BIOCHEMISTRY

 Volume 45, Number 5 February 7, 2006

Articles

Impact of Naturally Occurring Variants of HCV Protease on the Binding of Different Classes of Protease Inhibitors

Xiao Tong,[○] Zhuyan Guo,[§] Jacquelyn Wright-Minogue,[○] Ellen Xia,[○] Andrew Prongay,[§] Vincent Madison,[§] Ping Qiu,[‖] Srikanth Venkatraman,[⊥] Francisco Velazquez,[⊥] F. George Njoroge,[⊥] and Bruce A. Malcolm*[,○]

Antiviral Therapy, Structural Chemistry, Bioinformatics, and Medicinal Chemistry, Schering-Plough Research Institute, Kenilworth, New Jersey 07033

Received August 5, 2005; Revised Manuscript Received October 25, 2005

ABSTRACT: HCV drug discovery efforts have largely focused on genotype 1 virus due to its prevalence and relatively poor response to current therapy. However, patients infected with genotype 2 and 3 viruses account for a significant number of cases and would also benefit from new therapies. In vitro studies using two chemically distinct protease inhibitors with clinical potential showed that one, VX-950, was equally active on proteases from all three genotypes, whereas the other, BILN 2061, was significantly less active on genotype 2 and 3 proteases. Naturally occurring variation near the inhibitor binding site was identified based on sequence alignment of the protease region from genotype 1−3 sequences. Substitution of amino acids in genotype 1 based on genotype 2 and 3 has revealed residues which impact binding of BILN 2061. Substitution of residues 78−80, together with 122 and 132, accounted for most of the reduced sensitivity of genotype 2. The most critical position affecting inhibitor binding to genotype 3 protease was 168. Substitution of residues at positions 168, 123, and 132 fully accounted for the reduced sensitivity of genotype 3. Comparative studies of BILN 2061 and a closely related nonmacrocycle inhibitor suggested that the rigidity of BILN 2061, while conferring greater potency against genotype 1, rendered it more sensitive to variations near the binding site. Free energy perturbation analysis confirmed the experimental observations. The identification of naturally occurring variations which can affect inhibitor binding is an important step in the design of broad-spectrum, second generation protease inhibitors.

Hepatitis virus C (HCV)[1] is the etiologic agent for the majority of non-A, non-B hepatitis cases. It is estimated that ~170 million people worldwide are currently infected with the virus (*1*). HCV strains show extensive genome heterogeneity, and the natural variants are classified into six genotypes (1−6) and more than 50 subtypes (1a, 1b, 1c, and so forth) (*2*). The most prevalent genotypes are 1, 2, and 3. In industrialized countries (such as United States, Western Europe, and Japan), type 1 accounts for about 70% of the infections. Type 2 and 3 also have wide distributions in North America, Europe, and Japan. In particular, subtype 3a is common among intravenous drug users in the United States and Europe (*3*, *4*). The current standard of care is combination therapy with peginterferon-α and ribavirin. Although the sustained virological response (SVR) rate for genotype

* To whom correspondence should be addressed. Schering-Plough Research Institute, 2015 Galloping Hill Road, Kenilworth, NJ 07033. Tel: (908) 740-6738. Fax: (908) 740-3032. E-mail: bruce.malcolm@spcorp.com.

[○] Antiviral Therapy, Schering-Plough Research Institute.

[§] Structural Chemistry, Schering-Plough Research Institute.

[‖] Bioinformatics, Schering-Plough Research Institute.

[⊥] Medicinal Chemistry, Schering-Plough Research Institute.

[1] Abbreviations: HCV, hepatitis C virus; SVR, sustained viral response; FEP, free energy perturbation.

10.1021/bi051565g CCC: $33.50
Published on Web 01/10/2006

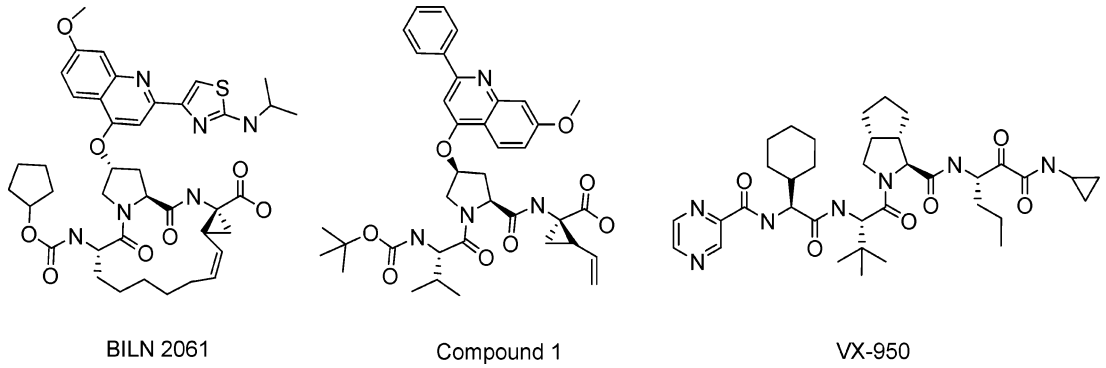

FIGURE 1: Chemical structures of protease inhibitors used in the study.

2 and 3 patients is relatively high (~80%), the treatment is much less efficacious in genotype 1 patients (SVR about 50%). In addition, the regimen is not well-tolerated due to toxic side effects associated with both drugs (*5*−*7*).

Research has focused initially on the HCV viral proteins to develop new antiviral therapies. The HCV genome encodes a single polyprotein of about 3000 amino acid residues which contains four structural proteins (Core−E1−E2−p7) followed by six nonstructural proteins (NS2−NS3−NS4A−NS4B−NS5A−NS5B) (*8*, *9*). The N-terminal third of NS3 is a serine protease, which, together with cofactor NS4A, is responsible for the cleavage of the nonstructural proteins NS3−NS5B (*10*). In the past several years, significant progress has been made in developing inhibitors of the HCV NS3/NS4A protease. The first HCV protease inhibitor to enter clinical trial was BILN 2061, a macrocyclic tripeptide (Figure 1, left panel) with an inhibition constant (K_i) for the genotype 1 enzymes in the sub-nanomolar range. It has also been shown to be highly potent in the genotype 1 HCV replicon system with an EC_{50} in the low nanomolar range (*11*). In phase I clinical trials, administration of BILN 2061 to genotype 1 patients resulted in a 2−3 log reduction of HCV RNA in the majority of patients after 2 days of dosing (*12*). However, the compound was less effective on genotype 2 and 3 patients; only about half of the group achieved more than a 1 log reduction in viral load (*13*) compared with 100% of the genotype 1 patients at the same dosing level (*12*). Previous studies have shown that BILN 2061 was less active on proteases cloned from genotype 2 and 3 viruses (*14*). On the basis of the alignment of a few sequences and structural analysis, the combination of positions 78−80 and 123, 132, and 168 in genotype 2 and 3 enzymes were suggested to be responsible for the loss of sensitivity to BILN 2061 (*14*).

