JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

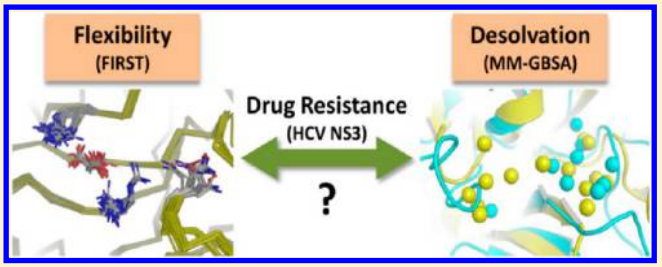
Structural Modeling of HCV NS3/4A Serine Protease Drug-Resistance Mutations Using End-Point Continuum Solvation and Side-Chain Flexibility Calculations

Hajira Ahmed Hotiana[§,†] and Muhammad Kamran Haider*[,§]

[†]Undergraduate Program in Science, [§]Department of Biology, Syed Babar Ali School of Science and Engineering, Lahore University of Management Sciences, Lahore 54792, Pakistan

Supporting Information

ABSTRACT: Computational methods of modeling protein−ligand interactions have gained widespread application in modern drug discovery. In continuum solvation-based methods of binding affinity estimation, limited description of solvent environment and protein flexibility is traded for a time scale that fits medicinal chemistry test cycles. The results of this speed-accuracy trade-off have been promising in terms of modeling structure−activity relationships of ligand series against protein targets. The potential of these approaches in recapitulating structural and energetic effects of resistance mutations, which involve large changes in binding affinity, remains relatively unexplored. We used continuum solvation binding affinity predictions and graph theory-based flexibility calculations to model thirteen drug resistance mutations in HCV NS3/4A serine protease, against three small-molecule inhibitors, with a 2-fold objective: quantitative assessment of binding energy predictions against experimental data and elucidation of structural/energetic determinants of resistance. The results show statistically significant correlation between predicted and experimental binding affinities, with R^2 and predictive index of up to 0.83 and 0.91, respectively. The level of accuracy was consistent with what has been reported for the inverse problem of binding affinity estimation of congeneric ligands against the same target. The quality of predictions was poor for mutations involving induced-fit effects, primarily, because of the lack of entropy terms. Flexibility analysis explained this discrepancy by indicating characteristic changes in side-chain mobility of a key binding site residue. The combined results from two approaches provide novel insights regarding the molecular mechanism of resistance. NS3/4A inhibitors, with large P2 substituents, derive high affinity with optimal van der Waals interactions in the S2 subsite, in order to overcome unfavorable desolvation and entropic cost of induced-fit effects. High-level resistance mutations tend to increase the desolvation and/or entropic barrier to ligand binding. The lead optimization strategies should, therefore, address the balance of these opposing energetic contributions in both the wild-type and mutant target.



INTRODUCTION

Computational structure-based approaches to estimate the affinity of protein−ligand binding are playing an increasingly important role in rational drug discovery. A wide range of approaches is now available that vary in their computational demand, accuracy, and scale of application.[1] These methods have been divided into two broad groups.[2] Class 1 methods are slow but accurate and incorporate detailed treatment of protein flexibility and explicit representation of solvent structure. These are usually considered as the benchmark methods showing very good correlation with experimental binding affinities and RMS errors ranging between 1 and 4 kcal/mol.[2,3] Two well-known formalisms are free energy perturbation (FEP) and thermodynamic integration (TI) in this class.[4] Class 2 methods, on the other hand, are fast but approximate in their treatment of solvent and protein flexibility. Further subcategories within this class include methods that use empirical functions, knowledge-based potentials, or force-field based expressions for calculating binding energy scores.[2,5] Although these scores are not true estimates of binding free energy, however, they are used for the evaluation of docked conformations of a ligand and/or for ranking of ligands within a series or a screening library. Various studies on comparative assessment of different docking and scoring approaches are present in the literature and highlight their pros and cons.[6−9] Force-field based methods have gained particular interest as they are based on well-founded assumptions regarding the physics of binding and have been shown to perform better than other scoring approaches.[6,10−12] Molecular mechanics force-field combined with an implicit solvent treatment of solvation is one of the most popular approaches in this class.[2,3] This method was pioneered by Kuhn and Kollman and is based on calculation of interaction energy from molecular mechanics (MM), polar solvation energy from Poisson−Boltzmann (PB) continuum solvation, and nonpolar solvation energy from a surface area (SA) model, hence

Received: October 4, 2012
Published: January 10, 2013

ACS Publications © 2013 American Chemical Society

referred to as MM-PBSA.[13]The more rigorous PB treatment of polar solvation free energy was later replaced by approximate but faster models such as analytical Generalized Born (GB) model and similar variants.[14,15] The MM-GBSA approaches have shown accuracy approaching that of MM-PBSA and are slightly more efficient due to less computational demand. The comparison between MM-GBSA and MM-PBSA has shown the former to be slightly more efficient.[11,12,16] Several recent studies have demonstrated application of these methods in virtual screening, rescoring of docking poses, and estimation of ligand binding energies in ligand series, using a variety of protein targets, and the results have shown better enrichment, improved ranking, and good correlation with experimental binding affinity data, respectively.[17−23] The comparison of MM-GBSA with benchmark methods such as free energy perturbation and thermodynamic integration has also shown encouraging results.[19,24]

Various limitations in continuum solvation-based binding energy predictions also exist owing to the approximations of an implicit solvent framework.[7,25−27] The errors in binding energy estimates originate from over/underestimation of solvent shielded electrostatic interactions, lack of entropic terms, incomplete sampling, explicit hydration effects, and poor estimation of protein desolvation.[2,7] In SAMPL3 blind challenge of predicting experimental binding affinities for a set of trypsin-fragment complexes, MM-PB/GBSA methods were unable to distinguish between inactive fragments from active ones.[28] The disappointing results were attributed partly to narrow range of experimental binding affinities, as the predictions were relatively better for ligand pairs with statistically significant difference in binding affinity. This pointed to the dynamic range (large spread) of binding affinity values produced by these calculations which could result from exaggerated enthalpic separation between ligands in narrow range of affinities. Consequently, the improvement in MM-GBSA methods is an ongoing area of research, and some of the recent efforts to address these issues have shown encouraging results.[29−31]

The performance of MM-GBSA based calculations in predicting changes in binding affinity upon mutations in protein structures has received relatively little attention. A detailed analysis is much needed as this problem is relatively more challenging. Furthermore, modeling energetic changes due to resistance mutations involves recapitulating dynamic ranges of experimental binding affinity which are traditionally reported as fold change in K_i or EC_{50}. Fold change values can span several orders of magnitude.[32] Here we attempt to provide a systematic analysis of MM-GBSA based binding energy calculations for a set of inhibitors against a series of mutants of a target protein. In doing so, we attempt to address the following main questions. First, can continuum solvation based binding energy calculations correctly reproduce the effects of resistance mutations on binding affinity of ligands? Second, what are the main limitations and potential directions of improvement? Finally, can we gain functional insights regarding molecular mechanism of resistance from such calculations?

The system chosen to perform this study is NS3/4A serine protease (NS3/4A) from hepatitis C virus (HCV) and three of its potent inhibitors at various stages of clinical development. NS3/4A is of huge pharmaceutical interest, and over the last ten years several inhibitors have been developed some of which are at various clinical development stages and recently two (VX-950 and boceprevir) have been approved by FDA.[33−36] HCV affects the lives of millions around the globe.[37] Its viral genome is encoded as a single polyprotein which is subsequently cleaved by viral (and host cell) proteases to yield functional proteins belonging to structural (C, E1, E2, and p7) or nonstructural groups (NS2, NS3, NS4A, NS4B, NS5A, and NS5B). NS3 is a serine protease which along with NS4A forms the fully functional enzyme (NS3/4A) that cleaves viral polyprotein to at least four sites (Figure 1S). In addition, it can target host cell proteins such as TRIF and MAVS and damages immune response.[38] The existing standard of care involves use of pegylated interferon-α alongside ribavirin which has a failure rate of up to 50% resulting from side effects, contraindications, resistance mutations, and genotype-specific responses.[33] The emergence of drug resistance is a major challenge in the use of a direct acting antiviral regimen, such as those targeting NS3/4A.[39] The quick emergence of resistance originates from the error-prone replication mechanism of viral RNA-dependent RNA polymerase (NS5B).[38]

The design of inhibitors against NS3/4A has been driven by the N-terminal product of its substrate.[35] The resulting inhibitors are mostly product-based analogs of peptidic nature and carry functional group moieties (P1′ to P4) that bind to corresponding binding pocket subsites (S1′ to S4) (Figure 1).

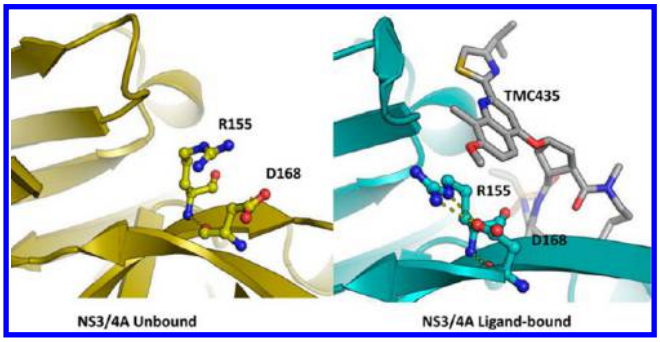


Figure 1. The role of R155 and D168 in inhibitor binding to NS3/4A. The relatively solvated orientations of R155 and D168 (left, PDB: 1JXP) undergo rearrangement upon inhibitor (TMC435) binding (right, PDB: 3KEE), resulting in optimal van der Waals overlap with a large aromatic P2 group.

Much of the effort in developing NS3/4A inhibitors has gone into optimizing these moieties, as demonstrated in several structure−activity relationship (SAR) studies.[34,40] Experimental and computational studies have also shed some light on the molecular mechanism of resistance against protease inhibitors. Romano et al. showed that some of the most widespread and high-level resistance mutations (A156 and R155) are located in the region of the binding site that is contacted by the inhibitors but not by the substrates, hence drugs designed to fit within the natural substrate envelope are expected to be less susceptible to these mutations. Inhibitors carrying a large P2 group have been shown to induce conformational rearrangements of R155 in the S2 subsite which is stabilized by D168 via a salt bridge.[41] The resulting extended conformation of R155 allows optimal hydrophobic contacts with P2 groups and results in a large gain in affinity.[42] Therefore, mutations in either of the two residues disrupt the salt bridge formation leading to suboptimal interactions with P2 and lead to a decrease in ligand binding affinity (Figure 1) .[42−44] The presence of P2-proline motif in all reported inhibitors has also been shown to cause increased susceptibility against mutations at the A156 position.[45,46] Some

dx.doi.org/10.1021/ci3004754 | J. Chem. Inf. Model. 2013, 53, 435−451

of the questions regarding the mechanism of resistance, however, have not been explicitly dealt with in these studies. Structural studies have shown that binding of the inhibitor P2 group can compensate for the conformational and solvation penalties incurred upon binding.[32,42,47] An important question in this regard is how the balance of these conflicting contributions is perturbed by mutations and what attributes of a ligand can tip the balance of these interactions toward favorable binding against mutant structures. Some crucial details such as role of side-chain flexibility and solvation changes accompanying resistance mutations also remain unclear. The energetic consequences of resistance mutations and the capacity of the MM-GBSA approach in recapitulating them prompted a thorough investigation and hence is the central theme of this study.

METHODS

Data Set. Several X-ray crystallographic structures of NS3/4A bound to various inhibitors are present in the Protein Data Bank (PDB).[48] In a majority of the complexes, ligands are bound in a conformation which is characteristic of the reversible covalent binding mode. In this study, we have focused only on noncovalent inhibitors of NS3/4A for the sake of simplicity. By the time of these experiments, three complexes of noncovalent inhibitors bound to NS3/4A were present in the PDB. The names of these inhibitors, corresponding PDB codes, and resolution of the complex structure are given in Table 1, and the inhibitor chemical structures are shown in Figure 2.

Table 1. NS3/4A−Inhibitor Complexes Selected from the PDB

inhibitor	PDB code	resolution (Å)	reference
BI201335	3P8N	1.90	42
ITMN-191	3M5L	1.25	41
TMC435	3KEE	2.40	45

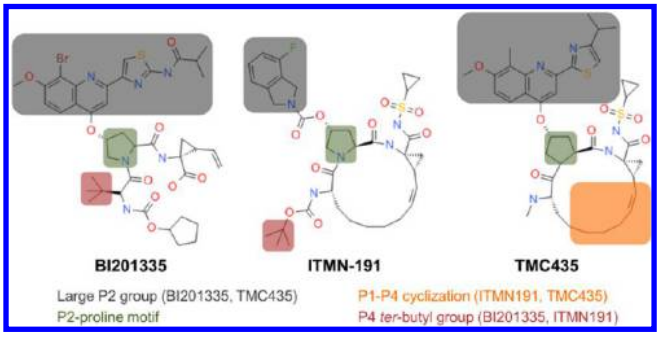


Figure 2. NS3/4A inhibitors considered in this study. All three inhibitors are under clinical development and unlike most reversible covalent inhibitors act via noncovalent binding mode. The functional groups in shaded boxes represent common and distinct features of inhibitors.

