

# Identification of HCV protease inhibitor resistance mutations by selection pressure-based method

Ping Qiu<sup>1,\*</sup>, Vincent Sanfiorenzo<sup>2</sup>, Stephanie Curry<sup>3</sup>, Zhuyan Guo<sup>4</sup>, Shaotang Liu<sup>3</sup>, Angela Skelton<sup>3</sup>, Ellen Xia<sup>3</sup>, Constance Cullen<sup>2</sup>, Robert Ralston<sup>3</sup>, Jonathan Greene<sup>1</sup> and Xiao Tong<sup>3,\*</sup>

<sup>1</sup>Molecular Design and Informatics, Schering-Plough Research Institute, 2015 Galloping Hill Road, Kenilworth, NJ 07033, <sup>2</sup>Preclinical and Clinical Bioanalytics, Schering-Plough Research Institute, 1011 Morris Avenue, Union, NJ 07083, <sup>3</sup>Department of Virology and <sup>4</sup>Department of Structural Chemistry, Schering-Plough Research Institute, 2015 Galloping Hill Road, Kenilworth, NJ 07033, USA

Received December 23, 2008; Revised March 18, 2009; Accepted April 2, 2009

## ABSTRACT

A major challenge to successful antiviral therapy is the emergence of drug-resistant viruses. Recent studies have developed several automated analyses of HIV sequence polymorphism based on calculations of selection pressure ( $K_a/K_s$ ) to predict drug resistance mutations. Similar resistance analysis programs for HCV inhibitors are not currently available. Taking advantage of the recently available sequence data of patient HCV samples from a Phase II clinical study of protease inhibitor boceprevir, we calculated the selection pressure for all codons in the HCV protease region (amino acid 1–181) to identify potential resistance mutations. The correlation between mutations was also calculated to evaluate linkage between any two mutations. Using this approach, we identified previously known major resistant mutations, including a recently reported mutation V55A. In addition, a novel mutation V158I was identified, and we further confirmed its resistance to boceprevir in protease enzyme and replicon assay. We also extended the approach to analyze potential interactions between individual mutations and identified three pairs of correlated changes. Our data suggests that selection pressure-based analysis and correlation mapping could provide useful tools to analyze large amount of sequencing data from clinical samples and to identify new drug resistance mutations as well as their linkage and correlations.

## INTRODUCTION

More than 170 million people worldwide are chronically infected with the hepatitis C virus (HCV). Combination therapy with pegylated interferon-alpha plus ribavirin achieves sustained virologic response rates of only ~50% in HCV genotype 1 infected patients (1–3), which emphasizes the need for new antiviral drugs. The viral serine protease NS3-4A is responsible for the cleavage of the viral polyprotein into functional individual proteins (4), and has become a promising drug target for specific antiviral treatment. Data for clinical efficacy drug resistance have been published for several protease inhibitors BILN-2061 (ciluprevir) (5,6), VX-950 (telaprevir) (7,8) and SCH 503034 (boceprevir) (9). A major challenge to successful antiviral therapy is the emergence of drug-resistant viruses. Due to the rapid viral dynamics and high mutation rate of HCV polymerase, large pools of genetically distinct but closely related variants are generated in infected patients (10). During treatment, the strong selective pressure allows the outgrowth of variants resistant to therapy which can result in viral load rebound and reduced clinical efficacy.

The protease inhibitor boceprevir (SCH 503034) binds to the enzyme active site and inhibits cleavage of the viral polyprotein (11). It has been under clinical investigation in mono- and combination therapy regimens. Resistance mutations to boceprevir have been identified in the HCV replicon system as well as in clinical trials. The resistance loci are V36, Q41, F43, T54, R155, A156 and V170, which are located near the inhibitor binding site (12–15).

Although the replicon system has proven to be a valuable tool to identify resistance mutations to protease inhibitors, comprehensive evaluation of clinical

\*To whom correspondence should be addressed. Tel: +1 980 740 4145; Fax: +1 908 740 4496; Email: ping.qiu@spcorp.com  
Correspondence may also be addressed to X. Tong. Tel: +1 980 740 7446; Fax: +1 908 740 3032; Email: xiao.tong@spcorp.com

sequencing data may provide further insight into the development of resistance and evolution of viral population under clinical settings. Several automated analyses of HIV sequence polymorphism to predict drug resistance mutations were published in the past few years (16,17). One very widely used metric of selection pressure that has been adapted in these studies is known as  $K_a/K_s$  (18,19) which measures the ratio of observed non-synonymous mutations over observed synonymous mutations, normalized by the ratio expected under a neutral model. Methods used for inferring  $K_a/K_s$  ratios are constantly being developed (20–23). A ratio around 1 indicates either neutral evolution or an averaging of positive and negative selective pressures. A ratio less than 1 indicates negative selection, with most amino-acid mutations being deleterious. A ratio significantly greater than 1 indicates positive selective pressure, which is less common and suggests that amino acid changes are favored, i.e. they reflect a change in the function of a gene or a change in environment conditions that forces the organism to adapt, thus increase organism's fitness.

*In silico* analysis of drug resistant mutation using a selection pressure-based method in these studies has provided a valuable resource for HIV research community. Similar studies for HCV are not currently available. In this study, taking advantage of the recently available sequence data of patient HCV samples from a boceprevir Phase II clinical study, we calculated the selection pressure for all the codons on the HCV protease region and by combining with frequency based filtering approach, we identified all major known resistant mutations as well as novel non-synonymous mutations that confer resistance to HCV protease inhibitor treatment.

