not compare the sensitivity of the endo- and exo-nuclease activities of the HBV RNaseH as the enzyme quality precluded generating 50% inhibitory concentration (IC_{50}) values of sufficient rigor to permit the comparison to be made.

Task #11. One mutant HBV RNaseH was assessed for its effects on inhibition by the HBV RNaseH inhibitors. G707D was designed to help improve solubility of the enzyme and underwent preliminary enzymatic characterization. It increased inhibition by 3 of 17 compounds tested, decreased inhibition by 5 compounds, and had normal inhibition for 10 compounds. There was no correlation of compound class with the changes in inhibition. The remaining 14 mutant HBV RNaseHs we created have not been analyzed because we lost confidence in the quality of the enzyme we could produce. This project will be redesigned once a reliable enzyme source is available.

Work done outside the approved Scope of Work.

The unanticipated difficulty in producing robust HBV RNaseH enzyme led us to focus on two new projects as described below under “Changes in approach and reasons for change”. Results from these new directions are:

Inhibition mechanism of human RNaseH 1 (Extension of Aim 2). The human enzyme most similar to the HBV RNaseH is ribonuclease H1 (huRH1). Consequently, we substituted huRH1 for the HBV RNaseH in the proposed studies. These data are valuable because antiviral studies employing similar chemotypes of inhibitors are ongoing against both the HIV and HBV RNaseHs, and inhibiting huRH1 would be a major source

of toxicity for RNaseH inhibitors targeting both viruses. In essence, we used huRH1 as a surrogate for the HBV RNaseH to advance as far as possible towards the goals enumerated in the Statement of Work.

These studies revealed that α-hydroxytropolones are noncompetitive inhibitors of huRH1 that bind to the active site and modulate substrate binding (see Fig. 2 In Ponzar et al., *Journal of Biological Chemistry* **298**:101790). This is an unusual inhibition pattern given the compounds' reversible binding to the active site, so we concluded that traditional noncompetitive and mixed inhibition mechanisms are unlikely. Instead, we propose a model in which, by binding to the active site, αHTs stabilize an inactive enzyme–substrate–inhibitor complex to block enzyme function (see Fig. 7 in Ponzar et al.). Computational docking the inhibitors into huRH1 active site revealed their binding poses, which will provide clear guidance when engineering the compound series to be selective for the viral RNaseHs (see Fig. 6 in Ponzar et al.). Details of this project were published in Ponzar et al., *Journal of Biological Chemistry* **298**:101790 and are provided in **Appendix 1** to this report.

Predicting and validating the structure of the HBV polymerase protein, including its RNaseH domain (De Novo studies to explain why we had such difficulty in Aims 1 and 2). We took advantage of the major advance in *ab initio* protein structure prediction provided by the program AlphaFold to predict the structure of the full four-domain HBV polymerase protein, of which the RNaseH is the C-terminal domain. We then extensively validated the model using a wide range of computational approaches and analyses employing historical wet-bench data of the enzyme's function. These studies revealed that the fold we predicted represented the initial conformation the polymerase adopts following its translation, before host HSP90 chaperones activate it for specific binding to the HBV pregenomic RNA. The most striking feature of the model was that the polymerase's N-terminal domain that primes DNA synthesis wraps around the catalytic core of the enzyme, placing the tyrosine residue that primes HBV DNA synthesis on a flexible loop over the DNA polymerase active site (See Fig. 4 in Tajwar et al., *Protein Science* **31**:e44210. This structure has clear mechanistic implications on polymerase function that will be pursued in a project for which we are seeking NIH R01 funding. It has also launched three new drug discovery efforts to identify non-nucleoside reverse transcriptase inhibitors (NNRTI) analogous to drugs that are analogous to HIV NNRTI drugs that are a backbone of HIV therapy. The drug discovery efforts have already garnered philanthropic seed funding. Details of this work were published in Tajwar et al., *Protein Science* **31**:e4421 and are provided in **Appendix 2** to this report.

The model also revealed why it is so difficult to express the HBV RNaseH as an active recombinant enzyme. The RNaseH active site has a "D-E-D-D" motif of acidic residues that chelate the two Mg^{++} ions that promote RNA cleavage. In all other RNaseHs, the last "D" residue is supported on an α-helix, but the analogous region is an unstructured loop in the HBV enzyme. This indicates that that region of the RNaseH must be held in its proper position by something else, likely the cellular chaperones that are bound to the polymerase. Mutating the prolines in the loop that prevent formation of the α-helix yields consistently active enzyme without altering the predicted position of the last "D" residue in the active site or altering how HBV RNaseH inhibitors dock into the RNaseH active site. We are defining the optimal mutant enzymes to pursue, and we will commence with the biochemical analyses of the HBV RNaseH that had been proposed in this project once the optimal construct is defined.

What opportunities for training and professional development has the project provided? This project supported the Ph.D. studies of my student Dr. Nathan Ponzar, post-doctoral studies by Drs. Juan Villa Torricella, Mafuza Akter, and Razia Tajwar, and undergraduate training of Ms. Jala Chalichama, Ms. Alaina Knier, Ms. Cassie Kukla, and Mr. Michael Lindt.

How were the results disseminated to communities of interest? Results were disseminated by presentations at meetings, seminars given by Dr. Tavis, and publications. See details below in the Products section of this report.

What do you plan to do during the next reporting period to accomplish the goals? Nothing to report.

Impact

What was the impact on the development of the principal discipline(s) of the project? The primary discipline of this work is HBV polymerase enzymology. Determining the structure of the enzyme is a major breakthrough

because technical barriers prevent traditional structural biology assessment of the enzyme. It vastly advances the sophistication with which mechanistic studies and drug discovery efforts can be conducted with the polymerase.

What was the impact on other disciplines? Secondary disciplines are huRH1 enzymology and inhibition of other viral RNaseHs. The impact is strong on huRH1 enzymology as our efforts greatly simplified the protein purification protocol for huRH1 and revealed the mechanism by which active site inhibitors block enzyme function. Implications from understanding how the α-hydroxytropolones inhibit huRH1 extend to the HIV and HBV RNaseHs as the enzymes share the same catalytic mechanism, so the inhibition mechanism of huRH1 can be extrapolated to the HBV and HIV RNaseHs.

What was the impact on technology transfer? The supplementary data in our *Protein Sciences* paper (**Appendix 2**) includes links to PDB files for the molecular models for polymerases from all nine of HBV's genotypes, homologs from five animal hepadnaviruses, and three relevant non-hepadnaviral polymerase proteins. All models have associated annotation data so that researchers can readily identify the various domains and active sites of the polymerase proteins. This will facilitate research on a wide range of viral and non-viral reverse transcriptases.

What was the impact on society beyond science and technology? Identification of the HBV polymerase structure has launched three new drug discovery efforts seeking NNRTI inhibitors. Successful development of the inhibitors into anti-HBV drugs will greatly improve health of people worldwide who are suffering from chronic HBV infections, including many of America's veterans, their dependents, and other loved ones.

Changes/Problems

Changes in approach and reasons for change. The HBV RNaseH is extremely difficult to produce as an active recombinant enzyme, and it cannot be studied biochemically in its native context. We felt we had these problems resolved when the grant was awarded, but the initial biochemical experiments revealed unacceptably high inter-assay and inter-protein preparation variability. Some of this was determined to be due to nucleic acid contamination in the preparations, but removing the contaminating nucleic acids was insufficient to yield enzyme suitable for the proposed studies despite intense optimization efforts. We consequently refocused our work onto two other major efforts, determining the mechanism of inhibition of huRH1 and structural prediction for the HBV polymerase protein. Results from these studies are described above under Accomplishments.

Actual or anticipated problems or delays and action or plans to resolve them. We faced three major sources of difficulty in this project. The first is the unanticipated variability in the quality of the HBV RNaseH that prevented progress as described above. The second was frustration of two postdoctoral scholars assigned to this project, Drs. Juan Villa Torrecilla and Mafuza Akter, that stemmed from the difficulties in producing enzyme of adequate quality. This led Drs. Villa Torrecilla and Akter to abandon the project and caused delays in the research. Finally, the COVID pandemic greatly slowed research for about 9 months and had lingering effects that have not yet fully resolved themselves due to supply chain disruptions and sporadic infections among Tavis lab personnel. Plans to resolve these issues are not applicable as this is a final report.

Changes that had a significant impact on expenditures. The staffing turmoil and the COVID-induced progress delays reduced the rate of expenditures on this project. These were managed as best possible, but it was impossible to fully compensate for the resulting inefficiencies.

Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents. Not applicable.

Products

Publications.

Journal publications.

Edwards, T.C., Ponzar, N.L., and Tavis, J.E. (2019). Shedding light on RNaseH: a promising target for hepatitis B virus (HBV). *Expert Opinion in Therapeutic Targets* **23:**559-563.

Ponzar, N.L., Tajwar, R., Pozzi, N., and Tavis, J.E. (2022). Alpha-hydroxytropolones are noncompetitive inhibitors of human RNase H1 that bind to the active site and modulate substrate binding. *J. Biol. Chem.* **298:**101790.

Tajwar, R., Bradley, D.P., Ponzar N.P., and Tavis, J.E. (2022). Predicted structure of the hepatitis B virus polymerase reveals an ancient conserved protein fold. *Protein Science* **31:**e4221.

Books or other non-periodical, one-time publications.

Clark D.N., Tajwar, R., Hu J., and Tavis, J.E. The Hepatitis B Virus Polymerase. (2021). *Viral Replication Enzymes and their Inhibitors, Vol. 50*. Cameron C.E. and Arnold, J.J. eds. Chapter 6, pg. 195-226. Academic Press, London.

Other publications, conference papers, and presentations.

Tavis, J.E. (September 3, 2021). *Webinar*. Inhibiting the HBV Ribonuclease H as an Antiviral and Antitumor Strategy. Brooklyn College of the City University of New York.

Tavis, J.E. (October 29, 2021). *Invited speaker*. The Hepatitis B Virus Ribonuclease as a Drug Target and Biochemical White Whale. Indiana University, Bloomington IN.

Ponzar, N., Kukla, C, Akter, M, Pozzi, and Tavis, J.E. (July 21, 2019). HBV Ribonuclease H and Mechanisms of Inhibition. (Oral). American Society for Virology Annual Meeting, Minneapolis, MN.

Ponzar NP, Kukla C, Akter M, Pozzi N, Tavis J.E. (Oct. 3, 2019). Inhibitors of the HBV ribonuclease H act by multiple Mg^{2+}-dependent mechanisms. (Poster). Molecular Biology of the Hepatitis B Viruses 2019, Melbourne, Australia.

Ponzar, N. P., Tajwar, R., and Tavis, J.E. (September 23, 2021). Complex inhibition characteristics of HBV and HIV ribonuclease H inhibitors against human ribonuclease H1. (Oral). 5th Innovative Approaches for Antiviral Agents Summer School. Virtual meeting.

Tajwar, R., Bradley, D.P., Ponzar, N.L., and Tavis, J.E. (March 23, 2022). Predicted Structure of the Hepatitis B Virus Polymerase Reveals an Ancient Conserved Protein Fold. (Poster). International Conference on Antiviral Research, Seattle, WA.

Tajwar, R., Bradley, D.P., Ponzar, N.L., and Tavis, J.E. (September 7, 2022). Predicted Structure of the Hepatitis B Virus Polymerase Reveals an Ancient Conserved Protein Fold. (Oral). International RNaseH Meeting. Cambridge, MD.

Tajwar, R., Bradley, D.P., Ponzar, N.L., Fallahee, I., and Tavis, J.E. (September 19, 2022). Predicted Structure of the Hepatitis B Virus Polymerase Reveals an Ancient Conserved Protein Fold (Oral). International HBV Meeting, Paris, France.

Website(s) or other internet site(s). None.

Technologies or techniques. The protein purification protocol for huRH1 was greatly simplified as described in **Appendix 1**, which makes work with this enzyme accessible to a wider range of labs around the world. A manuscript describing this work is in preparation.

Inventions, patent applications, and/or licenses. None.

Other products. PDB files for the molecular models and associated annotation for polymerases from all nine of HBV's genotypes, homologs from five animal hepadnaviruses, and three relevant non-hepadnaviral

polymerase proteins have been made freely available to the research community through the supplementary data associated with the paper in **Appendix 2**.

JBC RESEARCH ARTICLE

Alpha-hydroxytropolones are noncompetitive inhibitors of human RNase H1 that bind to the active site and modulate substrate binding

Received for publication, December 12, 2021, and in revised form, February 24, 2022 Published, Papers in Press, March 3, 2022,
https://doi.org/10.1016/j.jbc.2022.101790

Nathan L. Ponzar[1,2], **Razia Tajwar**[1,2], **Nicola Pozzi**[2,3], **and John E. Tavis**[1,2,4,*]

From the [1]*Department of Molecular Microbiology and Immunology,* [2]*Institute for Drug and Biotherapeutic Innovation, and* [3]*Department of Biochemistry and Molecular Biology, Saint Louis University, Saint Louis, Missouri USA;* [4]*Liver Center, Saint Louis, Missouri USA*

Edited by Craig Cameron

The ribonucleases H (RNases H) of HIV and hepatitis B virus are type 1 RNases H that are promising drug targets because inhibiting their activity blocks viral replication. Eukaryotic ribonuclease H1 (RNase H1) is an essential protein and a probable off-target enzyme for viral RNase H inhibitors. α-hydroxytropolones (αHTs) are a class of anti-RNase H inhibitors that can inhibit the HIV, hepatitis B virus, and human RNases H1; however, it is unclear how these inhibitors could be developed to distinguish between these enzymes. To accelerate the development of selective RNase H inhibitors, we performed biochemical and kinetic studies on the human enzyme, which was recombinantly expressed in *Escherichia coli*. Size-exclusion chromatography showed that free RNase H1 is monomeric and forms a 2:1 complex with a substrate of 12 bp. FRET heteroduplex cleavage assays were used to test inhibition of RNase H1 in steady-state kinetics by two structurally diverse αHTs, 110 and 404. We determined that turnover rate was reduced, but inhibition was not competitive with substrate, despite inhibitor binding to the active site. Given the compounds' reversible binding to the active site, we concluded that traditional noncompetitive and mixed inhibition mechanisms are unlikely. Instead, we propose a model in which, by binding to the active site, αHTs stabilize an inactive enzyme–substrate–inhibitor complex. This new model clarifies the mechanism of action of αHTs against RNase H1 and will aid the development of RNase H inhibitors selective for the viral enzymes.

Ribonuclease H1 (RNase H1) is a metallonuclease specific for RNA in RNA–DNA hybrids. It plays key roles in nucleic acid metabolism in the nucleus and mitochondria, where it processes R-loops and the RNA primer for mitochondrial replication (1). RNase H1 knockout is embryonically lethal in mice because of failure of the mitochondria to replicate (2).

RNase H1 requires two divalent metal ions, usually Mg^{2+} (Mg_A^{2+} and Mg_B^{2+}) to cleave RNA–DNA hybrids (3, 4). The divalent cations are bound to the enzyme adjacent to one another by the four carboxylate residues comprising the "aspartic acid, glutamic acid, aspartic acid, aspartic acid (DEDD) motif" motif in the active site (Fig. 1*A*). The Mg_A^{2+} ion is essential for recruitment, activation, and positioning of the nucleophilic water molecule that attacks the phosphodiester bond of the substrate, whereas Mg_B^{2+} stimulates product formation. Structural studies with *Bacillus halodurans* RNase H1 demonstrate that the Mg_A^{2+}-binding site is only occupied appreciably in the enzyme–substrate (ES) complex (5).

The overall protein folds and active sites of RNases H1 are well conserved in all branches of life, yet there are significant differences between the RNases H of prokaryotes, eukaryotes, and viruses. While most prokaryotic RNases H contain solely the RNase H domain, the eukaryotic counterpart carries an additional domain, called the RNA–DNA hybrid-binding domain (HBD) (3). The HBD is connected to the RNase H domain by a linker region (Fig. 1*B*). While RNA hydrolysis is catalyzed by the RNase H domain, the HBD enhances RNase H1 substrate-binding affinity and processivity and imparts dimerization upon substrate binding (6).

HIV and hepatitis B virus (HBV) replicate by reverse transcription. They encode homologous type 1 RNase H domains that are necessary for viral replication because they digest the viral genomic RNAs after they are copied into the first DNA strand to permit synthesis of the second DNA strand. Thus, targeting these viral RNases H is a promising therapeutic strategy (7). A potential problem with this strategy is off-target effects. Most inhibitors, including those developed by us targeting HBV (8, 9), could potentially bind to and inhibit human RNase H1, limiting therapeutic use. This is particularly important in the context of viruses that cause chronic infections, such as HIV and HBV, insofar as patients will require long-term treatment.

The α-hydroxytropolone (αHT) compound class is a promising scaffold for development of anti-HIV and anti-HBV RNase H inhibitors, many of which have nanomolar efficacy against the viral RNases H *in vitro* and/or against viral replication (9, 10) with varying degrees of selectivity for the viral enzymes over human RNase H1. Here, we adapted expression of catalytically active human RNase H1 in *Escherichia coli* and determined the mechanism of inhibition of two αHT compounds. In kinetic assays, the inhibitors display kinetic features

* For correspondence: John E. Tavis, john.tavis@health.slu.edu.

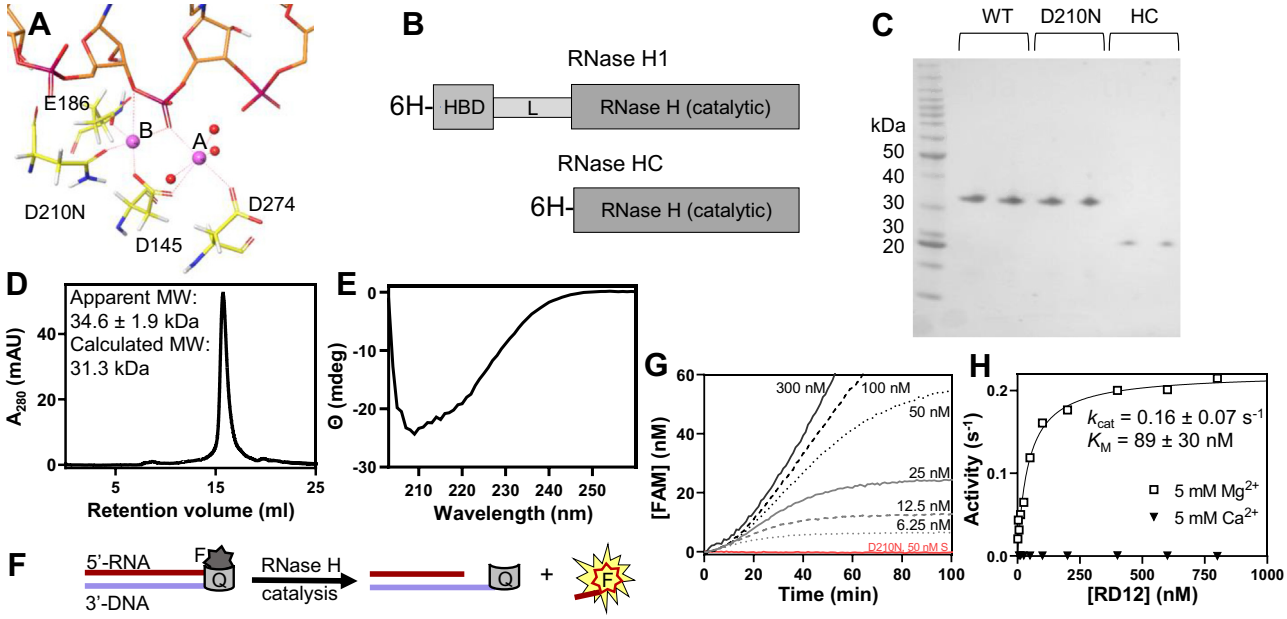

Figure 1. RNase H1 characterization. *A*, RNase H1 active site (PDB: 2QKK) (10) with Mg^{2+} ions *A* and *B* as *blue spheres*, water as *red spheres*, DEDD residues are in *yellow*, and the RNA backbone is in *orange* and *red*. *B*, recombinant gene structures. *C*, SDS-PAGE gel of purified proteins. About 1 μg of protein from two preparations of each protein was electrophoresed in a 4 to 16% gradient gel. Lane 1 is the BenchMark molecular weight marker. *D*, SEC chromatogram of 0.4 ml of 20 μM RNase H1. *E*, CD spectrum of 666 nM RNase H1. *F*, FRET heteroduplex cleavage assay. The *dark star* is quenched fluorescein (*F*), and the *cylinder* is the quencher (Q). Upon cleavage near the 3′ RNA terminus, the fluorescein-labeled fragment is released from proximity to the quencher and fluorescence increases (*yellow star*). *G*, example of RNase H1 (1 nM) reaction progress curves showing concentration of fluorescein (FAM) liberated from quencher over time. Substrate concentrations are indicated. The signal of catalytically inactive mutant RNase $H1^{D210N}$ (50 nM) is at the bottom in *red*. *H*, representative Michaelis–Menten plot for RNase H1 and RD12 fit with the Michaelis–Menten equation $\left(v_0 = V_{max}[S]/K_M+[S]\right)$ (15, 16). Parameters are reported as the mean ± SD of at least three independent measurements. PDB, Protein Data Bank; RNase H1, ribonuclease H1.

reminiscent of noncompetitive inhibition, suggesting that binding of substrate and compounds is not mutually exclusive. Since the compounds bind to the active site, we propose a model in which, by binding to the active site, αHTs stabilize an inactive enzyme–substrate–inhibitor (ESI) complex.