In this report, a comprehensive search for variable positions in close proximity to the BILN 2061 binding site was conducted based on the currently available sequences from GenBank as well as sequences obtained from patient samples. Several substitutions, as single mutations and in combinations, significantly reduced binding of BILN 2061. A closely related nonmacrocyclic inhibitor was found to be much less susceptible to changes near the binding site than BILN 2061. Another protease inhibitor, VX-950 (Figure 1, right panel), which has also shown good preclinical and clinical activity (K_i = 44 nM, replicon EC_{50} = 350 nM, genotype 1 patients showing a maximal drop in viral load >4 log) but a significantly different binding footprint (*15*, *16*), retained potency on most of the variants.

EXPERIMENTAL PROCEDURES

Identification of Naturally Occurring Variants in Genotype 2 and 3 Proteases. Serum samples of eight genotype 3 patients were purchased from Teragenix (Fort Lauderdale, FL). HCV RNA was isolated using QIAamp Viral RNA purification kit (Qiagen). The region encompassing the protease domain was reverse-transcribed and PCR-amplified using the Titan One Tube RT-PCR kit (Roche Diagnostics). The primers used for RT-PCR have been described (*17*): P3a2-3305-3324f, TCATCACCTGGGGTGCGGAT; and P3a3-4051-4070r, GTGCTCTTACCGCTGCCGGT. The RT-PCR products were sequenced following the manufacturer's instructions using the CEQ 2000 Cycle Sequencing kit (Beckman Coulter). In addition, serum samples of 10 genotype 2 patients were purchased from Teragenix and similarly processed to obtain sequence information. The primers used for RT-PCR were 19-2a425-447f, GTTTCCGCCCGGCTAGGTAGGGA; and 16-2a1152-1130r, TATGCGGCAGGGACCTTGGTGCT.

To construct an HCV database, HCV sequences were retrieved from NCBI GenBank (release 140) (http://www.ncbi.nlm.nih.gov/Genbank/), grouped into genotypes based on information provided by the depositor, or, when the 5′UTR sequence was available, genotyped as described (*18*) using an internally developed algorithm. A total of 3060 HCV related amino acid sequences were aligned by ClustalW (*19*), and the percentage of variation at each position was calculated (*20*). About 180 genotype 1 protease sequences, 50 genotype 2 sequences, and 5 genotype 3 sequences were available from the Genbank. To obtain more genotype 2 and 3 protease sequences, 10 genotype 2 and 8 genotype 3 patients' serum samples were purchased from commercial sources and sequenced.

Plasmid Construction. To generate mutant 1b proteases carrying naturally occurring variants from genotype 2 and 3 proteases, the nucleotide changes were introduced using QuikChange (Stratagene). The parental plasmid expressing His-tagged genotype 1b single chain NS3 protease, $NS4A_{21-32}$−GSGS−$NS3_{3-181}$ (plasmid p24), has been previously described (*21*); likewise, the construction of His-tagged genotype 3 single chain NS protease has been described (*22*). To generate His-tagged genotype 2 single chain NS protease, the region encompassing the protease domain was cloned from serum samples as described above. The protease domain (aa 3−181) was subsequently PCR-amplified from the RT-PCR product using the following primers.

The 5′ primer contains an NdeI cloning site (underlined) followed by residues 21−32 of NS4A (GCVSIIGRLHIN) from genotype 2 viruses as described (*23*), a tetrapeptide linker GSGS, and sequences encoding residues 3−9 of NS3: GATATACATATGGGTTGTGTTTCTATCATTGGTAGACTGCATATTAATGGAAGTGGTAGTATCACTGCTTATACTCAGCAG.

The 3′ primer contains an EcoRI clining site (underlined) and residues 175−181 of NS3: CTCAGCGAATTCTCATGACCGTGTAGCGACATCAAG.

The PCR fragment was digested and cloned into pET-28b vector (Novagen).

Expression and Purification of Recombinant Proteases. The expression and purification protocol has been described (*21*). Briefly, plasmid DNAs encoding mutant proteases were transformed into Novagen Rossette competent cells. Single colonies were used to initiate bacteria culture in 25 μg/mL kanamycin at 37 °C. When the cell density reached $OD_{600} \sim 1.5$, the culture was induced with 0.4 mM of IPTG and grown at 23 °C for 4 h. The cell pellet was resuspended in buffer A (25 mM HEPES, pH 7.3, 300 mM NaCl, 0.1% β-octylglucoside, 10% glycerol, 2 mM β-mercaptoethanol, or 0.2 mM DTT), and cells were lysed by passing through a microfluidizer (Microfluids Corp). The lysed supernatants were incubated with Ni−NTA beads (Qiagen) for 2 h at 4 °C and then packed into columns. The Ni-columns were washed with buffer A supplemented with 20 mM imidazole and 1 M NaCl. The bound His-tagged protease was eluted with buffer A supplemented with 250 mM imidazole. The eluted fractions were pooled and dialyzed at 4 °C for 18 h against 50 mM HEPES, 300 mM NaCl, 5 mM DTT, 0.1% β-octylglucoside, and 10% glycerol. The purified proteases were analyzed on 4−12% Novex NuPAGE gel (Invitrogen) and aliquoted for storage at −80 °C.

Protease Activity Assay. Recombinant proteases were tested using a chromogenic assay as described (*24*). The assays were performed at 30 °C in 96-well microtiter plates. Protease (100 μL) was added to 100 μL of assay buffer (25 mM MOPS, pH 6.5, 20% glycerol, 0.3 M NaCl, 0.05% lauryl maltoside, 5 μM EDTA, and 5 μM DTT) containing chromogenic substrate Ac-DTEDVVP(Nva)-*O*-4-phenylazophenyl ester. The reactions were monitored at intervals of 30 s for 1 h for change in absorbance at 370 nm using a Spectromax Plus microtiter plate reader (Molecular Devices). To determine enzyme concentration to be used in the assay, proteases were tested (final concentration: 1.6−100 nM) to achieve ∼12% substrate depletion over the course of the assay. To evaluate kinetic parameters of recombinant proteases, a range of substrate concentrations (0.29−150 μM) was used. Initial velocities were determined using linear regression and kinetic constants obtained by fitting the data to the Michaelis−Menten equation (PRISM, Graphpad Software, Inc.). To assay potency of VX-950, the inhibition constants were determined at fixed concentrations of enzymes and substrate (40 μM) and fitted to the two-step, slow-binding inhibition model using SAS (*25*):

$$P = v_s t + \frac{(v_0 - v_s)(1 - e^{-kt})}{k}$$

Where P is the product observed, v_0 is the initial velocity, and v_s the velocity at steady state. The overall inhibition constant K_i* was used as the measure of inhibitor potency:

$$v_s = \frac{V_{max}[S]}{K_m\left(1 + \frac{[I]}{K_i^*}\right) + [S]}$$

For BILN 2061, the tight-binding inhibition constant (K_i) was calculated as described (*26*):

$$K_i = \frac{[I] - [E]\left(1 - \frac{v}{v_0}\right)}{\frac{v_0}{v} - 1}$$

Where [E] is enzyme concentration, v is the initial velocity observed at inhibitor concentration [I], and v_0 is the initial velocity observed in the absence of inhibitor.