## MATERIALS AND METHODS

### Sample selections

Samples used in this study were obtained from a Phase II clinical study of HCV protease inhibitor boceprevir (SCH 503034) in genotype 1 non-responders. Patient blood samples at time points specified by study protocol were collected and HCV RNA was extracted, amplified and sequenced. Additional samples were selected by study clinicians based on virological response of individual patients. Because samples and protease sequences were not available from all subjects at all time points, it was not possible to select a single on-treatment time point for analysis. Therefore, the last on-treatment time point was used for each subject. Samples collected from subjects prior to boceprevir treatment were used as baseline controls.

### Amplification of HCV NS3 protease region and DNA sequencing

Viral RNA was extracted from human plasma samples using a commercially available silica-gel membrane based kit (QIAamp Virus BioRobot 9604 Kit, Qiagen, Valencia, CA) and processed on an automated BioRobot 9604 system (Qiagen). Reverse transcription of RNA was performed using a SuperScript III First Strand Synthesis Supermix kit (Invitrogen, Carlsbad,

CA), with random hexamers according to manufacturer's instructions. PCR was conducted with a Platinum PCR SuperMix kit (Invitrogen), using 3  $\mu$ l cDNA, 200 nM NS3 protease gene-specific primers (forward primer: GTAGAGCCCGTTCGTCTTCTC; reverse primer: GTGCTCTTGCCGCTGCCAGT), and 45  $\mu$ l of Platinum PCR Supermix (proprietary mix contains anti-*Taq* DNA polymerase antibody,  $Mg^{2+}$ , dNTPs and recombinant *Taq* DNA polymerase). Cycle sequencing reactions were performed using a BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA), gene-specific primer and 5–10 ng of purified DNA according to manufacturer's instructions. Reaction products were purified on a Biomek FX system (Beckman Coulter, Fullerton, CA) using a magnetic bead kit (Agencourt CleanSEQ Kit, Agencourt Bioscience Corporation, Beverly, MA). DNA sequencing of purified material was conducted on a 3730xl DNA Analyzer (Applied Biosystems).

Clonal sequencing was carried out on a subset of patient samples. Purified RT-PCR products were cloned using the TOPO TA Cloning Kit (pCR 2.1TOPO vector, Invitrogen). For each serum sample, 96 bacteria colonies were sent to Qiagen or Genewiz for sequencing, using M13 forward and reverse primers as well as two protease specific primers (56f, GACATCATCTTGGGTCTGCCGTCTC, 65r, GTGGGAGCGTGTAGGTGGGC). Sequence reads were aligned with HCV template sequence D90208 and mutations were analyzed.

### Sequencing data analysis

The sequenced region included codons 1–181 of the HCV protease NS3 region. Base calling was done using PHRED (24). Quality scores from PHRED output were extracted and used to select chromatograms with good quality for subsequent analysis. For mixture positions where the chromatogram indicated that a mixture of more than one nucleotide was present, only the major peak was called. For each sample, at least four sequencing reads with good quality were required to cover each nucleotide position and at least one of them was required to come from a different sequencing direction. ClustalW (25) was used to align sequences to a template HCV sequence (D90208). Consensus NS3 region sequences from both before- and after-treatment samples were generated for each subject from the ClustalW alignments and were compared at each nucleotide position for each subject. Nucleotide changes were recorded and mutation type (transition or transversion) were determined. Amino-acid changes before and after treatment for each codon position were also recorded. HCV NS3 region sequences from before and after treatment used in this study can be obtained from GenBank (FJ830936 to FJ831439).

### $K_a/K_s$ calculation and selection pressure analysis

Amino-acid selection pressure at each codon was calculated using the following formula developed by Chen *et al.* (16,17)

$$\frac{K_a}{K_s} = \frac{N_a/N_s}{(n_{a,t}/f_t + n_{a,v}/f_v)/(n_{s,t}/f_t + n_{s,v}/f_v)}$$

where  $N_a$  and  $N_s$  are the number of non-synonymous mutations and synonymous mutations observed at the codon,  $n_{a,t}$  is the number of possible transition mutations in the codon that would change the resulting amino acid,  $n_{s,t}$  is the number of possible transition mutations that are synonymous,  $n_{a,v}$  and  $n_{s,v}$  are the equivalent numbers for transversions, and  $f_t$  and  $f_v$  are the transition and transversion frequencies, respectively.  $f_t$  and  $f_v$  are measured using the entire dataset according to:  $f_t = N_t/(n_t S)$ ,  $f_v = N_v/(n_v S)$ , where  $S$  is the total number of samples,  $N_t$  and  $N_v$  are the number of observed transition and transversion respectively,  $n_t$  is the number of possible transitions in the region that was sequenced (equal to the length  $L$  of the sequence in nucleotides), and  $n_v$  is the number of possible transversions (equal to  $2L$ ).