Results

Purification and characterization of human RNase H1

Human RNase H1 (UniProtKB: O60930, amino acids 27–286), a catalytically inactive mutant RNase $H1^{D210N}$, and human RNase HC (amino acids 136–286) containing only the catalytic domain were cloned into a pET-15b vector and expressed in *E. coli* (Fig. 1*B*) (11). All proteins carry an N-terminal histidine tag and thrombin cleavage site. Cells carrying this plasmid exhibited slowed growth, leaky protein expression, and proteolysis of the poorly expressed RNase H1. Addition of 1% glucose to all media resolved these problems without interfering with IPTG-induced expression. Ni^{2+}-affinity purification from cells grown in this manner resulted in protein of >95% purity as determined by SDS-PAGE (Fig. 1*C*) that was free of nucleic acid contamination (absorbance at 260 nm/280 nm = 0.55–0.60). The high purity from one-step purification is due to the proteins' atypically strong binding to the Ni^{2+}–nitrilotriacetic acid resin, which allowed us to add 125 mM imidazole to the wash buffer, greatly reducing nonspecific binding. Size-exclusion chromatography (SEC) revealed a single elution peak corresponding to monomeric RNase H1 with an apparent molecular weight (MW) of 34.6 kDa ± 1.9 kDa (Fig. 1*D*; calculated MW = 31.4 kDa). The yield was ∼2.5 mg per liter of cell culture.

To determine if recombinant RNase H1 is properly folded, we next performed CD experiments in the far-UV range (Fig. 1*E*). The strong negative ellipticity obtained for the human enzyme was similar, yet not identical, to that of the homologous and well-characterized *E. coli* RNase H (12–14). This is consistent with RNase H1 being properly folded and containing ordered secondary structural elements. The difference between *E. coli* and human RNase H1 spectra is likely because the *E. coli* enzyme lacks the unstructured linker and HBD.

Finally, we performed a fluorogenic assay to evaluate the catalytic activity of the recombinant enzyme (Fig. 1*F*). Briefly, we incubated an RNA–DNA hybrid of 12 bp (RD12) labeled with fluorescein at the 3′ RNA terminus and a quencher at the 5′ DNA terminus with the enzyme before adding 5 mM Mg^{2+}. Addition of Mg^{2+} to a solution of 1 nM enzyme and substrate, but not substrate alone, resulted in a significant increase in fluorescence, consistent with separation of the two strands mediated by the enzyme (Fig. 1*G*). Importantly, enzyme activity toward the substrate obeyed Michaelis–Menten kinetics (Fig. 1*H*) (15, 16), implying saturation of enzyme at high substrate concentrations. While the presence of the N-terminal histidine tag did not affect catalytic activity of the enzyme, activity was specific to the presence of Mg^{2+} and residue D210,

since Ca^{2+} did not support catalysis and mutation of D210 to N abrogated its catalytic activity (Fig. 1, *G* and *H*). These data indicate that the recombinant human RNase H1 expressed in *E. coli* and is suitable for biochemical studies.

Inhibition profile of compounds 110 and 404 against human RNase H1

αHTs are RNase H inhibitors that bind to the active site *via* coordination of the divalent metal cofactors. This was determined by cocrystallization of HIV RNase H and the αHT β-thujaplicinol (Fig. 2*A*) (10), which inhibits both HIV and HBV RNase H (17). We focused on two compounds, 110 and 404 (Fig. 2*A*) that are structurally similar to β-thujaplicinol, yet diverse enough to assess the generality of the inhibition mechanism.

To confirm that 110 and 404 bind to metal ions in solution, we performed binding experiments using absorption spectroscopy to monitor complex formation between Mg^{2+} and Ca^{2+} with the compounds (Fig. S1, *A–D*). Mg^{2+} and Ca^{2+} altered the absorption spectra of both compounds substantially, and we determined the compounds' stability constants, *K*, which are related to the dissociation constant (K_D) by the relationship $\log K = -\log K_D$ (Table S1) (18). Two distinct equilibria were observed corresponding to low-affinity and high-affinity binding events, consistent with the compounds' dual metal-chelation motifs (Fig. S1, *E* and *F*). The high-affinity stability constant of Mg^{2+} for 110 ($\log K_1$) was 3.55 (K_D = 0.2 mM), whereas the low-affinity stability constant ($\log K_2$) was 1.61 (K_D = 30 mM). The high-affinity constant is similar to that of the one divalent cation-coordinating dihydroxytropylium ion ($\log K$ = 3.82; K_D = 0.15 mM) (19). Stability constants for 404 were similar to those of 110 ($\log K_1$ = 3.9 and $\log K_2$ = 1.7).

Next, to characterize the inhibitory profiles of the compounds against RNase H1, we titrated the compound and substrate against the enzyme in steady-state kinetics experiments. Kinetic and inhibition parameters are shown in

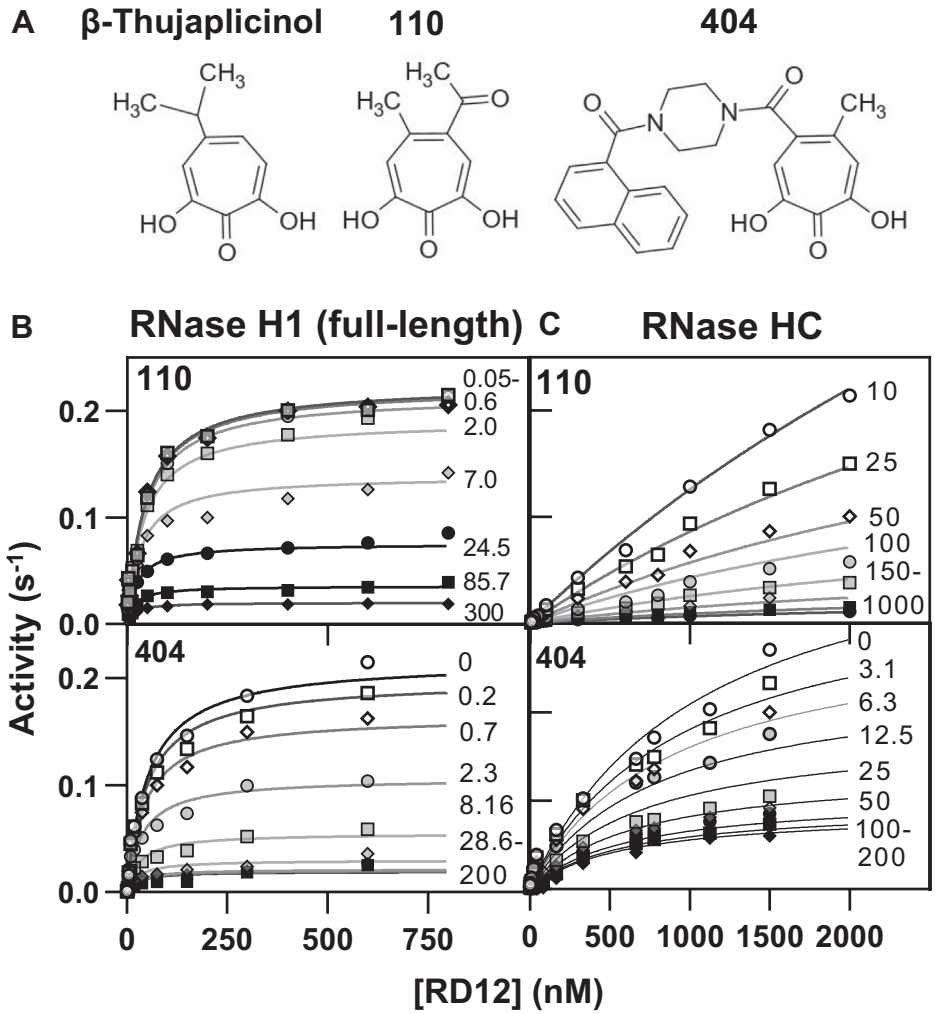

Figure 2. Inhibition kinetics of RNase H1 cleavage of RD12 substrate by compounds 110 and 404. *A*, structures of β-thujaplicinol and compounds 110 and 404. *B* and *C*, representative Michaelis–Menten plots of RNase H1 (*B*) and RNase HC (*C*) in the presence of compound 110 (*top panels*) or compound 404 (*bottom panels*). Data were fit with Equations 1–3 (20). Numbers to the *right* of the regression lines indicate inhibitor concentrations in micromolar. Parameters are shown in Table 1. Plots are representative of three or more independent experiments with additional data in Figs. S2–S5. RD12, RNA–DNA hybrid of 12 bp; RNase H1, ribonuclease H1.

Table 1
RNase H1 kinetic/inhibition parameters

Parameter	RNase H1-wt	RNase HC
k_{cat} (s^{-1})	0.16 ± 0.07	0.22 ± 0.04
K_M (nM)	89 ± 30	>500[a]
k_{cat}/K_M (M^{-1} s^{-1})	1,800,000	440,000
110 K_{iE} (μM)	17.9 ± 11	38 ± 12
110 α	0.61 ± 0.29	Unstable
110 K_{iES} (μM)	8.5 ± 0.0003	Unstable
110 I_{max} (%)	95 ± 2	∼100
404 K_{iE} (μM)	9.3 ± 8	∼15
404 α	0.39 ± 0.17	Unstable
404 K_{iES} (μM)	2.6 ± 0.5	Unstable
404 I_{max} (%)	94 ± 2	∼100

[a] Because of inability to saturate the RNase HC reaction, k_{cat} is highly extrapolated and K_M could not be reliably determined. Data for alternate substrates in Fig. S5. K_{iE} is the K_i of the inhibitor for the free enzyme and K_{iES} is the K_i for the enzyme–substrate complex. α is the proportionality constant for the two inhibition constants. α = K_{iES}/K_{iE} (21).

Table 1. Data were fit with Equations 1–3 (20). Both compounds showed no evidence of competitive inhibition (Figs. 2*B* and S2, *J* and *K* and Table S2). They instead reduced the turnover (V_{max}) while minimally enhancing the apparent affinity for the substrate (K_M) (Figs. 2*B* and S2 and Table S2). Importantly, saturating concentrations of 110 and 404 could not fully inhibit RNase H1 activity (Figs. 2*B* and S3, *A* and *B* and Table S2), both with maximum fractional inhibition (I_{max}) values of ∼95% (Table 1). Inhibition was reversible, as documented by recovery of enzymatic activity after dilution of the compound (Fig. S3, *E* and *F*) (21). These results argue for a reversible and noncompetitive mechanism of inhibition whereby binding of the substrate and compounds are not mutually exclusive.

Determining the impact of the HBD on inhibition

Human RNase H1 contains an HBD (Fig. 1*B*) that imparts high-affinity substrate binding (6, 22). In addition, binding of the substrate to the HBD could allosterically modulate active-site accessibility. To determine the impact of the HBD on inhibition, we assessed the compounds' inhibition profiles against a mutant enzyme that contains only the catalytic domain, referred to as RNase HC (11). Compounds 110 and 404 inhibited RNase HC in a dose-dependent manner (Figs. 2*C*, S3, *C* and *D* and S4). This confirms that the binding site is in the catalytic domain. Because of the much lower affinity of RNase HC for RD12, we could not saturate the reaction to accurately determine K_M and k_{cat}; however, high 110 and 404 concentrations reduced the K_M to within the measurable range, allowing us to fit the data to a global inhibition model (Equations 1–3) (Figs. 2*C* and S4). Like full-length RNase H1, RNase HC was inhibited in a noncompetitive manner. To further validate this observation, we repeated the same assays with 14-mer and 18-mer hybrid substrates for which RNase HC has a lower K_M. We observed clear noncompetitive inhibition with these substrates (Fig. S5). Assays with 404 using 14-mer (Fig. S5, *A*, *F*, and *G*) and 18-mer (Fig. S5, *B*, *C*, *F*, *H*, and *I*) heteroduplex substrates showed reductions of K_M, whereas the effect of compound 110 on K_M was unclear (Fig. S5, *D*, *F*, and *J*). Thus, both compounds inhibited turnover of RNase HC. This indicates that binding of substrate and compound to the catalytic domain is not mutually exclusive.

Two molecules of RNase H1 bind to one molecule of RD12 heteroduplex substrate

Next, we wanted to determine the effect of the compounds on substrate binding. First, we characterized the binding of RNase H1 to RD12 substrate in the absence of inhibitor. To determine the stoichiometry of human RNase H1, we performed SEC of 20 μM RNase H1 preincubated with 5, 10, and 15 μM of RD12 (Fig. 3*A*) in the presence of Ca^{2+} instead of Mg^{2+} to prevent degradation of the substrate (Fig. 1*H*). In all treatments, we observed a major peak with an apparent MW of 68 kDa, which corresponds to a 2:1 enzyme–substrate complex (E_2S; calculated MW = 71 kDa) (Figs. 3*A* and S6). In the 5 μM RD12/20 μM RNase H1 sample, there were two prominent peaks corresponding to the free enzyme and the E_2S complex.

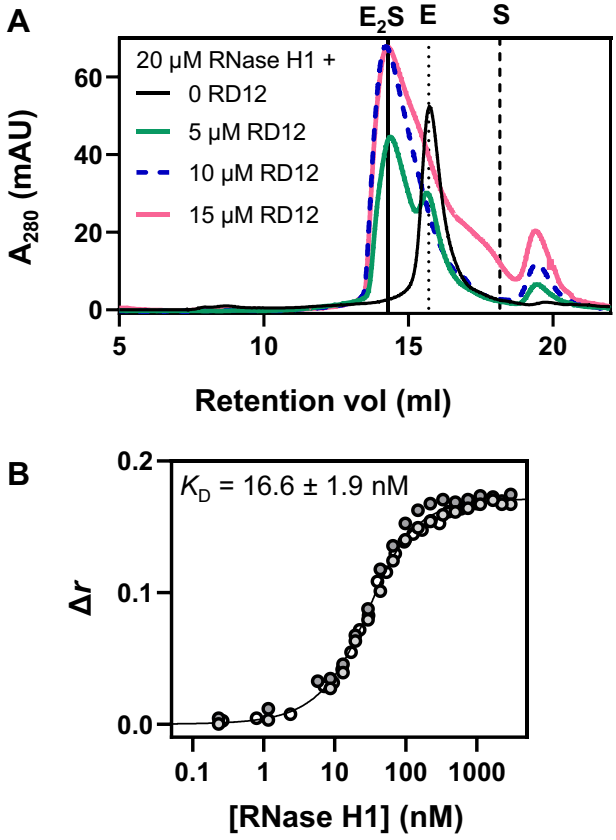

Figure 3. Characterization of the RNase H1–RD12 ES complex. *A*, size-exclusion chromatograms of RD12 substrate titrated against 20 μM RNase H1 in a Superdex 200 increase column with 500 mM NaCl. Superdex 75 chromatograms with 100 mM NaCl and MW standards are shown in Fig. S6*A*. *Vertical lines* indicate retention volumes of the E_2S complex (*solid*; 14.3 ml), the free enzyme (*dotted*; 15.7 ml), and the free substrate (*dashed*; 18.2 ml). Chromatograms are representative of at least two independent experiments. *B*, fluorescence polarization assay of 12.5 nM fluorescein-labeled 12-bp heteroduplex substrate (RD12) titrated with RNase H1. *Gray-filled*, *black-filled*, and *white-filled circles* represent data from three independent experiments. Data were fit with Equation 4 (23). ES, enzyme–substrate; MW, molecular weight; RD12, RNA–DNA hybrid of 12 bp; RNase H1, ribonuclease H1.

Addition of 10 μM RD12 (putatively 20 μM of binding sites) increased the absorption of the E_2S peak and eliminated the free enzyme peak. Finally, addition of 15 μM RD12 did not increase absorption in the E_2S peak, but the 5 μM excess eluted as a shoulder off the side of the ES peak rather than as free substrate, implying that binding is dynamic on the time-scale of the chromatographic run.

To determine substrate-binding affinity, complex formation between RNase H1 and RD12 substrate lacking a quencher on the DNA strand was monitored by fluorescence polarization (FP). Ca^{2+} was again used instead of Mg^{2+}. We titrated RNase H1 against 12.5 nM RD12 and fit the anisotropy data with the quadratic binding equation (Equation 4) (23), which accounts for ligand depletion, in this case RNase H1, when the receptor (RD12) concentration (R_{tot}) is near or above the K_D. R_{tot} was constrained to 25 nM because of the 2:1 stoichiometry (Fig. 3*A*) (6). This yielded a K_D of 16.6 ± 1.9 nM (Figs. 3*B* and S7, *A* and *B*). An approximately fivefold lower K_D relative to the K_M (Fig. 1*H*) is likely because of differences in assay conditions between the kinetics and binding experiments (*e.g.*, Mg^{2+} *versus* Ca^{2+}).

Finally, to independently confirm the stoichiometry of the complex determined by SEC, titrations were performed at 375 nM of RD12 substrate, which is saturating for the enzyme based on K_D = 16.6 nM. After fixing the K_D to 16.6 nM, the calculated concentration of binding sites was 724 nM, which, consistent with a 2:1 stoichiometry, is twice the concentration of the substrate (Fig. S7, *C* and *D*). Together, these results indicate that two molecules of RNase H1 bind one molecule of RD12 with high affinity.

Compounds 110 and 404 modulate substrate binding

Even though most of the inhibitory effects of the αHTs was on V_{max}, compounds 110 and 404 reduced K_M in steady-state kinetics (Figs. 2 and S2, S4 and S5). This is consistent with the observation that β-thujaplicinol binds to the HIV-1 reverse transcriptase (RT):substrate complex with higher affinity than to the free enzyme (10). However, if the assumption that k_{cat} << substrate release (k_{off}) is not true for the RNase H1 mechanism, then the modest reductions in K_M observed here could be due to inhibition of chemical cleavage, product release, or any other zero-order step in the mechanism. To directly assess the compounds' effects on substrate binding, we measured the binding of RNase H1 to substrate with FP as described previously (Fig. 3*B*) in the presence of saturating concentrations of 110 (200 μM), 404 (200 μM), or 2% dimethyl sulfoxide (DMSO) as a control. Compound 110 reduced the K_D approximately threefold, from 12.2 to 4.1 nM (Fig. 4), indicating enhanced affinity. By contrast, the presence of saturating concentrations of compound 404 (200 μM) increased the K_D value sevenfold from 12.2 to 84 nM, indicating loss of affinity (Fig. 4). This result is surprising as it is inconsistent with the K_M reduction observed with the treatment of compound 404 in the kinetics experiments (Figs. 2 and S2, S4 and S5). We suspect this is an artifact from the use of Ca^{2+} instead of Mg^{2+}. Hence, in the presence of Ca^{2+},

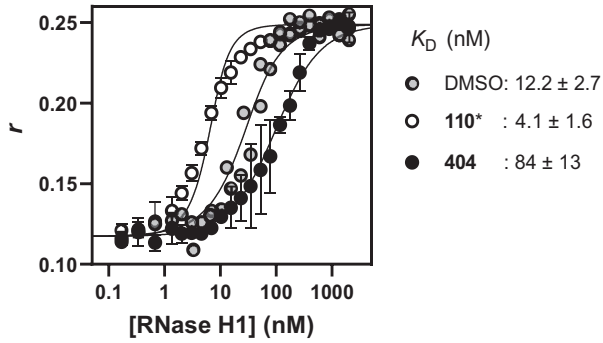

Figure 4. Impact of compounds 110 and 404 on RNase H1–substrate binding. Binding curves for RNase H1 against 12.5 nM RD12 are shown in the presence of 200 μM of compounds 110 and 404 or 2% DMSO. Data were fit with Equation 4 (23). K_D values are the mean and SD of at least four independent measurements. Error bars indicate SD of two independent experiments (more data are shown in Fig. S7). *The measured K_D of compound 110 is the upper limit of K_D. Additional data are shown in Fig. S7, *E* and *F*. DMSO, dimethyl sulfoxide; RD12, RNA–DNA hybrid of 12 bp; RNase H1, ribonuclease H1.

the two structurally similar compounds had opposite effects on substrate binding.