Synthesis of Inhibitors. BILN 2061 (*27*) and compound **1** (*28*, *29*) were prepared using procedures described in PCT publications. VX-950 was synthesized following similar procedures as described (*30*, *31*). The integrity of the compounds was confirmed by ^{1}H NMR and mass spectrometry.

Modeling Studies. The crystal structure of BILN 2061 complexed with the HCV protease NS3−NS4A was determined in-house. Compound **1** was modeled based on the observed binding mode of BILN 2061 by removing the P1, P3 macrocycle and modifying the corresponding side chains and the N-terminal cap. The inhibitor coordinates were then energy-minimized in the protein active site of BILN 2061. VX-950 was modeled based on the binding mode of a P1−P3 ketoamide inhibitor (SCH 503034) (Guo, et al., manuscript in preparation), and incorporating the crystal structure information of another ketoamide inhibitor, which has the same N-terminal cap (PDB code 1RGQ). The coordinates were then optimized in the active site of SCH 503034. In general, 500 steps of Steepest Descent followed by 500 steps of Conjugated Gradient energy minimization were performed using the SYBYL molecular modeling program (Tripos, Inc.). Protein side chains within 10 Å of the inhibitor were allowed to move during minimization.

In addition, free energy perturbation (FEP) calculations using the modeling program CHARMM (*32*) were carried out for the D168Q mutation involving compound **1** and BILN 2061. An all-atom representation of the protein, the inhibitor, and explicit water molecules was used. In these calculations, the wild-type enzyme was converted to the mutant through a series of intermediate unphysical states by using a coupling parameter λ ($0 < \lambda < 1$). Simulations with different values of λ were performed, and the free energy changes for all the λ intervals were added together to obtain the overall free energy change. Because of multiple molecular dynamics simulations that are involved in a pairwise comparison, the FEP method is rather computationally intensive (*33*). Details of the calculation will be published elsewhere (Guo et al., manuscript in preparation).

RESULTS

Identification of Variable Residues Near the Inhibitor Binding Site. To identify residues which may affect the

Table 1

	variable aa in genotype 2					variable in genotype 3		variable in 2 + 3	
genotype	78	79	80	122	160	123	168	132	no. of seq. analyzed
1	V: 100%	D: 100%	**Q: 92%** L: 8%	**S: 69%** G: 15.5% T: 8.% V: 0.5%	**T: 99.5%** N: 0.5%	**R: 99.5%** K: 0.5%	**D: 96%** E: 4%	**V: 62%** **I: 37.5%** L: 0.5%	186
2	V: 16% **A: 84%**	D: 19% **E: 81%**	Q: 14% L: 2% **G: 84%**	S: 5% **R: 61%** **K: 32%** N: 2%	T: 35% A: 28% S: 37%	R: 100%	D: 100%	V: 3.5% I: 5% **L: 84.5%**	57−63[a]
3	V: 100%	D: 100%	Q: 100%	S: 100%	T: 100%	T: 100%	**Q: 92%** R: 8%	I: 19% **L: 81%**	13
mutants	V78A	D79E	Q80G	S122K	T160A	R123T	D168Q	V132L	

[a] The number of sequences analyzed varied at different amino acid positions. The most prevalent sequences are in bold.

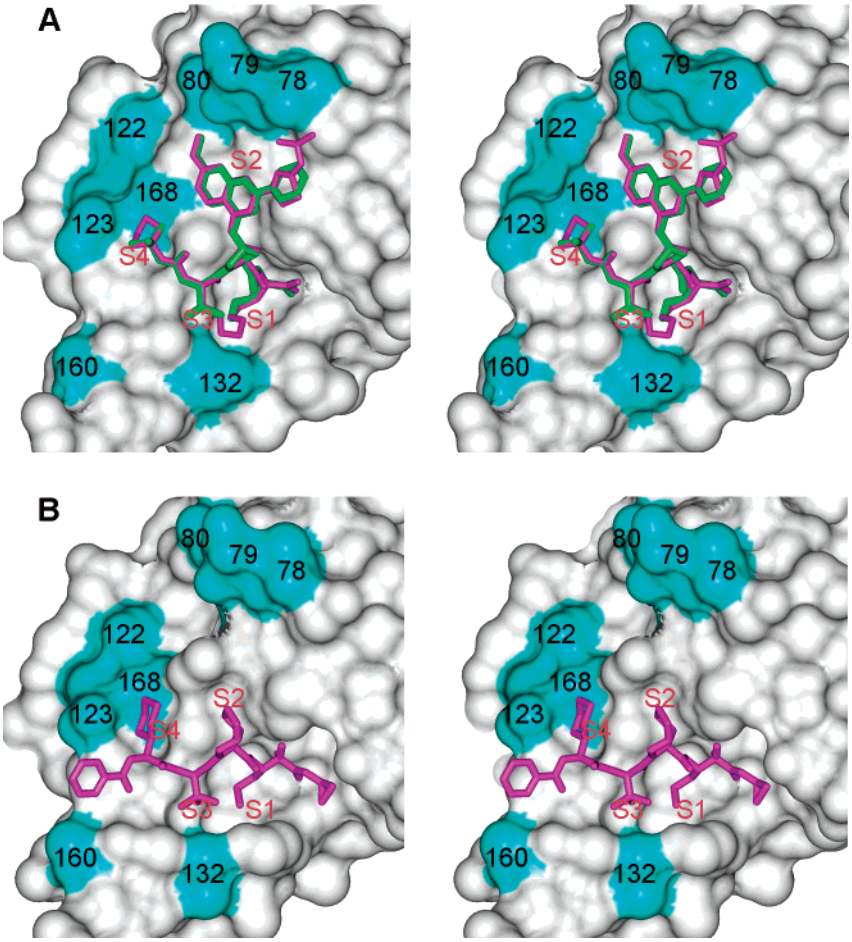

FIGURE 2: Structure of protease inhibitors bound to the active site of genotype 1b protease. (A) X-ray crystallographic structure of the NS3 protease and BILN 2061 (purple) and compound **1** (green) complexes. (B) An NS3 protease−VX-950 (purple) complex model generated by incorporating the binding mode information of a P1−P4 peptide inhibitor complexed with the HCV protease (PDB code 1RGQ). The naturally varying residues within 5 Å of the active site among genotype 1, 2, and 3 proteases are indicated.