An LOD confidence score for a codon to be under positive selection pressure was also calculated according to the following formula (16,17)

$$\text{LOD} = -\log_{10} \sum_{i=N_a}^N \binom{N}{i} q^i (1-q)^{N-i}$$

where  $N$  is the total number of mutations observed in the codon and  $q$  is calculated as follows:

$$q = \frac{n_{a,t}f_t + n_{a,v}f_v}{3f_t + 6f_v}$$

#### Confirmation of boceprevir resistance of novel mutations

To generate mutant proteases carrying resistance mutations, the nucleotide changes were introduced using the QuikChange mutagenesis kit (Stratagene). The parental plasmid expressing His-tagged single chain NS4A-NS3 protease domain, NS4A<sub>21-32</sub>-GSGS-NS3<sub>3-181</sub>, was described by Taremi *et al.* (26). The expression and purification protocol was described in detail by Taremi *et al.* (26). Recombinant proteases were tested using a chromogenic assay as described by Zhang *et al.* (27). The overall inhibition constant  $K_i^*$  (where  $v_s = V_{\max}S/(K_m(1 + I/K_i^*))$ ) was used to measure inhibitor potency.

The construction and sensitivity testing of mutant replicons have been described previously (12). The QuikChange (Stratagene) template was plasmid pUC18 1-5, which contained the 5' half of the subgenomic replicon sequence (T7 promoter to EcoRI site in NS5A). The PmeI-EcoRI fragment, which contained the coding sequence for NS3 to the N-terminal of NS5A, was sequenced to confirm the engineered mutations and used to replace the same region in pUC18 Bart, which contained the complete subgenomic replicon sequence (28). To measure anti-replicon activity of compounds, replicon cells (in triplicates) were dosed with protease inhibitors for 3 days. The replicon RNA level was measured using real-time PCR (Taqman assay) on the ABI PRISM 7900HT Sequence Detection System. The dCT values ( $CT_{5B} - CT_{GAPDH}$ ) were plotted against drug concentration and fitted to the sigmoid dose response model using Graphpad PRISM software (Graphpad Software Inc.).  $IC_{50}$  was the drug dose necessary to achieve dCT = 1 over the projected baseline.  $IC_{90}$  was the drug dose necessary to achieve dCT = 3.2 over the baseline.

#### Correlation map of selection pressure

The effect of amino-acid change at one position on amino-acid mutation at another position was also evaluated by calculating conditional  $K_a/K_s$ . The calculation was proposed by Pan *et al.* (17):

$$\left(\frac{K_a}{K_s}\right)_{Y|Xa} = \frac{N_{YaXa}/N_{YsXa}}{(n_{a,t}f_t + n_{a,v}f_v)/(n_{s,t}f_t + n_{s,v}f_v)}$$

where  $N_{YaXa}$  and  $N_{YsXa}$  are the numbers of non-synonymous mutations and synonymous mutations at site Y observed in the presence of non synonymous mutations at site X and all other variables retain their previous definitions. Calculation was done pairwise for all codons within 1-181 amino acids of NS3.

The 'conditional selection ratio' was defined as the ratio of the conditional  $K_a/K_s$  measured in the presence of non-synonymous mutation at X divided by the selection pressure at Y measured in the absence of any mutation at X:

$$\left(\frac{K_a}{K_s}\right)_{Y|X} = \frac{(K_a/K_s)_{Y|Xa}}{(K_a/K_s)_{Y|Xo}} = \frac{N_{YaXa}/N_{YsXa}}{N_{YaXo}/N_{YsXo}}$$

where  $N_{YaXo}$  and  $N_{YsXo}$  are the numbers of samples containing either a nonsynonymous mutation or synonymous mutation at Y and no mutation at X.

To evaluate the statistical significance of positive conditional selection pressure, a LOD score based on the likelihood of observing at least  $N_{YaXa}$  samples with non-synonymous mutations at codons Y and X, assuming neutral selection at codon Y:

$$\text{LOD} = -\log_{10} \sum_{i=N_{YaXa}}^N \binom{N}{i} q^i (1-q)^{N-i}$$

where  $N = N_{YaXa} + N_{YsXa}$ ; and  $q$  is as defined above.

#### Structural modeling

The crystal structure of boceprevir (SCH 503034) complexed with the NS3/4A protease was solved in house [PDB code 2OBO (29)]. To model the binding mode of telaprevir (VX-950) to the NS3/4A protease, the crystal structure of boceprevir-NS3/4A protease was used as a template. The backbone atoms and the ketoamide positions of boceprevir served as initial coordinates upon which the P1-P4 side chains and P1 cap of telaprevir were built. The P4 cap was built based on the binding mode of another ketoamide inhibitor which has the same N-terminal cap (PDB code 1RGQ) (30). The inhibitor molecule was then energy minimized in the protein environment. In addition, point mutations were also modeled into the three-dimensional structure of the protein. InsightII molecular modeling program (Accelrys Software Inc., San Diego, CA) was used in the modeling.

#### RESULTS AND DISCUSSION

During HCV antiviral therapy, viruses with drug-resistant mutations are expected to be positively selected.

**Table 1.** Positively selected mutations in HCV protease region (after filtering as described in the text)

Position	AA_SCRN	AA_Therapy	$K_a/K_s$	LOD	Number of Mutation_Therapy	Number of Mutation_SCRN	<i>P</i> -value	Mutation% in Genbank
36	V	M	18.27	90.47	48	0	6.0e-16	0
54	T	A	4.87	30.20	15	0	4.9e-5	0
54	T	S	47.15	160.07	66	1	<2.2e-16	0
155	R	K	80.80	227.17	113	1	<2.2e-16	0
156	A	S	19.30	49.00	15	0	4.9e-5	0
170	V	A	40.92	11.95	6	0	0.03	0.39
55	V	A	5.11	13.71	10	3	0.09	0
55	V	I	2.19	6.76	4	1	0.37	0
158	V	I	4.38	18.22	5	0	0.06	0.39
163	V	L	2.24	8.76	3	0	0.25	0

*P*-values are the mutation frequency differences in before and after treatment samples. Mutation % in GenBank are largely based on 1b (64%) and 1a (27%) sequences.