Several mutations in NS3/4A, reported from *in vitro* selection experiments and clinical trials, affect the activity of various inhibitors.[39] Experimental data on the impact of these mutations on the inhibitor activity are available in terms of fold-change in EC_{50} ($FC\text{-}EC_{50}$) or K_i ($FC\text{-}K_i$) of inhibitors.[32,47,49] A sufficiently large data set of experimental values measured in the same conditions is a prerequisite for proper assessment of predictions. We selected a set of fourteen mutations for which $FC\text{-}EC_{50}$ values were available for the three inhibitors shown in Figure 2 (Table 2). The data set was derived from two separate studies that used the same binding assay for EC_{50} measurements.[32,47] It should also be noted that variability in experimental data is reported slightly differently by the two studies (standard deviation and interquartile range), as shown in Table 2. Additionally, we also used a separate data set (Table 3) consisting of $FC\text{-}K_i$ values of the three inhibitors against 10 (out of 14) mutations in Table 2. Despite being a smaller data set, it serves as a useful additional reference as all the values come from the same study.[47] Furthermore, K_i values provide a direct thermodynamic assessment of binding, and $FC\text{-}K_i$ can be used to derive relative binding affinities in the following way:

$$\Delta\Delta G_{expt} = \Delta G_{mut} - \Delta G_{wt} = -RT\ln\left(\frac{K_{i,wt}}{K_{i,mut}}\right) \quad (1)$$

This can be directly compared to predicted relative binding affinities which we denote as $\Delta\Delta G_{bind}$. Some additional details about the choice of the data set are given in the Supporting Information (S1, Figure S2).

Preparation of the System. In order to prepare protein structures, all ligands and water molecules were removed. The cofactor, NS4A, and zinc ion were retained in the protein structures. When two or more copies of the NS3/4A-ligand complex were present in the asymmetric unit, only one was kept based on lower average B-factors of inhibitor and protein atoms in the active site. The binding site was defined as the set of residues with any atoms located at least 8 Å or less from any inhibitor atom.

The alternate side-chain conformers were not present in the binding site of any of the structures mentioned in Table 1. Where alternate conformers were present elsewhere in the protein structure, only reference conformation in the PDB file was retained for calculations. The active site of NS3/4A contains His-Asp dyad (H57, D81) for which assignment of protonation states is an important consideration. In each case we used evidence from the literature to assign protonation states of catalytic dyad and tautomeric state of H57.[32,42] In the case of TMC435 and ITMN-191, H57 and D81 were considered neutral, and H57 was assigned the δ1-NH tautomeric state (H-atom present on the δ1-nitrogen atom which faces the D80 carboxylate group, hence making them hydrogen bond donor and acceptor, respectively). In the case of BI201335, H57 was assigned the ε2-NH tautomeric state and the carboxylic group of D81 was protonated. The justification of these assignments is given in the Supporting Information (S2, Figure S3). After assigning the protonation states, protein atoms were typed with CHARMm Momany and Rone forcefield.[50,51] Hydrogen atoms were added, and their positions were optimized using CHARMm Version 3.5.1 using ABNR minimization to a gradient 0.1 kcal/molÅ. The minimized protein structures were then used as input to the docking procedure.

The parameters for ligand atoms were derived as described previously.[52] Briefly, ligand atoms were assigned atom-types from CHARMm Momany and Rone forcefield using an automated typing procedure in the Discovery Studio 3.1 package.[51] This procedure is based on a set of rules to assign types based on local topology such as nearest neighbor connectivity, hybridization, and aromaticity. Partial charges for ligand atoms were derived from MMFF94.[53]

Generation of Mutant Protein Structures. Point mutations (Table 1) were incorporated in each of the three

dx.doi.org/10.1021/ci3004754 | J. Chem. Inf. Model. 2013, 53, 435−451

Table 2. EC_{50} Values (nM) and Fold-Change in EC_{50} of Inhibitors against Wild-Type and Mutant NS3/4A

	BI201335		ITMN-191		TMC435	
	EC_{50}	FC ± SD[a]	EC_{50}	FC (IQR[b] or ± SD)	EC_{50}	FC (IQR or ± SD)
WT	4	1	0.5	1	11	1
V36M	8.4	2.1 ± 0.5	1.05	2.1 ± 0.4	20	2.0 (1.2−2.9)
T54A	3.6	0.9 ± 0.2	0.6	1.3	5.9	0.6 (0.4−0.7)
T54S	14	3.5 ± 0.7	0.6	1.2 ± 0.5	5.4	1.2 (0.9−1.5)
Q80K	8.8	2.2 ± 0.1	0.9	2.3 (2.1−2.8)	62	7.7 (6.2−8.7)
Q80L	4.8	1.2 ± 0.2	0.65	1.3 ± 0.2	12	2.1 (1.0−3.0)
Q80N	2.4	0.6 ± 0.1	0.35	0.7 ± 0.2	7.7	0.7 ± 0.4
Q80R	10.4	2.6 ± 0.5	1.6	4.5 (3.3−4.0)	49	6.9 (5.1−8.4)
R155K	1400	350 ± 14	135	447 (383−554)	260	30 (22−39)
R155Q	240	60 ± 9	13	6.8	21	1.6 (1.3−1.7)
A156T	1080	270 ± 98	4.8	41	377	44 (27−59)
A156V	600	150 ± 10	12	63 (58−69)	2149	177 (133−204)
D168A	2760	690 ± 332	31	153 (152−190)	6356	594 (473−949)
D168G	320	80 ± 30	2.2	8.1	42	4.4 (2.9−7.2)
D168V	3880	970 ± 276	15	50 ± 10	17917	2591 (1945−3137)

[a]Values with ± Standard Deviation (SD) were obtained from Lagacé et al.[47] SD where available was based on 3 experimental trials. [b]Values with Interquartile range (IQR) were obtained from Lenz et al.[32] IQR where available was based on 3 experimental trials.

Table 3. Fold Change in K_i and Calculated Relative Binding Affinity Values of Inhibitors against Wild-type and Mutant NS3/4A

	BI201335		ITMN-191		TMC435	
	FC ± SD[a]	$\Delta\Delta G_{expt}$	FC ± SD	$\Delta\Delta G_{expt}$	FC ± SD	$\Delta\Delta G_{expt}$
V36M	1	0	NA	NA	1.6	−0.28
T54A	NA	NA	NA	NA	0.9 ± 0.2	0.06 ± 0.95
Q80K	1.9 ± 0.1	−0.38 ± 1.36	1.3 ± 0.1	−0.15 ± 1.36	2.8 ± 0.4	−0.61 ± 0.54
Q80R	1.7 ± 0.1	−0.31 ± 1.36	1.3 ± 0.2	−0.15 ± 0.95	4 ± 0.2	−0.82 ± 0.95
R155K	210 ± 52	−3.15 ± 2.33	31 ± 2	−2.03 ± 0.41	44	−2.23
R155Q	430 ± 20	−3.58 ± 1.77	18 ± 1	−1.71 ± 0.00	13	−1.51
A156T	710	−3.87	4 ± 0.2	−0.82 ± 0.95	21	−1.80
A156V	1100 ± 43	−4.13 ± 2.22	7.3 ± 0.8	−1.17 ± 0.13	NA	NA
D168A	236 ± 27	−3.22 ± 1.94	NA	NA	68	−2.49
D168V	830 ± 106	−3.97 ± 2.75	5.2 ± 0.2	−0.97 ± 0.95	52	−2.33

[a]Values were obtained from reference. SD where available was based on 3 experimental trials.[47]

protein structures using Modeler.[54] The procedure involved mutation of the selected residue to the desired residue followed by molecular dynamics simulated annealing optimization of mutated residue and neighboring residues. A radius of 5 Å was used as a cutoff to select neighboring residues. As for ITMN-191, the X-ray structure of the enzyme−inhibitor complex already contained the Q80K mutation so we first created a mutant model K80Q from this structure, considering it the wild-type NS3/4A and then used this as input to create other mutant models. We evaluated the predicted orientations of mutant side-chains by comparison with reference X-ray structures. Unfortunately, the X-ray crystallographic data on NS3/4A mutants is very limited, and we could only obtain structures of two variants, R155K and V36M. For each protein−ligand complex, the positions of predicted side-chain and neighboring side-chains (within a 5 Å shell of mutated residue) were compared with a reference X-ray structure (Table 4).

Table 4. Comparison of Predicted Side-Chain Conformations in Mutant Structures with Available X-ray Crystallographic Data[c]

	V36M[a]		R155K[b]	
PDB	RMSD	$RMSD_{5\ Å}$	RMSD	$RMSD_{5\ Å}$
3KEE	0.61	0.96	0.71	1.25
3M5L	0.80	1.09	0.59	1.46
3P8N	0.50	1.03	0.54	1.20

[a]Reference PDB structure 2QV1. [b]Reference PDB structure 2OIN. [c]RMSD, root mean square deviation of mutated side-chain; $RMSD_{5\ Å}$, root mean square deviation of mutated side-chain of residues within 5 Å of mutated side-chain from reference structure.

Docking Protocol. Prior to binding energy calculations on mutated protein structures, ligands were optimally placed in the binding sites of mutant structures. This was done using CDOCKER which is a CHARMM-based docking algorithm.[51,55] The docking procedure consists of the following steps: The original conformation of each inhibitor in its PDB file was used as a starting point and subjected to simulated annealing molecular dynamics in the rigid binding site of mutant structure. The heating stage target temperature was set to 700 K in 2000 steps where cooling target temperature was set to 300 K using 5000 steps. The resulting ligand conformation was subjected to a final minimization consisting of 50 steps of steepest descent and 200 steps of conjugate gradient to a gradient 0.01 kcal/molÅ.

***In Situ* Ligand Minimization.** Once the ligands are optimally placed in the binding site, an implicit solvent minimization of the complex was performed to account for protein flexibility. During minimization the ligand itself and

dx.doi.org/10.1021/ci3004754 | J. Chem. Inf. Model. 2013, 53, 435−451

residues within 5 Å of any of the ligand atoms were allowed to move. Another set of residues within 3 Å of the 5 Å shell were harmonically restrained with a force constant of 50 kcal/mol/Å^2. The rest of the protein atoms were held fixed. The minimization was performed using ABNR protocol and to the gradient 0.01 kcal/mol/Å in GBSA implicit solvent model.

Binding Energy Calculations with MM-GBSA. *In situ* minimized complexes were subjected to binding energy calculation with MM-GBSA. The particular flavor of Generalized Born implicit solvent model used in this study,[56] referred to as GBSW (Generalized Born with a simple SWitching), uses a smoothing function to represent dielectric boundary and a van der Waals surface-based approach to calculate polar solvation energies. Despite being computationally less intensive, it is reported to produce stable molecular dynamics trajectories and has been frequently cited in protein−ligand binding energy estimation.[6,57−60] For each ligand, the binding energy is calculated as

$$\Delta G_{bind} = G_{complex} - G_{protein} - G_{ligand} \quad (2)$$

whereas the energy of each of the above species is calculated from the equation

$$G = E_{intra} + E_{elec} + E_{vdw} + G_{psolv} - G_{npsolv} - TS \quad (3)$$

E_{intra} represents gas-phase forcefield internal energy of the molecular species arising from covalent bonds, angles, torsions, and out of plane motions. E_{elec} is the electrostatic energy between charged atoms calculated from Coulomb's expression using a dielectric constant of 1. E_{vdW} is van der Waals energy calculated from the Lennard-Jones 6−12 potential. G_{psolv} represents the electrostatic component of solvation energy and was calculated using the GBSW model. G_{npsolv} is the nonpolar component of solvation free energy which was calculated from the surface area (SA) model which assumes a linear correlation between G_{npsolv} and solvent accessible surface area. A value of 5.0 cal/mol·Å^2 was used for the surface tension coefficient in the SA model. TS represents solute entropy. The binding energy of each inhibitor against NS3/4A wild-type and mutant structures, in this approach, is the sum of the following contributions:

$$\Delta G_{bind} = \Delta E_{elec} + \Delta E_{vdw} + \Delta G_{psolv} + \Delta G_{npsolv} \quad (4)$$

The calculation of ΔG_{bind} is based on a single-snapshot protocol where a suitable starting structure of the complex is subjected to an implicit solvent energy minimization. After minimization, energy of the complex is evaluated using eq 3. The energies for individual species (free protein and free ligand) are calculated in the same conformation as in the complex but in the absence of each other. An important aspect of this protocol (referred to as 1A-MMGBSA)[16] is that it leads to cancellation of energetic terms accounting for intramolecular strain of ligand and protein reorganization upon binding. It is noted, however, that this cancellation could be beneficial as it leads to a decrease in fluctuation of energetic terms and hence improves accuracy.[16] At the same time, mutational changes should also affect the binding pose of the ligand as it would be accommodated differently in the mutated binding site. In a variation of the protocol, referred to as 3A-MMGBSA, three separate energy minimizations were run for the complex, free protein, and free ligand. In calculation of relative binding energies for ligands in a series, TS is usually ignored.[61] In these calculations we also ignored the entropy terms; however, we note that configurational entropy can be an important factor contributing to binding affinity. Various issues arising from exclusion or inclusion of these terms in the binding energy calculation are discussed in detail in Results and Discussion sections. All of the calculations above were performed in Discovery Studio 3.1.[51]

Statistical Analysis. The performance of the MM-GBSA protocol was evaluated on three different measures, the coefficient of determination, R^2, predictive index, *PI*, and translated mean absolute deviation, *MADtr*. R^2 is the square of Pearson's correlation coefficient between experimental and predicted binding affinity estimates and gives an idea of the degree of linear correlation between two sets of values. *PI* is a measure of the rank correlation between the predicted and experimental sets of values and has been used to assess the quality of predictions from computational approaches for binding affinity prediction. Given a set of experimental values (e.g., EC_{50}, IC_{50}, K_i, or their log values), $E(i)$ and corresponding predicted scores, $P(i)$, *PI* is calculated as follows

$$PI = \frac{\sum_{j>i}\sum_i w_{i,j}C_{i,j}}{\sum_{j>i}\sum_i w_{i,j}} \quad (5)$$

where

$$w_{i,j} = |E(i) - P(i)| \quad (6)$$

and

$$C_{i,j} = \begin{cases} 1, \; if \dfrac{[E(j) - E(i)]}{[P(j) - P(i)]} < 0 \\ -1, \; if \dfrac{[E(j) - E(i)]}{[P(j) - P(i)]} > 0 \\ 0, \; if [P(j) - P(i)] = 0 \end{cases} \quad (7)$$