sensitivity of genotype 2 and 3 proteases to inhibitors, residues within a 5 Å sphere of the binding sites were located by using the crystal structure of the enzyme−inhibitor complex of BILN 2061 and modeled structure of VX-950. About 180 genotype 1 protease sequences were retrieved from GenBank. However, only 50 genotype 2 sequences and 5 genotype 3 sequences were available from the database. To obtain more sequence information on genotype 2 and 3 proteases, 10 genotype 2 and 8 genotype 3 patients' serum samples were purchased from commercial sources and sequenced. On the basis of the sequence alignment, most of the residues surrounding the inhibitor binding site were conserved among the three genotypes. When genotype 1 protease was used as the master sequence, the amino acids at four positions (aa 78, 79, 80, and 122) were found to vary between genotypes 1 and 2 (but were conserved between 1 and 3); two positions (aa 123 and 168) were found to vary between genotypes 1 and 3 (but were conserved between 1 and 2). One additional position (aa 132) was more conserved between genotype 2 and 3 but diverged from genotype 1 (Table 1 and Figure 2A, BILN 2061 shown in purple). The majority of genotype 2 and 3 isolates contains a leucine at

Table 2

enzyme	K_m (SE) (μM)	k_{cat} (SE) (min^{-1})	comments
1b	3.9 (0.4)	24 (1)	
2	2.6 (0.4)	26 (1)	wild-type
3	7.2 (0.5)	34 (4)	
V78A	5.9 (0.8)	18 (1)	single
D79E	5.8 (0.6)	13 (1)	mutants
Q80G	5.5 (0.8)	13 (1)	for
S122K	4.8 (1.4)	35 (3)	genotype
T160A	5 (0.9)	12 (1)	2
R123T	12 (2)	8 (1)	single mutants
D168Q	4.2 (0.9)	20 (1)	for genotype 3
V132L	3.4 (0.9)	5.7 (0.4)	single mutant for genotype 2 + 3
V78A D79E Q80G	7.3 (0.5)	10 (0.2)	combined
S122K V132L	4.4 (0.7)	15 (1)	mutants
V78A D79E Q80G S122K V132L	4.7 (1)	8 (1)	for genotype 2
R123T D168Q	5.1 (0.3)	15 (0.1)	combined mutants
R123T D168Q V132L	5 (2)	18 (2)	for genotype 3

position 132, but the majority of genotype 1 isolates contains a valine or isoleucine. The modeled structure of the protease with VX-950 is shown in Figure 2B for comparison. The amino acid at position 160 is variable in genotype 2 and may affect the binding of the P4 capping group of VX-950.

Impact of Variable Residues on Enzyme Activity. To investigate the effect of sequence variation on inhibitor binding, residues in the single chain genotype 1b protease (*21*) were replaced with the corresponding amino acids from genotype 2 or 3 (Table 1) and the mutant enzymes were analyzed for activity and sensitivity to BILN 2061 and VX-950. The kinetic parameters of the mutant enzymes were obtained using a substrate based on the NS5A-5B cleavage junction in which the P side peptide was esterified to an aromatic chromophore (*24*). As shown in Table 2, the proteases from the three genotypes had kinetic parameters within experimental error of each other, and the various mutations had little effect on substrate binding (K_m) or turnover number (k_{cat}). The only possible exception is mutant V132L, which had a marginally lower k_{cat} (5× fold).

Impact of Variable Residues on the Binding of BILN 2061 and Analogue. As previously reported (*14*), genotype 2 and 3 proteases were less sensitive to BILN 2061, resulting in a 30−40-fold increase in inhibition constants (K_i) compared to the parental genotype 1b enzyme (Figure 3A). Among the mutant 1b proteases generated from the genotype 2 sequence, the D79E, Q80G, and S122K mutations all gave rise to proteins with severalfold increases in K_i, whereas V78A and T160A had no impact (Figure 3A). Among the amino acid replacements based on the sequence of genotype 3, R123T had miminal impact (2-fold), whereas D168Q increased the K_i by 70-fold. Substitution of V132 with leucine (the dominant amino acid in genotype 2 and 3) also gave rise to a marginal increase in K_i (3-fold).

Having identified individual residues which affected BILN 2061 binding, combinations of these mutations were examined. The triple mutant, V78A/D79E/Q80G, was generated, as these mutations are linked in a majority of the genotype 2 sequences (Table 1) and their side chains line the active site. A double mutation, S122K/V132L, was also introduced into the 1b protease to see whether there was any synergy or interaction between the two apparently independent loci. Finally, all five mutations described above were combined, V78A/D79E/Q80G/S122K/V132L, to see if they could account for the full loss of sensitivity to BILN 2061. As shown in Figure 3B, the decrease in binding to the triple mutant V78A/D79E/Q80G (8-fold) was roughly what would be anticipated from combining the impact of the individual mutations (1-, 4-, and 2-fold, respectively), whereas the double mutation S122K/V132L gave rise to only a 3-fold decrease in binding, similar to either of the single mutations, suggesting that the impact of these two mutations cannot be combined. Importantly, the combination of all five mutations decreased the binding to BILN 2061 by 20-fold compared with 40-fold for the wild-type genotype 2 enzyme, suggesting that the five positions identified in the study were indeed responsible for the loss of potency of BILN 2061 on genotype 2 enzyme.

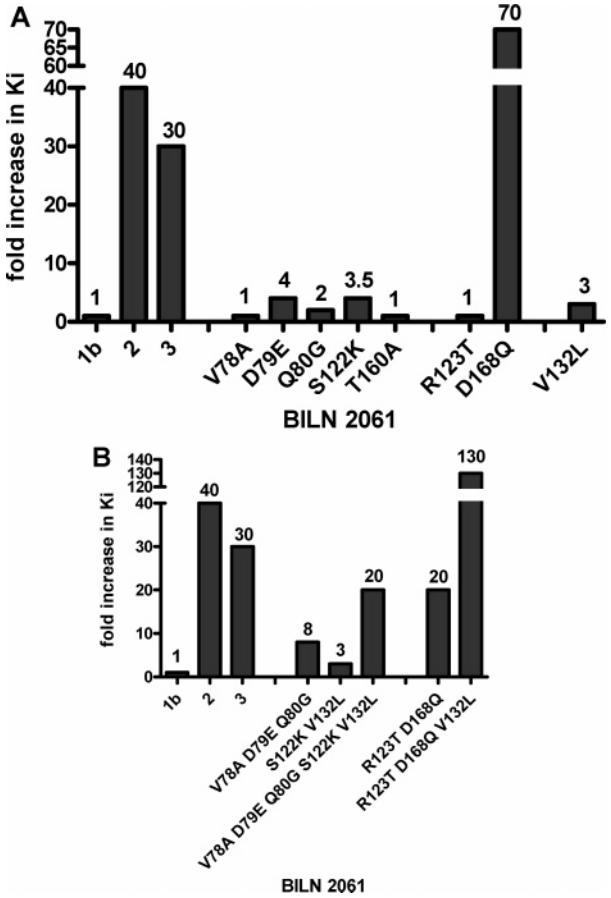

FIGURE 3: Effect of protease substitutions on the binding of BILN 2061. (A) Increases in K_i (fold) of single substitutions over the genotype 1b enzyme. (B) Increases in K_i (fold) of combinations of substitutions over the genotype 1b enzyme. Effect on genotype 2 and 3 proteases is shown for comparison. The 95% confidence interval for the inhibition assay is <3-fold.