To identify such positively selected amino acids, the  $K_a/K_s$  values were calculated for each position of the HCV NS3 protease region (amino acid 1–181). A total of 252 subjects from this HCV protease inhibitor clinical study with sequence data from before and on treatment samples were included in this analysis. Because the collecting schedule of serum samples varied with individual patients, it was difficult to choose a single fixed time point for all patients. Instead, the last on-treatment time was analyzed to maximize the time under selection pressure. For ‘mixture positions’ where the chromatogram indicated more than one nucleotide was present, only the major peak was analyzed. Out of the  $K_a/K_s$  values for 746 different non-synonymous mutations in the NS3 protease, 198 amino-acid changes have  $K_a/K_s$  value >1. In order to focus on major mutations which may have more impact on clinical outcome and to differentiate mutations selected specifically for drug resistance from ‘neutral drifts’, the following criteria were built into the mutation analysis process: (i) Low frequency non-synonymous mutations (in <1% treated subjects) were excluded from further analysis. (ii) Frequency differences of mutations in before and after treatment samples were compared using Fisher’s exact test. Mutations whose prevalence did not increase significantly during therapy ( $P > 0.1$ ) were excluded. Table 1 listed the resulting resistant mutations with  $K_a/K_s$  value >1 and meeting the criteria described above. As shown in Table 1, the major known boceprevir-resistant mutations (V36M, T54A/S, R155K/T, A156S, V170A) (12,13) were correctly identified by this analysis. V55A is a resistance mutation identified recently by clonal sequencing analysis of a phase I clinical trial of boceprevir, and has been shown to confer resistance to boceprevir and telaprevir in replicons (31). Structural analysis of V55 supports observed phenotypic resistance (31). In our study, two mutations at V55 (V55A and V55I) were positively selected based on  $K_a/K_s$  values, although V55A appears to be the more favored amino-acid change based on calculated *P*-values. In addition, our analysis identified a novel mutation V158I, which fit all the filtering criteria. Another mutation V163L was positively selected based on  $K_a/K_s$  value; although the *P*-value was greater than 0.1, it was also included in subsequent confirmation studies to evaluate the stringency of filtering criteria. The patient

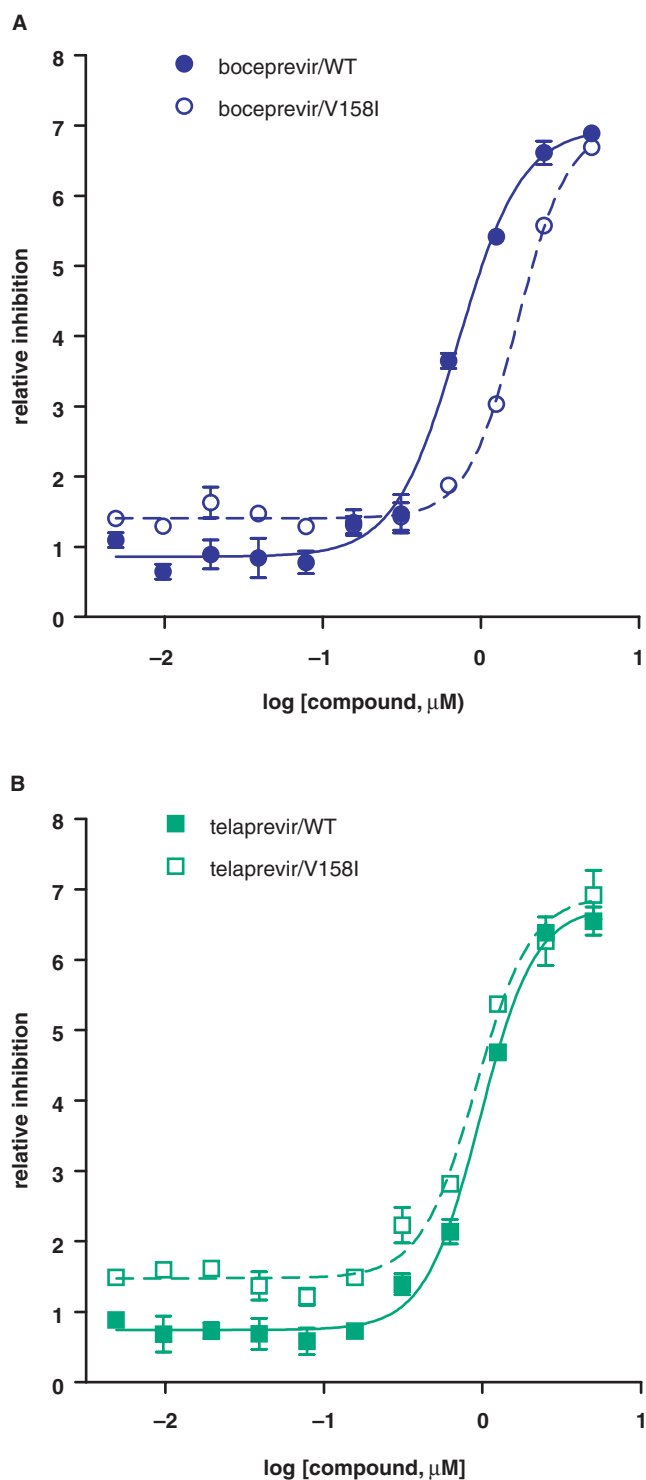
**Table 2.** Enzyme and replicon resistance of newly identified resistance mutations