To further explore this unexpected observation, we evaluated the effect of saturating compounds 110 and 404 (500 μM) on the distribution of free RNase H1 and RNase H1 complexed with RD12 using SEC. Ca^{2+} was used to prevent catalysis as in FP assays. We adjusted conditions such that the complex partially dissociated (∼50%) during elution. This involved reducing the enzyme and substrate concentrations to 2.5 and 1.25 μM, respectively. In 100 mM NaCl, the complex did not dissociate during elution, so 500 mM NaCl was added to the sample and elution buffers, which resulted in ∼50% dissociation of enzyme and substrate. Unexpectedly, when scouting conditions to cause partial ES dissociation during SEC, the peak of the complex shifted to an apparent MW of 43 kDa (Figs. 5 and S6, *B–D*) upon reduction of enzyme and substrate concentrations to or below 5 and 2.5 μM, respectively, corresponding closely to the calculated MW of a 1:1 RNase H1–RD12 complex (calculated MW = 39.5 kDa). This could be due to preference of a 1:1 complex at subsaturating concentrations.

To monitor the substrate's elution profile, we measured fluorescein's absorption (absorbance at 485 nm) from the labeled RD12 substrate simultaneously with absorbance at 280 nm, which detects both enzyme and substrate. The enzyme and substrate were preincubated with or without 500 μM of compound 110 or compound 404, and then the complexes were resolved by SEC. We observed a peak in both absorbances at 280 and 485 nm corresponding to the ES complex and two additional peaks and shoulders of 485 and 280 nm, respectively. One of the additional peaks aligns just ahead of the free RD12 retention volume and one just after (Fig. 5*A*). The middle peak of the 485 nm absorbance trace likely comes from free substrate released from the ES complex as it was diluted in the column, and the last peak may arise from single-stranded oligonucleotides, which dissociated as the 12 bp heteroduplex was diluted in the column. Addition of 500 μM of compound 110 to the sample buffer shifted the distribution of free and bound substrate almost entirely to the bound form

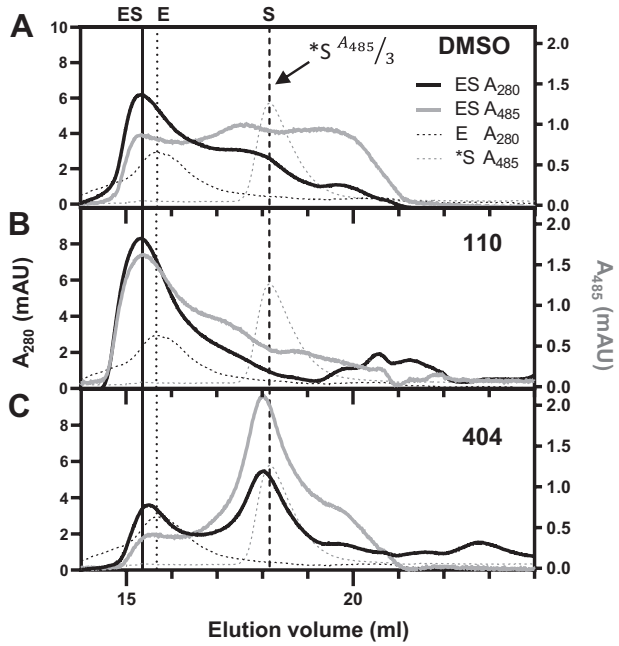

Figure 5. Size-exclusion chromatography (SEC) of ES complex pre-incubated with compound 110 or compound 404. *A*, SEC of 2.5 μM RNase H1 preincubated with 1.25 μM RD12 (ES). The absorbance at 280 nm is represented by *solid black lines*, and the absorbance at 485 nm (fluorescein) by *solid gray lines*. Both absorbances are measured simultaneously. Reference chromatograms of 2.5 μM enzyme (E; *dashed black*; absorbance at 280 nm) and 1.25 μM substrate (S; *dashed gray*; absorbance at 485 nm) assayed independently are shown. *Absorbance at 485 nm signal of this substrate-only control is expressed as absorbance at 485 nm divided by 3 to maintain scale. *B*, SEC of ES as in (*A*) with 500 μM compound 110 or (*C*) 500 μM compound 404. All buffers contain 2% DMSO. *Vertical lines* indicate the retention volumes of the putative 1:1 ES complex (*solid*, 15.3 ml), free RNase H1 (*dotted*; 15.7 ml), and free RD12 substrate (*dashed*; 18.2 ml). Chromatograms are representative of two or more experiments. DMSO, dimethyl sulfoxide; ES, enzyme–substrate; RD12, RNA–DNA hybrid of 12 bp; RNase H1, ribonuclease H1.

(Fig. 5*B*). The same experiment performed with 500 μM of compound 404 resulted in enhanced dissociation of the ES complex, indicated by reduced absorbance at 280 nm at the ES retention volume with the absorbance at 485 nm mostly shifted to the free-substrate position (Fig. 5*C*). Thus, results from SEC are consistent with the enhanced binding observed in the presence of compound 110 and inhibited binding in the presence of compound 404 (Fig. 4). These results support the finding that compound 110 enhances substrate binding in kinetics assays, and competition of compound 404 with substrate in binding assays is likely an artifact of using Ca^{2+} instead of Mg^{2+}.

Modeling of ESI ternary complexes with compounds 110 and 404

To understand how the compounds could produce the inhibition patterns and effects on substrate binding we observed, we employed induced-fit docking of the compounds to RNase H1 (Protein Data Bank [PDB] ID: 2QKK) using the Schrödinger software suite. We constrained our analyses to poses in which the compounds chelate both divalent ions *via* their oxygen trident because the metal-chelating binding mode is well established for the αHTs and RNase H (10). We observed multiple binding poses for compound 110 in the active site (Fig. S8*A*). The hydroxylated tropolone ring of compound 404 bound the Mg^{2+} ions in the active site in a single pose (Fig. S8*B*), but its large appendage docked along the substrate-binding groove outside the active site in different poses. The average predicted binding energy of all poses for compound 110 is −8.2 ± 0.6 kcal/mol and that of compound 404 is −9.2 ± 1 kcal/mol (Fig. S8). Superposition of the 14-mer RNA:DNA heteroduplex substrate from the original RNase H1-substrate cocrystal structure showed that the compounds and the RNA strand cannot simultaneously occupy the active site (Fig. 6), and that compound 404 much more substantially overlapped with the total substrate-binding interface than compound 110. This suggests that the substrate and/or enzyme change their conformation to accommodate the compounds in the active site while remaining bound *via* ES contacts that remain accessible. This seems to be the case with β-thujaplicinol and HIV RNase H (10), where the substrate is predicted to stay just above the active site, interacting with the compound. Compound 404 would require a larger accommodation on the part of the substrate. We suspect the binding poses could be substantially different in the presence of Ca^{2+}, with changes in the binding pose of compound 404 likely having greater effects than changes in the pose of compound 110, by virtue of its larger size.

Discussion

RNase H1 is increasingly being recognized as a central player in nuclear and mitochondrial genome maintenance and replication (1, 24). The consequences of RNase H1 knockout or loss-of-function mutations are severe (2, 25), and therefore, it is important that highly selective antiviral RNase H inhibitors be developed to avoid off-target human RNase H1 inhibition. In addition, αHTs and other metal-chelating compounds can inhibit other DEDD motif–containing viral nucleases, including pUL30/pUL42 of herpes simplex virus 1 and herpes simplex virus 2 (26), HIV integrase (27), the influenza and bunyavirus cap-snatching enzymes (28), and others. Therefore, understanding inhibition of RNase H1 by divalent metal-chelating compounds, including αHTs, will aid development of selective inhibitors of enzymes that share structural or enzymatic similarity with it.

Binding of RNA–DNA hybrid substrate by RNase H1

We showed that human RNase H1 binds to an RD12 with high affinity (K_D = 16.6; Fig. 3*B*) and 2:1 stoichiometry with SEC (Fig. 3*A*) and stoichiometric titration (Fig. S7, *C* and *D*). This is consistent with the 2:1 stoichiometry of the homologous murine RNase H1 (6), though the human enzyme may form 1:1 complexes in subsaturating enzyme and substrate concentrations (Figs. 5*A* and S6, *B–D*). Previous studies (6) show that the catalytic domain on its own binds substrate with lower affinity than the full-length protein, which is consistent with the elevated K_M of RNase HC observed here (Figs. 2*C*, S4

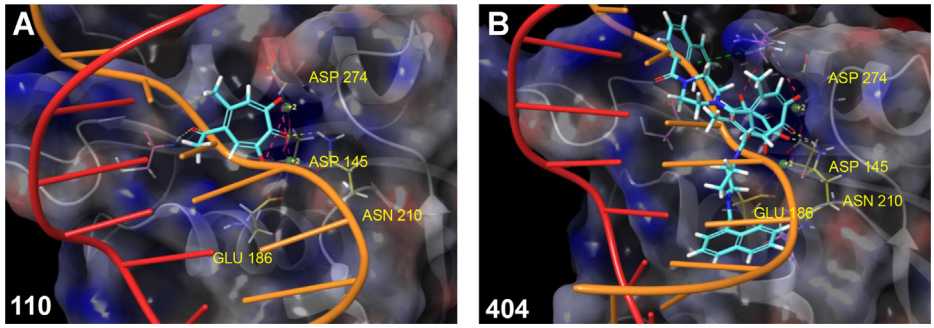

Figure 6. Docking of compound and substrate into the RNase H1 active site. *A* and *B*, induced-fit docking was used to dock compound 110 (*A*) and compound 404 (*B*) into the RNase H1 active with Mg^{2+} inserted where the Ca^{2+} ions were in the original structure (PDB ID: 2QKK). One pose of compound 110 and two poses of compound 404 are shown. The substrate from the original cocrystal structure was then superposed on the enzyme–inhibitor complex. DEDD residues are in *yellow* with D210 replaced with N. PDB, Protein Data Bank; RNase H1, ribonuclease H1.

and S5 and Table 1). RNase H1 contacts its substrate *via* the large binding interfaces of the RNase H domain and HBD (11, 22). Though no structures of the full eukaryotic RNase H1 exist, models have been proposed, most envisioning the enzyme as two balls connected by a string, corresponding to the globular RNase H domain and HBD and the putatively unstructured linker (Fig. 7) (22, 29). The HBD is thought to tightly anchor the enzyme to the substrate while the RNase H domain makes processive cleavages, repeatedly engaging and disengaging the substrate. Furthermore, the single-domain *E. coli* RNase H was recently shown to exhibit processivity (30), from which it follows that the RNase H active site must be empty for some time between processive cleavages while the rest of the enzyme remains bound. Thus, even in the context of the ES complex, there should be opportunities for the compounds to bind.

Compounds 110 and 404 are nontraditional noncompetitive inhibitors

Steady-state kinetics data for compounds 110 and 404 were inconsistent with competitive inhibition, showing strong reductions in V_{max} (Fig. 2). Both compounds also caused modest but significant reductions in K_M. Thus, a rigorous description of the kinetics data requires the general (mixed) model of inhibition (Equations 1–3) (20). In this model, α is the ratio of K_i for the ES complex (K_{iES}) to K_i for the free enzyme (K_{iE}) (21). A minimum V_{max} term was added in the modeling to account for incomplete inhibition in saturating compound and substrate concentrations (Equation 2). This model fits the data of both compounds significantly better than the simple competitive, uncompetitive, and noncompetitive models (Table S2 and Fig. S2, *J* and *K*), most often with $p < 0.01$. Data from RNase HC, which lacks the linker and HBD, also displayed noncompetitive inhibition, with both compounds reducing V_{max} (Figs. 2*C*, S4 and S5) and compound 404 also reducing K_M (Fig. S5, *A*–*C*). Finally, we ruled out irreversible inhibition (Fig. S3, *E* and *F*).

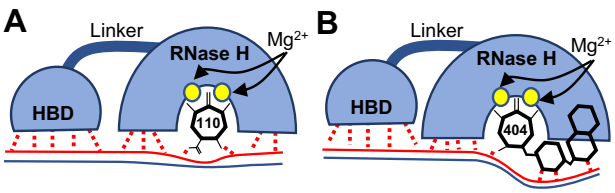

Figure 7. Model for the effects of compounds on substrate binding in the RNase H1 active site. *A* and *B*, models of compound 110 (*A*) and compound 404 (*B*) binding to the Mg^{2+}-bound enzyme–substrate complex. *Solid red line* = RNA, *solid blue line* = DNA, and *dashed red lines* = proposed interactions of substrate with enzyme and inhibitor within the ESI complex. ESI, enzyme–substrate–inhibitor; RNase H1, ribonuclease H1.

In noncompetitive inhibition, the compound may bind to the free enzyme and the ES complex. Noncompetitive kinetics have traditionally been seen as an evidence that the compound binds somewhere other than the active site, causing a conformational change in the enzyme that reduces the turnover rate and may also modulate substrate binding (21). However, active-site binding *via* metal chelation is well established for αHTs with HIV RNase H (10), and an intact two metal-chelating oxygen trident on the inhibitors is essential for their function against all RNases H (31–33). We demonstrated that compounds 110 and 404 bind Mg^{2+} and Ca^{2+} in solution using UV–visible absorption spectroscopy (Fig. S1 and Table S1). Finally, molecular modeling studies support the two-metal ion–bound active sites as the compound-binding site (Fig. S8). From these data, we conclude that compounds 110 and 404 do not work by the mechanism typically inferred from noncompetitive inhibition kinetics.

Instead, experimental data and docking results for both compounds are consistent with compounds binding in the active site as expected but forming stable and inactive ESI complexes rather than competing with the substrate (Fig. 7). Docking studies indicate that substrate and compound cannot bind in the active site at the same time (Fig. 6, (10)). With compound bound, the substrate may bind in an altered conformation, exchanging bonds with RNase H1 active-site residues and Mg^{2+} ions for bonds with the compound. This would account for the noncompetitive kinetics and the enhanced substrate binding. Compound 110 enhanced substrate binding in both kinetics and binding assays, whereas compound 404 was competitive in binding assays while reducing K_M in kinetics assays (Figs. 2, 4 and 5). The reason for this apparent discordant result is uncertain, though Occam's razor suggests it is related to the use of Mg^{2+} in kinetics assays and Ca^{2+} in binding assays. Ca^{2+} has a larger atomic radius

than Mg^{2+} (34), and because of this, the A-site Ca^{2+} ion is displaced from the position of the A-site Mg^{2+} ion, and the two Ca^{2+} ions are substantially further apart than are the two Mg^{2+} ions (35). Ca^{2+} is also more flexible in its number of coordination partners than Mg^{2+}. Finally, the stability constants of Ca^{2+} for compound 110 ($\log K_1$ = 2.56, $\log K_2$ = 0.35) were much lower than those of Mg^{2+} ($\log K_1$ = 3.46, $\log K_2$ = 1.61) (Table S1 and Fig. S1). This indicates differences in how the compounds interact with the two cations that may be relevant to their effects on substrate binding.

It is not clear why saturating concentrations of compound 110 or compound 404 did not cause 100% inhibition of catalysis (Figs. 2, S2 and S3 and Tables 1 and S2). One possibility is that there may be one or more step(s) that occur in assembly of the ESI complexes that have some probability of substrate cleavage. Mg^{2+} stimulates binding of αHTs to the RNase H type 1 active site (10), but the A-site Mg^{2+} ion ($Mg_A{}^{2+}$) does not bind efficiently in the absence of bound substrate (5). Thus, it may be necessary for the substrate to enter the active site to recruit $Mg_A{}^{2+}$ and then partially dissociate to create the two Mg^{2+} ion–bound active site to which the compounds bind most efficiently. After $Mg_A{}^{2+}$ recruitment, the substrate may be cleaved and released from the active site or released uncleaved. We speculate that the compounds may bind immediately after the substrate is released from the active site and form an ESI complex. However, if recruitment of $Mg_A{}^{2+}$ by substrate precedes compound binding to the ES complex, then there would be residual activity related directly to the chance of cleavage when $Mg_A{}^{2+}$ is recruited and inversely to the stability of the ESI complex.

Comparison of human RNase H1 inhibition to viral RNase H inhibition

The αHT β-thujaplicinol (Fig. 2*A*) inhibits HIV-1 RT–associated RNase H in a reversible noncompetitive manner (10), and the compound also exhibited improved binding in the presence of substrate. This is consistent with our finding that inhibition of αHTs 110 and 404 against RNase H1 is greatest in saturating substrate concentrations (*i.e.*, $K_{iES} < K_{iE}$) (Fig. S3, *A–D* and Table 1). The RNases H of both HIV and HBV are domains of larger polymerase proteins with much of their substrate-binding capacity located in their adjacent RT domains (36, 37), recapitulating through a different structural mechanism the overall situation observed with RNase H1. Thus, the mostly noncompetitive behavior of β-thujaplicinol against HIV RT–associated RNase H is because the small β-thujaplicinol molecule does not block a substantial fraction of the total ES contacts. While the mechanism of inhibition of HBV RNase H by αHTs has not been determined because of difficulties in producing suitable recombinant enzyme, the HBV enzyme may also be inhibited in a noncompetitive manner.

Experimental procedures

Protein expression and purification

Full-length RNase H1 (UniProtKB: O60930, amino acids 27–286) and a catalytic domain–only mutant (RNase HC; UniProtKB: O60930, amino acids 136–286) were purified by Ni^{2+}-affinity chromatography as before with modifications (11). Transformed *E. coli* (LOBSTR) were grown and induced in LB media with 1% glucose. 2-Mercaptoethanol was replaced with 1 mM Tris(2-carboxyethyl)phosphine (TCEP) in purification buffers. After Ni^{2+}–nitrilotriacetic acid purification, the buffer was exchanged *via* centrifugal desalting columns into storage buffer (50 mM Hepes, pH 7.5, 400 mM NaCl, 10% glycerol, and 2 mM TCEP) and stored at −80 °C. Proteins were quantified by UV spectroscopy with absorbance at 280 nm molar extinction coefficients of 47,440 M^{-1} cm^{-1} for full-length RNase H1 and 31,970 M^{-1} cm^{-1} for the catalytic domain–only mutant (RNase HC). These values were determined by ExPASy ProtParam using the amino acid sequences of RNase H1 and RNase HC.

CD

Three milliliters of 666 nM RNase H1 in CD buffer (6.25 mM Tris, pH 7.5, 100 mM NaF, 10% glycerol, and 2 mM TCEP) was read in a 10 mm × 10 mm quartz cuvette with an Applied Photophysics CD spectrophotometer at 20 °C. Four sets of 10 reads were taken and averaged. The buffer signal was read and subtracted from the protein sample's signal.

SEC of RNase H1 and ligands

RNase H1, substrate, and compound were run alone and in different combinations. All SEC experiments were conducted in a Superdex 200 Increase 10/300 GL column at room temperature, except in Fig. S6*A*, in which a Superdex 75 10/300 GL was used. The final composition of the elution buffer was 50 mM HEPES, 500 mM NaCl, 10% glycerol, and 2 mM TCEP. The pH was adjusted with NaOH to 7.5. NaCl was reduced to 100 mM in Fig. S6*A*. Buffer stocks were prepared and stored at room temperature without TCEP. A freshly thawed aliquot of HEPES-buffered TCEP (pH 7.5) was added to fresh buffer stock every 2 to 3 h. Each day, ≥40 ml fresh buffer was used to equilibrate the column. Runs were conducted at a flow rate of 0.5 ml/min. Samples were prepared and incubated at room temperature for 15 min prior to starting the run. The sample (0.4 ml) was loaded into the sample loop for each run. Concentrations of enzyme and substrate employed are indicated in the figures. In all experiments using inhibitors, 500 μM of the compound was added to the sample buffer and 2% of DMSO to vehicle controls. Apparent MWs were determined by resolving a set of Bio-Rad gel filtration standards in the same conditions. Average apparent MWs of elution peaks were determined from three or more independent experiments.

Heteroduplex substrate preparation

Fluorescein-labeled RNA oligonucleotides and complementary Iowa Black–labeled DNA oligonucleotides were purchased from Integrated DNA Technologies with HPLC purification. RNAs and DNAs were dissolved in nuclease-free water and combined in a RNA:DNA ratio of 1:1.1 in 50 mM HEPES, pH 7.5, 100 mM NaCl, and 2 mM TCEP. The

substrate solution was heated to 90 °C for 10 min and slowly cooled to 4 °C to promote annealing. For RD12, the RNA sequence is GACACCUGAGUC/36-FAM and the DNA sequence is 5IABkFQ/GACTCAGGTGTC. Additional sequences are shown in Fig. S5*E*.