In genotype 3 protease, the two variants R123T and D168Q appeared to be genetically linked (Table 1). A double mutation R123T/D168Q was therefore introduced into the 1b protease. The double mutant R123T/D168Q gave rise to a 20-fold increase in K_i, less than the single mutant D168Q (70-fold) (Figure 3). However, when the double mutation R123T/D168Q was combined with V132L, the triple mutation R123T/D168Q/V132L resulted in an increase of the K_i of more than 100-fold (Figure 3B), suggesting that these three mutations in combination could account for the full loss of potency of BILN 2061 on the genotype 3 enzyme.

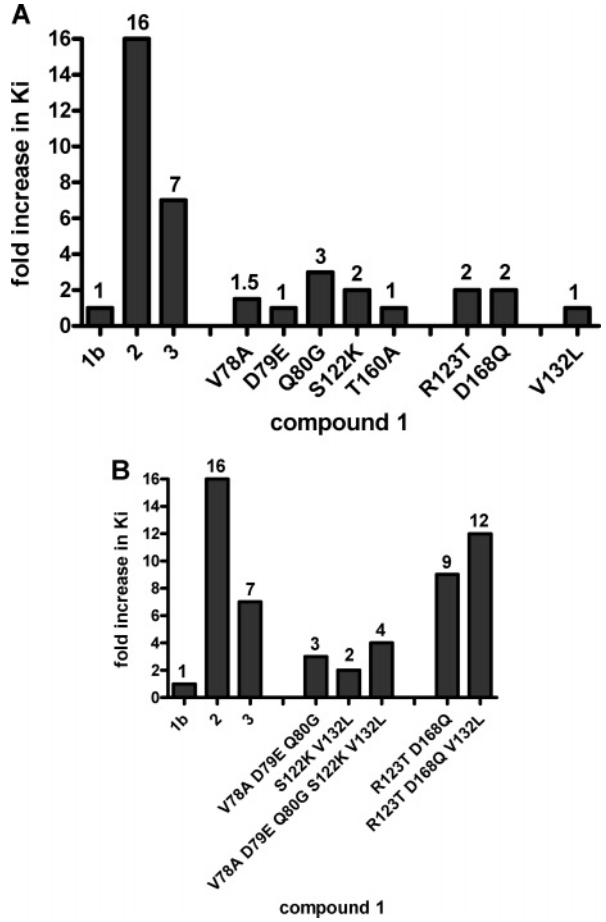

FIGURE 4: Effect of protease substitutions on the binding of compound **1**. (A) Increases in K_i (fold) of single substitutions over the genotype 1b enzyme. (B) Increases in K_i (fold) of combinations of substitutions over the genotype 1b enzyme. Effect on genotype 2 and 3 proteases is shown for comparison. The 95% confidence interval for the inhibition assay is <3-fold.

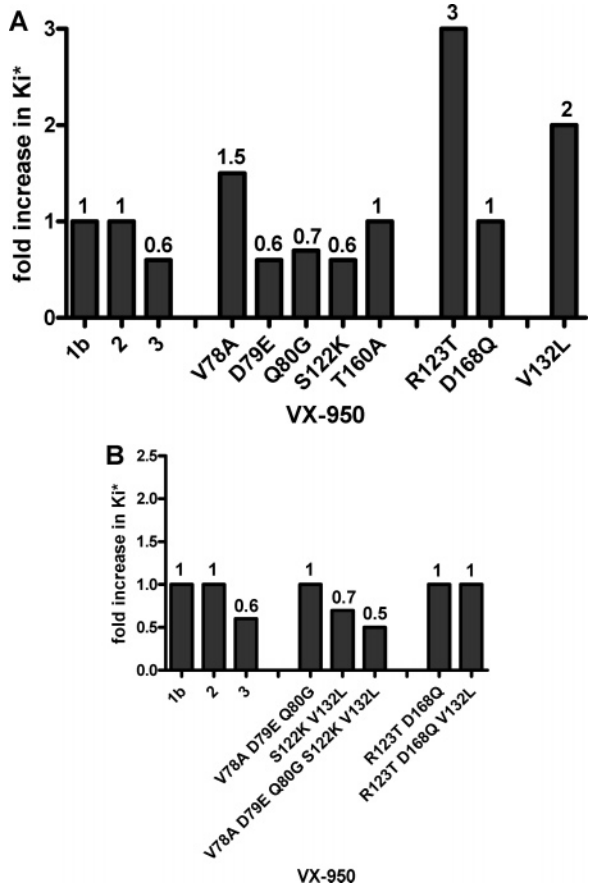

FIGURE 5: Effects of protease substitutions on the binding of VX-950. (A) Increases in K_i (fold) of single substitutions over the genotype 1b enzyme. (B) Increases in K_i (fold) of combinations of substitutions over the genotype 1b enzyme. Effect on genotype 2 and 3 proteases is shown for comparison. The 95% confidence interval for the inhibition assay is <3-fold.

To better understand the loss of potency of BILN 2061 against genotype 2 and 3 proteases, a closely related inhibitor, compound **1** (*34*), was studied. Compound **1** has a similar P2 group to BILN 2061 (Figure 1), and treatment of replicon cells with the inhibitor yielded resistant clones with mutations at residue 168 (*34*). Structural modeling also suggested that compound **1** binds to the enzyme in a manner similar to BILN 2061 (Figure 2A, compound **1** shown in green). The major difference between compound **1** and BILN 2061 is that compound **1** is not a macrocyclide and is therefore more flexible. Although compound **1** was less potent than BILN 2061 in inhibiting genotype 1 protease (K_i = 20 nM and 1 nM, respectively), it was less affected by almost all the variations near the active site (Figure 4). For example, compound **1** lost potency on genotype 2 and 3 proteases by 16- and 7-fold, respectively, compared with 38- and 29-fold loss, respectively, for BILN 2061. The D168Q mutation, which increased the K_i of BILN 2061 by nearly 70-fold, had only a small effect on compound **1** (2-fold). However, the overall profile of the various single and multiple mutations on compound **1** was similar to that of BILN 2061, although the magnitude of the impact was smaller.