	Mutations	Fold change over WT ± SD	
		Enzyme ( $K_i^*$ )	Replicon (EC50)
Boceprevir	V158I	2.5 ± 1.1	3.3 ± 0.5
	V163L	1 ± 0.3	not done
Telaprevir	V158I	0.5 ± 0.2	0.9 ± 0.1
	V163L	1 ± 0.3	not done

Averages and SD of two to four experiments are reported.

population of this trial consists of approximately equal mixture of genotype 1a and 1b subjects. Mutation V55A/I was found in five 1a subjects, seven 1b subjects and two subjects with unknown subtypes. Mutation V158I was found in five 1b subjects, and mutation V163L in two 1b subjects and one 1a subject. The number of subjects is too small to show any statistically significant difference in mutation distribution in genotype 1a and 1b subjects. Mutations V55A/I, V158I and V163L mutations are very rare among natural isolates based on the amino-acid frequency database constructed from GenBank data as described previously (32).

To test whether V158I and V163L confer resistance to boceprevir, each mutation was introduced into the single-chain genotype 1b protease (26). As shown in Table 2, V158I conferred low level resistance (as shown by a 2.5-fold increase in inhibition constant  $K_i^*$ ) to boceprevir but not to telaprevir, another ketoamide inhibitor under clinical evaluation. To confirm *in vitro* enzyme results, mutation V158I was cloned into a genotype 1b replicon and shown to decrease replicon sensitivity (as shown by a ~3-fold increase in replicon EC50 value) to boceprevir but not to telaprevir (Figure 1 and Table 2). In contrast, mutation V163L did not affect boceprevir inhibition of the protease in biochemical assays (Table 2). Structural analysis (Figure 2) of the NS3 protease active site showed that V158 is part of the substrate binding pocket and is in direct contact with the *tert*-butyl P3 capping group of boceprevir. Change of this residue to a slightly bulkier Ile could interfere with boceprevir binding. Telaprevir has a more extended interaction with the active site via



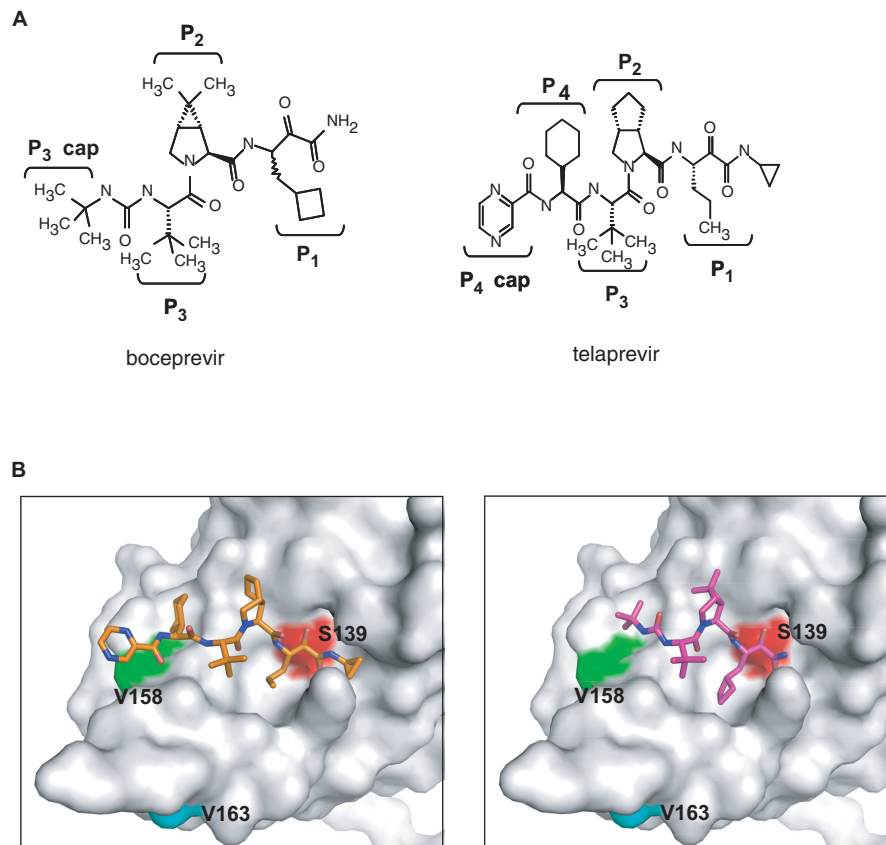
**Figure 1.** Inhibition of mutant replicons by protease inhibitors. Replicon cells carrying wild-type or mutant V158I subgenomes were treated with boceprevir (A) or telaprevir (B) for 3 days, and replicon RNA levels were measured by Taqman assay. The relative reduction in replicon RNA level (dCT) was plotted against compound concentration.

its P4 group which could compensate the potency loss caused by the V158I mutation. V163 is about 9 Å away from the *tert*butyl P3 capping group of boceprevir and is not expected to be involved in inhibitor binding.