PI values vary between +1 and −1, indicating high and weak correlation with experimental rank order, respectively. Both R^2 and *PI* are helpful in performance comparison of different methods and rely on relative binding affinity estimates. Additionally, we also calculated *MADtr* for evaluation of absolute estimates. For a series of experimental binding affinity estimates, $E(i)$ and corresponding predicted values, $P(i)$, *MADtr* is calculated from mean absolute deviation (MAD) after removal of mean signed deviation (signed value of MAD) which is given as

$$MADtr = \frac{\sum \left| E(i) - P(i) - \frac{\sum (E(i) - P(i))}{N} \right|}{N} \quad (8)$$

For each of these metrics, the standard error was determined using a parametric bootstrap resampling test. For each set of predicted binding energy values, iterative resampling was performed from a Gaussian distribution whose parameters are derived from the distribution of original predicted binding energy estimates. We performed 10,000 iterations where in each iteration all three metrics (R^2, *PI*, *MADtr*) were recalculated and the standard deviation of each metric was reported as its statistical uncertainty. This approach of statistical evaluation was similar to the one reported by Genheden and Ryde and is very useful for error analysis of binding energy estimates and for comparison of different protocols.[16,29]

dx.doi.org/10.1021/ci3004754 | J. Chem. Inf. Model. 2013, 53, 435−451

Table 5. Predicted Binding Energy Values of Inhibitors against NS3/4A from MM-GBSA Calculations[a]

	BI201335		ITMN-191		TMC435	
	$\Delta\Delta G_{bind}$	RMSD	$\Delta\Delta G_{bind}$	RMSD	$\Delta\Delta G_{bind}$	RMSD
V36M	9.49	0.72	12.76	1.18	4.49	0.85
T54A	−0.38	1.12	10.73	0.86	18.29	1.12
T54S	6.84	0.36	9.17	1.34	2.42	0.69
Q80K	3.19	1.15	9.79	1.40	14.63	1.22
Q80L	10.23	1.15	10.71	1.82	14.53	1.52
Q80N	4.18	1.07	8.60	1.69	18.68	1.69
Q80R	9.75	1.42	8.84	1.34	11.20	1.41
Mean ± SD	6.19 ± 4.01		10.09 ± 1.45		12.03 ± 6.40	
R155K	22.80	1.22	19.14	2.00	21.86	1.47
R155Q	12.01	1.00	23.06	2.07	13.41	1.59
A156T	27.90	1.75	43.82	2.65	24.98	2.56
A156V	22.85	2.14	30.34	2.14	23.49	2.03
D168A	3.94	0.91	18.62	1.76	9.83	0.69
D168G	−0.36	1.03	26.09	1.47	8.21	1.30
D168V	−8.27	1.21	17.91	1.54	4.45	0.83
mean ± SD	11.15 ± 3.63		25.27 ± 9.23		15.18 ± 8.22	

[a] $\Delta\Delta G_{bind}$, relative binding energy values (with reference to wild-type) for each inhibitor against each mutant structure. RMSD, root mean square deviation of predicted ligand conformation in mutant structure from corresponding conformation in wild-type structure.

Covariance Analysis of Ligand-Induced Side-Chain Motions. In order to assess ligand-induced correlated side-chain motions in the active site, we performed a covariance analysis of their relative positions in a set of X-ray crystallographic structures. A set of 27 NS3/4A-ligand complex structures was downloaded from the PDB, and binding sites were extracted based on the same criterion as defined for docking calculations above. A reference structure of the active site was obtained by averaging over three-dimensional coordinates in all structures. Each side-chain was represented as a centroid (over side-chain atoms only), and its distance from the reference structure was calculated. Based on the fluctuations in side-chains from reference position ($x_{i,ref}$), an $N \times N$ pairwise covariance matrix was calculated, where N is the number of side-chains. Each element of the matrix represents pairwise covariance between fluctuations of residue i and j and is given as

$$Cov_{i,j} = \langle (x_i - x_{i,ref})(x_j - x_{j,ref}) \rangle \quad (9)$$

Typically, pairwise covariance analysis of residue positions is applied on molecular dynamics trajectories, and it provides a good understanding of correlated dynamics in protein structures. Instead of using trajectory of molecular dynamic simulation, we used a set of experimentally determined structures of NS3/4A bound with different ligands. The changes in side-chain positions observed in this manner represent an ensemble which consists of experimentally observed conformations and hence provides a relatively more 'realistic' sample of conformational space accessible to binding site residues.

Binding Site Flexibility Analysis. The FIRST (Floppy Inclusion and Rigid Substructure Topology)[62,63] package was used to perform flexibility analysis. In FIRST framework, protein structure is represented as a topological network consisting of vertices and edges that correspond to atoms and covalent/noncovalent interactions, respectively. A constraint-counting algorithm is then applied in order to classify parts of the structure as rigid or flexible depending on independent degrees of freedom in the network. The accessible degrees of freedom are based on the number and strength of hydrogen bonds or hydrophobic interactions. A flexibility index is reported at bond level which can be used to analyze flexibility of a local region of the structure such as the binding site. Flexibility index ranges from +1 (completely flexible) to −1 (completely rigid). A Python script was used to calculate a residue-based flexibility index which identifies rotatable bonds in a residue and reports a residue-based flexibility measure which is an average over bond-level flexibility indices.

The calculations were applied on two sets of structures. The unbound state of the protein was represented by apo-NS3/4A (1JXP) and mutant models derived from it (prepared as described earlier). The bound state of the protein was represented by ligand bound protein structures (3P8N, 3M5L, and 3KEE) and mutant models derived from them. The residue-based flexibility index, F_i, was calculated in each case. F_i is a normalized density of accessible degrees of freedom in the binding site residue which can be used for pairwise comparison. The only adjustable parameter in FIRST calculation is hydrogen bond energy cutoff. This was set to its default value of −1.0 kcal/mol. This value has been used in other studies and is considered to be an optimal choice.[62,64,65]

Conformational Entropy Estimation. Side-chain conformational entropies were estimated using a simple accessible surface area model.[66] This model estimates conformational entropy from changes in accessible surface area upon binding. The relationship is based on the fact that fixing of a side-chain in a single rotamer upon ligand binding decreases its conformational degrees of freedom. The burial of side-chain can, therefore, be an indirect way of estimating the change in conformational entropy. Hence, the contribution of an individual side-chain to conformational entropy is given by

$$\Delta S_{conf} = \sum S'_{sc} \frac{\Delta ASA_{mut-wt}}{ASA_{max}} \quad (10)$$

$\acute{S}_{sc}$ is maximum conformational entropy of the side-chain, and $\Delta ASA_{mut\text{-}wt}$ is the difference in accessible surface area of the side-chains between mutant and wild-type bound conformation of NS3/4A. ASA_{max} is the change in accessible surface area of the side-chain if it was fully exposed. The values of $\acute{S}_{sc}$ were

dx.doi.org/10.1021/ci3004754 | J. Chem. Inf. Model. 2013, 53, 435−451

Table 6. Quality Measures for Assessment of Predicted Binding Affinity Estimates from MM-GBSA Calculations[b]

	BI201335		ITMN-191		TMC435	
$\Delta\Delta G_{bind}$ vs[a]	$\Delta\Delta G_{expt}$	pEC_{50}	$\Delta\Delta G_{expt}$	pEC_{50}	$\Delta\Delta G_{expt}$	pEC_{50}
SE	4.67	2.73	2.01	1.48	2.16	2.08
MADtr	6.97 ± 0.06	7.98 ± 0.05	3.84 ± 0.04	5.19 ± 0.03	4.14 ± 0.037	5.97 ± 0.04
R^2	0.71 ± 0.00	0.83 ± 0.00	0.83 ± 0.00	0.Assignment	0.23 ± 0.007	0.38 ± 0.00
PI	0.96 ± 0.01	0.91 ± 0.01	0.97 ± 0.01	0.87 ± 0.01	0.41 ± 0.015	0.65 ± 0.01

[a]These values are reported for truncated data set excluding outliers noticed in Figure 4. [b]*SE*, standard error; *MADTr*, translated mean absolute deviation; R^2, coefficient of determination; *PI*, predictive index.

obtained from Lee et al.[67] The use of $\Delta ASA_{mut\text{-}wt}$ ensures that ΔS_{conf} is an estimate of relative entropic cost of binding between wild-type and mutant protein. We used ΔS_{conf} values independently in our analyses because a combined framework for MM-GBSA based binding energy predictions and surface area based entropy estimation is not well-established.

RESULTS

Comparison of Predicted and Experimental Resistance Profiles. The relative binding energy values, $\Delta\Delta G_{bind}$, (the difference of binding energy between wild-type and mutant NS3/4A for each inhibitor against each mutation) were compared against the experimentally observed level of resistance (which is typically divided into four categories,[68] based on the range of $FC\text{-}EC_{50}$; high-level resistance: $FC\text{-}EC_{50}$ > 100; high-intermediate-level resistance: $FC\text{-}EC_{50}$ 80−100; intermediate-level resistance: $FC\text{-}EC_{50}$ 10−80; and low-level resistance: $FC\text{-}EC_{50}$ < 10). In general, large values of $\Delta\Delta G_{bind}$ were associated with high-level resistance mutations (positions 155, 156, and 168), whereas little or negligible decrease in binding energy values was associated with intermediate or low-level resistance mutations (positions 80, 54, and 36). The mean $\Delta\Delta G_{bind}$ of high-level resistance mutations was separated from that of the intermediate/low resistance mutations. However, for BI201335 and TMC435 the mean values of the two groups were within standard deviation of each other (Table 5). The anomalous predictions for D168 and A156 mutants were the major source of spread in $\Delta\Delta G_{bind}$ observed in high-level resistance mutations.

A qualitative comparison between $\Delta\Delta G_{bind}$ and pEC_{50}, at first, gave R^2 values of 0.10, 0.30, and 0.02 for BI201335, ITMN-191, and TMC435, respectively, which did not indicate any meaningful correlation. The presence of common and, in some cases, distinct outliers was noticed for each inhibitor (D168A/G/V for BI201335, D168A/V for TMC435 and D168G, A156T/V for ITMN-191) (Figure 4). As a result of excluding outliers, R^2 improved to 0.82, 0.77, and 0.37 for BI201335, ITMN-191, and TMC435, respectively. The chances of getting similar correlation coefficients on a set of random data points with corresponding size are 0.09%, 0.03%, and 11%, respectively, which indicate that the correlations obtained are indeed statistically significant, except for TMC435. Table 6 shows additional performance evaluation metrics and associated uncertainties from bootstrap resampling test. The standard error of binding affinity measurement, for each inhibitor, was calculated by averaging ΔG_{bind} values over all mutations. The values of these metrics are based on data points excluding outliers and can be compared with other studies using similar approaches but different systems. For example, a recent study of MM-GBSA-based predictions for binding of seven biotin analogues to avidin reported *MADTr, PI, and* R^2 values of 15.2, 0.85, and 0.60,[16] respectively. Another study of 13 sulfonamide inhibitors of carbonic anhydrase reported an R^2 value of 0.71 with MM-GBSA protocol.[19] Several other studies have reported a similar success rate, and hence the comparable performance reported in this study is encouraging.[10,19,24,29,30,69] The results from the 3A-MMGBSA protocol showed poor correlation with experimental data and a higher number of outliers (Table S2). The most probable explanation for the poor performance is the inaccurate calculation of changes in internal energy terms. This result was not unexpected as other studies have also shown that even small fluctuations in calculating internal energy terms can lead to inconsistent results.[16] This issue is further elaborated in the Discussion session. All of the results described below are based on 1A-MMGBSA binding energy predictions.

The calculated $\Delta\Delta G_{bind}$ values were also compared with $\Delta\Delta G_{expt}$ values which were derived from $FC\text{-}K_i$ data.[47] Although the size of this comparison set is smaller due to limited available experimental data, it reflects a more direct comparison between two thermodynamic estimates. The R^2 and *PI* values show a similar trend as observed for $\Delta\Delta G_{bind}$ and pEC_{50}. The exclusion of outliers, belonging to high-level resistance sites, led to better accuracy (R^2 values, 0.71, 0.83, and 0.23) (Figure 4).

It should also be noted that the experimental data for high-level resistance sites in general showed much more variability as compared to the intermediate/low-level resistance sites (Table 2). This uncertainty was even higher for TMC435 for which relatively weaker correlation was obtained. This reflects that at least part of the weaker correlation with these mutations stems from variability in reference data.

The experimentally determined resistance profiles also show that mutations at various sites exhibit residue-specific trends (Tables 2 and 3).[32,47] For instance, based on the experimental data used in this study, the following trends can be noted:

1. R155K is more resistant than R155Q (except for ITMN-191).
2. A156V is more resistant than A156T for all three inhibitors (for BI201335 $FC\text{-}K_i$ and $FC\text{-}EC_{50}$ data were in conflict).
3. D168 mutants show a somewhat size-dependent trend where large nonpolar residues show greater resistance and vice versa.
4. Q80K and Q80R exhibit almost comparable resistance, which is slightly higher than Q80L and Q80N, whereas Q80N exhibits the least resistance in this group.

From the calculated binding energy values (Figure 3 and Table 5), it can be noticed that trends 1 and 4 were completely or partially reproduced, whereas predictions did not agree with trends 2 and 3.