Computational Analysis of the Impact of Mutation D168Q. As discussed earlier, among the mutations studied, D168Q showed the largest change in the fold resistance for BILN 2061. From Figures 3 and 4, the differential fold resistance between BILN 2061 and compound **1** (increase in K_i for BILN 2061/increase in K_i for compound **1**) is 35, or a 2.1 kcal/mol differential impact of the mutation on binding free energy ($\Delta\Delta G$). To correlate the shift in experimental binding energy with structural differences in the mutant protein, free energy perturbation (FEP) was used. FEP has been successfully applied to calculate the relative binding free energy of protein−ligand complexes and to study the effect of mutation on inhibitor binding (*35*). For both compound **1** and BILN 2061, the calculations showed an increase in binding free energy upon introduction of the D168Q mutation, with a differential impact ($\Delta\Delta G$) of 2.4 kcal/mol, in good agreement with the experimental value of 2.1 kcal/mol.

Impact of Variable Residues on the Binding of VX-950. Inhibition constants were obtained for inhibitor VX-950 on all the mutant proteases. Unlike BILN 2061, the potency of VX-950 was minimally affected (<3-fold) by any of the single or multiple mutations (including substitution at position 160, which is in close proximity to its P4 capping group). This result indicates that the two inhibitors are substantially different in their tolerance of protease sequence variation (Figure 5).

DISCUSSION

Much of the drug discovery effort for new anti-HCV therapies has been targeted toward genotype 1 virus because

of the significant unmet medical need in this infected population. However, HCV genotypes 2 and 3 viruses account for a large proportion of infections worldwide (*36*). In the past 10 years, the proportion of genotype 3a has been increasing because intravenous drug use is now the principal route of transmission, accounting for about 70% of new cases (*3*). Although the current treatment is relatively effective in this group of patients, safer therapies are needed to avoid the side effects associated with ribavirin and interferon administration.

In this study, the sensitivities of genotype 2 and 3 proteases to two distinct classes of protease inhibitors, represented by BILN 2061 (and an analogue) and VX-950, were compared to that of genotype 1b. All proteases had similar kinetic parameters, reflecting conservation of the P side of the active site across the three genotypes ((*22*) and as described in Table 1); the peptide substrate used in the assay interacts only with the S subsites (*24*). However, the genotype 2 and 3 proteases were less sensitive to the inhibition by BILN 2061, suggesting that the inhibitor interacts with different residues than the substrate (i.e., that certain residues only affect inhibitor binding) and/or that the peptide substrate used to evaluate the activity is more accommodating to changes in the active site. Residues near the inhibitor binding site that are variable between the three genotypes were identified based on a comprehensive alignment of the available HCV protease sequences. The substitutions of variable residues had little effect on substrate binding or enzyme turn over rate, consistent with the notion that these variants are not expected to have any major impact on protease function as they have survived the rigor of natural selection.

In contrast, several of these substitutions were found to significantly reduce the binding of BILN 2061, consistent with a previous report (*14*). In genotype 2, residues at positions 78−80 are different from those in genotype 1 and 3 proteases. These residues are located in close proximity to the quinoline moiety of BILN 2061 (Figure 2A), and mutations at positions 79 and 80 indeed reduced inhibitor binding. The impact of the combination of substitutions at all three positions was multiplicative and accounted for about one-third of the total loss of potency. Residue 122 is located near the charged residues D168 and R123 (Figure 2). On the basis of the X-ray structure, the side chain of R155 is observed to be pushed toward amino acid 122 upon the binding of BILN 2061 in order to accommodate the P2 quinoline group. Consequently, the guanidinium group of R155 would be in close proximity to the side chain of S122. Replacement of the neutral serine (as seen in genotype 1 and 3) by a charged amino acid (lysine or arginine in genotype 2) would increase the positive charge in the region, which would not be favored electrostatically. The larger side chains of lysine and arginine may also interfere with the conformation of nearby residues. The experimental results agree well with such structural analysis; both substitutions (S122K and S122R) reduced inhibitor binding by a few fold (Figure 3A and ref *14*). Amino acid 132 is part of the S1−S3 pocket which interacts with the macrocycle ring of BILN 2061. The dominant residue at 132 is valine in genotype 1 and leucine in genotype 2 and 3. When introduced into the genotype 1 protease, the substitution (V132L) reduced inhibitor binding by a few fold, which may be caused by the loss of tight binding of the P1 group to the S1 pocket when V132 is replaced by a more flexible leucine. Interestingly, the combination of S122K and V132L substitutions were not additive, unlike the combination of mutations 78, 79, and 80. Nevertheless, the combination of all five substitutions discussed above accounted for most of the loss of potency of BILN 2061 on the genotype 2 enzyme. Additional studies are needed to identify other residues, possibly further away from the inhibitor binding site, which may play a role in conferring resistance to BILN 2061.

In the genotype 3 enzyme, position 123 is a threonine instead of arginine as in genotype 1 and 2; and position 168 has a glutamine in place of the aspartic acid. The mutation D168Q increased the inhibition constant by over 60-fold, whereas the R123T substitution had no effect. Residue 168 was of particular interest, since it was previously identified as a major determinant conferring resistance to BILN 2061 and its analogues (*16*, *34*). This observation is also consistent with the crystal structure which shows that the P2 quinoline moiety partially overlaps with a chain of residues with alternating charges that include R155, D168, and R123. The salt bridge between D168 and R123 may facilitate binding of the P2 group, and this interaction would be disrupted by the D168Q mutation. We note that an earlier study (*22*) comparing genotype 1b and 3a proteases reported that residue 123, not residue 168, was primarily responsible for the differences observed in sensitivity to hexapeptide inhibitors with substitutions in the P4 and P5 positions. These results are not inconsistent with the current findings, as the interaction sites of the two classes of inhibitors are different. Furthermore, computational analysis showed that for both compounds the increase in binding free energy upon the D168Q mutation is mainly due to the breaking of the salt bridges of D168 with R155 and R123. The movement of these residues would destabilize the inhibitor binding conformation of the protein. The impact of this conformational change should be greater on BILN 2061 than compound **1** as a result of its increased rigidity. This analysis is consistent with the experimental results and provides a rationale for the observed differences.

