Supporting Information

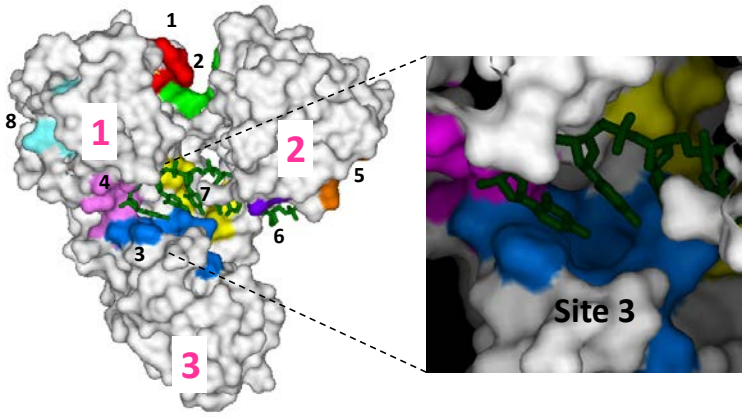
Integrated Strategies for Identifying Leads that Target the NS3 Helicase of the Hepatitis C Virus

Steven R. LaPlante*,[1] Anil K. Padyana,[2] Asitha Abeywardnae,[2] Pierre Bonneau,[1] Mireille Cartier,[1] René Coulombe,[1] Araz Jakalian,[1] Jessi W. Jones,[3] Xiang Li,[2] Shuang Liang,[2] Ginette McKercher,[1] Peter White,[1] Qiang Zhang,[2] Steven J. Taylor[2]

Table of contents:

S1: Bioinformatics and pocket variant analysis.

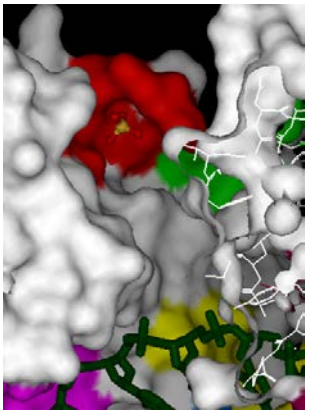
In Silico PASS analysis of NS3 helicase domains to predict potential pockets and binding sites



See the main text for an explanation of the data and columns.

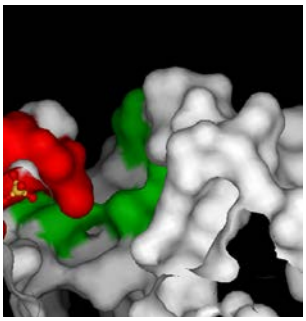
PASS site 1
ATP phosphate binding site

PASS Site	Residue	All genotype	Genotype 1 only	flexibility
1	T206	361/361	216/216	+
	G207	361/361	216/216	+
	S208	360/361	215/216	+
	G209	361/361	216/216	+
	K210	359/361	216/216	-
	S211	359/361	214/216	+
	T212	359/361	214/216	-



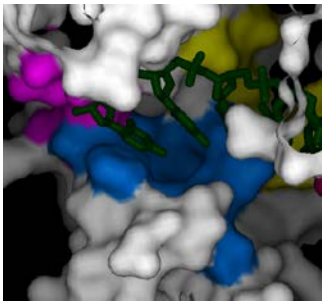
PASS site 2
Domain 1 and 2 cleft

PASS Site	Residue	All genotype	Genotype 1 only	flexibility
2	A323	359/361	216/216	-
	T324	359/361	215/216	+
	V331	315/361	215/216	-
	T459	261/361	171/216	-
	R462	361/361	216/216	-



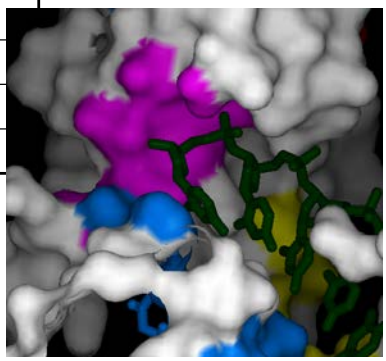
PASS site 3
RNA binding site on domain 3