Compound sourcing

RNase H inhibitors were synthesized by Dr Ryan Murelli (Brooklyn College, City University of New York) (17, 38). Compounds were >95% pure, dissolved at 25 mM in 100% DMSO, and stored in small aliquots at −25 °C in opaque tubes.

RNase H heteroduplex cleavage assay

RNase H reactions were assembled by combining enzyme and substrate in a buffer of 50 mM HEPES, 100 mM NaCl, 2 mM TCEP, and bringing the mixture to 90% final volume with nuclease-free water. Reactions were initiated by adding 2.5 μl of 50 mM $MgCl_2$ to 22.5 μl of the 90% reaction mixture for a final $MgCl_2$ concentration of 5 mM in 25 μl. Assays were assembled at room temperature and conducted at 28 °C in 384-well black plates, and fluorescence was detected at 10 to 60 s intervals in a Biotek Synergy HTX plate reader using 485/20 nm and 528/20 nm filters. Fluorescence of substrate concentration–matched no-catalysis controls lacking either RNase H1 or Mg^{2+} was subtracted from each reaction progress curve at each time point. Maximum rates of fluorescence increase in relative fluorescence units were determined from five or more data points by the Biotek Gen5 3.10 or 3.11 software (BioTek). A linear standard curve was used to convert fluorescence units to nanomolar of released fluorescein by plotting the plateaus of reaction progress curves against RD12 concentration.

Steady-state kinetics inhibition assays

A binary titration was conducted using 12 substrate concentrations spanning the enzyme's K_M and eight inhibitor concentrations spanning the compounds' IC_{50}s, along with a DMSO-matched control and a no-enzyme control for background subtraction. These assays employed RNase H1 at 0.1 to 1 nM or 20 to 40 nM RNase HC. Rates were plotted as a function of substrate concentration for each inhibitor concentration. The compounds do not absorb in the range of fluorescein's excitation or emission; so, no corrections for inner-filter effect were necessary. Data were fit to global models of noncompetitive, uncompetitive, and mixed inhibition in GraphPad Prism (GraphPad Software, Inc). The fits were statistically compared with the extra sum-of-squares *F* test in GraphPad Prism to determine the best-fitting model (Supplementary information and Tables S2–S4). Models were also compared when a minimum V_{max} term ($V_{max_{min}}$) was introduced to account for incomplete inhibition in saturating compound concentrations. We tested whether this term significantly improved the fit by the extra sum-of-squares *F* test. Final data analysis was carried out with the mixed inhibition model (Equations 1–3) (20). All parameters were left free but shared between all datasets within an experiment.

$$v_0 = \frac{V_{max_{app}}[S]}{K_{Mapp} + [S]} \tag{1}$$

$$V_{max_{app}} = V_{max_{min}} + \left[\frac{V_{max_{max}} - V_{max_{min}}}{\left(1 + \frac{[I]}{\alpha * K_i}\right)}\right] \tag{2}$$

$$K_{M_{app}} = K_M * \left[\left(1+\frac{[I]}{K_i}\right) \Big/ \left(1+\frac{[I]}{\alpha * K_i}\right)\right] \tag{3}$$

Where $V_{max_{app}}$ is the V_{max} in the presence of a particular inhibitor concentration, $V_{max_{max}}$ is the V_{max} in the absence of inhibitor, $V_{max_{min}}$ is the V_{max} when inhibition reaches saturation, K_M is the Michaelis constant, $K_{M_{app}}$ is the K_M in a particular inhibitor concentration, K_i is the inhibition constant, α is the proportionality constant between the inhibition constant for the free enzyme relative to the ES complex ($\alpha = K_{iES}/K_{iE}$) (21), v_0 is the initial velocity, and $[I]$ and $[S]$ are inhibitor and substrate concentrations, respectively. Maximum rates of fluorescein release are reported as k_{cat} ($k_{cat} = V_{max}/[E]$). Equations are from Ref. (20) as in GraphPad Prism.

FP substrate-binding assays

RNase H1 was titrated against a fixed concentration of RD12 substrate consisting of fluorescein-labeled RNA and unlabeled DNA in 50 mM HEPES, pH 7.5, 100 mM NaCl, 10% glycerol, 2 mM TCEP, and 5 mM $CaCl_2$. The RD12 concentration was 12.5 nM unless indicated otherwise. After setup, reactions were incubated for 15 min at 28 °C before reading. Polarized fluorescence was read in black 384-well plates with nonbinding surface treatment with a BioTek Synergy H1. Anisotropy was computed in BioTek Gen5 3.11 software. Raw or background-subtracted anisotropy values (r) were fit with Equation 4 (23). R_{tot} was constrained to twice the RD12 concentration because the stoichiometry of RNase H1 for RD12 is 2:1 (Fig. 3*A*, (6)). For K_D determination with inhibitors, 200 μM of compound 110 or compound 404 or 2% DMSO as a vehicle control were added to the binding assay, and data were analyzed as described previously.

$$\Delta r = r_0 + \Delta r_{max}\left(\frac{[R]_{tot} + [L]_{tot} + K_D - \sqrt{\left([R]_{tot} + [L]_{tot} + K_D\right)^2 - \left(4 \cdot [R]_{tot} \cdot [L]_{tot}\right)}}{2 \cdot [R]_{tot}}\right) \tag{4}$$

where Δr is the anisotropy change at a particular enzyme concentration, Δr_{max} is the maximum change at saturation, r_0 is the anisotropy of free RD12, K_D is the dissociation constant, $[R]_{tot}$ is the total receptor binding-site concentration (*i.e.*, [RD12] × 2), and $[L]_{tot}$ is the total ligand concentration (RNase H1) (Equation 6 in Ref. (23)).

Compound-docking studies

The induced-fit docking protocol of Schrödinger suite (Schrödinger 2021-4 LLC) was used to predict binding conformations of compounds 110 and 404 within the active site of RNase H1. Ligands were prepared with LigPrep (Schrödinger LLC) by the following steps: (1) energy minimization with OPSL4 force field and different deprotonation states of the ligands were generated using Epik; (2) metal-binding sites were defined; and (3) compounds were desalted and tautomerized while retaining chirality. We removed the 14 bp DNA–RNA heteroduplex from the crystal structure of RNase HC (PDB ID: 2QKK) and replaced Ca^{2+} ions present in metal coordination site of RNase HC with Mg^{2+} ions by superposition of the HIV RNase H domain (PDB ID: 1RTD). The RNase H1 structure containing Mg^{2+} ions in the metal-binding sites was prepared with protein preparation wizard in Maestro (Schrödinger LLC). Water molecules that were close to the active-site residues were retained while the remaining molecules were removed, the protein was protonated at pH 7.5 ± 2, hydrogen bonds were assigned with PROPKA (Schrödinger LLC) at pH 7.5, and energy minimization was done with OPSL4 force field. β-Thujaplicinol was placed into the active site of RNase H1 by superposition of the DEDD motif of the HIV RNase H–β-thujaplicinol cocrystal structure (PDB ID: 3K2P) onto that of RNase H1. A receptor grid of 10 Å was generated around the centroid of the bound ligand, which was then used for docking of compounds. Protein refinement was carried out at a Van der Waals radius scaling factor of 0.7 for the protein and 0.5 for the ligand. Twenty poses were retained in the initial docking, residues were refined within 5.0 Å of the ligand poses, and redocking was performed with the best structures within 30.0 kcal/mol and the top 20 overall structures. We then superposed the 14-mer RNA–DNA substrate from the original structure (PDB ID: 2QKK) onto the inhibitor-bound enzyme.

Data analysis and statistics

All nonlinear curve fitting, statistical analyses, and graph creation were performed in GraphPad Prism 9.1.1. Outlying values were omitted based on a 1% Q ROUT test (39).

Data availability

All data are either in the main document or in the supporting information.

Supporting information—This article contains supporting information (18, 20, 21, 40).

Acknowledgments—We thank Alex Berkowitz and Ryan P. Murelli for the compounds used in this study. We also thank Sahiti Kuppa and Edwin Antony for assistance with CD spectroscopy.

Author contributions—N. L. P., N. P., and J. E. T. conceptualization; N. L. P. methodology; N. L. P. and R. T. formal analysis; N. L. P. and R. T. investigation; N. L. P. writing–original draft; R. T., N. P., and J. E. T. writing–review and editing; N. L. P. and R. T. visualization; N. P. and J. E. T. supervision; J. E. T. project administration.

Funding and additional information—This work was funded by the US Department of Defense grant (grant no.: W81XWH-18-1-0307 [to J. E. T.]).

Conflict of interest—The authors declare that they have no conflicts of interest with the contents of this article.

Abbreviations—The abbreviations used are: αHT, α-hydroxytropolone; DMSO, dimethyl sulfoxide; ES, enzyme–substrate; ESI, enzyme–substrate–inhibitor; FP, fluorescence polarization; HBD, hybrid-binding domain; HBV, hepatitis B virus; MW, molecular weight; PDB, Protein Data Bank; RD12, RNA–DNA hybrid of 12 bp; RNase H, ribonuclease H; RNase H1, ribonuclease H1; RT, reverse transcriptase; SEC, size-exclusion chromatography; TCEP, Tris(2-carboxyethyl)phosphine.

References

1. Holt, I. J. (2019) The Jekyll and Hyde character of RNase H1 and its multiple roles in mitochondrial DNA metabolism. *DNA Repair* **84**, 102630
2. Cerritelli, S. M., Frolova, E. G., Feng, C., Grinberg, A., Love, P. E., and Crouch, R. J. (2003) Failure to produce mitochondrial DNA results in embryonic lethality in Rnaseh1 null mice. *Mol. Cell* **11**, 807–815
3. Hyjek, M., Figiel, M., and Nowotny, M. (2019) RNases H: Structure and mechanism. *DNA Repair (Amst)* **84**, 102672
4. Samara, N. L., and Yang, W. (2018) Cation trafficking propels RNA hydrolysis. *Nat. Struct. Mol. Biol.* **25**, 715–721
5. Nowotny, M., Gaidamakov, S. A., Crouch, R. J., and Yang, W. (2005) Crystal structures of RNase H bound to an RNA/DNA hybrid: Substrate specificity and metal-dependent catalysis. *Cell* **121**, 1005–1016
6. Gaidamakov, S. A., Gorshkova, , II, Schuck, P., Steinbach, P. J., Yamada, H., Crouch, R. J., and Cerritelli, S. M. (2005) Eukaryotic RNases H1 act processively by interactions through the duplex RNA-binding domain. *Nucleic Acids Res.* **33**, 2166–2175
7. Edwards, T. C., Ponzar, N. L., and Tavis, J. E. (2019) Shedding light on RNaseH: A promising target for hepatitis B virus (HBV). *Expert Opin. Ther. Targets* **23**, 559–563
8. Edwards, T. C., Mani, N., Dorsey, B., Kakarla, R., Rijnbrand, R., Sofia, M. J., and Tavis, J. E. (2019) Inhibition of HBV replication by N-hydroxyisoquinolinedione and N-hydroxypyridinedione ribonuclease H inhibitors. *Antiviral Res.* **164**, 70–80
9. Li, Q., Lomonosova, E., Donlin, M. J., Cao, F., O'Dea, A., Milleson, B., Berkowitz, A. J., Baucom, J.-C., Stasiak, J. P., Schiavone, D. V., Abdelmessih, R. G., Lyubimova, A., Fraboni, A. J., Bejcek, L. P., Villa, J. A., *et al.* (2020) Amide-containing α-hydroxytropolones as inhibitors of hepatitis B virus replication. *Antivir. Res.* **177**, 104777
10. Himmel, D. M., Maegley, K. A., Pauly, T. A., Bauman, J. D., Das, K., Dharia, C., Clark, A. D., Jr., Ryan, K., Hickey, M. J., Love, R. A., Hughes, S. H., Bergqvist, S., and Arnold, E. (2009) Structure of HIV-1 reverse transcriptase with the inhibitor beta-Thujaplicinol bound at the RNase H active site. *Structure* **17**, 1625–1635
11. Nowotny, M., Gaidamakov, S. A., Ghirlando, R., Cerritelli, S. M., Crouch, R. J., and Yang, W. (2007) Structure of human RNase H1 complexed with

an RNA/DNA hybrid: Insight into HIV reverse transcription. *Mol. Cell* **28**, 264–276
12. Spudich, G. M., Miller, E. J., and Marqusee, S. (2004) Destabilization of the Escherichia coli RNase H kinetic intermediate: Switching between a two-state and three-state folding mechanism. *J. Mol. Biol.* **335**, 609–618
13. Yamasaki, K., Ogasahara, K., Yutani, K., Oobatake, M., and Kanaya, S. (1995) Folding pathway of Escherichia coli ribonuclease HI: A circular dichroism, fluorescence, and NMR study. *Biochemistry* **34**, 16552–16562
14. Dabora, J. M., and Marqusee, S. (1994) Equilibrium unfolding of Escherichia coli ribonuclease H: Characterization of a partially folded state. *Protein Sci.* **3**, 1401–1408
15. Leonor Michaelis, M. M. (1913) Die kinetik der Invertinwirkung. *Biochemische Z.* **49**, 333–369
16. Johnson, K. A., and Goody, R. S. (2011) The original Michaelis constant: Translation of the 1913 michaelis–menten paper. *Biochemistry* **50**, 8264–8269
17. Hu, Y., Cheng, X., Cao, F., Huang, A., and Tavis, J. E. (2013) β-Thujaplicinol inhibits hepatitis B virus replication by blocking the viral ribonuclease H activity. *Antivir. Res.* **99**, 221–229
18. Kocyła, A., Pomorski, A., and Krężel, A. (2017) Molar absorption coefficients and stability constants of Zincon metal complexes for determination of metal ions and bioinorganic applications. *J. Inorg. Biochem.* **176**, 53–65
19. Hirai, M., and Oka, Y. (1970) Stability of tropolone chelates of the Bi- and tervalent metal ions. *Bull. Chem. Soc. Jpn.* **43**, 778–782
20. Copeland, R. A. (2013) Reversible modes of inhibitor interactions with enzymes *Evaluation of Enzyme Inhibitors in Drug Discovery. A Guide for Medicinal Chemists and Pharmacologists*, 2 Ed., Wiley, Hoboken, NJ
21. Strelow J, D. W., Dewe, W., Iverson, P. W., Brooks, H. B., Radding, J. A., McGee, J., and Weidner, J. (2012) Mechanism of action assays for enzymes. In: Markossian., S., Grossman, A., Brimacombe, K., Arkin, M., Auld, D., Austin, C. P., Baell, J., Chung, T. D. Y., Coussens, N. P., Dahlin, J. L., Devanarayan, V., Foley, T. L., Glicksman, M., Hall, M. D., Haas, J. V., *et al*. eds. *Assay Guidance Manual [Internet]*, Eli Lilly & Company and the National Center for Advancing Translational Sciences, Bethesda, MD
22. Nowotny, M., Cerritelli, S. M., Ghirlando, R., Gaidamakov, S. A., Crouch, R. J., and Yang, W. (2008) Specific recognition of RNA/DNA hybrid and enhancement of human RNase H1 activity by HBD. *Embo J.* **27**, 1172–1181
23. Pollard, T. D. (2010) A guide to simple and informative binding assays. *Mol. Biol. Cell* **21**, 4061–4067
24. Parajuli, S., Teasley, D. C., Murali, B., Jackson, J., Vindigni, A., and Stewart, S. A. (2017) Human ribonuclease H1 resolves R-loops and thereby enables progression of the DNA replication fork. *J. Biol. Chem.* **292**, 15216–15224
25. Lima, W. F., Murray, H. M., Damle, S. S., Hart, C. E., Hung, G., De Hoyos, C. L., Liang, X.-H., and Crooke, S. T. (2016) Viable RNaseH1 knockout mice show RNaseH1 is essential for R loop processing, mitochondrial and liver function. *Nucleic Acids Res.* **44**, 5299–5312
26. Tavis, J. E., Wang, H., Tollefson, A. E., Ying, B., Korom, M., Cheng, X., Cao, F., Davis, K. L., Wold, W. S. M., and Morrison, L. A. (2014) Inhibitors of nucleotidyltransferase superfamily enzymes suppress herpes simplex virus replication. *Antimicrob. Agents Chemother.* **58**, 7451–7461
27. Didierjean, J., Isel, C., Querré, F., Mouscadet, J.-F., Aubertin, A.-M., Valnot, J.-Y., Piettre, S. R., and Marquet, R. (2005) Inhibition of human immunodeficiency virus type 1 reverse transcriptase, RNase H, and integrase activities by hydroxytropolones. *Antimicrob. Agents Chemother.* **49**, 4884–4894
28. Kumar, G., Cuypers, M., Webby, R. R., Webb, T. R., and White, S. W. (2021) Structural insights into the substrate specificity of the endonuclease activity of the influenza virus cap-snatching mechanism. *Nucleic Acids Res.* **49**, 1609–1618
29. Lima, W. F., Wu, H., Nichols, J. G., Prakash, T. P., Ravikumar, V., and Crooke, S. T. (2003) Human RNase H1 uses one tryptophan and two lysines to position the enzyme at the 3′-DNA/5′-RNA terminus of the heteroduplex substrate. *J. Biol. Chem.* **278**, 49860–49867
30. Lee, H., Cho, H., Kim, J., Lee, S., Yoo, J., Park, D., and Lee, G. (2021) RNase H is an exo- and endoribonuclease with asymmetric directionality, depending on the binding mode to the structural variants of RNA:DNA hybrids. *Nucleic Acids Res* **50**, 1801–1814
31. Edwards, T. C., Lomonosova, E., Patel, J. A., Li, Q., Villa, J. A., Gupta, A. K., Morrison, L. A., Bailly, F., Cotelle, P., Giannakopoulou, E., Zoidis, G., and Tavis, J. E. (2017) Inhibition of hepatitis B virus replication by N-hydroxyisoquinolinediones and related polyoxygenated heterocycles. *Antivir. Res.* **143**, 205–217
32. Cao, L., Song, W., De Clercq, E., Zhan, P., and Liu, X. (2014) Recent progress in the research of small molecule HIV-1 RNase H inhibitors. *Curr. Med. Chem.* **21**, 1956–1967
33. Tramontano, E., Corona, A., and Menéndez-Arias, L. (2019) Ribonuclease H, an unexploited target for antiviral intervention against HIV and hepatitis B virus. *Antivir. Res.* **171**, 104613
34. Pauling, L. (1947) Atomic Radii and interatomic distances in metals. *J. Am. Chem. Soc.* **69**, 542–553
35. Nowotny, M., and Yang, W. (2006) Stepwise analyses of metal ions in RNase H catalysis from substrate destabilization to product release. *EMBO J.* **25**, 1924–1933
36. Zhang, Z., and Tavis, J. E. (2006) The duck hepatitis B virus reverse transcriptase functions as a full-length monomer. *J. Biol. Chem.* **281**, 35794–35801
37. Tian, L., Kim, M.-S., Li, H., Wang, J., and Yang, W. (2018) Structure of HIV-1 reverse transcriptase cleaving RNA in an RNA/DNA hybrid. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 507
38. Lu, G., Lomonosova, E., Cheng, X., Moran, E. A., Meyers, M. J., Le Grice, S. F., Thomas, C. J., Jiang, J. K., Meck, C., Hirsch, D. R., D'Erasmo, M. P., Suyabatmaz, D. M., Murelli, R. P., and Tavis, J. E. (2015) Hydroxylated tropolones inhibit hepatitis B virus replication by blocking viral ribonuclease H activity. *Antimicrob. Agents Chemother.* **59**, 1070–1079
39. Motulsky, H. J., and Brown, R. E. (2006) Detecting outliers when fitting data with nonlinear regression – a new method based on robust nonlinear regression and the false discovery rate. *BMC Bioinformatics* **7**, 123
40. Jarmoskaite, I., AlSadhan, I., Vaidyanathan, P. P., and Herschlag, D. (2020) How to measure and evaluate binding affinities. *Elife* **9**, e57264

Received: 18 May 2022 | Revised: 31 July 2022 | Accepted: 2 August 2022

DOI: 10.1002/pro.4421

FULL-LENGTH PAPER

THE PROTEIN SOCIETY WILEY

Predicted structure of the hepatitis B virus polymerase reveals an ancient conserved protein fold

Razia Tajwar | **Daniel P. Bradley** | **Nathan L. Ponzar** | **John E. Tavis**

Department of Molecular Microbiology and Immunology, School of Medicine and Institute for Drug and Biotherapeutic Innovation, Saint Louis University, Saint Louis, Missouri, USA

Correspondence
John E. Tavis, Department of Molecular Microbiology and Immunology, Saint Louis University School of Medicine, 1100 S. Grand Boulevard, Saint Louis, MO 63104 USA.
Email: john.tavis@health.slu.edu

Funding information
Department of Defense, Grant/Award Number: W81XWH-18-1-0307; National Institutes of Health, Grant/Award Numbers: R01 AI148362, R01 AI150610

Review Editor: Nir Ben-Tal

Abstract

Hepatitis B virus (HBV) chronically infects >250 million people. It replicates by a unique protein-primed reverse transcription mechanism, and the primary anti-HBV drugs are nucleos(t)ide analogs targeting the viral polymerase (P). P has four domains compared to only two in most reverse transcriptases: the terminal protein (TP) that primes DNA synthesis, a spacer, the reverse transcriptase (RT), and the ribonuclease H (RNase H). Despite being a major drug target and catalyzing a reverse transcription pathway very different from the retroviruses, HBV P has resisted structural analysis for decades. Here, we exploited computational advances to model P. The TP wrapped around the RT domain rather than forming the anticipated globular domain, with the priming tyrosine poised over the RT active site. The orientation of the RT and RNase H domains resembled that of the retroviral enzymes despite the lack of sequences analogous to the retroviral linker region. The model was validated by mapping residues with known surface exposures, docking nucleic acids, mechanistically interpreting mutations with strong phenotypes, and docking inhibitors into the RT and RNase H active sites. The HBV P fold, including the orientation of the TP domain, was conserved among hepadnaviruses infecting rodent to fish hosts and a nackednavirus, but not in other non-retroviral RTs. Therefore, this protein fold has persisted since the hepadnaviruses diverged from nackednaviruses >400 million years ago. This model will advance mechanistic analyses into the poorly understood enzymology of HBV reverse transcription and will enable drug development against non-active site targets for the first time.