Interestingly, some inhibitors-specific trends across the set of mutations were reproduced in the predicted resistance profiles. For example, it has been observed that Q80 variants exhibit a much lower level of resistance against BI201335 than against

dx.doi.org/10.1021/ci3004754 | J. Chem. Inf. Model. 2013, 53, 435−451

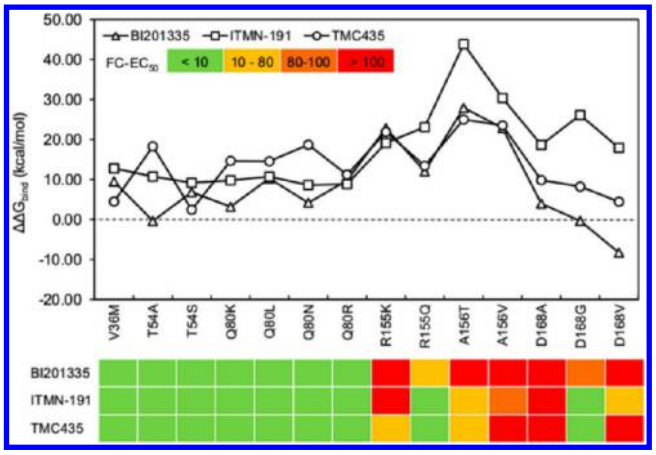


Figure 3. A qualitative assessment of binding energy predictions. Predicted relative binding affinities are plotted for each inhibitor against each mutation. *In vitro* susceptibility data are represented, at the bottom, as a color coded scheme reflecting the level of fold-change in EC_{50}. High $FC\text{-}EC_{50}$ values correspond to high levels of resistance and vice versa.

macrocyclic inhibitors such as TMC435.[47] This was supported from the predictions as TMC435 showed the highest $\Delta\Delta G_{bind}$ for Q80 variants (except Q80R) among other inhibitors. Similarly, R155 and A156 variants are known to be particularly resistant against inhibitors containing a large P2 moiety attached to the P2-proline group, which is the case for all three inhibitors in this study. The comparable $\Delta\Delta G_{bind}$ for three inhibitors on these positions also reproduce this trend, with ITMN-191 predicted to be more resistant toward R155Q and A156V/T.

In a nutshell, calculated $\Delta\Delta G_{bind}$ values generally discriminated between high-level and intermediate/low-level resistance mutations and showed good agreement with $FC\text{-}EC_{50}$ in most cases. However, residue-specific changes in the resistance level were captured somewhat less efficiently by the protocol. This was particularly the case for high-level resistance sites (A156 and D168).

Balance of van der Waals Interactions and Desolvation in High-Level Resistance Mutations. *D168A/G/V.* In order to investigate the impact of high-level resistance mutations in the S2 subsite, we analyzed the deviation in inhibitor binding mode (RMSD of ligand pose from wild-type conformation) in the active site and consequent changes in various energetic contributions to ΔG_{bind} (Table 5). Various energetic contributions, ΔG_{bind}, ΔE_{vdw}, and ΔG_{desolv} (which is equal to ΔE_{elec} plus ΔG_{psolv} and represents the desolvation penalty) were plotted for each inhibitor against each mutation (Figure 5). This energetic profile analysis was based on the observation that ΔE_{vdw} and ΔG_{desolv} were predominant opposing contributions whose balance determined the ΔG_{bind} in these calculations. It was expected that high-level resistance mutations would manifest as large deviations in inhibitor binding mode (high RMSD) and significant changes in energetic profiles.

Taking D168A as an example, we noticed that ITMN-191 showed the largest RMSD (Table 5) and the most distinct energetic profile from wild-type conformation, as compared to BI201335 and TMC435 (Figure 5). This was consistent with the correlation observed between $\Delta\Delta G_{bind}$ and pEC50 where D168A/V were not outliers in the case of ITMN191 (Figure 4). The energetic profile in D168A indicated ITMN-191 to be associated with a decrease in ΔE_{vdw} and increase in ΔG_{desolv}. At a structural level, the main source of high RMSD of ITMN-191

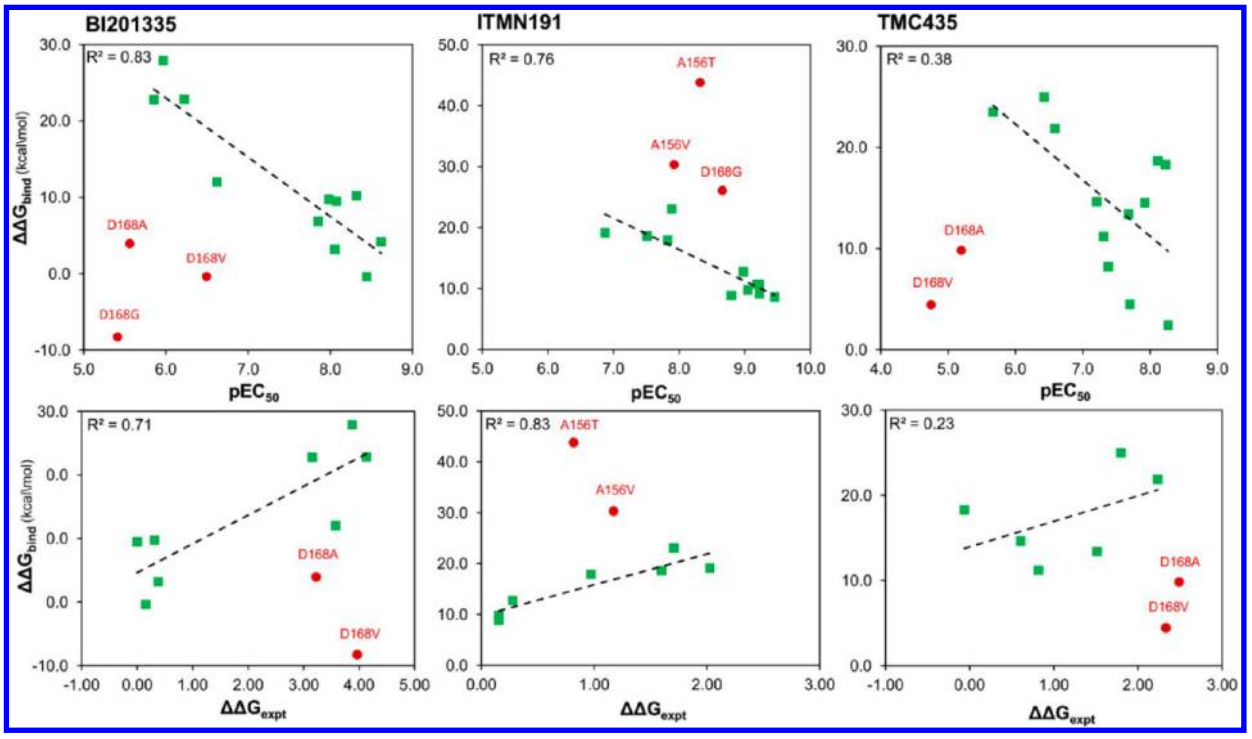


Figure 4. Correlation between experimental and predicted binding affinity measurements. $\Delta\Delta G_{bind}$ correspond to predicted relative binding affinities ($\Delta G_{bind,wt} - \Delta G_{bind,mut}$), $\Delta\Delta G_{expt}$ values were derived from K_i data using eq 1, and pEC_{50} values represent negative log conversions of nanomolar EC_{50} values from refs 47 and 48.

dx.doi.org/10.1021/ci3004754 | J. Chem. Inf. Model. 2013, 53, 435−451

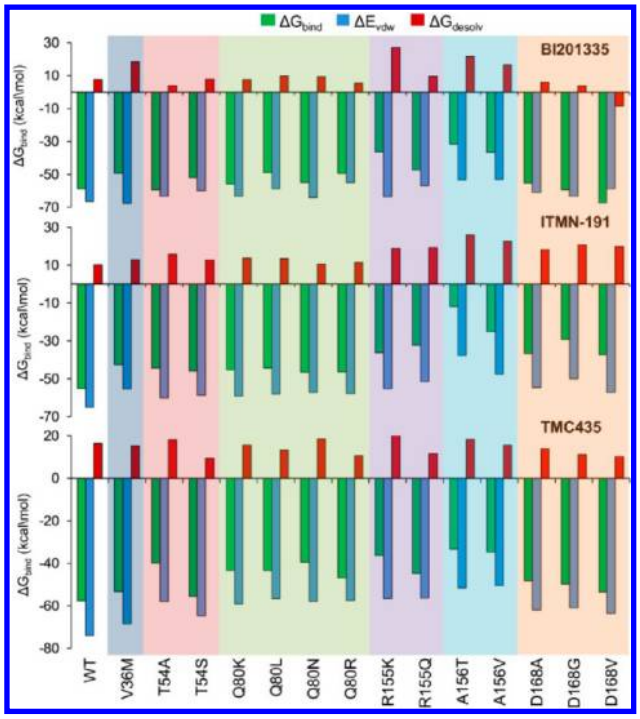


Figure 5. Profiles of various binding energy components for each inhibitor across mutations. Energetic profiles were produced by decomposing ΔG_{bind} from eq 10 to its constituents. ΔE_{elec} and ΔG_{psolv} were combined together to give ΔG_{desolv} (energetic cost of desolvation). ΔG_{npsolv} was omitted as its contribution was constant for a given inhibitor against all mutations. ΔG_{bind} in this plot is, therefore, the sum of ΔE_{vdw} and ΔG_{desolv}. For the ease of comparison across inhibitors, mutations in the same residue are represented with the same shaded background.

was the deviation in the position of the P2 group (the RMSD of P2 heavy atoms from wild-type position was 2.85 Å, followed by BI201335, 0.96 Å and TMC435, 0.70 Å) (Figure 6). A

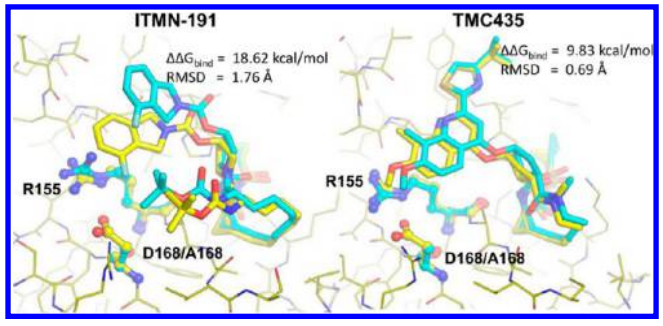


Figure 6. Predicted ligand conformations in wild-type and D168A NS3/4A binding sites. Binding interfaces for wild-type (yellow) and D168A mutant (cyan) are superimposed. In each case, deviation in inhibitor binding mode and impact on affinity is indicated by RMSD and $\Delta\Delta G_{bind}$ values. Unlike ITMN-191, TMC435 predicted conformation shows minimal impact on the position of R155 in D168A mutant.

similar trend was not repeated for BI201335 and TMC435. It can be seen from RMSD values that both of these inhibitors exhibited minimal deviation from the wild-type conformation (Table 5) and insignificant changes in the energetic profile (Figure 5). Unlike ITMN-191, both BI201335 and TMC435 have large P2 groups with relatively more polar character (Figure 2). The energetic profiles indicate stabilization of these electrostatic interactions between P2 and R155. In reality, these electrostatic interactions are most probably solvent-shielded and should have marginal contribution to BI201335 and TMC435 affinity. In MM-GBSA scoring, however, these interactions were rewarded such that the concomitant increase in ΔG_{desolv} was almost always compensated. This resulted in only small RMSD values and comparable van der Waals interactions, eventually leading to very similar energetic profiles of these two inhibitors in wild-type and D168 mutants. These results indicated that one reason for anomalous predictions at D168 was the overestimation of electrostatic interactions.

Although the energetic profile of ITMN-191 for D168 variants apparently conformed to our expectations (decrease in ΔE_{vdw} and increase ΔG_{desolv}), the residue-specific trend (i.e, the increasing order of resistance for D168V, D168A followed by D168G) was not observed. Moreover, it was interesting to note that the position of R155 deviated in each variant by more than about 1 Å from wild-type conformation (1.07 Å, 0.98 Å and 1.23 Å for D168A, D168G, and D168V, respectively). Similar deviation in R155 was not observed in predicted binding modes of BI201335 and TMC435 which is consistent with the observation that polar interactions with large P2 groups stabilized R155 in its induced-fit conformation (Figure 6). These results strongly suggest that changes in side-chain flexibility also play a significant role in D168 mutations, and their correct treatment is essential for high quality predictions. *A156V/T*: A156 mutations have been associated with the P2 proline motif which was common among all three inhibitors used in this study (Figure 2). The results consistently showed large deviations in inhibitor binding modes and differences in energetic profiles (Table 5, Figure 5). Both mutations involved decrease in ΔE_{vdw}, and for A15T, in particular, decrease in ΔG_{desolv} was also observed. These trends were relatively more pronounced in the case of ITMN-191. The additional decrease in van der Waals interactions came from the displacement of P4 *tert*-butyl moiety (Figure 7). Although BI201335 also contains the same P4 functional group (Figure 2), its energetic profile did not show a similar decrease in ΔE_{vdw} (Figure 5). This is because BI201335 has a linear structure, as opposed to a macrocyclic structure of ITMN-191, and can adapt more easily to the changed binding site and minimize the impact of suboptimal van der Waals interactions. This was also partly supported by smaller deviations in the BI201335 binding mode than that of ITMN191 (Table 5). TMC435 and ITMN-191 are both macrocyclic inhibitors but did not show a similar decrease in ΔE_{vdw}. This can be explained by the observation that unlike ITMN-191, TMC435 does not contain a *tert*-butyl group (Figure 2) and does not make extra van der Waals contacts in the S4 subsite. It is likely that the combination of a rigid P2−P4 linker and a large alkyl P4 side-chain makes ITMN-191 more susceptible to A156T/V mutations (Figure 7).