PASS Site	Residue	All genotype	Genotype 1 only	flexibility
3	E493	360/361	215/216	+
	D496	360/361	215/216	-
	A497	346/361	214/216	-
	W501	361/361	216/216	-
	E555	149/361	51/216	++
	N556	356/361	214/216	+
	F557	296/361	196/216	-
	P558	308/361	216/216	-



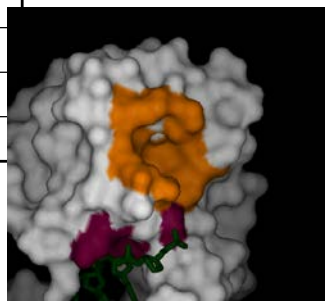
PASS site 4
RNA binding on domain 1

PASS Site	Residue	All genotype	Genotype 1 only	flexibility
4	G255	361/361	216/216	-
	T269	359/361	215/216	+
	G271	357/361	215/216	-
	K272	360/361	216/216	+
	A275	361/361	216/216	-
	Y502	359/361	215/216	-



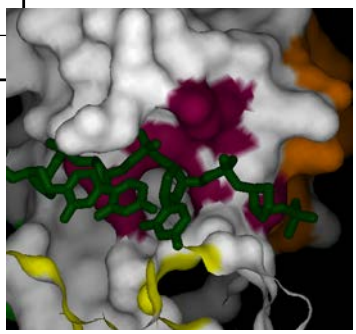
PASS site 5
domain 2 pocket

PASS Site	Residue	All genotype	Genotype 1 only	flexibility
5	P348	359/361	215/216	-
	F349	360/361	216/216	-
	Y350	361/361	216/216	-
	K373	361/361	216/216	+
	E376	357/361	214/216	-
	L377	355/361	214/216	-
	K380	303/361	212/216	+



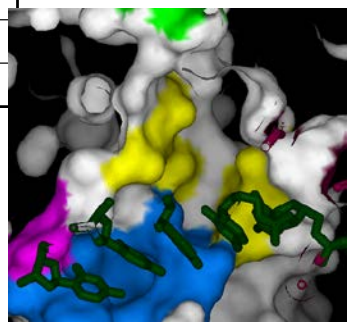
PASS site 6
RNS binding on domain 2

PASS Site	Residue	All genotype	Genotype 1 only	flexibility
6	H369	359/361	214/216	+
	K371	314/361	174/216	++
	T411	360/361	216/216	+
	D412	359/361	214/216	-



PASS site 7:
RNA binding at the domain 1 & 2 interface

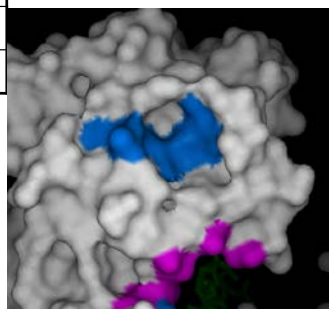
PASS Site	Residue	All genotype	Genotype 1 only	flexibility
7	T295	297/361	203/216	-
	D296	359/361	215/216	-
	T433	342/361	208/216	++
	Q434	360/360	216/216	++
	V490	356/361	215/216	-



PASS site 8
pocket on domain 1

PASS Site	Residue	All genotype	Genotype 1 only	flexibility
8	N251	346/361	206/216	-
	S263	94/361	42/216	-
	T266	361/361	216/216	-
	A283	361/361	216/216	+
	Y284	361/361	216/216	-