KEYWORDS
hepatitis B virus, polymerase, predicted structure, reverse transcription

1 | INTRODUCTION

Hepatitis B virus (HBV) is a small, partially double-stranded DNA virus in the *Hepadnaviridae* virus family that replicates by protein-primed reverse transcription.[1] HBV has nine genotypes differing by >8% at the nucleotide level.[2] Over 250 million people are chronically infected with the virus,[3] leading to >850,000 deaths annually.[4] The primary drugs for HBV infection are nucleos(t)ide analogs that block reverse transcription, but treatment is not curative and is life-long for most patients.[5,6]

The HBV polymerase (P) that catalyzes reverse transcription has two additional domains, the terminal protein (TP) and spacer domains, that are absent in the much better understood retroviral reverse transcriptases

(RT). Together, these unique domains account for almost half of P's sequence (Figure 1a). The TP domain contains about 180 amino acids (aa), including Y63 that primes DNA synthesis. This is followed by the poorly conserved ~175 aa spacer with little known function.[1] The catalytic core of the enzyme contains the RT and ribonuclease H (RNase H) domains and is homologous to retroviral RTs despite sharing only about 20% aa sequence homology and lacking sequences analogous to the linker region between the RT and RNase H domains. The RT domain (~335 aa) carries the A-E reverse transcription motifs,[7] including the YMDD motif that chelates 2 Mg^{2+} ions essential for DNA synthesis. The ~155 aa RNase H domain contains the D-E-D-D motif that chelates 2 Mg^{2+} ions essential for RNA hydrolysis.[8–10] The C-terminal ~35 residues of P are dispensable for RNase H activity in vitro[11] and are presumed to be unstructured.

P is 843 aa long in most HBV genotypes, with the genotypes sharing 86–88% identity. Genotype A has a two-residue insertion between positions 16 and 17 in the terminal protein (genotype B numbering employed unless indicated otherwise), genotype D lacks 11 aa in the spacer domain (residues 184–194), and genotypes E and G lack aa 184 in the spacer.

HBV reverse transcription[12,13] begins with binding of P to the ε RNA stem-loop on the viral pregenomic RNA (pgRNA) via an HSP90-mediated reaction.[14,15] ε binding depends on a bi-partite binding element comprised of the T3 motif in the TP domain and the RT1 motif in the RT domain.[16] The RT activity initiates DNA synthesis using Y63 in the TP domain as a primer and a bulge in ε as the template.[17] This covalently links P to the viral minus-polarity DNA strand, but it is unknown how Y63 accesses the RT active site. The RNase H degrades the pgRNA after it has been copied into minus-polarity DNA. Studies with duck hepatitis B virus (DHBV), a distant homolog of HBV, reveal that the RNase H leaves a 15–18 nt capped RNA that primes plus-polarity DNA strand synthesis due to the distance between the RT and RNase H active sites.[18] Reverse transcription employs three strand transfers to make the mature partially double-stranded viral DNA in viruses. P is a monomer and remains covalently attached to the viral DNA within mature virions.[19] Only the enzymology of HBV pgRNA encapsidation and DNA chain elongation are understood in any depth.

Structural analyses of P have been stymied for >30 years despite its importance as a drug target, the presence of two unique domains, and the unusual reverse transcription pathway it catalyzes. P is extremely difficult to produce in an active, soluble form due to its instability in most protein production systems and aggregation issues. Both problems likely stem from P's existence in a complex with HSP90 chaperones[14,15] in which the chaperones are believed to cover hydrophobic regions on the protein surface. Second, P has at least five known conformations, identified primarily using DHBV P (Figure 1b).[16,20–23] Flexibility is promoted by HSP90 in the holoenzyme complex and is necessary to permit reverse transcription to occur while P is covalently attached to one end of its product DNA. A globular ab initio predicted model exists for the TP domain based on secondary structure predictions.[24,25] Multiple homology models generated against retroviral RTs exist for the RT domain of HBV P.[26,27] The RT models are accurate enough in the active site to permit interpretation of the mechanisms of resistance to nucleoside analog drugs, but little is known of their accuracy

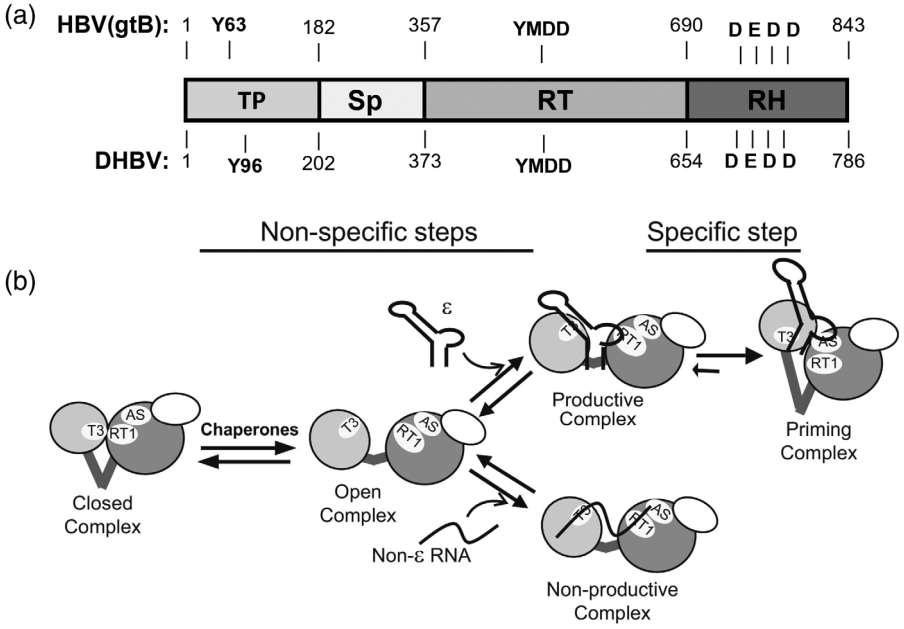

FIGURE 1 Gene organization for HBV P and conformations adopted by the enzyme. (a) Gene organization. Genotype B numbers are used for the domain boundaries, and the Y63 priming residue, YMDD RT active site motif, and D-E-D-D RNase H active site motif are indicated. (b) Conformations adopted by P identified by prior analyses. The TP domain is in light gray, the RT domain in dark gray, the RNase H domain as an unshaded oval, and the spacer domain is indicated by thick lines. AS, RT active site; T3, T3 motif in the TP domain; RT1, RT1 motif in the RT domain; ε, the HBV ε stem loop needed for specific RNA encapsidation and priming. Reprinted with permission from[16]

outside of the active site. Three homology models exist for the RNase H domain.[28–30] None of them contain the full domain[31] and only one[28] correctly predicts the last D-E-D-D active site residue.[8,32]

AlphaFold is an ab initio protein structure prediction program.[33] It uses correlated sequence variations among homologous sequences to define intra-protein distance constraints plus structural information from homologous proteins (when available) to predict the structure of the α-carbon backbone of a protein. Side chain orientations are then refined by energy minimization. Structures were predicted by AlphaFold for 10,795 proteins with experimentally determined structures, and >50% of the models had α-carbon chain root-mean-square deviation (RMSD) values ≤2 Å compared to experimentally determined structures.[33] Accuracy of the predictions is measured in part by the lowest distance difference test (lDDT).[34] The predicted lDDT scores (pLDDT) generated by AlphaFold correlate well with lDDT scores for known protein structures.[33] These studies validate AlphaFold's utility, with the understanding that the predicted structures need to be experimentally validated.

Here, we predicted the structure of HBV P and validated the model using data from prior analyses of P and molecular docking. Structures of P proteins from animal hepadnaviruses and nackednaviruses were predicted and compared to the HBV P. These studies reveal the fold of the enzyme, permit approximation of its antiquity, generate predictions for how P primes reverse transcription, and enable drug discovery against non-active sites for the first time.

2 | RESULTS AND DISCUSSION

2.1 | Modeling

HBV P models were primarily generated using genotype B, Genbank AB554017. Models were generated for each domain independently, the catalytic core of the enzyme, and the full-length enzyme using AlphaFold2. Models with the highest pLDDT scores among the five primary models for each fold were energy-minimized and used for all analyses. The domain boundaries were from Donlin et al.,[35] with the N-terminus of the RT domain moved to residue 357 to include a β-sheet at the boundary of the spacer and RT domains conserved in all models. The individual domain models are in Figures S1–S4.

2.2 | Catalytic core

A model of the HBV catalytic core containing the RT and RNase H domains folded with high confidence (pLDDT >80) for the bulk of the model, with lower confidence at the N-terminus of the RT domain, ~80 aa in the RT that includes the YMDD motif, and the extreme C-terminus (Figure S5).

The RT domain adopted the expected right-hand shape of a DNA polymerase with fingers, palm, and thumb subdomains (Figure 2a). The conserved A-E DNA polymerase motifs[7] were in the palm, similar to their locations in other DNA polymerases. The model accommodated two Mg^{2+} ions separated by 3.23 Å in the RT active site that were coordinated by the YMDD motif in positions analogous to those in the HIV RT. The RT domain within the predicted HBV catalytic core model could be superimposed with the RT domain model from Das et al.[26] with an RMSD = 3.38 Å, with the AlphaFold structure predicting more of the conserved α-helixes and β-sheets in the fingers subdomain (Figure 3a). The HBV RT domain from the catalytic core model aligned to the HIV RT domain (PDB: 5XN1) with an RMSD = 3.34 Å (Figure 3b). The orientations of the α-helixes and β-sheets in the fingers subdomain from the AlphaFold HBV RT model were more similar to the HIV structure than the Das et al. model,[26] although they were rotated outward more than expected.

The RNase H domain adopted the canonical RNase H fold[9,10] with a β-sheet platform overlaid by α-helices arranged in an H (Figures S5 and S24b–d). The model accommodated 2 Mg^{2+} ions 3.96 Å apart that were coordinated by the D-E-D-D motif that binds divalent cations in other RNases H, as expected because HBV RNase H activity requires Mg^{2+} or Mn^{2+} and mutating the D-E-D-D residues yields an RNase H-deficient phenotype.[8,32] The RNase H domain from the catalytic core model aligned well with the HIV RNase H (PDB: 3K2P, RMSD = 2.84 Å; Figure S24b) and human RNase H1 (PDB: 2QKK, RMSD = 2.66 Å; Figure S24c). The AlphaFold model is likely superior to our prior RNase H model[28] due to AlphaFold's better performance[33] compared to the Phyre2 modeling software[36] used for the older model, and because it predicted an α helix and a β sheet strand that is conserved among known RNase H structures[9,10] that were previously poorly formed.

The relative orientation of the two domains in the HBV catalytic core model was similar to the HIV RT-RNase H (Figures 3b and S24f) despite the HIV enzyme being a heterodimer that carries a second copy of an enzymatically inactive RT domain. The HBV genotype B catalytic core aligned with the HIV enzyme with an RMSD of 3.75 Å (Figure S24f) and with the monomeric Ty3 RT-RNase H (PDB: 4OL8) with an RMSD of 3.60 Å (Figure S24h), but alignment qualities were suppressed because the HBV model lacks the linker sequences between the two catalytic domains which helps

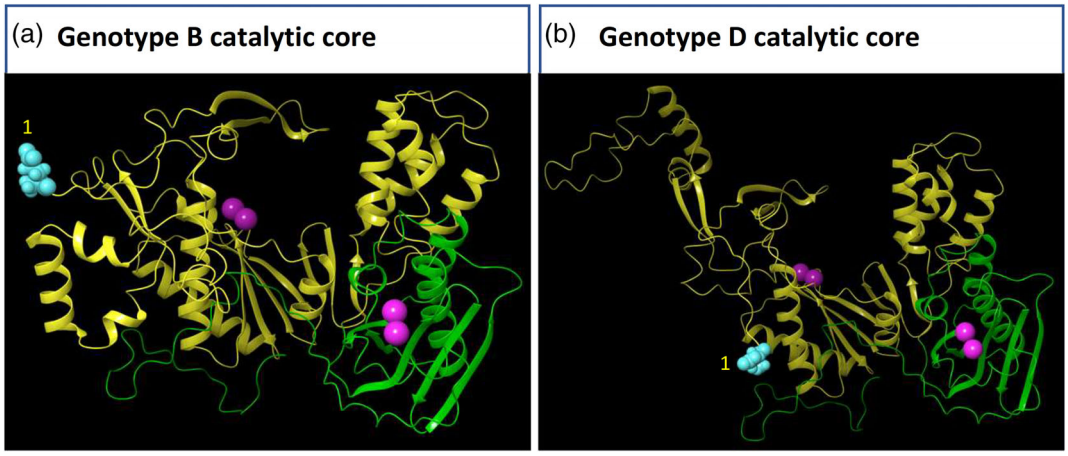

FIGURE 2 Models for the catalytic core of HBV P. (a) Genotype B RT and RNase H domains. (b) Genotype D RT and RNase H domains. Yellow, RT domain; Green, RNase H domain; Magenta spheres, Mg^{2+} ions. The amino termini are marked with cyan spheres

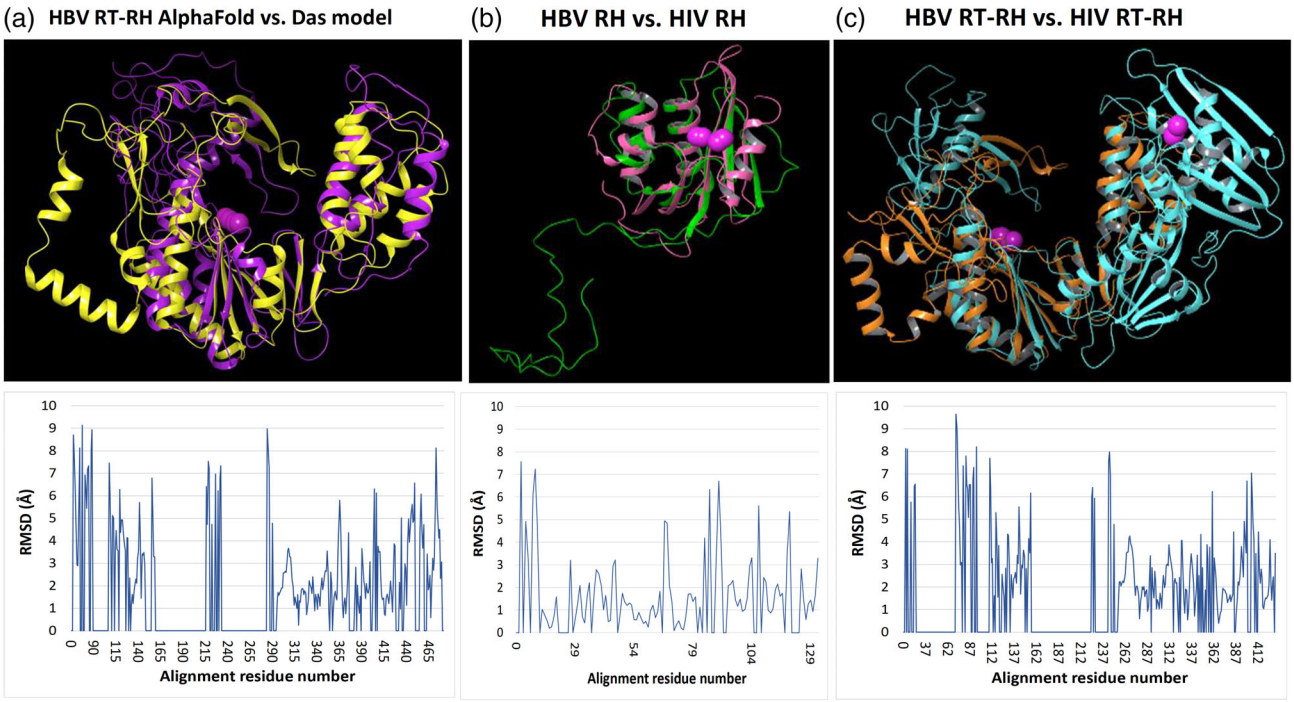

FIGURE 3 Superpositions of HBV P models. (a) HBV genotype B RT domain from the catalytic core model (yellow) versus the HBV RT domain model (purple) from Das et al.[26] (b) HBV RNase H (green) versus HIV RNase H (magenta). (c) HBV genotype B catalytic core model (orange) versus HIV RT-RNase H structure (cyan). Top panels, superpositions; Bottom panels, plots of RMSD values by residue number in the superposition alignment. RT, RT domain; RH, RNase H domain; Magenta spheres, Mg^{2+} ions

accommodate the heteroduplex in HIV and other retroviral RT-RNases H.[37,38] A plausible nucleic acid binding channel between the two active sites featuring positively charged and neutral residues was present, although the channel was not as positively charged as in HIV. The reduced affinity for nucleic acids implied by the relatively low positive charge of the binding channel in the HBV enzyme is plausible because reverse transcription occurs within a capsid particle with an interior diameter of 250 Å, so the effective concentration of the nucleic acids is very high. This consideration also applies to the HIV enzyme which is similarly active within a viral capsid during viral replication.

We also predicted a two-domain model for HBV genotype D as it is commonly used for inhibitor screening (Figure S6). The genotype B and D RT-RNase H models aligned with an RMSD = 2.15 Å (Figure S24e). The orientation of the fingers subdomain in the genotype D model

is more plausible than in the genotype B model because it resembles the orientation commonly seen in DNA polymerases.

2.3 | Full-length HBV polymerase

Full-length P folded with high confidence (pLDDT >80), with the exception of the extreme N-terminus, spacer, the C-terminal ~50 residues, and short regions within each domain (Figures 4a and S8). A full-length P model for genotype D was also predicted that aligned well with the genotype B model (RMSD = 1.96 Å; Figures S10 and S25c).

Intriguingly, the TP domain formed a novel C-shaped structure that cupped around the catalytic core of P rather than the expected globular fold (Figure 4a). Y63 in the TP domain that primes DNA synthesis and covalently links P to the viral minus-polarity DNA strand[17] was predicted to be on a loop above the domain YMDD RT active site motif. Y63 is near the N-terminal edge of a region of high pLDDT confidence that was flanked by regions with pLDDT values of 50–60 that imply poorly structured and/or flexible sequences. The putative flexibility of the loop is consistent with the need for Y63 to bend downwards to prime DNA synthesis in the RT active site and then be displaced from the active site by the growing DNA. It is also consistent with the conformational changes that must occur in the priming loop during the three DNA strand transfers during reverse transcription[12,13] while P remains attached to the 5′ end of the minus-polarity DNA strand via Y63.