Current research has largely focused on the impact of single mutations on drug resistance, and the potential of interactions among mutations has received little attention. The impact of an individual mutation may vary depending on other mutations present. The interactions could be positive (synergistic), leading to higher levels of drug resistance than would be expected from the individual mutations alone; or negative (antagonistic), with some mutations effectively reducing the effects of others. Compensatory mutations also can increase the replication fitness of the mutant virus, increasing the apparent effects of resistance mutations. The lack of understanding of coordinations between mutations is at least partly due to the limitation of commonly used sequencing methods. Clonal sequencing can provide linkage information for multiple mutations, but the method is labor intensive and can only be applied to relatively small number of samples. Population sequencing is more cost effective in evaluating large number of samples, but can not evaluate directly the linkage of mutations at different positions. The multiple mutations observed in a patient by population sequencing could be linked in a single genome or could arise from two or more individual virus strains.

Correlated mutations may occur at both short range positions (33) and distant positions (34). Several bioinformatics methods for detecting correlated mutations have been developed over the years (35–37). One of the key problems that all the methods share is distinguishing signal from noise, such as those arise from evolutionary processes related to common ancestry. In this study, to extract correlated mutation information from population sequencing, we calculated  $K_a/K_s$  values for mutation at one position (Y) conditioned on the mutation on another position (X), a method reported by Pan *et al.* (17). A total of 627 pairs of amino-acid mutations show conditional  $K_a/K_s$  values  $>1$ . Similar filters as described for single mutations were used to reduce false positives and the resulting amino-acid pairs were shown in Table 3. All three pairs of double mutations have been identified by clonal sequencing of samples from 19 patients from this clinical study. V36M + R155K was observed in 15/19 patients, T54S + R155K in 10/19 patients (a related mutation T54A + R155K in 1/19 patient), T54S + A156S in 1/19 patient (a related mutation T54A + A156S in 1/19 patient). Double mutations (e.g. V36M + R155K) have also been observed in patient treated with telaprevir (8). The level of resistance conferred by double mutations is generally multiplicative or greater (13,38). For example, double mutant V36M + 155K confers 14-fold resistance to boceprevir, while single mutants V36M and R155K confer only 2- and 3-fold resistance, respectively (13). The mechanism of interaction between those mutations on drug resistance remains to be analyzed.

In summary, five major known HCV boceprevir drug resistance mutations were successfully identified using the selection pressure-based method described here, two additional mutations V55A and V158I were identified that were previously not revealed in subjects from this clinical study. Use of more sensitive sequencing method and inclusion of more data set from future studies will increase the likelihood of identifying low



**Figure 2.** (A) Structures of boceprevir and telaprevir. (B) Three-dimensional structure of NS3 protease-inhibitor complex. The structures of boceprevir (purple, right panel) and telaprevir (gold, left panel) are presented as stick models. Side chains of the active site Ser (Ser139) and two mutation sites (V158 and V163) are depicted on the Connolly surface of the NS3 protease.

**Table 3.** Interaction of mutations in HCV protease region (after filtering as described in text)

Position 1	AA1	Position 2	AA2	Number of double mutation <sup>a</sup>	$K_a/K_s$	LOD	<i>P</i> -value
36	M	155	K	33	4.44	57.84	$7.6e-10$
54	S	156	S	12	2.69	28.25	$4.3e-4$
155	K	36	M	33	2.79	57.11	$7.6e-10$
155	K	54	S	31	3.49	53.64	$3.5e-10$
156	S	54	S	12	10.67	36.04	$4.3e-4$

*P*-values are the double mutation frequency differences in before and after treatment samples.

<sup>a</sup>The number of double mutation observed on therapy, no double mutations were detected in SCRN samples.

prevalence mutations. Although the data set included in this study is relatively small compared to datasets used in similar HIV studies and other frequency based mutation detection methods have been reported for HCV clonal sequence analysis (8,31,39), the high success rate of our method suggests that  $K_a/K_s$  analysis and correlated mutation mapping combined with proper filtering mechanism described above could provide useful tools to identify new drug resistant mutations in HCV and shed light on potential interactions between mutations at different sites.

## FUNDING

Funding for open access charge: Schering-Plough Research Institute.

*Conflict of interest statement.* None declared.

## REFERENCES

- Fried, M.W., Shiffman, M.L., Reddy, K.R., Smith, C., Marinos, G., Goncalves, F.L. Jr, Haussinger, D., Diago, M., Carosi, G., Dhumeaux, D. *et al.* (2002) Peginterferon alpha-2a plus ribavirin for chronic hepatitis C virus infection. *N. Engl. J. Med.*, **347**, 975–982.
- Hadziyannis, S.J., Sette, H. Jr, Morgan, T.R., Balan, V., Diago, M., Marcellin, P., Ramadori, G., Bodenheimer, H. Jr, Bernstein, D., Rizzetto, M. *et al.* (2004) Peginterferon-alpha2a and ribavirin combination therapy in chronic hepatitis C: a randomized study of treatment duration and ribavirin dose. *Ann. Intern. Med.*, **140**, 346–355.
- Manns, M.P., McHutchison, J.G., Gordon, S.C., Rustgi, V.K., Shiffman, M., Reindollar, R., Goodman, Z.D., Koury, K., Ling, M. and Albrecht, J.K. (2001) Peginterferon alpha-2b plus ribavirin compared with interferon alpha-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomised trial. *Lancet*, **358**, 958–965.
- Reed, K.E. and Rice, C.M. (2000) Overview of hepatitis C virus genome structure, polyprotein processing, and protein properties. In Hagedorn, C.H. and Rice, C.M. (eds), *Current Topics in Microbiology and Immunology: The Hepatitis C Viruses*, Vol. 242. Springer, Heidelberg, pp. 55–84.
- Hinrichsen, H., Benhamou, Y., Wedemeyer, H., Reiser, M., Sentjens, R.E., Calleja, J.L., Forn, X., Erhardt, A., Cronlein, J., Chaves, R.L. *et al.* (2004) Short-term antiviral efficacy of BILN