A263 206/361 158/216



S2: More on chemotype series S3 and S4

Chmeotype Series S4 - Detection and Explorations. HTS data was revisited and a short list of fragment-like compounds were selected based on low molecular weight, weak activity based on percentage of control (POC) data, drug-like properties (DLP), and high predicted solubility. Eighty-five compounds from our collection were then screened by NMR for binding to helicase using ligand-based DLB and STD experiments (data not shown) and protein-based ^{15}N TROSY experiments. Twenty compounds were clearly found to exhibit strong STD signals as a result of binding to helicase, and were confirmed by SPR including the arylacid **3 (**Figure S1**)**. Direct interaction of **3** with helicase was verified by a fast and simple methyl-shift NMR assay as shown in Figure S1B. Additionally, the more time-consuming ^{15}N TROSY experiment also was useful for detecting ligand binding and it revealed details of protein folding and flexibility. Interestingly, the ^{15}N TROSY data acquired during several hours after the addition of compound **3** resulted in the observation of very few peaks (displayed as red and overlaid onto the blue spectrum in Figure S1D suggesting that the protein was unfolding and precipitated. Thus, ligand binding resulted in significant changes (e.g. stability, fold, aggregation) to HCV helicase over time.

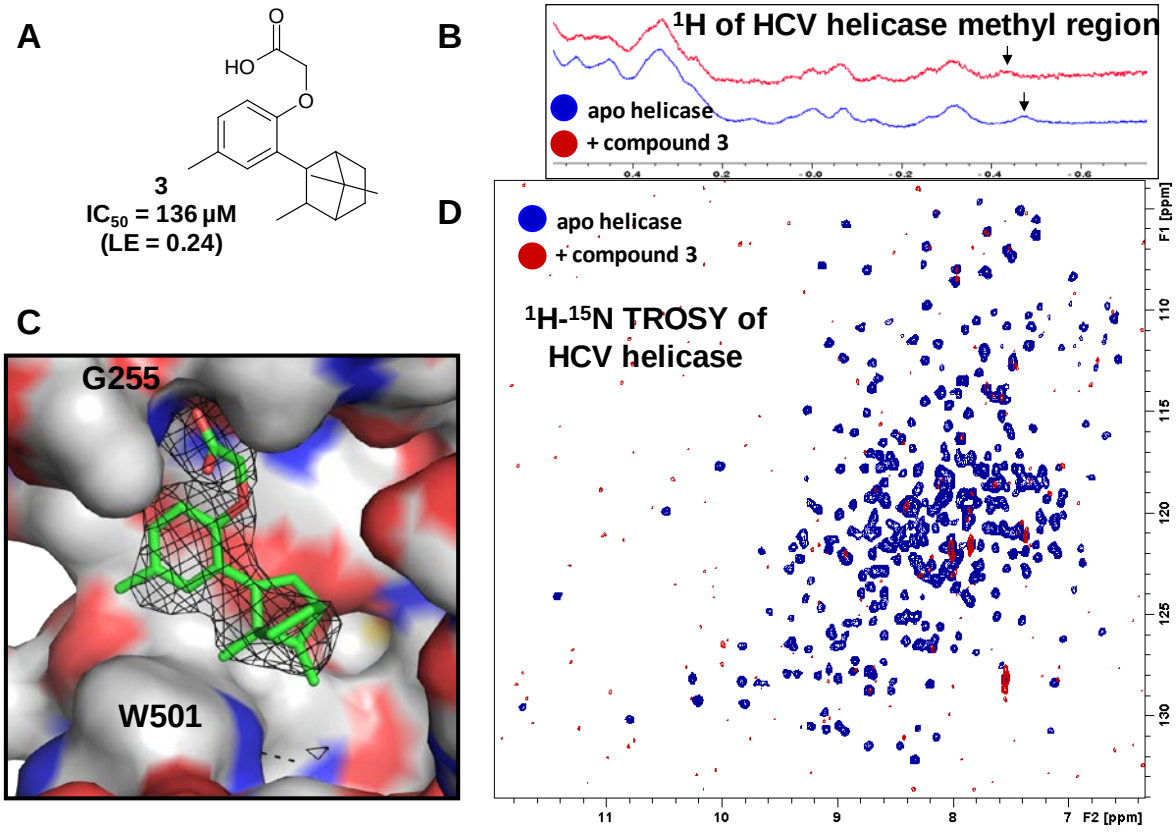


Figure S1. (A) The primary structure, activity and LE for compound **3**. (B) ^{1}H NMR methyl region of apo helicase (blue) and complex with compound **3** (blue). (C) X-ray structure of compound **3** bound to site 3+4. (D) ^{1}H-^{15}N TROSY of apo HCV helicase (blue) and complex with **3** (1:5 ratio of helicase to ligand).