In contrast to the reduced sensitivity of BILN 2061, none of the naturally occurring variants affected binding of VX-950. One advantage of VX-950 over BILN 2061 is that the P2 moiety of VX-950 is smaller and the inhibitor aligns closer to the conserved substrate binding pocket. Amino acid 122 (near R155 as discussed earlier) is one of the few variable residues that is located near the binding site of VX-950. Unlike BILN 2061, the conformation of R155 upon binding of VX-950 is similar to its conformation when bound to substrate (*37*). As a result, the guanidine group of R155 is further away from the side chain of S122, consistent with the observation that replacement of S122 to lysine had no impact on VX-950 binding. Residue 160 is located near the P4 capping group of VX-950 (∼6 Å) but is not predicted to make direct contact. As expected, mutation (T160A) at this position does not seem to have any effect.

The importance of designing inhibitors with flexibility to minimize drug resistance has been examined with HIV reverse transcriptase inhibitors (*38*, *39*) and is corroborated by our studies of BILN 2061 and its analogue compound **1**. The two inhibitors are closely related and bind similarly to the protease; the major difference is that compound **1** is not a macrocycle and is more flexible, similar to a natural peptide

substrate. Although, in general, the two compounds are affected by the same mutations near the active site (as expected based on their similar "footprints"), compound **1** is affected to a lesser extent due to its flexibility. Likewise, the flexibility of VX-950 must also contribute to its relative insensitivity to the various changes that affect interaction. Therefore, although inhibitors such as compound **1** and VX-950 bind to the active site with less affinity than BILN 2061, their flexibility allows the inhibitors to potentially have a broader spectrum. The comparative study of these inhibitors has underlined the paradox of optimizing affinity while retaining tolerance to natural viral variants in drug design (*40*).

As discussed earlier, for both compounds, the increase in binding free energy upon the D168Q mutation is mainly due to the breaking of the salt bridges of D168 with R155 and R123. The movement of these residues would destabilize the inhibitor binding conformation. BILN 2061 gives a larger increase in fold resistance than compound **1** due to its increased rigidity. Together, the structural and energetic analyses confirm the experimental values and further our understanding of the observed differences.

In summary, these studies have shown that naturally occurring variants close to the HCV protease active site can have significant impact on the potency of protease inhibitors. The reduced sensitivity of genotype 3 proteases to BILN 2061 correlates with the clinical observation that the inhibitor was less efficacious in genotype 3 patients. Even within a given genotype, sequence variations at critical positions as identified in this study may affect clinical outcome. For example, the most prevalent amino acid at position 79 in genotype 2 is glutamate (~80%), but a significant number of the isolates (~20%) have aspartic acid at this position, similar to genotype 1 and 3. Since the D79E substitution was shown to reduce susceptibility to BILN 2061 by 4-fold, the majority of the genotype 2 patients would be predicted to have reduced sensitivity to the compound; only those patients with aspartate at position 79 would be expected to show a better antiviral response. Taken together, these findings raise the possibility that choosing a more clinically effective inhibitor may be possible using sequence information from a patient prior to treatment. As more protease inhibitors enter clinical trials, the relationship between protease sequence and in vivo efficacy will be better defined.

REFERENCES

1. Wasley, A., and Alter, M. J. (2000) Epidemiology of hepatitis C: geographic differences and temporal trends, *Semin. Liver Dis. 20*, 1−16.
2. Bukh, J., Miller, R. H., and Purcell, R. H. (1995) Genetic heterogeneity of hepatitis C virus: quasispecies and genotypes, *Semin. Liver Dis. 15*, 41−63.
3. Pawlotsky, J. M. (2003) Hepatitis C virus genetic variability: pathogenic and clinical implications, *Clin. Liver Dis. 7*, 45−66.
4. Nolte, F. S. (2001) Hepatitis C virus genotyping: clinical implications and methods, *Mol. Diagn. 6*, 265−277.
5. Di Bisceglie, A. M., and Hoofnagle, J. H. (2002) Optimal therapy of hepatitis C, *Hepatology 36*, S121−127.
6. 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 alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomised trial, *Lancet 358*, 958−965.
7. Fried, M. W., Shiffman, M. L., Reddy, K. R., Smith, C., Marinos, G., Goncales, F. L., Jr., Haussinger, D., Diago, M., Carosi, G., Dhumeaux, D., Craxi, A., Lin, A., Hoffman, J., and Yu, J. (2002) Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection, *N. Engl. J. Med. 347*, 975−982.
8. Reed, K. E., and Rice, C. M. (1998) in *Hepatitis C Virus* (Reesink, H. W., Ed.) pp 1−37, Karger, Basel, Switzerland.
9. Bartenschlager, R., and Lohmann, V. (2000) Replication of hepatitis C virus, *J. Gen. Virol. 81*, 1631−1648.
10. Reed, K. E., and Rice, C. M. (2000) in *Current Topics in Microbiology and Immunology: The Hepatitis C Viruses* (Hagedorn, C. H., and Rice, C. M., Eds.) pp 55−84, Springer, Heidlberg, Germany.
11. Lamarre, D., Anderson, P. C., Bailey, M., Beaulieu, P., Bolger, G., Bonneau, P., Bos, M., Cameron, D. R., Cartier, M., Cordingley, M. G., Faucher, A. M., Goudreau, N., Kawai, S. H., Kukolj, G., Lagace, L., LaPlante, S. R., Narjes, H., Poupart, M. A., Rancourt, J., Sentjens, R. E., St George, R., Simoneau, B., Steinmann, G., Thibeault, D., Tsantrizos, Y. S., Weldon, S. M., Yong, C. L., and Llinas-Brunet, M. (2003) An NS3 protease inhibitor with antiviral effects in humans infected with hepatitis C virus, *Nature 426*, 186−189.
12. Hinrichsen, H., Benhamou, Y., Wedemeyer, H., Reiser, M., Sentjens, R. E., Calleja, J. L., Forns, X., Erhardt, A., Cronlein, J., Chaves, R. L., Yong, C. L., Nehmiz, G., and Steinmann, G. G. (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.
13. Reiser, M., Hinrichsen, H., Benhamou, Y., Sentjens, R. E., Wedemeyer, H., Calleja, J. L., Cronlein, J., Yong, C. L., Nehmiz, G., and Steinmann, G. (2003) Antiviral effect of BILN 2061, a novel HCV serine protease inhibitor, after oral treatment over 2 days in patients with chronic hepatitis C, non-genotype 1, *Hepatology 38*, Abstract No. 136.
14. Thibeault, D., Bousquet, C., Gingras, R., Lagace, L., Maurice, R., White, P. W., and Lamarre, D. (2004) Sensitivity of NS3 serine proteases from hepatitis C virus genotypes 2 and 3 to the inhibitor BILN 2061, *J. Virol. 78*, 7352−7359.
15. Perni, R. B., Chandorkar, G., Chaturvedi, P. R., Courtney, L. F., Decker, C. J., Gates, C. A., Harbeson, S. L., Kwong, A. D., Lin, C., Lin, K., Luong, Y. P., Markland, W., Rao, B. G., Tung, R. D., and Thompson, J. A. (2003) VX-950: the discovery of an inhibitor of the hepatitis C NS3 4A protease and a potential heptitis C virus therapeutic, *Hepatology 38*, Abstract No. 972.
16. Lin, C., Lin, K., Luong, Y. P., Rao, B. G., Wei, Y. Y., Brennan, D. L., Fulghum, J. R., Hsiao, H. M., Ma, S., Maxwell, J. P., Cottrell, K. M., Perni, R. B., Gates, C. A., and Kwong, A. D. (2004) In vitro resistance studies of hepatitis C virus serine protease inhibitors, VX-950 and BILN 2061: structural analysis indicates different resistance mechanisms, *J. Biol. Chem. 279*, 17508−17514.
17. Lodrini, S., Bagaglio, S., Canducci, F., De Mitri, M. S., Andreone, P., Loggi, E., Lazzarin, A., Clementi, M., and Morsica, G. (2003) Sequence analysis of NS3 protease gene in clinical strains of hepatitis C virus, *J. Biol. Regul. Homeostatic Agents 17*, 198−204.
18. Smith, D. B., Mellor, J., Jarvis, L. M., Davidson, F., Kolberg, J., Urdea, M., Yap, P. L., and Simmonds, P. (1995) Variation of the hepatitis C virus 5′ non-coding region: implications for secondary structure, virus detection and typing. The International HCV Collaborative Study Group, *J. Gen. Virol. 76* (Pt 7), 1749−1761.
19. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Res. 22*, 4673−4680.
20. 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.
21. Taremi, S. S., Beyer, B., Maher, M., Yao, N., Prosise, 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.
22. Beyer, B. M., Zhang, R., Hong, Z., Madison, V., and Malcolm, B. A. (2001) Effect of naturally occurring active site mutations on hepatitis C virus NS3 protease specificity, *Proteins 43*, 82−88.
23. Wright-Minogue, J., Yao, N., Zhang, R., Butkiewicz, N. J., Baroudy, B. M., Lau, J. Y., and Hong, Z. (2000) Cross-genotypic interaction between hepatitis C virus NS3 protease domains and NS4A cofactors, *J. Hepatol. 32*, 497−504.