- 2061, a hepatitis C virus serine protease inhibitor, in hepatitis C genotype 1 patients. *Gastroenterology*, **127**, 1347–1355.
6. Herrmann, E., Zeuzem, S., Sarrazin, C., Hinrichsen, H., Benhamou, Y., Manns, M.P., Reiser, M., Reesink, H., Calleja, J.L., Fornis, X. *et al.* (2006) Viral kinetics in patients with chronic hepatitis C treated with the serine protease inhibitor BILN 2061. *Antivir. Ther.*, **11**, 371–376.
  7. Reesink, H.W., Zeuzem, S., Weegink, C.J., Forestier, N., van Vliet, A., van de Wetering de Rooij, J., McNair, L., Purdy, S., Kauffman, R., Alam, J. *et al.* (2006) Rapid decline of viral RNA in hepatitis C patients treated with VX-950: a phase Ib, placebo-controlled, randomized study. *Gastroenterology*, **131**, 997–1002.
  8. Sarrazin, C., Kieffer, T.L., Bartels, D., Hanzelka, B., Muh, U., Welker, M., Wincheringer, D., Zhou, Y., Chu, H.M., Lin, C. *et al.* (2007) Dynamic hepatitis C virus genotypic and phenotypic changes in patients treated with the protease inhibitor telaprevir. *Gastroenterology*, **132**, 1767–1777.
  9. Sarrazin, C., Rouzier, R., Wagner, F., Forestier, N., Larrey, D., Gupta, S.K., Hussain, M., Shah, A., Cutler, D., Zhang, J. *et al.* (2007) SCH 503034, a novel hepatitis C virus protease inhibitor, plus pegylated interferon alpha-2b for genotype 1 nonresponders. *Gastroenterology*, **132**, 1270–1278.
  10. Soriano, V., Perelson, A.S. and Zoulim, F. (2008) Why are there different dynamics in the selection of drug resistance in HIV and hepatitis B and C viruses? *J. Antimicrob. Chemother.*, **62**, 1–4.
  11. Malcolm, B.A., Liu, R., Lahser, F., Agrawal, S., Belanger, B., Butkiewicz, N., Chase, R., Gheys, F., Hart, A., Hesk, D. *et al.* (2006) SCH 503034, a mechanism-based inhibitor of hepatitis C virus NS3 protease, suppresses polyprotein maturation and enhances the antiviral activity of alpha interferon in replicon cells. *Antimicrob. Agents Chemother.*, **50**, 1013–1020.
  12. Tong, X., Chase, R., Skelton, A., Chen, T., Wright-Minogue, J. and Malcolm, B.A. (2006) Identification and analysis of fitness of resistance mutations against the HCV protease inhibitor SCH 503034. *Antiviral Res.*, **70**, 28–38.
  13. Tong, X., Bogen, S., Chase, R., Girijavallabhan, V., Guo, Z., Njoroge, F.G., Prongay, A., Saksena, A., Skelton, A., Xia, E. *et al.* (2008) Characterization of resistance mutations against HCV ketoamide protease inhibitors. *Antiviral Res.*, **77**, 177–185.
  14. Zeuzem, S., Sarrazin, C., Rouzier, R., Tarral, A., Brion, N., Forestier, N., Gupta, S., Deckman, D., Fellows, K., Hussain, M. *et al.* (2005) Anti-viral activity of SCH 503034, a HCV protease inhibitor, administered as monotherapy in HCV-1 patients refractory to pegylated interferon (PEG-IFN- $\alpha$ ). *Hepatology*, **42** (Suppl. 1), 233.
  15. Zeuzem, S., Sarrazin, C., Wagner, F., Rouzier, R., Forestier, N., Gupta, S., Hussain, M., Shah, A., Cutler, D. and Zhang, J. (2005) Combination therapy with HCV protease inhibitor SCH 503034, plus PEG-intron in hepatitis C genotype-1 PEG-intron non-responders: phase Ib results. *Hepatology*, **42** (Suppl. 1), 276.
  16. Chen, L., Perlina, A. and Lee, C.J. (2004) Positive selection detection in 40,000 human immunodeficiency virus (HIV) type 1 sequences automatically identifies drug resistance and positive fitness mutations in HIV protease and reverse transcriptase. *J. Virol.*, **78**, 3722–3732.
  17. Pan, C., Kim, J., Chen, L., Wang, Q. and Lee, C. (2007) The HIV positive selection mutation database. *Nucleic Acids Res.*, **35**, D371–D375.
  18. Li, W.H. (1993) Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *J. Mol. Evol.*, **36**, 96–99.
  19. Nei, M. and Gojobori, T. (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.*, **3**, 418–426.
  20. Li, W.H., Wu, C.I. and Luo, C.C. (1985) A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Mol. Biol. Evol.*, **2**, 150–174.
  21. Wong, W.S., Yang, Z., Goldman, N. and Nielsen, R. (2004) Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics*, **168**, 1041–1051.