Nonetheless, with binding activity confirmed, a co-structure of **3** with helicase was determined revealing binding of the compound at site 3+4. Compound **3** fits in the wide groove over W501 and exhibits hydrogen bonds to G255 and T269 similar to the carboxylic acid group of **2** (shown in Figure S1C). No significant interactions with the lipophilic pocket or polar residues are apparent in the structure. Subsequently synthesized analogues from this series

showed two- to four-fold loss in activity and similar values (see Supporting Information). The lack of productive SAR led to the deprioritization of this series.

Chemotype Series S3 (pyrimido-pyrimidines) - Detection and Explorations. Series S3 was identified using our virtual screening tool PharmShapeCC to search a database of 10^{12} virtual compounds (Qiang Z.; Muegge, I. PharmShapeCC: 3D pharmacophore searching against ten trillion combinatorially accessible compounds. 9th International Conference on Chemical Structures, June 5-9, 2011, Noordwijkerhout, The Netherlands). It uses pharmacophore-based queries defined from a ligand co-crystal structure or a docked model to search for features such as donor, acceptor, aromatic and hydrophobic groups as well as positive or negative ionizable groups in 3D space. In this case, the ligand template structure was derived from an X-ray structure of HCV helicase co-complex with **2** (Figure S2). In this structure, the ligand carboxylic acid group formed key hydrogen-bond interactions with Thr269 and Gly255. Given that this interaction did not involve charged groups on the protein, a general acceptor feature (Ac) was defined as one of the query pharmacophores for the virtual screen. The X-ray structure also showed that the indole core was in close contact with a flat grooved surface of the pocket; therefore an aromatic feature was included. A second aromatic feature was defined given another potentially important interaction involving the phenyl group of compound **2** with Trp501 via a pi-pi interaction. Finally, the shape query was simply defined as the shape of the binding pocket occupied by the indole compound in the crystal structure (not shown). The goal was to identify libraries of compounds whose low energy conformations match all the required pharmacophore features. Low energy conformations were typically derived using OMEGA from Openeye with windows < 10 kcal/mol.

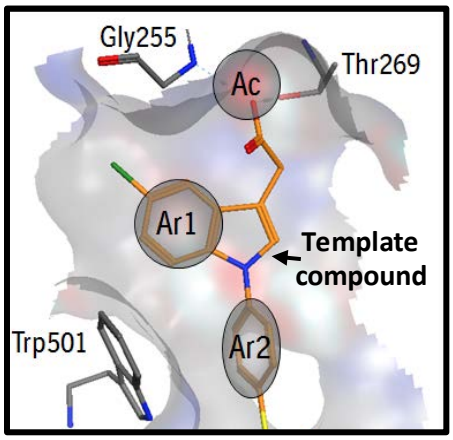


Figure S2. The X-ray structure of compound **2** (template) bound to site 3+4 of HCV helicase. Also provided are the definitions of pharmacophore and shape queries.

The virtual screening approach yielded interesting analogs from our corporate library including a pyrimido-pyrimidine chemotype. Thirty-six analogues were made and tested at high concentrations. Inhibitory activities were observed in the range of 300-500 μM and poor LE scores (e.g. <0.2). X-ray co-structures of hits from this series showed overlap with the query and satisfied the pharmacophore features (data not shown). While early SAR revealed that potency of this series could be improved, some unattractive aspects were also uncovered. X-ray crystal structure of representative compounds showed stacking of multiple compounds on top of one another) within the helicase 3+4 site (data not shown). This undesirable feature likely reflected the tolerable wide grooves of site 3+4, and the relative "stickiness" and low solubility of compounds from this series. Figure S3 shows an example where the same compound had self-attractive tendencies using the NMR aggregation assay (i.e. resonances shifting left at higher concentrations). Therefore, this class of compounds was deprioritized due to poor

physicochemical properties, activity in counter screens, non-linear SAR trends, and poor LE values.