Overall, these results are consistent with experimental studies. The interactions of A156 with P2-proline scaffold and additional close contacts with nonpolar atoms in P2 and P3 groups make this residue crucial for binding as it facilitates optimal interactions of both inhibitor P2 and P3 moieties. It has been reported that resistance-causing mutations at this site should result in a steric clash with P2-proline followed by weak van der Waals interactions of inhibitor at S2 and S3 subsites.[45,46] The predicted energetic profiles confirmed this mode of resistance. One notable difference was that the experimental trend shows that generally A156V causes more

dx.doi.org/10.1021/ci3004754 | J. Chem. Inf. Model. 2013, 53, 435−451

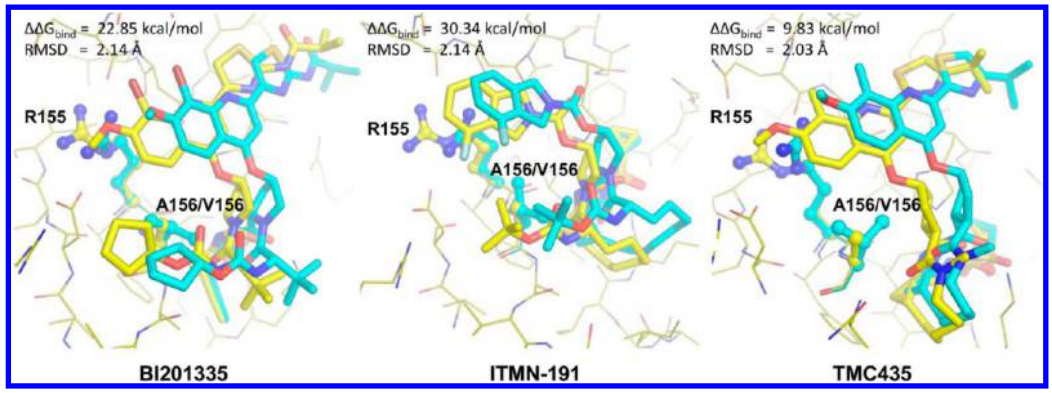


Figure 7. Predicted ligand conformations in NS3/4A wild-type and A156V mutant. Binding interfaces for wild-type (yellow) and D168A mutant (cyan) are superimposed. Key residues that participate in resistance mechanism are shown in ball-and-stick. In each case, deviation in inhibitor binding mode and impact on affinity is indicated by RMSD and $\Delta\Delta G_{bind}$ values. The largest deviation and decrease in binding affinity can be observed for ITMN-191.

resistance than A156T (Table 2). It is likely that the extent of repulsion between T156 and P2-proline is mitigated by a potential hydrogen bond between threonine hydroxyl and proline N-atom. However, the solvent shielding of this interaction would make its contribution negligible as opposed to the huge negative contribution by the displacement of P2 and P3 moieties. As mentioned above, a relatively higher desolvation cost was noticed for A156T which along with a decrease in ΔE_{vdw} made it more resistant than A156V (Figure 5). It is difficult to ascertain whether the contradiction arises from variability in the experimental data and/or incorrect estimation of solvent-shielded interactions.

R155K/Q. The induced-fit conformation adopted by R155 and its stabilization by an array of hydrogen bonds involving D168, and potentially Q80, is crucial for binding of inhibitors with large P2 groups. Consequently, mutations at this position are known to cause high-level resistance.[41,45] The energetic profiles for all three inhibitors BI201335 and TMC435 indicate a decrease in ΔE_{vdw} for both R155K and R155Q. For BI201335 and TMC435, ΔG_{desolv} for R155K was significantly higher than wild-type and R155Q variant (Figure 5). This originated from a greater desolvation cost of burying a lysine side-chain. The increased preference of K155 for solvation predicted in this way is consistent with the experimental structure of the R155K mutant (PDB: 2OIN) in which the lysine side-chain extends much further out of the S2 subsite than the arginine side-chain in wild-type (PDB: 1JXP). The calculations also predicted R155Q to be less resistant than R155K which was consistent with *EC-FC*$_{50}$ values. Unlike R155K, no increase in desolvation cost was observed, in the case of R155Q, as compared to wild-type, hence the main factor contributing to the decrease in binding energy was the decrease in van der Waals interaction energy. The predicted binding modes of inhibitors indicated that although Q155 was able to adopt an R155-like conformation, it was unable contribute as much nonpolar surface area to the P2 group due to its smaller size, which was reflected in ΔE_{vdw} (Figure 5).

Analysis of Intermediate/Low-level Resistance Mutations. The binding energy prediction for intermediate/low resistance mutations correlated fairly well with experimental data. For instance, a relatively larger decrease in ΔG_{bind} against Q80 mutants was observed for TMC435 as compared to other inhibitors, which is consistent with the experimental observations. However, the energetic profiles did not coincide well with expected residue-specific effects. For example, contrary to the expectations, ΔG_{desolv} for Q80K/R was lower (or comparable) than the corresponding value for wild-type NS3/4A. This could result from the poor estimation of desolvation penalty. Furthermore, ΔE_{vdw} values in Q80K/R mutants were comparable to wild-type, which probably means that substitutions at Q80 positions do not hugely impact the optimal interaction of the P2 group with S2 subsite residues, particularly R155. Nevertheless, in the wild-type structure, it is clear that induced-fit conformation of the S2 subsite in the case of TMC435 involves a hydrogen bond between Q80 and R155.[45] This formation of this hydrogen bond should result in additional stabilization for R155 which undergoes conformational changes and desolvation upon binding. The contribution of Q80 might not be as much as D168, as explained above, which explains the lower levels of resistance observed for these mutations.[32]

It appears that the requirement of Q80 in stabilization of extended S2 conformation depends on the type of the P2 group. For example, BI201335 and TMC435 share a very similar P2 group, and for both Q80-R155 interactions have been reported.[42,45] In unbound NS3/4A, Q80 adopts a more solvated conformation, whereas in the inhibitor-bound enzyme it moves toward R155 to form a hydrogen bond. From the experimental data, it appears that this role of Q80 is important for TMC435 binding despite the fact that there is additional conformational entropy loss and desolvation associated with such an interaction. In fact, Q80K and Q80R causes comparable but intermediate-level resistance at this site which is probably due to the higher desolvation penalty for a charged arginine or lysine side-chain that might preclude it from participating in the network of coordinated H-bonds involving Q80, R155, D168, and R123.

For V36 M and T54A/S mutations, the predicted binding modes of ITMN-191 and TMC435 indicated deviation in the P1′ cyclopropyl moiety, which in the case of TMC435 was quite significant against T54A. This manifested in a decrease in ΔE_{vdw} and an increase in ΔG_{desolv} (Figure 5). BI201335 lacks such a P1′ functional group; therefore, binding energy predictions showed no significant change in binding affinity against T54A/S mutants. Additionally, V36 M in the case of BI201335 showed a decrease in binding affinity, resulting mainly from an increase in ΔG_{desolv}. This could be the result of burying a polar side-chain (methionine) in protein interior as opposed to a nonpolar side-chain (valine). The low-level resistance mutations, T54A/S and V36M, are located further

dx.doi.org/10.1021/ci3004754 | J. Chem. Inf. Model. 2013, 53, 435−451

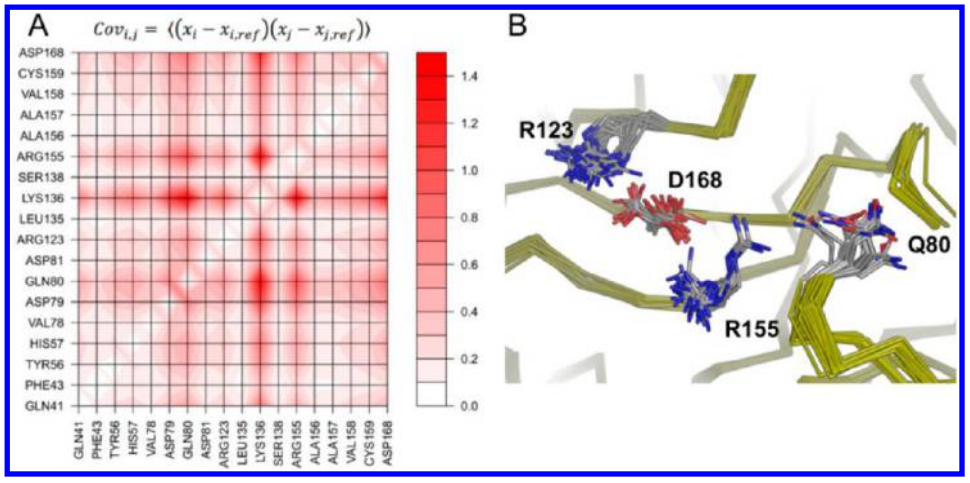


Figure 8. Pair-wise covariance analysis of side-chain motions in the NS3/4A binding site from X-ray crystallographic structures. A. Covariance matrix is shown as two-dimensional contour plot for side-chains located within 5 Å of the ligand. B. A superimposition of available NS3/4A ligand bound structures. Residue pairs with some of the most pronounced correlated motions are shown in stick representation. These side-chains can come within hydrogen bonding distance of each other depending on the nature of bound ligand and participate in induced-fit effects.

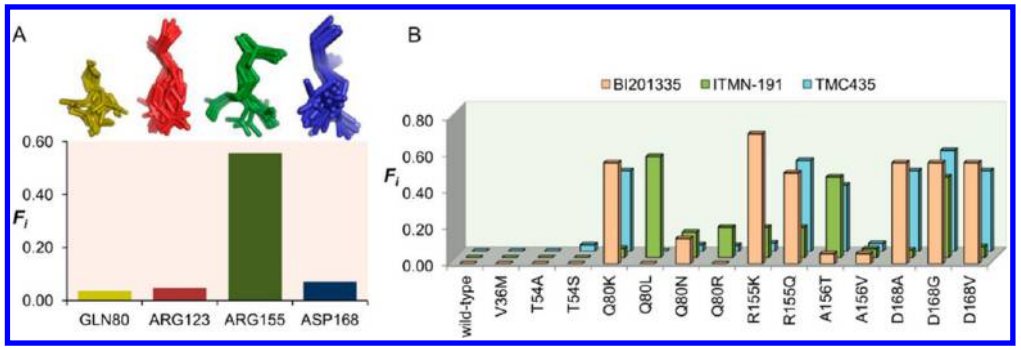


Figure 9. Side-chain flexibility measurements based on FIRST analysis. A. Flexibility indices (F_i) plotted for Q80, R123, R155, and D168. 3-D spatial distribution of each side-chain in superimposed X-ray structures is shown on top for qualitative assessment of side-chain mobility estimated by F_i. B. F_i values for R155 in wild-type and mutant NS3/4A in the presence of three different ligands used in this study.

away from the binding site and are not in direct contact with the inhibitor. However these residues have been shown to interact with a set of residues lining a hydrophobic cavity in the binding site (Q41, F43, G58, and A59).[70] This cavity provides interaction sites for P1′ cyclopropyl groups in ligands such as TMC435 and ITMN-191.

Conformational Flexibility at the S2 Subsite. In order to assess changes in flexibility of one or more residues in the S2 subsite, we used a stepwise approach that consisted of pairwise covariance analysis of side-chain motions, graph-theory based flexibility analysis, and conformational entropy estimation using solvent accessible surface area model. The pairwise covariance matrix is shown as a contour plot in Figure 8A. The most prominent correlated motions were observed for K136 and residues in and around the S2 subsite (R155, D168, Q80, and R123). However, K136 is not in direct contact with these residues. The correlation in their motion results from two different parts ligands resulting in simultaneous side-chain rearrangements. The functional significance of K136 movement is not known; however, from visual inspection of the structures it appears that ketoamide or sulfonamide groups in inhibitors could engage in solvent-shielded polar interactions with K136, hence causing it to shift upon ligand binding. Since mutations in the S2 subsite are not likely to affect the position of K136, therefore it can be safely ignored. This leaves correlated motions between R155, D168, Q80, and R123. In some structures it was observed that these side-chains come close to form an array of hydrogen bonds potentially stabilizing the induced-fit conformation (Figure 8B). However, the variability in the positions of these residues also indicates that the extent of side-chain rearrangements is related to the nature of the bound ligand. Similarly, mutation in one of these residues is also expected to cause changes in the flexibility which would in turn affect conformational entropy of binding. We further studied these ligand-induced flexibility changes using a FIRST-based analysis.[63]

The results from flexibility analysis for these four residues are shown in Figure 9A. The residue-based flexibility indices, F_i, were calculated from the wild-type apo-structure (1JXP). The results indicated that R155 was significantly more flexible than the other residues in the extended S2 subsite (Figure 9A). In order to check if F_i values are indeed useful measures of side-chain mobility, we compared these values with NMR data in the literature. Barbato et al. analyzed NMR relaxation data for NS3/4A and obtained order parameter, S_2, values for side-chains in the binding site.[71] S_2 values vary from 0 to 1, with lower values indicating higher mobility. Their analysis showed a much lower S_2 for R155 (0.56) than for other binding site residues which were in the range 0.79 to 0.97, hence showing higher mobility for R155.[71] The predicted flexibility values (Figure 9A), therefore, appear to be consistent with

dx.doi.org/10.1021/ci3004754 | J. Chem. Inf. Model. 2013, 53, 435−451

experimental observations and could be used for further analysis.

The results of flexibility calculations on the ligand-bound state of the receptor conformations are shown in Figure 9B. The flexibility profiles of all three ligands show that in wild-type and low resistance mutations, R155 loses its flexibility upon ligand binding, which is the hallmark of induced-fit binding mode (with optimal P2−S2 contacts).[42,45] For all D168 mutants, R155 regained flexibility particularly in the case of BI201335 and TMC435. For ITMN-191, this was not the case except for D168G (Figure 9B). Interestingly, the flexibility of R155 also increased for Q80 mutants, for instance in Q80K, R155 had a flexibility change almost of the same magnitude as observed for D168 mutants (for BI201335 and TMC435). Q80K exhibits relatively higher resistance among Q80 variants, and this effect is relatively more profound with ligands that engage in direct interactions with Q80 (BI201335 and TMC435) (Table 2 and 5). High R155 flexibility in Q80L for ITMN-191 was somewhat anomalous. Finally, in R155 mutants, a varying degree of increase in flexibility of the substituted residue was observed for all three ligands. The F_i values do not lend themselves to a direct thermodynamic interpretation, and hence their comparison with K_d or EC_{50} values is merely qualitative. Nevertheless, the comparison provides results that are fairly well-connected with the experimental observations and explain discrepancy in continuum-solvation based binding energy predictions (Figure 4).