  22. Tang, H. and Wu, C.I. (2006) A new method for estimating nonsynonymous substitutions and its applications to detecting positive selection. *Mol. Biol. Evol.*, **23**, 372–379.
  23. Doron-Faigenboim, A. and Pupko, T. (2007) A combined empirical and mechanistic codon model. *Mol Biol Evol.*, **24**, 388–397.
  24. Ewing, B., Hillier, L., Wendl, M.C. and Green, P. (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.*, **8**, 175–185.
  25. Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W. improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.
  26. Taremi, S.S., Beyer, B., Maher, M., Yao, N., Prosis, W., Weber, P.C. and Malcolm, B.A. (1998) Construction, expression, and characterization of a novel fully activated recombinant single-chain hepatitis C virus protease. *Protein Sci.*, **7**, 2143–2149.
  27. Zhang, R., Beyer, B.M., Durkin, J., Ingram, R., Njoroge, F.G., Windsor, W.T. and Malcolm, B.A. (1999) A continuous spectrophotometric assay for the hepatitis C virus serine protease. *Anal. Biochem.*, **270**, 268–275.
  28. Lohmann, V., Körner, F., Koh, J.-O., Herian, U., Theilmann, L. and Bartenschlager, R. (1999) Replication of subgenomic hepatitis C virus RNAs in a hepatoma cell line. *Science*, **285**, 110–113.
  29. Prongay, A.J., Guo, Z., Yao, N., Pichardo, J., Fischmann, T., Strickland, C., Myers, J. Jr, Weber, P.C., Beyer, B.M., Ingram, R. *et al.* (2007) Discovery of the HCV NS3/4A protease inhibitor (1R,5S)-N-[3-amino-1-(cyclobutylmethyl)-2,3-dioxopropyl]-3-[2(S)-[[[(1,1-dimethylethyl)amino]carbonyl]amino]-3,3-dimethyl-1-oxobutyl]-6,6-dimethyl-3-azabicyclo[3.1.0]hexan-2(S)-carboxamide (Sch 503034) II. Key steps in structure-based optimization. *J. Med. Chem.*, **50**, 2310–2318.
  30. Liu, Y., Stoll, V.S., Richardson, P.L., Saldívar, A., Klaus, J.L., Molla, A., Kohlbrenner, W. and Kati, W.M. (2004) Hepatitis C NS3 protease inhibition by peptidyl-alpha-ketoamide inhibitors: kinetic mechanism and structure. *Arch. Biochem. Biophys.*, **421**, 207–216.
  31. Susser, S., Welker, M., Zettler, M., Wohnsland, A., Hughes, E., Ralston, R., Tong, X., Zeuzem, S. and Sarrazin, C. (2008) Clonal analysis of mutations selected in the HCV NS3 protease domain of genotype 1 non-responders treated with boceprevir (SCH503034). *J. Hepatol.*, **48** (Suppl 2), 29A.
  32. Qiu, P., Cai, X.Y., Wang, L., Greene, J.R. and Malcolm, B. (2002) Hepatitis C virus whole genome position weight matrix and robust primer design. *BMC Microbiol.*, **2**, 29.
  33. Altschuh, D., Lesk, A.M., Bloomer, A.C. and Klug, A. (1987) Correlation of co-ordinated amino acid substitutions with function in viruses related to tobacco mosaic virus. *J. Mol. Biol.*, **193**, 693–707.
  34. Kass, I. and Horovitz, A. (2002) Mapping pathways of allosteric communication in GroEL by analysis of correlated mutations. *Proteins*, **48**, 611–617.
  35. Neher, E. (1994) How frequent are correlated changes in families of protein sequences? *Proc. Natl Acad. Sci. USA*, **91**, 98–102.
  36. Ortiz, A.R., Kolinski, A., Rotkiewicz, P., Ilkowski, B. and Skolnick, J. (1999) Ab initio folding of proteins using restraints derived from evolutionary information. *Proteins*, **37** (Suppl. 3), 177–185.
  37. Noivirt, O., Eisenstein, M. and Horovitz, A. (2005) Detection and reduction of evolutionary noise in correlated mutation analysis. *Protein Eng. Des. Sel.*, **18**, 247–253.
  38. Zhou, Y., Muh, U., Hanzelka, B.L., Bartels, D.J., Wei, Y., Rao, B.G., Brennan, D.L., Tigges, A.M., Swenson, L., Kwong, A.D. *et al.* (2007) Phenotypic and structural analyses of HCV NS3 protease ARG155 variants: sensitivity to telaprevir (VX-950) and interferon alpha. *J. Biol. Chem.*, **282**, 22619–22628.
  39. Kieffer, T.L., Sarrazin, C., Miller, J.S., Welker, M.W., Forestier, N., Reesink, H.W., Kwong, A.D. and Zeuzem, S. (2007) Telaprevir and pegylated interferon-alpha-2a inhibit wild-type and resistant genotype 1 hepatitis C virus replication in patients. *Hepatology*, **46**, 631–639.