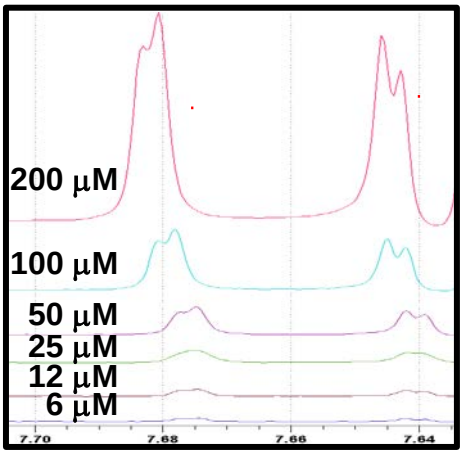
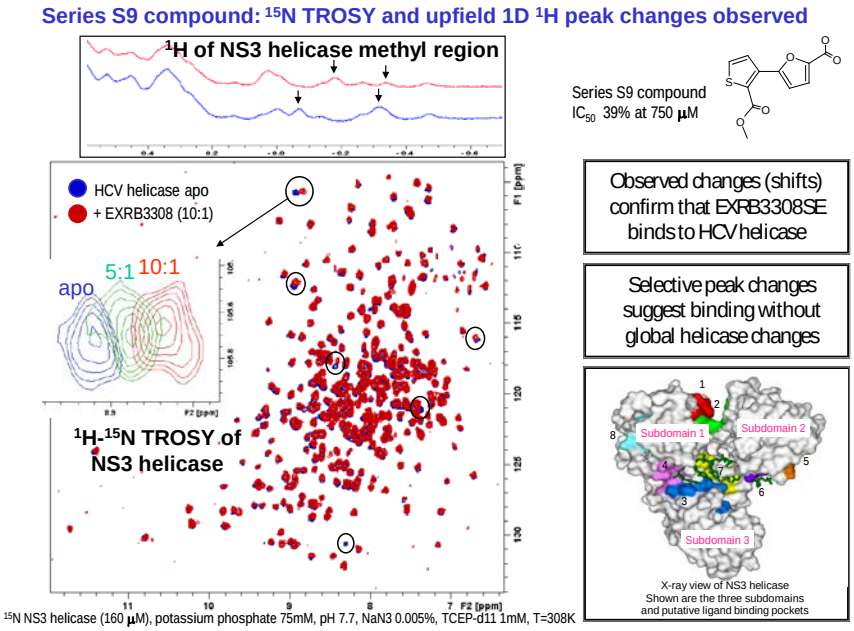
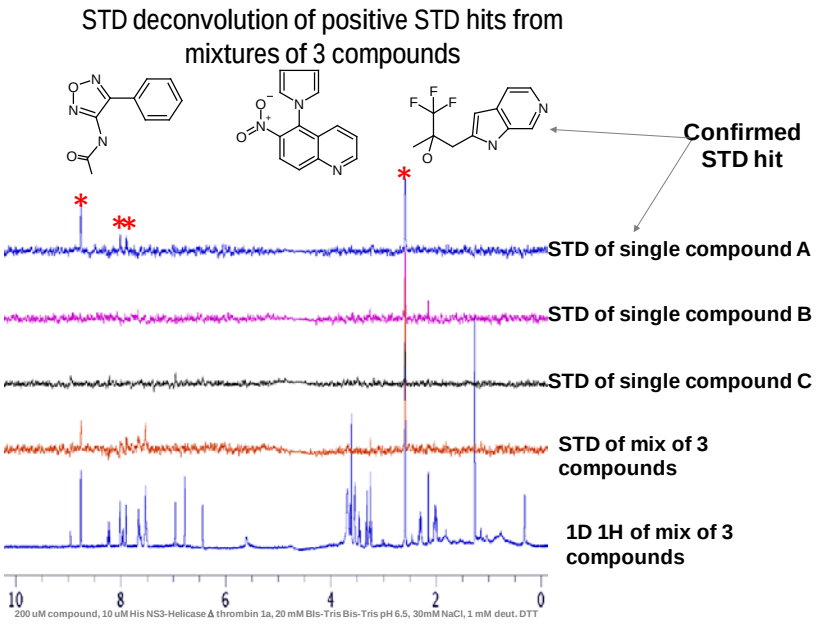


Figure S3. (A) The X-ray structure of a compound from the S3 series where electron density indicating the stacking of compounds at site 3+4. (B) Shown are ^{1}H NMR spectra of the same representative compound from series S3. Spectra are from free compound in solution as a function of dilution.