The spacer domain was predicted to form an unstructured loop (Figures 4a and S2) that was different in all five primary models generated by AlphaFold. This is consistent with its low sequence conservation, the ability of the TP and RT domains to complement each other in *trans*,[39–41] and with the ability to alter sequences in the spacer without affecting reverse transcription.[19,42–44]

The RT and RNase H domain predictions in the two- and four-domain genotype B HBV models aligned with an RMSD of 2.37 Å (Figure S26e). The fingers subdomain of the AlphaFold HBV P model was more similar to the HIV RT than the fingers predicted by the Das et al. HBV RT domain model[26] and had α-helixes and β-sheets of similar lengths and orientations as in the HIV RT (Figures S26d and S27a). Both catalytic domains in the full-length model accommodated Mg^{2+} ions coordinated by the YMDD and D-E-D-D motifs in positions very similar to those in the HIV RT and RNase H active sites (Figure 4b,c).

2.4 | Nucleic acid docking

RNA:DNA heteroduplexes (21 and 24 bp) were extracted from HIV RT-RNase H co-crystal structures (PDB: 4B3Q and 6BSH) and docked into the HBV RT-RNase H catalytic core and full-length P models for genotypes B and D employing the Schrödinger BioLuminate module. The

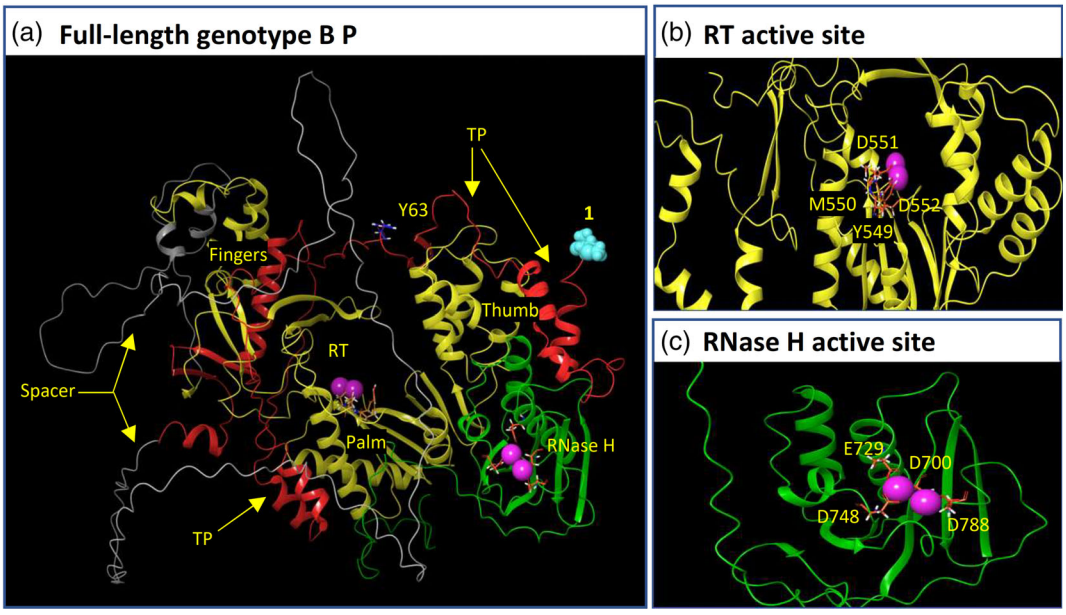

FIGURE 4 Predicted model of genotype B P. (a) Full-length P. The amino terminus is marked with cyan spheres. (b) RT active site. (c) RNase H active site. Red, TP domain; Gray, Spacer domain; Yellow, RT domain; Green, RNase H domain. Magenta spheres, Mg^{2+} ions. Y63 and the residues in the YMDD and D-E-D-D motifs are labeled and shown as sticks

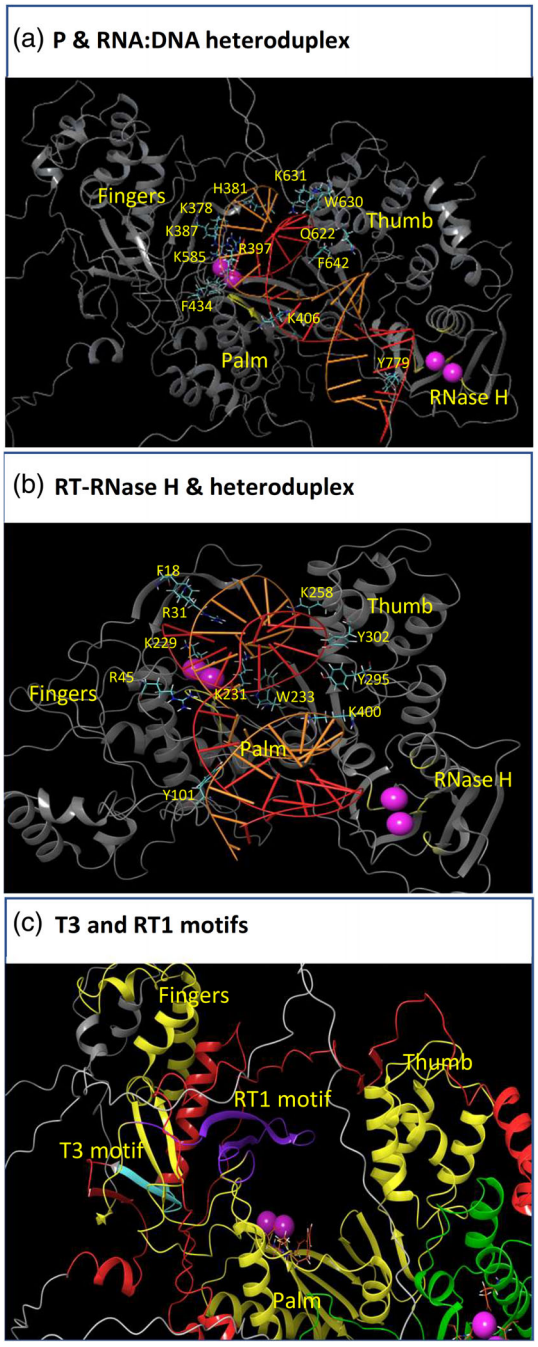

FIGURE 5 Computational heteroduplex nucleic acid docking into the HBV genotype B catalytic core and full-length P models. Docking was conducted using PIPER program in the BioLuminate module within the Schrödinger molecular analysis suite. (a and b) Representative binding poses between the indicated model and a heteroduplex. The fingers, palm, and thumb subdomains of the RT domain are labeled. Residues in the RT and RNase H domains contacting the RNA:DNA heteroduplex are shown as sticks and labeled. Locations of residues that interact with ε among the various binding poses are indicated: Violet, RT1 motif; Cyan, T3; Yellow, all other interacting residues. (c) Positions of the T3 and RT1 motifs. Magenta sphere, Mg^{2+} ions

heteroduplexes docked into the RT active site but were usually located just outside the RNase H active site (Figures 5a,b and S28b,d). One pose for the genotype D catalytic core model simultaneously aligned the DNA strand near the Mg^{2+} ions in the RT active site and the RNA strand opposite the Mg^{2+} ions in the RNase H active site (Figure S28e–g), with the heteroduplex following the nucleic acid binding channel predicted by the apo models. The 19 bp spanning the RT and RNase H active sites is similar to the 15–18 bp measured with DHBV P.[18] The infrequency of binding poses that engage both active sites is similar to what has been seen with HIV RT-RNase H co-crystals where the heteroduplexes were located outside the RNase H active site.[45–47]

The HBV ε RNA stem-loop that templates DNA priming [PDB: 6VAR[48]] was also docked into P (Figure S29a–c). Three classes of binding poses were found. Two classes wrapped end-to-end around the side of P, and the third class docked perpendicularly to the other poses. All poses contacted the palm subdomain, the TP domain, and the spacer, but none placed ε in the RT active site. These binding poses may represent the non-specific early phase in binding between ε and P (Figure 1b).[16] Overall, the ε docking studies were less informative than the heteroduplex dockings due to the diversity of binding poses, possibly due to inability of the docking algorithm to recapitulate the dynamic, chaperone-driven ε binding mechanism.

2.5 | Assessing plausibility of the predicted HBV P model

The novelty of the HBV P fold necessitated careful validation. This was done computationally and using the wealth of experimental data generated by us and others over the last 30 years.

Plausibility of the TP interface with the catalytic core of the enzyme was evaluated computationally by assessing binding interactions between the TP and the catalytic core of P comprised of the RT and RNase H domains. The model predicted 31 hydrogen bonds and one salt bridge between residues in the TP and RT domains, and three hydrogen bonds and two salt bridges between the TP and RNase H domains (Table 1). The model predicted 184 residues with extensive hydrophobic regions that were buried within the structure at the interface between the TP domain and the catalytic core (76 in the TP, 91 in the RT, and 17 in the RNase H domains; Tables 2 and 3). The large majority of these hydrophobic residues formed patches between the TP and RT

TABLE 1 Molecular bonds predicted between the TP domain and the rest of P

TP-RT H-bonds			TP-RT salt bridges		
TP residue	**Distance (Å)**	**RT residue**	**TP residue**	**Distance (Å)**	**RH residue**
L41	1.77	S670	K130	2.60	D391
T53	1.93	P627	TP-RNase H H-bonds		
T53	2.14	D629	TP residue	Distance (Å)	RH[a] residue
T60	1.85	R635	E22	3.70	R739
L62	2.18	D629	N34	2.25	L763
N71	1.93	T383	F95	1.84	R829
W74	1.89	D377	TP-RNase H salt bridges		
Q75	2.15	S386	TP residue	Distance (Å)	RH residue
I82	1.74	K514	E22	2.96	R739
Q85	1.98	L577	E103	3.45	R539
V96	2.09	R539			
E103	1.37	S535			
R106	2.13	Q528			
A113	1.90	H440			
A113	1.97	P438			
T120	1.85	H446			
K121	1.93	Q394			
Y122	2.15	P445			
L123	1.78	V369			
D126	2.32	R460			
G128	1.80	V373			
K130	1.91	V373			
K130	1.07	D391			
H146	2.00	Y504			
I156	2.03	K514			
I156	2.33	R513			
R160	1.89	T365			
R164	2.90	T362			
F168	2.05	H358			
F168	2.08	E357			
Q180	2.71	R359			

[a]RH, RNase H domain.

domains. These observations are consistent with binding of the TP domain to the rest of P being strong enough to support folding of the enzyme into this conformation.

The T3 motif in the TP domain and the RT1 motif in the RT domain were previously found to form two parts of a discontinuous RNA binding motif in HBV and DHBV that is obscured when P is initially translated but that promotes binding of ε to P upon being exposed by cellular chaperones (Figure 1b).[16,20,21,49,50] Despite being separated by 207 residues, the T3 and RT1 motifs in the models formed a nearly contiguous motif wrapping over the fingers subdomain and into the interior face of the RT domain (Figure 5c). All the binding poses for ε (Figure S29) had at least one contact between ε and RT1, but few contacts were found with T3. A strand of the TP domain overlaid the RT1 motif near where it contacts T3, and one of the few structured portions of the spacer domain obscured most of T3. Comparing the full P model to previously identified conformations for P implies the AlphaFold model may approximate the *closed complex*[16]

TABLE 2 TP domain residues with buried hydrophobic regions at the interface with the RT or RNase H domains

Residue	Solvent accessibility (%)	Residue	Solvent accessibility (%)	Residue	Solvent accessibility (%)
S4	9.7	T76	15.4	H136	46.1
F8	9.3	F79	6.9	V137	8.6
L11	23.5	I82	0.2	V138	15.4
L12	2.6	L84	0.1	Y141	0
L13	8.4	Q85	39.4	F142	0
L21	40.3	I88	0.8	Q143	10.6
L25	0	V89	22.7	T144	0
P26	8.6	C92	0	R145	1.9
L28	14.5	F95	32.5	L148	0
A29	6.1	L99	0	H149	7.3
L33	34.8	T100	19.7	L151	0
V37	17.5	Q103	1.5	W152	0.7
N42	46.4	R106	23.6	K153	47
L 43	3.9	L107	0.3	I156	0.1
L46	48.4	I110	0.3	L157	0
V48	35.9	M111	5.1	Y158	0.1
W52	37.5	A113	0	K159	11.5
T53	10.6	F115	0.9	A166	0
H54	34.4	Y116	45.3	F168	0
V56	0.3	T120	5.5	C169	6
G57	26.4	Y122	28.8	G170	25.1
F59	11.1	L123	27.3	Y173	1.2
G61	4.8	L125	12	W175	12
L62	46.4	D126	21.5	E176	0.4
F70	1.8	I129	0.3	L179	37.3
N71	28.8	Y132	37.3		
W74	0.9	Y133	3.6		

initially formed upon translation of P that cannot bind specifically to ε (Figure 1b).

Further support for the hepadnaviral P fold comes from the DHBV P model (Figure S16; see below). Partial proteolysis revealed that E17 or E18, E369, E370, E371, D468, E555, E565, E571, and E572 are protease accessible, and that D468 is inaccessible.[21,23] The location of these 11 residues on the DHBV P model is consistent with these observations. E164, E176, and E199 are resistant to proteolysis in the absence of a chaperone-mediated conformational change but become sensitive upon chaperone action.[21] E164 is partially obscured in a groove but could be exposed by minor structural changes, and E199 is buried but could be exposed by shifting the helix in which it is located. These observations are consistent with the model reflecting the closed complex (Figure 1b). E176 is at the end of an α-helix that could reduce its proteolytic sensitivity. Remodeling that helix upon chaperone activation is plausible because E176 is part of the T3 motif that participates in chaperone-dependent ε binding. E106 and E124 are protease resistant[21] but were exposed in the model. E124 is in an α-helix that could reduce its sensitivity to proteolysis, but E106 is in a loop. The E106 result indicates either a weakness in the model or that the site is obscured by the spacer or chaperones in the native holoenzyme complex. The epitope for monoclonal antibody 11 (residues 53–59) is constitutively exposed, and the epitopes for monoclonal antibodies 5 (residues 141–147) and 6 (residues 191–197) are obscured before the chaperone-mediated conformational change leading to formation of the *open complex* (Figure 1b).[16,21] The antibody 11 epitope is fully exposed in the DHBV P model, the antibody 6 epitope is in a deep groove that

TABLE 3 Catalytic core residues with buried hydrophobic regions at the interface with the TP domain

Residue	Solvent accessibility (%)	Residue	Solvent accessibility (%)	Residue	Solvent accessibility (%)
RT domain		RT domain (*continued*)		RT domain (*continued*)	
I360	1.0	R460	31.5	P623	22.7
T362	0	L461	27.8	P627	4.7
P366	28.1	L486	4.9	I628	0
G371	0.4	S489	26.8	W630	15
G372	2.4	L493	8.3	L631	34.4
V373	0	Y497	7.3	VV632	0
F374	13.9	G498	32.1	I636	0
L375	1.0	L501	0	Q648	17.4
V376	29	L503	17.3	C649	0
S386	0.7	S505	0	P656	5.3
L388	3.1	I508	7	L657	0
V389	38.9	I509	7.4	C660	5.7
V390	2.8	L510	0	Q665	45.6
Q394	19	L514	1.3	A666	11.4
F395	0.8	I515	0.6	T668	10
V402	27.4	P516	0.2	F676	3.2
W404	16.7	M517	34.3	N678	38.7
S431	13.2	L521	12	RNase H domain	
F434	16.6	S522	1.6	P719	0.6
Y435	5.8	L525	6.9	L723	13.5
L437	0	L 526	0.1	T727	2.7
P438	0.3	Q528	23.2	L731	0
L439	1.0	F529	0.4	C734	0
H440	24.1	A532	0	F735	0
P441	0	I533	1.3	P761	28
A443	0	V536	0.1	W762	11.1
M444	0	R539	8.6	L763	5.4
P445	7.1	A540	1.2	L764	0
L447	0.3	F541	4.2	C766	1.3
V449	0	V570	0	A 767	0
G450	0	F573	0.4	A768	0
S452	0.6	L574	0.6	W770	21.3
G453	0	L577	0.3	I771	10.5
L454	0.6	I579	1.3	S819	18.9
Y457	0.7	H580	34.3	R829	32.6
V458	12.5	C618	0.1		
A459	48.1	L622	0		

could be exposed by shifting the C-terminus of the spacer, and the spacer domain may overlay the antibody 5 epitope and be removed by chaperone activation.

We next evaluated the ability of the P model to provide mechanistic explanations for the effects of mutations affecting HBV RNA binding, DNA priming, or RNase H activity. We introduced these mutations into the full-

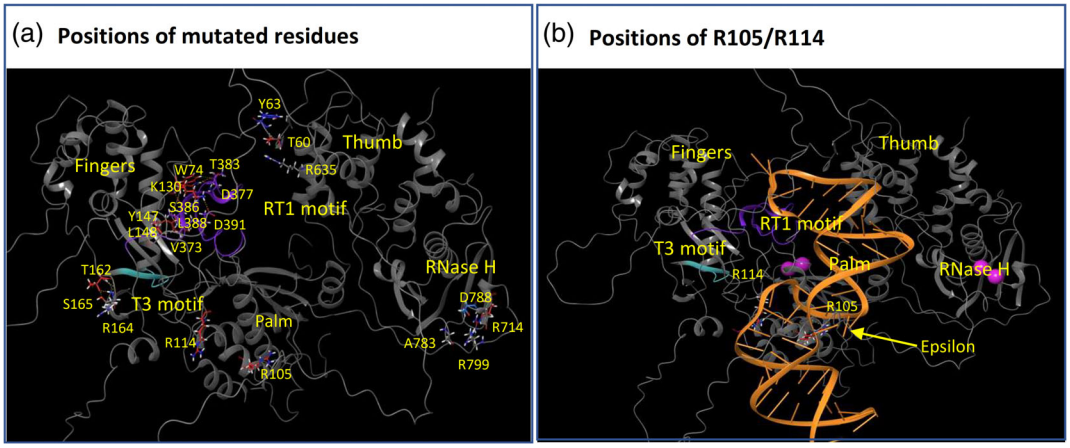

FIGURE 6 Location of motifs and residues used for assessing validity of the predicted model for full-length P. (a) Positions of the mutated residues in the context of the full-length genotype B model. (b) Positions of R105 and R114 in the context of the P model docked to the ε RNA stem loop. This image reflects only one of the three docking poses proposed for ε (Figure S29) and these interactions should be interpreted conservatively

length genotype B model, re-energy minimized the wild-type and mutant models using the Schrödinger OPSL4 force field, and identified polar interactions involving the residues. Limitations to the interpretation of the mutant phenotypes were anticipated because the extent of the functional limitations imposed on P by the closed conformation that the model appears to reflect is unknown.

Four sets of mutations in the TP that impair RNA binding and/or RNA packaging into viral capsids can be explained by the positions of the mutated residues (Figure 6). First, R105A is deficient in RNA packaging, R114E reduces RNA binding and packaging, and both mutations ablate protein priming and DNA synthesis [R105[50,51]; R114[24]]. The model places these residues close to each other on the outside of the fingers subdomain (Figure 6a,b), and docking studies predict that they may form salt bridges with the sugar-phosphate backbone of ε (Figures 7a and S30d,e). Mutating these residues removes these interactions, which would impair ε binding and subsequent DNA priming. Second, K130 in the TP domain is predicted to form hydrogen bonds with D391 and V373 in the RT1 motif, and to form a salt bridge with D391 (Figures 6a and 7b). It also binds to ε in docking pose 2 (Figure S29). K130L ablates these interactions (Figure S30f) and is defective in packaging RNA into viral capsids.[52] This would be expected if the orientation, flexibility, and/or exposure of RT1 were changed by disrupting K130:D391 binding, or alternatively, if K130 works together with RT1 to promote ε binding. Third, Y147 and L148 in the TP domain are adjacent to the T3 motif, and Y147 hydrogen bonds to S386 and L388 in the RT1 motif (Figure 7c). Y147A/L148A removes these contacts (Figure S30g) and is deficient in RNA packaging into capsids for both HBV and DHBV,[16,49] as would be anticipated if disrupting these interactions between the TP and RT1 impaired T3 and/or RT1 function. Finally, T162 is two residues C-terminal to the T3 motif in the TP domain. It is predicted to be the N terminal residue of a turn between two strands of a β-sheet. It hydrogen bonds with R164 (the last residue in the turn) and S165 (Figure 7d). T162P is defective in RNA packaging into capsids,[53] which requires T3-dependent binding to ε.[16] Inserting a proline disrupts the hydrogen bonds with R164 and S165 (Figure S30h), which may disrupt the local structure and/or impede conformational changes required for the specific binding of ε to P.