In order to assess the energetic consequences of flexibility changes, we calculated side-chain conformational entropy using a structural-energetic framework. The conformational entropy (ΔS_{conf}) of a side-chain in this context is related to changes in accessible surface area upon binding. ΔS_{conf} reflects the relative change in conformational entropy of the binding upon mutation. In Figure 10, ΔS_{conf} (derived from only binding site

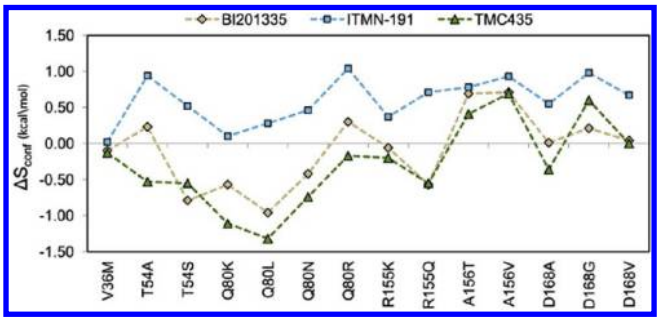


Figure 10. Side-chain entropy estimation from a solvent accessible surface area model. ΔS_{conf} represents a relative conformational entropy estimate (with reference to wild-type bound conformation of side-chains in the binding site).

residues) is plotted for all three inhibitors against all mutations. All three ligands showed somewhat similar profiles, but for ITMN191 the extent of change was higher than its counterparts. It should be noted that BI201335 and TMC435 have an almost equally sized P2 group which is most probably the reason behind their similar profiles. Both ligands showed an increase in ΔS_{conf} only for A156T/V and D168G mutations. For the low-level resistance mutations ΔS_{conf} was either comparable or even lower (more favorable) than the corresponding value for wild-type. ITMN191 showed an increase in ΔS_{conf} to varying extents for all mutations. An increasing trend is observed for high-level resistance mutations, but a similar increase was also found for low-level resistance mutations (Q80R and T54A). These results therefore do not indicate a very clear trend. The sensitivity of this approach toward surface area calculation makes it less accurate. Any perturbation in a small-sized P2 group would affect accessibility of the S2 subsite more profoundly than a large P2 group. Additionally, as D168 mutants did not change R155 conformation significantly in docked complexes (Figure 6); therefore, significant changes in ΔS_{conf} were not observed. Overall, the results indicate that flexibility calculations point toward change in flexibility of R155 in the S2 subsite, which should result in an increase in entropy of binding. A surface−area based estimation of conformational entropy seemed insufficient to model the energetic consequences of these flexibility changes. Nevertheless, these results strongly indicate that D168 and Q80 mutants lead toward an increase in flexibility of R155 which should raise the entropic cost of binding. As a consequence of the increase in flexibility, interactions made by the P2 group with the extended S2 subsite become suboptimal and cannot compensate for unfavorable entropic and desolvation changes. It is also likely that the size of the P2 group in ligands plays an important role in determining the extent to which these changes take place.

DISCUSSION

Strengths and Pitfalls of MM-GBSA Binding Energy Predictions for Resistance Mutations. The rationale behind the speed-accuracy trade-off in MM-GBSA approaches is that their application remains relevant to the time-scale of typical medicinal chemistry optimization campaigns while retaining as much accuracy as possible. In the following, we discuss strengths and pitfalls of this approach.

One particular aspect of the trade-off is the exact cancellation of protein and ligand internal energies which represent the energetic cost of protein reorganization and ligand strain on binding, respectively. These contributions are usually ignored in the calculation of ligand binding affinities in congeneric series due to similar effects induced by ligands with very similar scaffolds.[16] In order to incorporate such effects, we applied a variation of the protocol, where three separate implicit solvent minimizations of each molecule (protein, ligand, and complex) were used to calculate ΔG_{bind}. Surprisingly, the inclusion of internal energy terms led to deterioration of the performance. The resulting R^2 values did not show any significant correlation and contained more outliers (Table S1). Such a decrease in performance of MM-GBSA protocols upon inclusion of internal energy terms has been reported by others as well.[18,22] Mostly, the reason behind such behavior is the fact that calculating changes in protein internal energy upon ligand binding involves calculating small differences in large energetic contributions, which tend to accumulate errors resulting from small fluctuations. These small fluctuations arise from approximations of the theoretical framework of calculation and/or problems inherent in forcefields.

It can also be argued that a single-snapshot energy minimization protocol is limited in terms of sampling, and hence it is less likely to get better estimates of these energy terms. Alternatively, ensemble averages of energetic terms obtained from molecular dynamics trajectories should be more suitable. Two recent studies, investigating ITMN-191 and TMC435 binding to wild-type and a small set of NS3/4A mutants (R155K, A156V, and D168A), used MD-based sampling prior to binding energy calculations.[43,72] Unfortunately, the common data points between ours and these two

dx.doi.org/10.1021/ci3004754 | J. Chem. Inf. Model. 2013, 53, 435−451

studies are too few for any useful comparison. Nevertheless, if all six $\Delta\Delta G_{bind}$ values from these two studies are combined and compared with experimental data, the correlation is as insignificant as observed in our studies for these high-level resistance mutations. Interestingly, if D168A is removed from the comparison, MD-based protocols show much improved correlation (R^2, 0.92) which is slightly higher than the maximum correlation obtained in our calculations (R^2, 0.83). The observation that both protocols (MD-based vs single-snapshot minimization) produce outliers and can potentially show similar performance prompts a broader and more detailed comparison between the two variations of MM-GBSA. Some authors have pointed out that MD-based ensembles might suffer from sampling of irrelevant conformations which could lead to noise.[10,30,61]

Capturing the Role of Residue Flexibility in NS3/4A Resistance Mutations. The results from flexibility analysis point out that in some cases changes in conformational entropy possibly play a significant role. The most unexpected result in our calculations was related to binding energy predictions against D168 mutants. The calculated ΔG_{bind} (mostly close to wild-type ΔG_{bind}) values were in stark contrast with the high levels of resistance caused by these mutations. It has been previously proposed that the primary effect of D168 mutations is the ablation of the D168-R155 salt bridge that is required to stabilize R155 in induced-fit conformation.[43,45,46,72] The results from our calculations suggest that despite the absence of salt-bridge, R155 remains in an induced-fit conformation and exhibits only slight variations from its wild-type conformation. It can be argued that this resulted from incomplete sampling of R155 side-chain offered by the minimization based protocol. However, an almost similar weak correlation for D168A by MD-ensemble based studies of ITMN-191 and TMC435, as mentioned above, prompted us to do a deeper analysis of the role of R155 in D168 mutants. The formation of the D168 and R155 salt bridge, that stabilizes induced-fit conformation required for binding, incurs a desolvation and entropic cost which is more than compensated by favorable hydrophobic interactions between the P2 group and R155 (and possibly by favorable displacement of unstable water molecules in the S2 subsite). It is, therefore, reasonable to ask if R155 can exist in induced-fit conformation in the absence of this salt bridge, as predicted in our results. The set of apo-NS3/4A structures available in the PDB suggests that in an unbound state R155 does exist in a conformation similar to that observed in a bound state. Additionally, SAR studies of various inhibitors carrying modifications of P1 and P1′ groups showed that some ligands could bind to D168A mutants without any loss of affinity.[46] All of these modified ligands carried large P2 groups (similar in size and nature to the P2 groups of inhibitors used in this study), and yet, even in the absence of R155-D168 salt-bridge, they were able to induce extended conformation of R155. Consequently, the stability of the R155 induced-fit conformation does not entirely depend on the D168-R155 salt-bridge and can be driven by the hydrophobic interactions between the ligand P2 group and the S2 subsite. We, therefore, propose that the ablation of the salt-bridge is perhaps not directly responsible for the resistance. Our results from flexibility analysis indicate that the flexibility of R155 and hence the entropic cost of restricting it in induced-fit conformation increases in D168 mutants. These results suggest that the resistance caused by D168 mutants is primarily mediated by a rise in the entropic barrier imposed by a relatively more flexible R155 side-chain. Interestingly, an NMR study of NS3/4A inhibitor potency shift across genotype 1b and 3A supports this argument.[73] Genotype 3b NS3/4A carries two important substitutions with respect to 1b, D168Q, and R123T and is less susceptible to most protease inhibitors. The mobility of side-chains was studied using Hβ-Cβ cross-peak intensities and indicated increased flexibility for R123, Q168, and R155 in genotype 3b. The study proposed that the decreased affinity in the case of genotype 3b enzyme could be attributed to entropic effects resulting from D168 and R123 substitutions.[73]

In a nutshell, the flexibility of R155 and other side-chains responsible for salt-bridge network formation is an important piece of information in the understanding of molecular mechanism of resistance. The fact, that our MM-GBSA based calculations combined with a qualitative assessment of side-chain flexibility were able to highlight this facet of inhibitor binding, shows that these approaches can provide complementary pieces of information. However, correct predictions for changes in binding affinity due to mutations that affect flexibility of R155 would rely on accurate estimation of conformation entropy of binding. The inclusion of an entropy calculation framework in the MM-GBSA protocol, that is consistent with its time-scale, should make this technique more promising. Indeed, surface area based entropy calculations models are under development that can be used with MM-GBSA energy calculations.[74]

Implicit Solvent and Estimation of Nonpolar Salvation Energy. The decomposition of binding energy to its constituent terms revealed that hydrophobic interactions between the inhibitor P2 and the S2 subsite are a crucial aspect of inhibitor binding. Our predictions and available experimental data consistently show that, in most high-level resistance mutations, these interactions are perturbed significantly. In MM-GBSA framework, the estimation of van der Waals interactions largely captured these changes very well. However, another closely related and important term is the nonpolar solvation energy, ΔG_{npsolv}. The contribution of ΔG_{npsolv} to ΔG_{bind} values calculated in this study was more or less constant. It is recognized that the calculation of nonpolar solvation energy in continuum solvation approaches is not theoretically as well-grounded as that of polar solvation energy component.[14] Indeed, it was recently reported that continuum solvation based methods failed to give accurate estimates for nonpolar solvation free energy on at least four protein−ligand complexes with varying degrees of solvent exposure of the binding site.[75] Consequently, it was argued that explicit consideration of water molecules is very important in correct estimation of nonpolar solvation energies. It is pertinent to ask whether explicit solvent treatment could lead to better results, in this particular system, particularly improved estimation of ΔG_{npsolv}. Several approaches have recently emerged that explicitly evaluate thermodynamics of water molecules in the binding site.[76] One such approach is based on inhomogeneous fluid solvation theory and is implemented as WaterMap.[77] WaterMap scoring function gives free energy of displacement of water molecules from the binding site and has been validated in a variety of targets.[78] MM-GBSA protocols combined with WaterMap have also been reported and seemed to increase the accuracy of binding energy estimates.[30] In one such study, nonpolar solvation energy terms were replaced with WaterMap score, and the binding energies were calculated for a set of blood coagulation factor serine proteases.[61] We were interested in assessing as to what extent explicit consideration of water

dx.doi.org/10.1021/ci3004754 | J. Chem. Inf. Model. 2013, 53, 435−451

molecules could improve nonpolar solvation energy calculation. To this end, a brief analysis of conserved water position between thrombin and NS3/4A protease can provide useful information. Thrombin represents a good reference for comparison because of its structural similarity with NS3/4A, and the thermodynamics of water displacement from its binding site has been studied experimentally[79] and computationally.[61]

Thrombin and NS3/A share very little sequence similarity, but structurally they belong to the same chymotrypsin-like fold and certain regions of the protein show remarkable structural similarity. Such regions can be quickly identified by a sequence-based structure alignment (Figure 11A) using DALI.[80] Upon

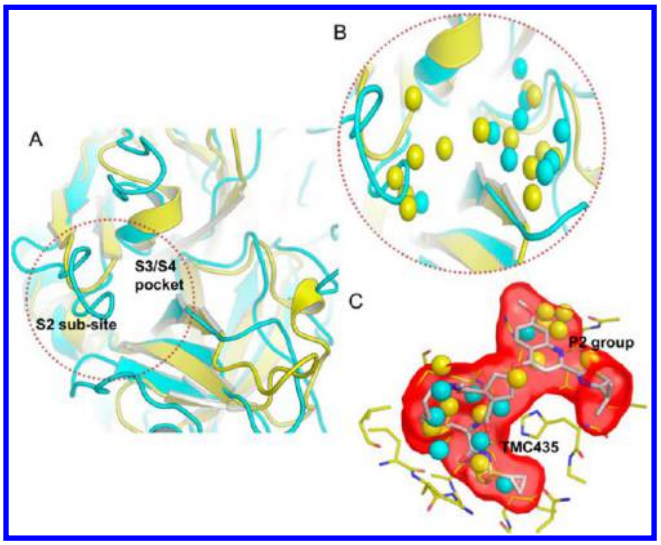


Figure 11. Comparison of binding site hydration between NS3/4A and thrombin. A. A superimposition of NS3/4A (yellow) (PDB: 1JXP) and thrombin (cyan) (PDB: 2UUF) shows local regions of high similarity (S3/S4) and variation (S2 subsite). B. Structural variation in the S2 subsite is reflected in the differences in the hydration pattern between NS3/4A, bound water molecules taken from X-ray structures are shown as spheres (NS3/4A: yellow, thrombin: cyan). C. Ligands carrying large P2 functional groups (e.g., TMC435, shown in gray) displace water molecules unique to the NS3/4A S2 subsite.