S3: Additional NMR experiments

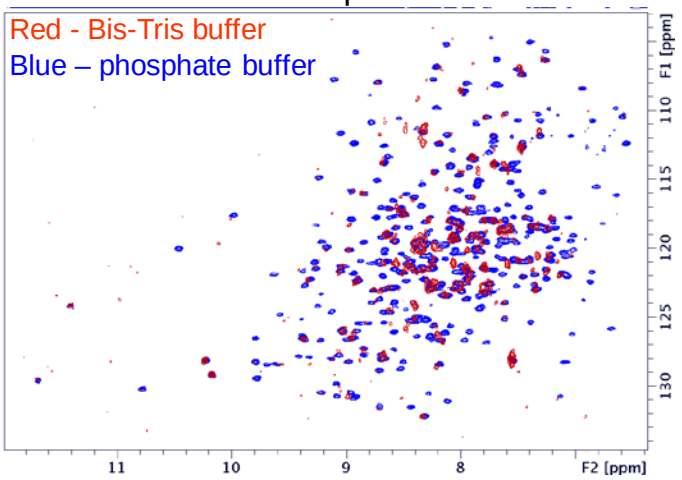


The primary structure and activity of a series S9 compound are shown. X-ray structure and full surface view of apo HCV helicase displayed on the bottom right. ^{1}H NMR methyl region of apo helicase (blue) and complex with compound (red) shown at top left. ^{1}H-^{15}N TROSY of apo HCV helicase (blue) and complex with compound (red, 1:5 ratio of helicase to ligand) shown bottom left.



Shown above is an example of STD deconvolution of positive STD hits from a mixture of three compounds. Here, single compound sample were made and spectral comparisons made.

TROSY of ^{15}N HCV helicase in two buffers show important differences

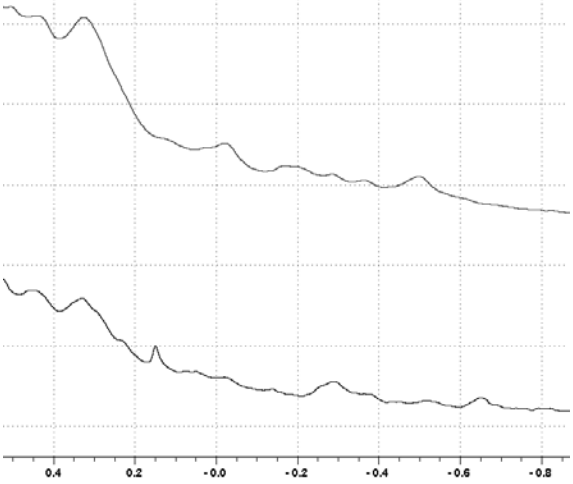


^{15}N

^{1}H

Phosphate buffer: 120 µM, 75 mM K_iPO4 pH 7.7, 2 mM DTT
BIs-Tris buffer: 110 mM, 20mM deut Bis-Tris pH 6.5, 30mM NaCl, 1 mM deut. DTT

Buffer-dependent differences in the 1D NMR spectrum (methyl region at natural abundance) of HCV helicase were observed



Other buffer:
110 uM helicase 1a
20 mM BIs-Tris pH 6.5
30 mM NaCl
1 mM TCEP (deuterated)
300 mM NaCl

Assay-like conditions:
23.3 uM helicase 1a
25 mM Tris-HCl (deuterated), pH 7.5
5 mM MgCl2
0.5 mM DTT (deuterated)
3% glycerol (deuterated)
1 mM EDTA (deuterated)
300 mM NaCl
0.01% IGEPAL
5% DMSO (deuterated)