Two interactions predicted by the HBV P model can explain mutations defective in DNA priming. T60 is in the priming loop in the TP domain and is predicted to form hydrogen bonds with S64 in the TP domain and R653 in the RT domain (Figure 7e). T60E is defective in protein priming, but T60A has no phenotype.[24] The T60A mutation would lessen its affinity with R653 in the RT domain and retain flexibility of the priming loop (Figure S30a). In contrast, T60E would create a salt bridge with R653 (Figure S30b) which would increase affinity for R653 and impair folding of the priming loop downward, reducing Y63's ability to access the RT active site during priming. Second, W74 at the C-terminus of the priming loop hydrogen bonds with T383 and D377 within the RT1 motif to help place Y63 above the RT active site (Figure 7f). W74A removes these interactions (Figure S30c) and is deficient in protein priming but not RNA packaging, as expected if the mutation disrupted a network of interactions anchoring the priming loop in a position where Y63 can shift downward into the RT active site during priming.[50,54]

Finally, the RNase H-deficient R714A mutation[32] was assessed. R714 hydrogen bonds to the backbone of R799.

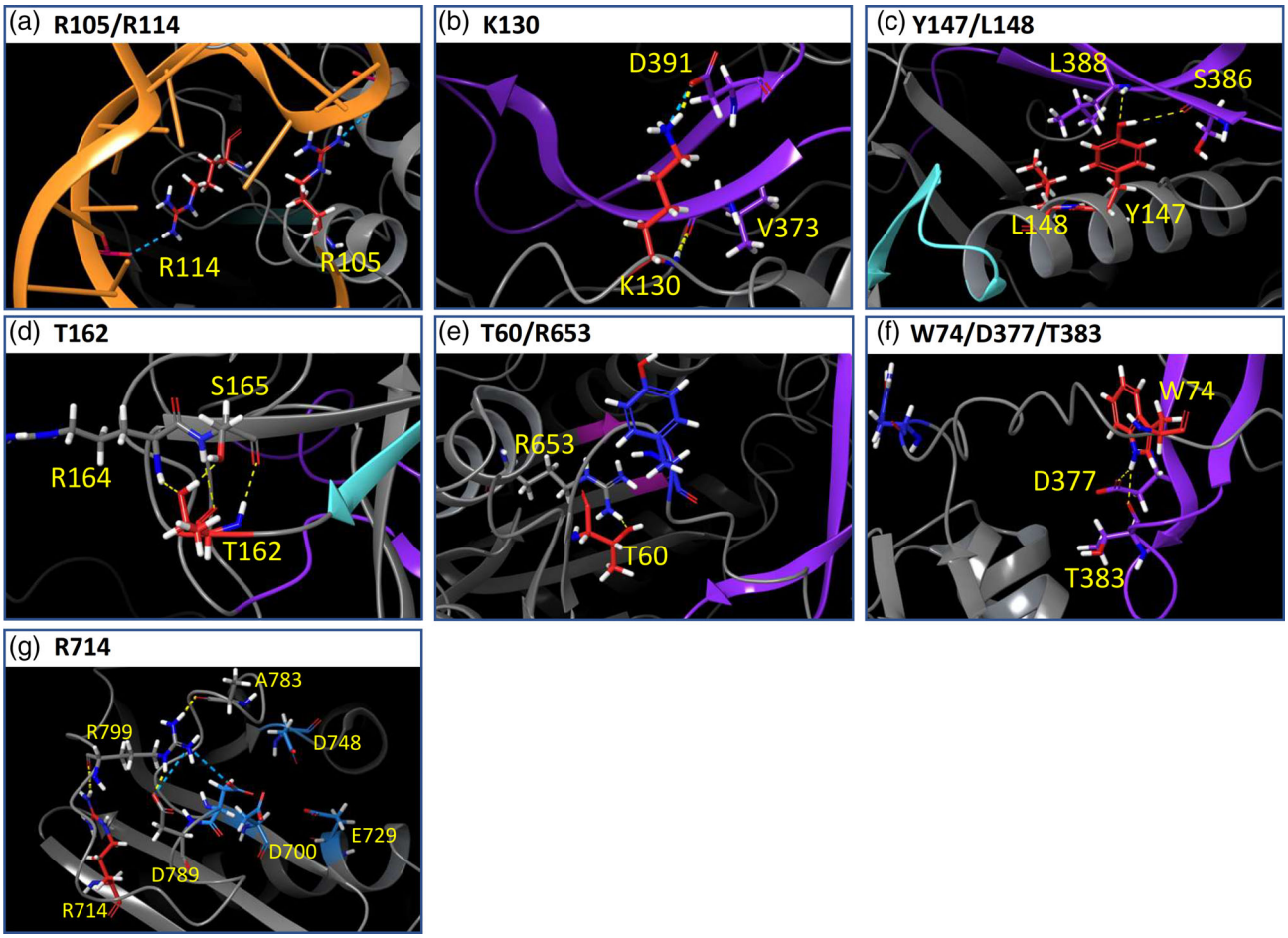

FIGURE 7 Interactions among motifs and residues used for assessing validity of the model for P. Interactions among HBV motifs and key residues are shown in the genotype B model. (a–g) Networks of polar interactions involving the wild-type residues at the indicated positions are indicated as colored dashed lines. Mutating the indicated residues yielded strong phenotypes as described in the text. Comparisons between the wild-type and mutant interaction networks are in Figure S30. Red, sites that were mutated and used for validation analyses; Orange, ε RNA with phosphates interacting with P shown; Violet, RT1 motif; Cyan, T3 motif; Blue, Y63 priming residue; Magenta, YMDD RT active site motif; Light blue, D-E-D-D motif. Blue dashed lines, salt bridges; Yellow dashed lines, hydrogen bonds

R799 also hydrogen bonds with A783, and D789. It also forms salt bridges with D788 and D789, the last residue in the RNase H D-E-D-D motif (Figure 7g). This region forms a loop in the HBV model but is an α-helix in most other RNase H enzymes. R714A removes a hydrogen bond at the base of this interaction network (Figure S30i) which would increase the flexibility of the loop. Shifting the loop containing D788 would inhibit RNase H activity by altering the location or conformation of the DEDD motif.

2.6 | Inhibitor docking

2.6.1 | Reverse transcriptase active site

Plausibility of the RT active site in the models was evaluated by docking the active triphosphate form of three nucleos(t)ide analog HBV reverse transcriptase inhibitors, Entecavir, Tenofovir, and Lamivudine[5] into the active site in the genotypes B and D catalytic core and full-length P models. A double-stranded DNA duplex from the HIV: substrate co-crystal (PDB: 1RTD) was superimposed on the HBV RT active site in the four models, and the inhibitors were docked using Glide XP within the Schrödinger Maestro suite. The inhibitors bound in the expected poses in all four models, with the α and β phosphates coordinated by the active site Mg^{2+} ions, and the inhibitor pairing with the template strand of the primer-template (Figure 7a). Docking energies ranged from −8.86 to −13.3 kCal/mol (Table 4), correspond to K_d values of 0.32 to 0.00018 μM. The near identity of the AlphaFold model to the prior RT domain models (e.g., Das et al.[26]) in the RT active site indicates that the ability of the older RT models to reveal mechanisms of drug resistance mutations is retained by the AlphaFold model of P.

TABLE 4 Docking scores for the most stable poses of compounds into the RT and RNase H active sites

Compound[a]	Class[b]	gtB RT-RNase H	gtB P	gtD RT-RNase H	gtD P
RT active site[c]					
Entecavir-TP	NA	−13.31	−13.01	−12.92	−10.42
Tenofovir-TP	NA	−11.05	−8.86	−9.27	−12.96
Lamivudine-TP	NA	−11.76	−8.94	−9.61	−11.31
RNase H active site[c]					
208	HPD	−10.82	−9.45	−9.58	−9.38
A25		−10.69	−9.71	−9.92	−10.74
110	αHT	−7.88	−7.78	−7.64	−7.80
404		−8.64	−8.72	−8.31	−8.47
12	HNO	−7.44	−8.12	−5.10	−7.46
1,073		−7.72	−7.73	−7.39	−7.64

[a]Compounds described in.[8,28,55–58]
[b]NA, nucleos(t)ide analog; HPD, *N*-hydroxypyridinedione; αHT, α-hydroxytropolone; HNO, *N*-hydroxynapthyridinone.
[c]Values in kCal/mol.

2.6.2 | Ribonuclease H active site

The HBV RNase H active site was validated by docking inhibitors into the active site in the genotype B and D catalytic core and full-length models using Glide XP. Inhibitors were α-hydroxytropolones (compounds **110** and **404**), *N*-hydroxypyridinediones (**208** and **A25**), and *N*-hydroxynapthyridinones (**12** and **1,073**).[8,28,55–58] All four models docked HBV RNase H inhibitors in poses where the compounds coordinated the Mg^{2+} ions via their trident of metal chelating atoms (Figure 7b), as predicted from the experimentally determined binding poses of similar compounds against the HIV RNase H[59–63] and the need for an intact metal chelating trident for the inhibitors to work against HBV. The compounds adopted many binding poses as anticipated from their structural diversity, usually along the nucleic acid binding groove, with docking energies of −5.1 to −10.8 kCal/mol (Table 4), corresponding to K_d values of 182 to 0.012 μM.

2.7 | Effects of HBV's genetic diversity on the models

HBV has nine genotypes[2] with P proteins being 832–845 aa long and differing by 11–16% at the aa level (Table S1). To evaluate how these variations may affect P's structure, models for P proteins from all genotypes were generated using AlphaFold (Figures S7–S15) and compared to the genotype B model. All models had similar pLDDT score profiles, with the C-terminal half of the TP domain and the catalytic core having high pLDDT values (~70–90), with much lower pLDDT values for the spacer domain and the N- and C-termini. Superpositioning all full-length models against the genotype B model revealed the same overall fold, with an average pairwise RMSD of 1.98 Å (1.67–2.16 Å) outside the highly variable spacer domain (Figure S25a–h). The key features of the genotype B model were conserved in all models, including cupping of the TP domain around the catalytic core, positioning the priming tyrosine residue above the YMDD active site motif, the orientation of the T3 and RT1 motifs, and the orientation of the RT and RNase H domains (Figures S7–S15).

2.8 | Models of animal P proteins and non-retroviral reverse transcriptases

To evaluate phylogenetic conservation of the HBV P fold, models were constructed for P proteins from animal hepadnaviruses[2,64] including woodchuck hepatitis virus (WHV, rodent), DHBV (bird), skink HBV (SkHBV, reptile), Tibetan frog HBV (TFHBV, amphibian), and tetra metahepadnavirus (TMDV, fish; Figure 8, Table S1, and Figures S16–S20). These viruses share the same four-domain structure of P and replicate by protein-primed reverse transcription.[64] We also generated a model for P from the rockfish nackednavirus (RNDV; Figure S21). Nackednaviruses and hepadnaviruses diverged about 400 million years ago, before the lineage that became the hepadnaviruses acquired its surface glycoproteins.[64] Consequently, nackednavirus P lacks the spacer domain that encodes part of the surface protein gene in a different reading frame. The priming mechanism is conserved between HBV, DHBV, and RNDV because their P

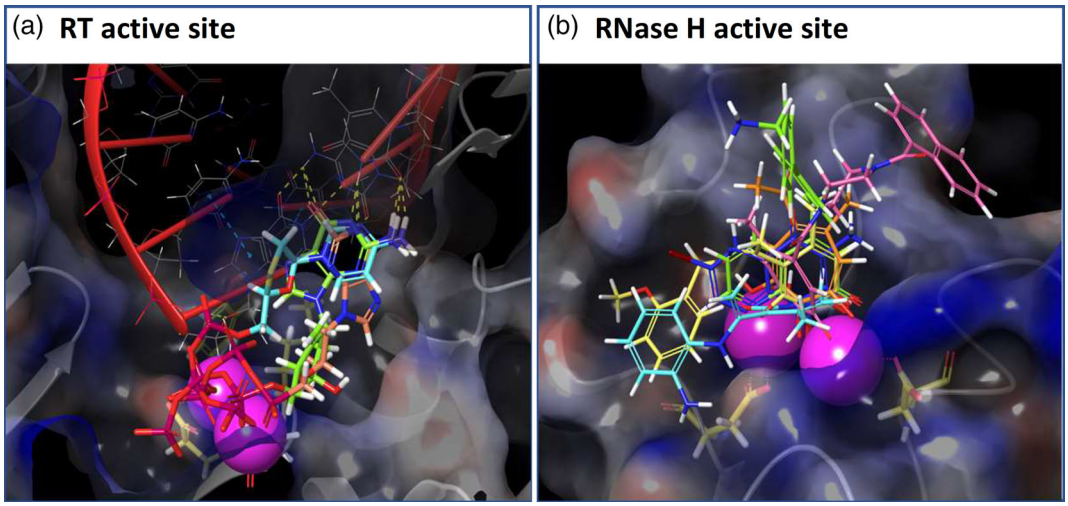

FIGURE 8 Computational docking of compounds into the RT and RNase H active sites. Compounds were docked using the Glide XP program within the Schrödinger molecular analysis suite. Magenta spheres, Mg^{2+} ions. (a) Triphosphate form of nucleos(t)ide analog drugs docked into the RT active site. Red; DNA duplex; Orange, Tenofovir triphosphate; Green, Entecavir triphosphate; Cyan, Lamivudine triphosphate. (b) RNase H inhibitors docked into the RNase H active site. Yellow, Compound **A25**; Cyan, **208**; Pink, **404**; Orange, **110**; Blue, **1,073**; Green, **12**

proteins can prime DNA synthesis using ε RNAs from the other viruses with only minor nucleotide substitutions in ε.[65]

All animal virus P proteins shared the same basic predicted structure, with the catalytic core of the RT and RNase H domain forming a globular unit featuring a contiguous nucleic acid binding groove, the TP domain cupping around the catalytic core, and the priming tyrosine residue poised over the RT active site (Figures 9 and S16–S21). All models had the A-E DNA polymerase active site motifs and the D-E-D-D RNase H motif in positions analogous to their locations in the HBV P model. The T3 and RT1 motifs in all six animal P protein models were in the same relative position as in HBV P. As expected, the spacer domain in the animal hepadnaviruses was unstructured, and RNDV had only a very short sequence linking the TP to the RT domain (Figure 9d). The genotype B HBV P model superimposed well with the RNDV RT model (RMSD = 2.76 Å; Figure S27d), with the largest differences being in the flexible priming loop, the tip of the fingers subdomain, and position of the RNase H domain (Figure S27d).

These predicted structures imply that the hepadnaviral P fold existed before the split between the nackednaviruses and hepadnaviruses 400 million years ago.[64] The models also imply that the acquisition of an envelope in the hepadnaviruses included insertion of sequences for part of the surface glycoproteins into sequences encoding residues corresponding to 174–192 of the modern nackednavirus RNDV.

Finally, we compared the predicted hepadnaviral P fold to models we generated for reverse transcriptases from cauliflower mosaic virus (CaMV)[66,67] and the mitochondrial retrotransposon pFOXC3 from the fungus *Fusarium oxysporum* f. sp. Matthiolae[68,69] (Figures S22 and S23). CaMV primes reverse transcription using a tRNA similar to the retroviruses, whereas the pFOXC3 RT primes reverse transcription using an unidentified tyrosine residue in a mechanism similar to the hepadnaviruses. The predicted folds for the CaMV and pFOXC3 RTs revealed a nucleic acid binding groove and YVDD (CaMV) and YADD (pFOXC3) motifs analogous to the hepadnaviral YMDD RT active site motif in their expected positions. Neither the CaMV nor pFOXC3 models had domains analogous to the TP or spacer domains. The pFOXC3 model aligned to the HBV RT-RNase H domains in the full-length model with an RMSD = 4.52 Å (Figure S27e) and positioned Y35 in a plausible position for priming reverse transcription. The pFOXC3 model did not have a readily identifiable RNase H domain or an identifiable D-E-D-D RNase H active site motif. The HBV RT and RNase H domains aligned with the CaMV RT-RNase H domains with an RMSD = 3.96 Å (Figure S27g). The CaMV RNase H active site was predicted to contain a D-E-E-D motif rather than D-E-D-D, and the CaMV N-terminal aspartic proteinase formed a globular domain absent in the hepadnaviral proteins. Overall, these models imply that the hepadnaviral P fold is a feature of the nackednavirus/hepadnavirus lineages rather than being widespread among non-retroviral eukaryotic RTs.

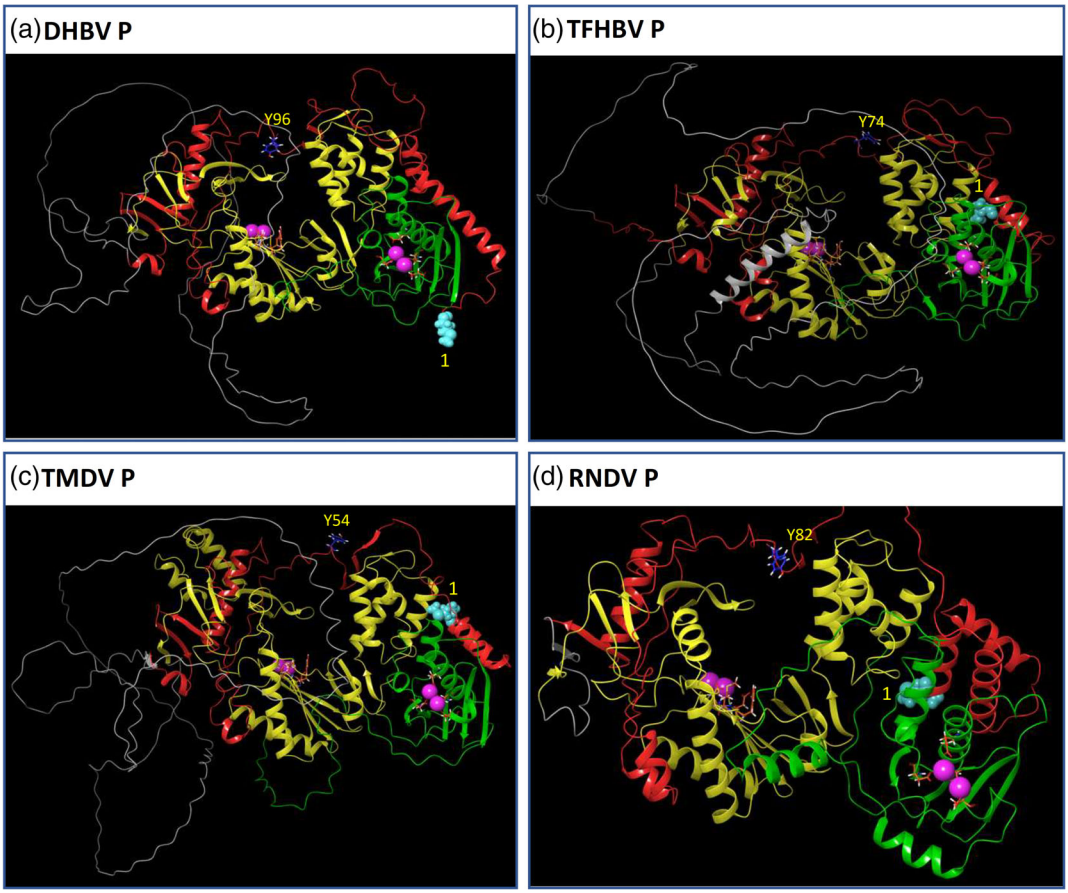

FIGURE 9 Predicted folds of animal hepadnavirus and nackednavirus P proteins. (a) DHBV (avian hepadnavirus). (b) TFHBV (frog hepadnavirus). (c) TMDV (fish hepadnavirus). (d) RNDV (fish nackednavirus). Red, TP domain; Gray, spacer domain; Yellow, RT domain; Green, RNase H domain. The priming tyrosine residues, YMDD RT active site motif residues, and D-E-D-D RNase H active site residues are shown as sticks. Putative priming residues are labeled. Magenta spheres, Mg^{2+} ions. The amino termini are marked with cyan spheres

2.9 | Limitations

There are two limitations to this analysis. First, all models are predictions. The models are well supported by the available biological and biochemical data, but they are not based on experimental structural data. Second, P is structurally dynamic, so there is no single structure that can represent all of its biologically relevant conformations. The full-length P models correspond best to the *closed conformation*,[16] but they could represent an average of multiple conformations adopted by P.

2.10 | Utility of the models

HBV P has resisted empirical structural analyses for decades due to intractable protein production problems, so these models provide the first structural approximation for the whole enzyme. They reveal an orientation of the TP domain relative to the rest of the enzyme that makes clear predictions for ε binding, DNA priming, and conformational dynamics of P during viral replication. They enable formulation of detailed hypotheses regarding the mechanisms of DNA elongation during reverse transcription and how the RT and RNase H domains coordinate during reverse transcription. Finally, the models can support structure-guided drug design against the RT and RNase H active sites, and they open the door to rational drug discovery against targets other than the enzymatic active sites.