visual inspection, one can identify several ordered water molecules (based on four available NS3/4A apo structures). In the NS3/4A binding site, we identified at least eight water positions for which a corresponding water molecule (located at ≤2.0 Å) was available in the superimposed thrombin binding site. All of these water molecules were located in S1, S3/S4 subsites (Figure 11B). The displacement of these water molecules is considered highly favorable leading to entropy gain and increased van der Waals contact between the protein and the inhibitor.[61,79] Interestingly, the extended S2 subsite of NS3/4A differs largely from corresponding region in thrombin and is relatively more solvated. The ligand P2 moieties bind to the extended S2 subsite with highly favorable van der Waals interactions (Figure 11C). Based on the solvation pattern of S2, it is very likely that part of the affinity gain originates from favorable entropy of displacing these water molecules. Based on this analysis, it appears that displacement of most of these water molecules should result in an increased van der Waals contact between the ligand and the protein. In a combined implementation of MM-GBSA with WaterMap, it was shown that when van der Waals interactions play a dominant role in binding, the WaterMap score and van der Waals interaction energies were highly correlated.[30] Hence, in such cases, the explicit consideration of water molecules does not necessarily lead to improvement in accuracy of binding affinity estimate. This was evident from the modest improvement in accuracy with the WM/MM approach as compared to MM-GBSA.[30,61] We, therefore, argue that for those NS3/4A active site mutations that primarily perturb the van der Waals interactions (e.g., P2−S2 interactions), the continuum solvation framework can avoid the limitations related to the calculation of nonpolar solvation energies. This argument, however, is applicable only to this scenario where the same ligand binds to the protein with active-site mutations. Where ligands displace different water molecules (due to different substituents) an explicit water treatment of nonpolar solvation free energies might be required for higher accuracy. These findings indicate that a detailed thermodynamic study of hydration sites in wild-type and mutant NS3/4A would elaborate their role in affinity gain and possibly in drug resistance. It is quite likely that a change in the solvation pattern of the extended S2 subsite upon mutation could play a role in resistance. We noticed that the R155K mutation, which is a conservative substitution, accompanies an increase in desolvation cost upon ligand binding (Figure 5, Table 5). The experimentally determined solvation free energy of a fully exposed lysine side-chain was found to be higher than that of an arginine side-chain in a set of host−guest pentatpepides.[81] Although lysine could make polar interactions with D168 and adopt R155-like induced conformation, MD-based studies have, however, reported the instability of the K155-D168A salt-bridge in R155K mutants.[43,72] The origin of this instability could be the relatively increased solvation preference of lysine which was captured correctly by MM-GBSA predictions (Figure 5).

Additional insights regarding the mechanism of resistance can also be gained from these calculations. For instance, the combination of a P4 *tert*-butyl moiety and P1−P4 cyclization in ITMN-191 was predicted to be more susceptible to A156T/V mutants. The calculations suggest that the decrease in van der Waals contact for ITMN-191 was more profound because of suboptimal interactions of two hydrophobic moieties in the ligand, as opposed to one in the case of BI201335 and TMC435. The results from modeling also pointed out the significance of the inhibitor P2 group in determining response to some R155 and D168 mutations. A smaller sized P2 group that is optimized to provide maximum affinity gain from van der Waals interactions but robust to substitution in the S2 subsite would be a major step forward for development of next generation NS3/4A inhibitors. In a recent study, a combination of a smaller sized P2 group along with P2−P4 cyclization led to a ligand with unabated affinity against these mutants.[82,83]

■ CONCLUSION

The precise information regarding the structural and energetic determinants of resistance can prove valuable when dealing with mutating drug targets. We demonstrated the use of fast and approximate methods of structural modeling in generating this information. Our findings suggest that continuum solvation based approaches provide good quality predictions for changes in binding affinity of ligands against most of the resistance mutations in NS3/4A. There were however notable discrepancies as well, particularly when mutations involved changes in flexibility of active-site residues. We attempted to address this by combining MM-GBSA with a FIRST flexibility analysis

approach. Although the combination does not lend itself to a quantitative application, however, it provides valuable information regarding the molecular mechanism of resistance. The information can be used, as such, in a qualitative way to assist ligand design. Indeed, with additional fitting of parameters based on a suitable and large reference data set, a quantitatively robust MM-GBSA/FIRST framework is conceivable. We also used structural information available for NS3/4A-ligand complexes, to test a solvent-accessible surface area-based entropy approximation. The isolated results again pointed out that with good sampling quality the approximation could be a useful way of accounting for entropic effects. It should, however, be noted that these results are system-specific and might not be generally applicable. Additionally, further improvements could come from correct parametrization of implicit solvent forcefields to reproduce the correct balance of electrostatic interactions, and solvation preferences at binding interface.[84] Finally, the significance of sampling protein and/or ligand conformations for MM-GBSA calculations needs a detailed evaluation. Overall, these findings support the notion that development of fast but approximate modeling approaches for binding affinity prediction should be a challenging but fruitful pursuit.

■ ASSOCIATED CONTENT

Supporting Information

Figures of structure features of HCV NS3/4A serine protease (Figure S1), comparison of NS3 sequences from experimental and crystallographic data (Figure S2), and a schematic of assignment of protonation states (Figure S3). Additional data regarding absolute binding affinity measurements from MM-GBSA calculations and decomposition of various binding energy components (Table S1) and relative binding energy estimates from a triple snapshot energy minimization protocol (Table S2). This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: (+92) 42 3560 8000 Ext: 8352. Fax: (+92) 42 35608303. E-mail: kamran.haider@lums.edu.pk.

Funding

Funding Sources: This work was funded in part by Faculty Initiative Grant, Office of Graduate Studies and Sponsored Research, Lahore University of Management Sciences.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Authors would like to thank Roderick E. Hubbard for helpful advice in conceiving this study and reviewing the manuscript.

■ REFERENCES

(1) Gohlke, H.; Klebe, G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angew. Chem., Int. Ed. Engl.* **2002**, *41*, 2645−2676.

(2) Guvench, O.; MacKerell, A. D., Jr. Computational evaluation of protein-small molecule binding. *Curr. Opin. Struct. Biol.* **2009**, *19*, 56−61.

(3) Mobley, D. L.; Dill, K. A. Binding of small-molecule ligands to proteins: "what you see" is not always "what you get". *Structure* **2009**, *17*, 489−498.

(4) Wereszczynski, J.; McCammon, J. A. Statistical mechanics and molecular dynamics in evaluating thermodynamic properties of biomolecular recognition. *Q. Rev. Biophys.* **2012**, *45*, 1−25.

(5) Fenu, L. A.; Lewis, R. A.; Good, A. C.; Bodkin, M.; Essex, J. W. Scoring functions: from free-energies of binding to enrichment in virtual screening. In *Structure-based Drug Discovery*; Jhoti, H., Leach, A., Eds.; Springer: 2007; pp 223−245.

(6) Ferrara, P.; Gohlke, H.; Price, D. J.; Klebe, G.; Brooks, C. L. Assessing scoring functions for protein-ligand interactions. *J. Med. Chem.* **2004**, *47*, 3032−3047.

(7) Graves, A. P.; Shivakumar, D. M.; Boyce, S. E.; Jacobson, M. P.; Case, D. A.; Shoichet, B. K. Rescoring docking hit lists for model cavity sites: predictions and experimental testing. *J. Mol. Biol.* **2008**, *377*, 914−934.

(8) Wang, R.; Lu, Y.; Wang, S. Comparative evaluation of 11 scoring functions for molecular docking. *J. Med. Chem.* **2003**, *46*, 2287−2303.

(9) Warren, G. L.; Andrews, C. W.; Capelli, A. M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. A critical assessment of docking programs and scoring functions. *J. Med. Chem.* **2006**, *49*, 5912−5931.

(10) Guimarães, C. R.; Cardozo, M. MM-GB/SA rescoring of docking poses in structure-based lead optimization. *J. Chem. Inf. Model.* **2008**, *48*, 958−970.

(11) Hou, T.; Wang, J.; Li, Y.; Wang, W. Assessing the performance of the molecular mechanics/Poisson Boltzmann surface area and molecular mechanics/generalized Born surface area methods. II. The accuracy of ranking poses generated from docking. *J. Comput. Chem.* **2011**, *32*, 866−877.

(12) Hou, T.; Wang, J.; Li, Y.; Wang, W. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J. Chem. Inf. Model.* **2011**, *51*, 69−82.

(13) Kuhn, B.; Kollman, P. A. Binding of a diverse set of ligands to avidin and streptavidin: an accurate quantitative prediction of their relative affinities by a combination of molecular mechanics and continuum solvent models. *J. Med. Chem.* **2000**, *43*, 3786−3791.

(14) Onufriev, A. Continuum Electrostatics Solvent Modeling with the Generalized Born Model. In *Modeling Solvent Environments Applications to Simulations of Biomolecules*; Feig, M., Ed.; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, 2010; Chapter 6, pp 127−165.

(15) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.

(16) Genheden, S.; Ryde, U. Comparison of end-point continuum-solvation methods for the calculation of protein-ligand binding free energies. *Proteins: Struct., Funct., Bioinf.* **2012**, *80*, 1326−1342.

(17) Bernacki, K.; Kalyanaraman, C.; Jacobson, M. P. Virtual ligand screening against Escherichia coli dihydrofolate reductase: improving docking enrichment using physics-based methods. *J. Biomol. Screening* **2005**, *10*, 675−681.

(18) Foloppe, N.; Hubbard, R. Towards predictive ligand design with free-energy based computational methods? *Curr. Med. Chem.* **2006**, *13*, 3583−3608.

(19) Guimarães, C. R. W. A direct comparison of the MM-GB/SA scoring procedure and free-energy perturbation calculations using carbonic anhydrase as a test case: strengths and pitfalls of each approach. *J. Chem. Theory Comput.* **2011**, *7*, 2296−2306.

(20) Huang, N.; Jacobson, M. P. Physics-based methods for studying protein-ligand interactions. *Curr. Opin. Drug Discovery Dev.* **2007**, *10*, 325−331.

(21) Huang, N.; Kalyanaraman, C.; Bernacki, K.; Jacobson, M. P. Molecular mechanics methods for predicting protein-ligand binding. *Phys. Chem. Chem. Phys.* **2006**, *8*, 5166−5177.

(22) Huang, N.; Kalyanaraman, C.; Irwin, J. J.; Jacobson, M. P. Physics-based scoring of protein-ligand complexes: enrichment of known inhibitors in large-scale virtual screening. *J. Chem. Inf. Model.* **2006**, *46*, 243−253.