24. 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.
25. Morrison, J. F., and Walsh, C. T. (1988) The behavior and significance of slow-binding enzyme inhibitors, *Adv. Enzymol. Relat. Areas Mol. Biol. 61*, 201−301.
26. Kuzmic, P., Sideris, S., Cregar, L. M., Elrod, K. C., Rice, K. D., and Janc, J. W. (2000) High-throughput screening of enzyme inhibitors: automatic determination of tight-binding inhibition constants, *Anal. Biochem. 281*, 62−67.
27. Tsantrizos, Y. S., Cameron, D. R., Faucher, A. M., Ghiro, E., Goudreau, N., Halmos, T., and Llinas-Brunet, M. (2000) WO patent 0059929.
28. Llinas-Brunet, M., Bailey, M., Cameron, D. R., Ghiro, E., Goudreau, N., Poupart, M. A., Rancourt, J., and Tsantrizos, Y. S. (2000) WO patent 0009558.
29. Llinas-Brunet, M., Bailey, M., Cameron, D. R., Faucher, A. M., Ghiro, E., Goudreau, N., Halmos, T., Poupart, M. A., Rancourt, J., Tsantrizos, Y. S., Wernic, D. M., and Simoneau, B. (2000) WO patent 0009543.
30. Babine, R. E., Chen, S. H., Lamar, J., Snyder, N. J., Sun, X., Tebbe, M. J., Victor, F., Wang, Q. M., Yip, Y., Collado, I., Garcia-Paredes, C., Parker, R. S. I., Jin, L., Guo, D., and Glass, J. I. (2002) WO patent 0218369.
31. Yip, Y., Victor, F., Lamar, J., Johnson, R., Wang, Q. M., Glass, J. I., Yumibe, N., Wakulchik, M., Munroe, J., and Chen, S. H. (2004) P4 and P1′ optimization of bicycloproline P2 bearing tetrapeptidyl alpha-ketoamides as HCV protease inhibitors, *Bioorg. Med. Chem. Lett. 14*, 5007−5011.
32. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) CHARMM: a program for macromolecular energy, minimization, and dynamics calculations, *J. Comput. Chem. 4*, 187−217.
33. Beveridge, D. L., and DiCapua, F. M. (1989) Free energy via molecular simulation: applications to chemical and biomolecular systems., *Annu. Rev. Biophys. Biophys. Chem. 18*, 431−492.
34. Trozzi, C., Bartholomew, L., Ceccacci, A., Biasiol, G., Pacini, L., Altamura, S., Narjes, F., Muraglia, E., Paonessa, G., Koch, U., De Francesco, R., Steinkuhler, C., and Migliaccio, G. (2003) In vitro selection and characterization of hepatitis C virus serine protease variants resistant to an active-site peptide inhibitor, *J. Virol. 77*, 3669−3679.
35. Wang, D.-P., Rizzo, R. C., Tirado-Rives, J., and Jorgensen, W. L. (2001) Antiviral drug design: computational analyses of the effects of the L100I mutation for HIV-RT on the binding of NNRTIs, *Bioorg. Med. Chem. Lett. 11*, 2799−2802.
36. Mondelli, M. U., and Silini, E. (1999) Clinical significance of hepatitis C virus genotypes, *J. Hepatol. 31* (Suppl. 1), 65−70.
37. Yao, N., Reichert, P., Taremi, S. S., Prosise, W. W., and Weber, P. C. (1999) Molecular views of viral polyprotein processing revealed by the crystal structure of the hepatitis C virus bifunctional protease-helicase, *Struct. Fold Des. 7*, 1353−1363.
38. Hopkins, A. L., Ren, J., Tanaka, H., Baba, M., Okamato, M., Stuart, D. I., and Stammers, D. K. (1999) Design of MKC-442 (emivirine) analogues with improved activity against drug-resistant HIV mutants, *J. Med. Chem. 42*, 4500−4505.
39. Hsiou, Y., Das, K., Ding, J., Clark, A. D., Jr., Kleim, J. P., Rosner, M., Winkler, I., Riess, G., Hughes, S. H., and Arnold, E. (1998) Structures of Tyr188Leu mutant and wild-type HIV-1 reverse transcriptase complexed with the non-nucleoside inhibitor HBY 097: inhibitor flexibility is a useful design feature for reducing drug resistance, *J. Mol. Biol. 284*, 313−323.
40. Ohtaka, H., and Freire, E. (2005) Adaptive inhibitors of the HIV-1 protease, *Prog. Biophys. Mol. Biol. 88*, 193−208.

BI051565G