3 | MATERIALS AND METHODS

3.1 | Sequences and protein structures

Protein sequences are in Dataset S1. Table S1 contains Genbank numbers or literature references for the sequences plus their pairwise identities with the genotype B reference sequence (AB554017). Appendix S1 lists the protein domain and motif boundaries employed. The Tibetan frog hepatitis B virus (TFHBV) sequence lacks an

in-frame ATG for P, so the start site was set to match the N-terminus of HBV genotype B P.

3.2 | Molecular modeling with AlphaFold

Sequences were folded using AlphaFold2 Advanced (https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/beta/AlphaFold2_advanced.ipynb) with default parameters. Five models were generated for each sequence, and side chain positions for the model having the highest pLDDT value were refined by the Amber-Relax module. The relaxed structures were used for all analyses. Predicted protein database (PDB) files are in Dataset S2.

Hydrophobic surfaces on HBV P gtB model were analyzed using the Protein Surface Analyzer program within the Schrödinger suite. Residues at the interfaces between the TP domain and the catalytic core of HBV P were manually identified. Solvent accessibility of each residue was measured using the web server GETAREA (http://curie.utmb.edu/getarea.html). Residues having ratio values >50% were considered to be solvent exposed, whereas residues with the ratio value <20% were defined as buried. The Schrödinger Protein Preparation Wizard was used to identify hydrogen bonds and salt bridges. Hydrogen bonds and salt bridges at the interface of TP domain and catalytic core of HBV polymerase were then manually identified.

3.3 | Superpositions

Superpositioning was done using Protein Structure Alignment in the Maestro module of the Schrödinger suite (Schrödinger LLC, New York, NY). Superpositioning employed no constraints except that the HBV P and catalytic core superpositions with HIV RT-RNase H, Ty3 RT, and CaMV RT were done by aligning selected residues in fingers, palm, and thumb subdomains of HBV RT domain due to the lack of the linker region in HBV P. Graphs for the RMSD distributions were determined using Stamp Structural Alignment utility in the Visual Molecular Dynamics (VMD) software followed by plotting data each residue's RMSD values against the amino acid alignment generated by the superpositioning. In some cases, structure alignments were performed by deleting nonaligned residues to improve the alignment of similar regions between two structures.

3.4 | Nucleic acid docking

A 21 bp heteroduplex (PDB: 4B3Q) and a 24 bp heteroduplex (PDB: 6BSH) were extracted from HIV RT-RNase H co-crystals and docked into HBV P using the PIPER program in the BioLuminate module of the Schrödinger suite. Protonation state was set using Epik at pH 7.5 ± 2, hydrogen bonds were assigned employing PROPKA at pH 7.5, and energy minimization was done with the OPSL4 force field. Models were docked with both substrates separately without constraints. Final poses were selected based on substrate placement into the binding channel between active sites of RT and RNase H domains while considering the engagement of RNA or DNA strands with active site residues. Similarly, an NMR structure for ε (PDB: 6VAR) was docked into HBV P. Protein and nucleic acid structures were prepared with the Protein Preparation wizard in the BioLuminate module.

3.5 | Compound docking

Compound docking into the models employed Glide XP in the Schrödinger Maestro module. Triphosphates for the nucleos(t)ide analogs Entecavir, Lamivudine, and Tenofovir were extracted from HIV-1 RT-RNase H:inhibitor co-crystals (PDB: 5XN1, 6KDJ, and 3JSM) for docking. RNase H inhibitors were selected for structural diversity. Ligands were prepared with LigPrep (Schrödinger LLC), where compound energy minimizations were done with the OPSL4 force field, protonation states of the ligands were defined using Epik at pH 7.5 +/−2 to set the ionization state of the metal binding motifs, and compounds were desalted and tautomerized while retaining chirality. The RT-RNase H and full-length P models containing Mg^{2+} ions in the appropriate ionization states were prepared with the Schrödinger Protein Preparation wizard as described above. A 20 bp double-stranded DNA was placed into the RT active site of the models for nucleos(t)ide analog docking by superposition of HIV RT-RNase H:dsDNA co-crystal (PDB: 1RTD). The RT active site docking grid was defined by placing Entecavir-triphosphate from the HIV RT-RNase H structure co-crystal (PDB: 5XN1) into the RT active site of the HBV models by superposing the YMDD motifs. The RNase H docking grid was defined by placing β-thujaplicinol into the active site by superposing the D-E-D-D motif from the HIV RNase H:β-thujaplicinol co-crystal (PDB: 3K2P) onto the HBV model. The centroids of the ligands in the active sites were used to create 10 Å receptor grids for docking. Docking employed Schrödinger Glide XP with default settings.

3.6 | Mutations employed for model validation

Mutations to HBV P were identified from the literature, especially.[24] Mutations were mapped onto the full-length

HBV genotype B P model (genotype B numbers are reported, which may differ from the source if a different genotype was used). Hydrogen bonds were identified using PyMOL's *Find polar contacts* function (h-bond cutoff center = 4.0 Å); all predicted polar interactions assessed were ≤3.0 Å, implying moderate to strong interactions.

AUTHOR CONTRIBUTIONS

Razia Tajwar: Conceptualization (equal); data curation (equal); formal analysis (equal); investigation (equal); methodology (lead); project administration (equal); resources (equal); software (lead); supervision (supporting); validation (equal); visualization (lead); writing – original draft (supporting); writing – review and editing (equal). **Daniel P. Bradley:** Conceptualization (supporting); data curation (supporting); formal analysis (equal); methodology (supporting); resources (supporting); software (supporting); validation (equal); visualization (supporting); writing – original draft (supporting); writing – review and editing (equal). **Nathan L. Ponzar:** Data curation (supporting); formal analysis (supporting); methodology (supporting); resources (supporting); visualization (supporting); writing – original draft (supporting). **John Tavis:** Conceptualization (lead); data curation (equal); formal analysis (equal); funding acquisition (lead); investigation (equal); methodology (equal); project administration (lead); resources (equal); software (equal); supervision (lead); validation (equal); visualization (equal); writing – original draft (lead); writing – review and editing (lead).

ACKNOWLEDGEMENTS

This work was supported by NIH grants R01 AI150610 and R01 AI148362 and Department of Defense grant W81XWH-18-1-0307 to John E. Tavis. We thank Dr John Kennell for input on the pFOX-C3 reverse transcriptase, Dr Juan Villa for early modeling work on the HBV RNase H, and the AlphaFold team for enabling this study. A version of this article is posted on the BioRxiv preprint server under the Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) license.

CONFLICT OF INTEREST

The authors have no conflicts to declare.

ORCID

John E. Tavis https://orcid.org/0000-0002-8711-4240

REFERENCES

1. Seeger C, Zoulim F, Mason WS. Hepadnaviridae. Fields virology, Vol 2: DNA viruses. Philadelphia, PA: Wolters Kluwer, 2021; p. 640–682.
2. Glebe D, Goldmann N, Lauber C, Seitz S. HBV evolution and genetic variability: Impact on prevention, treatment and development of antivirals. Antiviral Res. 2021;186:104973.
3. Polaris Observatory Collaborative. Global prevalence, treatment, and prevention of hepatitis B virus infection in 2016: A modelling study. Lancet Gastroenterol Hepatol. 2018;3:383–403.
4. Trepo C, Chan HL, Lok A. Hepatitis B virus infection. Lancet. 2014;384:2053–2063.
5. Pierra Rouviere C, Dousson CB, Tavis JE. HBV replication inhibitors. Antiviral Res. 2020;179:104815.
6. Ghany MG. Current treatment guidelines of chronic hepatitis B: The role of nucleos(t)ide analogues and peginterferon. Best Pract Res Clin Gastroenterol. 2017;31:299–309.
7. Poch O, Sauvaget I, Delarue M, Tordo N. Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. EMBO J. 1989;8:3867–3874.
8. Tavis JE, Cheng X, Hu Y, et al. The hepatitis B virus ribonuclease H is sensitive to inhibitors of the human immunodeficiency virus ribonuclease H and integrase enzymes. PLoS Pathog. 2013;9:e1003125.
9. Yang W, Steitz TA. Recombining the structures of HIV integrase, RuvC and RNase H. Structure. 1995;3:131–134.
10. Nowotny M. Retroviral integrase superfamily: The structural perspective. EMBO Rep. 2009;10:144–151.
11. Villa JA, Pike DP, Patel KB, et al. Purification and enzymatic characterization of the hepatitis B virus ribonuclease H, a new target for antiviral inhibitors. Antiviral Res. 2016;132:186–195.
12. Tavis JE, Badtke MP. Hepadnaviral genomic replication. In: Cameron CE, Goette M, Raney KD, editors. Viral genome replication. New York: Springer Science+Business Media, LLC, 2009; p. 129–143.
13. Beck J, Nassal M. Hepatitis B virus replication. World J Gastroenterol. 2007;13:48–64.
14. Hu J, Toft DO, Seeger C. Hepadnavirus assembly and reverse transcription require a multi-component chaperone complex which is incorporated into nucleocapsids. EMBO J. 1997;16:59–68.
15. Hu J, Flores D, Toft D, Wang X, Nguyen D. Requirement of heat shock protein 90 for human hepatitis B virus reverse transcriptase function. J Virol. 2004;78:13122–13131.
16. Badtke MP, Khan I, Cao F, Hu J, Tavis JE. An interdomain RNA binding site on the hepadnaviral polymerase that is essential for reverse transcription. Virology. 2009;390:130–138.
17. Lanford RE, Notvall L, Lee H, Beams B. Transcomplementation of nucleotide priming and reverse transcription between independently expressed TP and RT domains of the hepatitis B virus reverse transcriptase. J Virol. 1997;71:2996–3004.
18. Loeb DD, Hirsch RC, Ganem D. Sequence-independent RNA cleavages generate the primers for plus strand DNA synthesis in hepatitis B viruses: Implications for other reverse transcribing elements. EMBO J. 1991;10:3533–3540.
19. Zhang Z, Tavis JE. The duck hepatitis B virus reverse transcriptase functions as a full-length monomer. J Biol Chem. 2006;281:35794–35801.
20. Cao F, Badtke MP, Metzger LM, et al. Identification of an essential molecular contact point on the duck hepatitis B virus reverse transcriptase. J Virol. 2005;79:10164–10170.
21. Stahl M, Beck J, Nassal M. Chaperones activate hepadnavirus reverse transcriptase by transiently exposing a C-proximal

region in the terminal protein domain that contributes to epsilon RNA binding. J Virol. 2007;81:13354–13364.
22. Wang X, Hu J. Distinct requirement for two stages of protein-primed initiation of reverse transcription in hepadnaviruses. J Virol. 2002;76:5857–5865.
23. Lin L, Wan F, Hu J. Functional and structural dynamics of hepadnavirus reverse transcriptase during protein-primed initiation of reverse transcription: Effects of metal ions. J Virol. 2008;82:5703–5714.
24. Clark DN, Flanagan JM, Hu J. Mapping of functional subdomains in the terminal protein domain of hepatitis B virus polymerase. J Virol. 2017;91:e01785-16.
25. Buhlig TS, Bowersox AF, Braun DL, et al. Molecular, evolutionary, and structural analysis of the terminal protein domain of hepatitis B virus polymerase, a potential drug target. Viruses. 2020;12:570.
26. Das K, Xiong X, Yang H, et al. Molecular modeling and biochemical characterization reveal the mechanism of hepatitis B virus polymerase resistance to lamivudine (3TC) and emtricitabine (FTC). J Virol. 2001;75:4771–4779.
27. Xu X, Thai H, Kitrinos KM, et al. Modeling the functional state of the reverse transcriptase of hepatitis B virus and its application to probing drug-protein interaction. BMC Bioinformatics. 2016;17(Suppl 8):280.
28. Li Q, Lomonosova E, Donlin MJ, et al. Amide-containing alpha-hydroxytropolones as inhibitors of hepatitis B virus replication. Antiviral Res. 2020;177:104777.
29. Potenza N, Salvatore V, Raimondo D, et al. Optimized expression from a synthetic gene of an untagged RNase H domain of human hepatitis B virus polymerase which is enzymatically active. Protein Expr Purif. 2007;55:93–99.
30. Hayer J, Rodriguez C, Germanidis G, et al. Ultradeep pyrosequencing and molecular modeling identify key structural features of hepatitis B virus RNase H, a putative target for antiviral intervention. J Virol. 2014;88:574–582.
31. Hyjek M, Figiel M, Nowotny M. RNases H: Structure and mechanism. DNA Repair (Amst). 2019;84:102672.
32. Ko C, Shin YC, Park WJ, Kim S, Kim J, Ryu WS. Residues Arg703, Asp777, and Arg781 of the RNase H domain of hepatitis B virus polymerase are critical for viral DNA synthesis. J Virol. 2014;88:154–163.
33. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596:583–589.
34. Mariani V, Biasini M, Barbato A, Schwede T. lDDT: A local superposition-free score for comparing protein structures and models using distance difference tests. Bioinformatics. 2013;29:2722–2728.
35. Donlin MJ, Szeto B, Gohara DW, Aurora R, Tavis JE. Genome-wide networks of amino acid covariances are common among viruses. J Virol. 2012;86:3050–3063.
36. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc. 2015;10:845–858.
37. Das D, Georgiadis MM. The crystal structure of the monomeric reverse transcriptase from Moloney murine leukemia virus. Structure. 2004;12:819–829.
38. Sarafianos SG, Marchand B, Das K, et al. Structure and function of HIV-1 reverse transcriptase: Molecular mechanisms of polymerization and inhibition. J Mol Biol. 2009;385:693–713.
39. Lanford RE, Kim YH, Lee H, Notvall L, Beames B. Mapping of the hepatitis B virus reverse transciptase TP and RT domains by transcomplementation for nucleotide priming and by protein-protein interaction. J Virol. 1999;73:1885–1893.
40. Beck J, Nassal M. Reconstitution of a functional duck hepatitis B virus replication initiation complex from separate reverse transcriptase domains expressed in Escherichia coli. J Virol. 2001;75:7410–7419.
41. Boregowda RK, Adams C, Hu J. TP-RT domain interactions of duck hepatitis B virus reverse transcriptase in cis and in trans during protein-primed initiation of DNA synthesis in vitro. J Virol. 2012;86:6522–6536.
42. Tavis JE, Ganem D. Expression of functional hepatitis B virus polymerase in yeast reveals it to be the sole viral protein required for correct initiation of reverse transcription. Proc Natl Acad Sci USA. 1993;90:4107–4111.
43. Beck J, Nassal M. Efficient Hsp90-independent in vitro activation by Hsc70 and Hsp40 of duck hepatitis B virus reverse transcriptase, an assumed Hsp90 client protein. J Biol Chem. 2003;278:36128–36138.
44. Hu J, Toft D, Anselmo D, Wang X. In vitro reconstitution of functional hepadnavirus reverse transcriptase with cellular chaperone proteins. J Virol. 2002;76:269–279.
45. Huang H, Chopra R, Verdine GL, Harrison SC. Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: Implications for drug resistance. Science. 1998;282:1669–1675.
46. Ha B, Larsen KP, Zhang J, et al. High-resolution view of HIV-1 reverse transcriptase initiation complexes and inhibition by NNRTI drugs. Nat Commun. 2021;12:2500.
47. Sarafianos SG, Das K, Tantillo C, et al. Crystal structure of HIV-1 reverse transcriptase in complex with a polypurine tract RNA:DNA. EMBO J. 2001;20:1449–1461.
48. LeBlanc RM, Kasprzak WK, Longhini AP, et al. Structural insights of the conserved "priming loop" of hepatitis B virus pre-genomic RNA. J Biomol Struct Dyn. 2021;1–13. https://doi.org/10.1080/07391102.2021.1934544
49. Cao F, Jones S, Li W, et al. Sequences in the terminal protein and reverse transcriptase domains of the hepatitis B virus polymerase contribute to RNA binding and encapsidation. J Viral Hepat. 2014;21:882–893.
50. Jones SA, Clark DN, Cao F, Tavis JE, Hu J. Comparative analysis of hepatitis B virus polymerase sequences required for viral RNA binding, RNA packaging, and protein priming. J Virol. 2014;88:1564–1572.
51. Shin YC, Park S, Ryu WS. A conserved arginine residue in the terminal protein domain of hepatitis B virus polymerase is critical for RNA pre-genome encapsidation. J Gen Virol. 2011;92:1809–1816.
52. Roychoudhury S, Faruqui AF, Shih C. Pregenomic RNA encapsidation analysis of eleven missense and nonsense polymerase mutants of human hepatitis B virus. J Virol. 1991;65:3617–3624.
53. Blum HE, Galun E, Liang TJ, von Weizsacker F, Wands JR. Naturally occurring missense mutation in the polymerase gene terminating hepatitis B virus replication. J Virol. 1991;65:1836–1842.
54. Shin YC, Ko C, Ryu WS. Hydrophobic residues of terminal protein domain of hepatitis B virus polymerase contribute to

distinct steps in viral genome replication. FEBS Lett. 2011;585: 3964–3968.
55. Edwards TC, Mani N, Dorsey B, et al. Inhibition of HBV replication by N-hydroxyisoquinolinedione and N-hydroxypyridinedione ribonuclease H inhibitors. Antiviral Res. 2019;164:70–80.
56. Edwards TC, Lomonosova E, Patel JA, et al. Inhibition of hepatitis B virus replication by N-hydroxyisoquinolinediones and related polyoxygenated heterocycles. Antiviral Res. 2017;143: 205–217.
57. Lu G, Lomonosova E, Cheng X, et al. Hydroxylated tropolones inhibit hepatitis B virus replication by blocking the viral ribonuclease H activity. Antimicrob Agents Chemother. 2015;59: 1070–1079.
58. Chauhan R, Li Q, Woodson ME, Gasonoo M, Meyers MJ, Tavis JE. Efficient inhibition of hepatitis B virus (HBV) replication and cccDNA formation by HBV ribonuclease H inhibitors during infection. Antimicrob Agents Chemother. 2021;65: e0146021.
59. Lansdon EB, Liu Q, Leavitt SA, et al. Structural and binding analysis of pyrimidinol carboxylic acid and N-hydroxy quinazolinedione HIV-1 RNase H inhibitors. Antimicrob Agents Chemother. 2011;55:2905–2915.
60. Kirschberg TA, Balakrishnan M, Squires NH, et al. RNase H active site inhibitors of human immunodeficiency virus type 1 reverse transcriptase: Design, biochemical activity, and structural information. J Med Chem. 2009;52:5781–5784.
61. Himmel DM, Maegley KA, Pauly TA, et al. Structure of HIV-1 reverse transcriptase with the inhibitor beta-Thujaplicinol bound at the RNase H active site. Structure. 2009;17:1625–1635.
62. Chung S, Himmel DM, Jiang JK, et al. Synthesis, activity, and structural analysis of novel alpha-hydroxytropolone inhibitors of human immunodeficiency virus reverse transcriptase-associated ribonuclease H. J Med Chem. 2011;54:4462–4473.
63. Billamboz M, Bailly F, Lion C, et al. 2-hydroxyisoquinoline-1,3 (2H,4H)-diones as inhibitors of HIV-1 integrase and reverse transcriptase RNase H domain: Influence of the alkylation of position 4. Eur J Med Chem. 2011;46:535–546.
64. Lauber C, Seitz S, Mattei S, et al. Deciphering the origin and evolution of hepatitis B viruses by means of a family of non-enveloped fish viruses. Cell Host Microbe. 2017;22:387–399.e6.
65. Beck J, Seitz S, Lauber C, Nassal M. Conservation of the HBV RNA element epsilon in nackednaviruses reveals ancient origin of protein-primed reverse transcription. Proc Natl Acad Sci USA. 2021;118:e2022373118.
66. Haas M, Bureau M, Geldreich A, Yot P, Keller M. Cauliflower mosaic virus: Still in the news. Mol Plant Pathol. 2002;3: 419–429.
67. Volovitch M, Modjtahedi N, Yot P, Brun G. RNA-dependent DNA polymerase activity in cauliflower mosaic virus-infected plant leaves. EMBO J. 1984;3:309–314.
68. Walther TC, Kennell JC. Linear mitochondrial plasmids of *F. oxysporum* are novel, telomere-like retroelements. Mol Cell. 1999;4:229–238.
69. Galligan JT, Marchetti SE, Kennell JC. Reverse transcription of the pFOXC mitochondrial retroplasmids of *Fusarium oxysporum* is protein primed. Mob DNA. 2011;2:1.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Tajwar R, Bradley DP, Ponzar NL, Tavis JE. Predicted structure of the hepatitis B virus polymerase reveals an ancient conserved protein fold. Protein Science. 2022; 31(10):e4421. https://doi.org/10.1002/pro.4421