(23) Lyne, P. D.; Lamb, M. L.; Saeh, J. C. Accurate prediction of the relative potencies of members of a series of kinase inhibitors using molecular docking and MM-GBSA scoring. *J. Med. Chem.* **2006**, *49*, 4805−4808.
(24) Genheden, S.; Nilsson, I.; Ryde, U. Binding affinities of factor Xa inhibitors estimated by thermodynamic integration and MM/GBSA. *J. Chem. Inf. Model.* **2011**, *51*, 947−958.
(25) Pearlman, D. A.; Charifson, P. S. Are free energy calculations useful in practice? A comparison with rapid scoring functions for the p38 MAP kinase protein system. *J. Med. Chem.* **2001**, *44*, 3417−3423.
(26) Barril, X.; Gelpí, J. L.; López, J. M.; Orozco, M.; Luque, F. J. How accurate can molecular dynamics/linear response and Poisson−Boltzmann/solvent accessible surface calculations be for predicting relative binding affinities? Acetylcholinesterase huprine inhibitors as a test case. *Theor. Chem. Acc.* **2001**, 2−9.
(27) Kawatkar, S.; Wang, H.; Czerminski, R.; Joseph-McCarthy, D. Virtual fragment screening: an exploration of various docking and scoring protocols for fragments using Glide. *J. Comput.-Aided Mol. Des.* **2009**, 527−539.
(28) Mikulskis, P.; Genheden, S.; Rydberg, P.; Sandberg, L.; Olsen, L.; Ryde, U. Binding affinities in the SAMPL3 trypsin and host-guest blind tests estimated with the MM/PBSA and LIE methods. *J. Comput.-Aided Mol. Des.* **2011**, *26*, 527−541.
(29) Genheden, S.; Ryde, U. How to obtain statistically converged MM/GBSA results. *J. Comput. Chem.* **2010**, *31*, 837−846.
(30) Guimarães, C. R.; Mathiowetz, A. M. Addressing limitations with the MM-GB/SA scoring procedure using the WaterMap method and free energy perturbation calculations. *J. Chem. Inf. Model.* **2010**, *50*, 547−559.
(31) Ravindranathan, K.; Tirado-Rives, J.; Jorgensen, W. L.; Guimarães, C. R. W. Improving MM-GB/SA Scoring through the Application of the Variable Dielectric Model. *J. Chem. Theory Comput.* **2011**, *7*, 3859−3865.
(32) Lenz, O.; Verbinnen, T.; Lin, T. I.; Vijgen, L.; Cummings, M. D.; Lindberg, J.; Berke, J. M.; Dehertogh, P.; Fransen, E.; Scholliers, A.; Vermeiren, K.; Ivens, T.; Raboisson, P.; Edlund, M.; Storm, S.; Vrang, L.; de Kock, H.; Fanning, G. C.; Simmen, K. A. In vitro resistance profile of the hepatitis C virus NS3/4A protease inhibitor TMC435. *Antimicrob. Agents Chemother.* **2010**, *54*, 1878−1887.
(33) Gordon, C. P.; Keller, P. A. Control of hepatitis C: a medicinal chemistry perspective. *J. Med. Chem.* **2005**, *48*, 1−20.
(34) Kwong, A. D.; Kauffman, R. S.; Hurter, P.; Mueller, P. Discovery and development of telaprevir: an NS3−4A protease inhibitor for treating genotype 1 chronic hepatitis C virus. *Nat. Biotechnol.* **2011**, *29*, 993−1003.
(35) Njoroge, F. G.; Chen, K. X.; Shih, N. Y.; Piwinski, J. J. Challenges in modern drug discovery: a case study of boceprevir, an HCV protease inhibitor for the treatment of hepatitis C virus infection. *Acc. Chem. Res.* **2008**, *41*, 50−59.
(36) U.S. Food and Drug Administration. http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm256299.htm (accessed May 12, 2012).
(37) Shepard, C. W.; Lyn, F.; Alter, M. J. Global epidemiology of hepatitis C virus infection. *Lancet Infect. Dis.* **2005**, *5*, 558−67.
(38) Qureshi, S. A. Hepatitis C virus−biology, host evasion strategies, and promising new therapies on the horizon. *Med. Res. Rev.* **2007**, *27*, 353−373.
(39) Lopez-Labrador, F. X.; Moya, A.; Gonzalez-Candelas, F. Mapping natural polymorphisms of hepatitis C virus NS3/4A protease and antiviral resistance to inhibitors in worldwide isolates. *Antiviral Ther.* **2008**, *13*, 481−494.
(40) Bennett, F.; Huang, Y.; Hendrata, S.; Lovey, R.; Bogen, S. L.; Pan, W.; Guo, Z.; Prongay, A.; Chen, K. X.; Arasappan, A.; Venkatraman, S.; Velazquez, F.; Nair, L.; Sannigrahi, M.; Tong, X.; Pichardo, J.; Cheng, K. C.; Girijavallabhan, V. M.; Saksena, A. K.; Njoroge, F. G. The introduction of P4 substituted 1-methylcyclohexyl groups into boceprevir: a change in direction in the search for a second generation HCV NS3 protease inhibitor. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 2617−2621.
(41) Romano, K. P.; Ali, A.; Royer, W. E.; Schiffer, C. A. Drug resistance against HCV NS3/4A inhibitors is defined by the balance of substrate recognition versus inhibitor binding. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 20986−20991.
(42) Lemke, C. T.; Goudreau, N.; Zhao, S.; Hucke, O.; Thibeault, D.; Llinas-Brunet, M.; White, P. W. Combined X-ray, NMR, and kinetic analyses reveal uncommon binding characteristics of the hepatitis C virus NS3-NS4A protease inhibitor BI 201335. *J. Biol. Chem.* **2011**, *286*, 11434−11443.
(43) Xue, W.; Pan, D.; Yang, Y.; Liu, H.; Yao, X. Molecular modeling study on the resistance mechanism of HCV NS3/4A serine protease mutants R155K, A156V and D168A to TMC435. *Antiviral Res.* **2012**, *93*, 126−137.
(44) 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.; Lin, C. Phenotypic and structural analyses of hepatitis C virus NS3 protease Arg155 variants: sensitivity to telaprevir (VX-950) and interferon alpha. *J. Biol. Chem.* **2007**, *282*, 22619−22628.
(45) Cummings, M. D.; Lindberg, J.; Lin, T. I.; de Kock, H.; Lenz, O.; Lilja, E.; Fellander, S.; Baraznenok, V.; Nystrom, S.; Nilsson, M.; Vrang, L.; Edlund, M.; Rosenquist, A.; Samuelsson, B.; Raboisson, P.; Simmen, K. Induced-fit binding of the macrocyclic noncovalent inhibitor TMC435 to its HCV NS3/NS4A protease target. *Angew. Chem., Int. Ed. Engl.* **2010**, *49*, 1652−1655.
(46) Ortqvist, P.; Vema, A.; Ehrenberg, A. E.; Dahl, G.; Ronn, R.; Akerblom, E.; Karlen, A.; Danielson, U. H.; Sandstrom, A. Structure-activity relationships of HCV NS3 protease inhibitors evaluated on the drug-resistant variants A156T and D168V. *Antiviral Ther.* **2010**, *15*, 841−852.
(47) Lagacé, L.; White, P. W.; Bousquet, C.; Dansereau, N.; Do, F.; Llinas-Brunet, M.; Marquis, M.; Massariol, M. J.; Maurice, R.; Spickler, C.; Thibeault, D.; Triki, I.; Zhao, S.; Kukolj, G. In vitro resistance profile of the hepatitis C virus NS3 protease inhibitor BI 201335. *Antimicrob. Agents Chemother.* **2012**, *56*, 569−572.
(48) Berman, H. M.; Bhat, T. N.; Bourne, P. E.; Feng, Z. K.; Gilliland, G.; Weissig, H.; Westbrook, J. The Protein Data Bank and the challenge of structural genomics. *Nat. Struct. Biol.* **2000**, *7*, 957−959.
(49) Lim, S. R.; Qin, X.; Susser, S.; Nicholas, J. B.; Lange, C.; Herrmann, E.; Hong, J.; Arfsten, A.; Hooi, L.; Bradford, W.; Najera, I.; Smith, P.; Zeuzem, S.; Kossen, K.; Sarrazin, C.; Seiwert, S. D. Virologic escape during danoprevir (ITMN-191/RG7227) monotherapy is hepatitis C virus subtype dependent and associated with R155K substitution. *Antimicrob. Agents Chemother.* **2012**, *56*, 271−279.
(50) Rone, R.; Momany, F. A.; Dygert, M. Conformational studies on vancomycin using QUANTA-CHARMM. In Smith, J. A., Rivier, J. E. 1992; pp 299−301.
(51) *Discovery Studio 3.1*; Accelrys: San Diego, CA, USA, 2010.
(52) Haider, M. K.; Bertrand, H. O.; Hubbard, R. E. Predicting fragment binding poses using a combined MCSS MM-GBSA approach. *J. Chem. Inf. Model.* **2010**, *51*, 1092−1105.
(53) Halgren, T. A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490−519.
(54) Sali, A.; Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **1993**, *234*, 779−815.
(55) Wu, G.; Robertson, D. H.; Brooks, C. L., III; Vieth, M. Detailed analysis of grid-based molecular docking: a case study of CDOCKER-A CHARMm-based MD docking algorithm. *J. Comput. Chem.* **2003**, *24*, 1549−1562.
(56) Im, W.; Lee, M. S.; Brooks, C. L., III Generalized born model with a simple smoothing function. *J. Comput. Chem.* **2003**, *24*, 1691−1702.
(57) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.;

York, D. M.; Karplus, M. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545−1614.
(58) Feig, M.; Onufriev, A.; Lee, M. S.; Im, W.; Case, D. A.; Brooks, C. L. Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. *J. Comput. Chem.* **2003**, *25*, 265−284.
(59) Gan, W.; Roux, B. Binding specificity of SH2 domains: insight from free energy simulations. *Proteins: Struct., Funct., Bioinf.* **2008**, *74*, 996−1007.
(60) Tamamis, P.; López de Victoria, A.; Gorham, R. D.; Bellows-Peterson, M. L.; Pierou, P.; Floudas, C. A.; Morikis, D.; Archontis, G. Molecular dynamics in drug design: new generations of compstatin analogs. *Chem. Biol. Drug Des.* **2012**, *79*, 703−718.
(61) Abel, R.; Salam, N. K.; Shelley, J.; Farid, R.; Friesner, R. A.; Sherman, W. Contribution of explicit solvent effects to the binding affinity of small-molecule inhibitors in blood coagulation factor serine proteases. *ChemMedChem* **2011**, *6*, 1049−1066.
(62) Fulle, S.; Gohlke, H. Flexibility analysis of biomacromolecules with application to computer-aided drug design. *Methods Mol. Biol.* **2012**, *819*, 75−91.
(63) Jacobs, D. J.; Rader, A. J.; Kuhn, L. A.; Thorpe, M. F. Protein flexibility predictions using graph theory. *Proteins* **2001**, *44*, 150−165.
(64) Ahmed, A.; Gohlke, H. Multiscale modeling of macromolecular conformational changes combining concepts from rigidity and elastic network theory. *Proteins: Struct., Funct., Bioinf.* **2006**, *63*, 1038−1051.
(65) Wells, S.; Menor, S.; Hespenheide, B.; Thorpe, M. F. Constrained geometric simulation of diffusive motion in proteins. *Phys. Biol.* **2005**, *2*, S127−S136.
(66) Murphy, K. P. Predicting binding energetics from structure: looking beyond dG. *Med. Res. Rev.* **1999**, *19*, 333−339.
(67) Lee, K. H.; Xie, D.; Freire, E.; Amzel, L. M. Estimation of changes in side chain configurational entropy in binding and folding: general methods and application to helix formation. *Proteins* **1994**, *20*, 68−84.
(68) Thompson, A. J.; McHutchison, J. G. Antiviral resistance and specifically targeted therapy for HCV (STAT-C). *J. Viral Hepat.* **2009**, *16*, 377−387.
(69) Genheden, S.; Ryde, U. Comparison of the efficiency of the LIE and MM/GBSA methods to calculate ligand-binding energies. *J. Chem. Theory Comput.* **2011**, *7*, 3768−3778.
(70) Welsch, C.; Domingues, F. S.; Susser, S.; Antes, I.; Hartmann, C.; Mayr, G.; Schlicker, A.; Sarrazin, C.; Albrecht, M.; Zeuzem, S.; Lengauer, T. Molecular basis of telaprevir resistance due to V36 and T54 mutations in the NS3−4A protease of the hepatitis C virus. *Genome Biol.* **2008**, *9*, R16.
(71) Barbato, G.; Cicero, D. O.; Cordier, F.; Narjes, F.; Gerlach, B.; Sambucini, S.; Grzesiek, S.; Matassa, V. G.; De Francesco, R.; Bazzo, R. Inhibitor binding induces active site stabilization of the HCV NS3 protein serine protease domain. *EMBO J.* **2000**, *19*, 1195−1206.
(72) Pan, D.; Xue, W.; Zhang, W.; Liu, H.; Yao, X. Understanding the drug resistance mechanism of hepatitis C virus NS3/4A to ITMN-191 due to R155K, A156V, D168A/E mutations: a computational study. *Biochim. Biophys. Acta, Gen. Subj.* **2012**, *1820*, 1526−1534.
(73) Gallo, M.; Bottomley, M. J.; Pennestri, M.; Eliseo, T.; Paci, M.; Koch, U.; Bazzo, R.; Summa, V.; Carfi, A.; Cicero, D. O. Structural characterization of the hepatitis C virus NS3 protease from genotype 3a: the basis of the genotype 1b vs. 3a inhibitor potency shift. *Virology* **2010**, *405*, 424−438.
(74) Wang, J.; Hou, T. Develop and test a solvent accessible surface area-based model in conformational entropy calculations. *J. Chem. Inf. Model.* **2012**, *52*, 1199−1212.
(75) Genheden, S.; Mikulskis, P.; Hu, L.; Kongsted, J.; Soderhjelm, P.; Ryde, U. Accurate predictions of nonpolar solvation free energies require explicit consideration of binding-site hydration. *J. Am. Chem. Soc.* **2011**, *133*, 13081−13092.
(76) Li, Z.; Lazaridis, T. Water at biomolecular binding interfaces. *Phys. Chem. Chem. Phys.* **2007**, *9*, 573−581.
(77) Abel, R.; Young, T.; Farid, R.; Berne, B. J.; Friesner, R. A. Role of the active-site solvent in the thermodynamics of factor Xa ligand binding. *J. Am. Chem. Soc.* **2008**, *130*, 2817−2831.
(78) Beuming, T.; Che, Y.; Abel, R.; Kim, B.; Shanmugasundaram, V.; Sherman, W. Thermodynamic analysis of water molecules at the surface of proteins and applications to binding site prediction and characterization. *Proteins* **2012**, *80*, 871−883.
(79) Biela, A.; Sielaff, F.; Terwesten, F.; Heine, A.; Steinmetzer, T.; Klebe, G. Ligand binding stepwise disrupts water network in thrombin: enthalpic and entropic changes reveal classical hydrophobic effect. *J. Med. Chem.* **2012**, *55*, 6094−6110.
(80) Holm, L.; Park, J. DaliLite workbench for protein structure comparison. *Bioinformatics* **2000**, *16*, 566−567.
(81) Wimley, W. C.; Creamer, T. P.; White, S. H. Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides. *Biochemistry* **1996**, *35*, 5109−5124.
(82) Harper, S.; McCauley, J. A.; Rudd, M. T.; Ferrara, M.; DiFilippo, M.; Crescenzi, B.; Koch, U.; Petrocchi, A.; Holloway, M. K.; Butcher, J. W.; Romano, J. J.; Bush, K. J.; Gilbert, K. F.; McIntyre, C. J.; Nguyen, K. T.; Nizi, E.; Carroll, S. S.; Ludmerer, S. W.; Burlein, C.; DiMuzio, J. M.; Graham, D. J.; McHale, C. M.; Stahlhut, M. W.; Olsen, D. B.; Monteagudo, E.; Cianetti, S.; Giuliano, C.; Pucci, V.; Trainor, N.; Fandozzi, C. M.; Rowley, M.; Coleman, P. J.; Vacca, J. P.; Summa, V.; Liverton, N. J. Discovery of MK-5172, a macrocyclic hepatitis C virus NS3/4a protease inhibitor. *ACS Med. Chem. Lett.* **2012**, *3*, 332−336.
(83) Summa, V.; Ludmerer, S. W.; McCauley, J. A.; Fandozzi, C.; Burlein, C.; Claudio, G.; Coleman, P. J.; Dimuzio, J. M.; Ferrara, M.; Di Filippo, M.; Gates, A. T.; Graham, D. J.; Harper, S.; Hazuda, D. J.; McHale, C.; Monteagudo, E.; Pucci, V.; Rowley, M.; Rudd, M. T.; Soriano, A.; Stahlhut, M. W.; Vacca, J. P.; Olsen, D. B.; Liverton, N. J.; Carroll, S. S. MK-5172, a selective inhibitor of hepatitis C virus NS3/4a protease with broad activity across genotypes and resistant variants. *Antimicrob. Agents Chemother.* **2012**, *56*, 4161−4167.
(84) Wimley, W. C.; Gawrisch, K.; Creamer, T. P.; White, S. H. Direct measurement of salt-bridge solvation energies using a peptide model system: implications for protein stability. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 2985−2990.
(85) Boyce, S. E.; Mobley, D. L.; Rocklin, G. J.; Graves, A. P.; Dill, K. A.; Shoichet, B. K. Predicting ligand binding affinity with alchemical free energy methods in a polar model binding site. *J. Mol. Biol.* **2009**, *394*, 747−763.