Near-Neighbor Interactions in the NS3-4A Protease of HCV Impact Replicative Fitness of Drug-Resistant Viral Variants

Nadezhda T. Doncheva [1,2], Francisco S. Domingues [3], David R. McGivern [4], Tetsuro Shimakami [5], Stefan Zeuzem [6], Thomas Lengauer [1], Christian M. Lange [6], Mario Albrecht [7] and Christoph Welsch [1,6]

1 - *Department of Computational Biology and Applied Algorithmics,* Max Planck Institute for Informatics, Saarbrücken, Germany
2 - *Graduate School of Computer Science,* Saarland University, Saarbrücken, Germany
3 - *Center for Biomedicine,* European Academy Bozen/Bolzano (EURAC), Bolzano, Italy
4 - *Division of Infectious Diseases,* Department of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
5 - *Department of Gastroenterology,* Kanazawa University Hospital, Kanazawa, Japan
6 - *Department of Internal Medicine 1,* Goethe University Hospital Frankfurt, Frankfurt a.M., Germany
7 - *Institute for Knowledge Discovery,* Graz University of Technology, Graz, Austria

Correspondence to Christoph Welsch: Department of Computational Biology and Applied Algorithmics, Max Planck Institute for Informatics, Saarbrücken, Germany. christoph.welsch@kgu.de
https://doi.org/10.1016/j.jmb.2019.04.034
Edited by John Johnson

Abstract

A variety of amino acid substitutions in the NS3-4A protease of the hepatitis C virus lead to protease inhibitor (PI) resistance. Many of these significantly impair the replication fitness of the resistant variants in a genotype- and subtype-dependent manner, a critical factor in determining the probability with which resistant variants will persist. However, the underlying molecular mechanisms are unknown. Here, we present a novel residue-interaction network approach to determine how near-neighbor interactions of PI resistance mutations in NS3-4A can impact protease functional sites dependent on their genomic background. We constructed subtype-specific consensus residue networks for subtypes 1a and 1b from protease structure ensembles combined with biological properties of protein residues and evolutionary amino acid conservation. By applying local and global network topology analysis and visual exploration, we characterize PI resistance-associated sites and outline differences in near-neighbor interactions. We find local residue-interaction patterns and features at protease functional sites that are subtype specific. The noncovalent bonding patterns indicate higher fitness costs conferred by PI resistance mutations in a subtype 1b genomic background and explain the prevalence of Q80K and R155K in subtype 1a. Based on local residue interactions, we predict a subtype-specific role for the protease residue NS3–Q80 in molecular mechanisms related to the assembly of infectious virus particles that is supported by experimental data on the capacity of Q80K variants to replicate and produce infectious virus in subtype 1a and 1b cell culture.

© 2019 Elsevier Ltd. All rights reserved.

Introduction

Graph-based approaches provide a convenient means for characterizing protein structure information [1]. A comprehensive method for comparative analysis of molecular details from multiple protein structure models, however, is a difficult task. Here, we present a novel consensus residue network approach from ensembles of experimental protein structures that we applied to characterize near neighbor interactions of drug resistance mutations and their impact on the replicative fitness of hepatitis C virus (HCV) variants. Resistance-associated amino acid (AA) substitutions in drug target sites of direct antivirals frequently cause deficits in the replicative fitness of HCV resistance-associated viral variants (RAVs) [2]. Interestingly, RAVs from patients that failed on inhibitors against the NS3-4A protease (PI) disappear significantly more

Table 1. Subtype-specific RAV persistence

Residue	Median persistence (months)			Reference
	RAV	Subtype 1a	Subtype 1b	
V36	A	2.8	3.6	Sullivan *et al.* [3]
	M	9.3	6.3	
T54	A	2.3	2.8	Sullivan *et al.* [3]
	S	11.5	1.8	
V55	A	ND[a]	57	Susser *et al.* [4]
Q80	K	36	24	Lenz *et al.* [5]
R155	K	9.8	–[b]	Sullivan *et al.* [3]
A156	S	6.7	8.2	Sullivan *et al.* [3]
	T	1.8	2.7	
D168	A	5	ND[a]	Manns *et al.* [6]
				Krishnan *et al.* [7]
	V	ND[a]	>9	Zeuzem *et al.* (LEAGUE-1) [8]
				Lenz *et al.* [5]
				Krishnan *et al.* [7]
	N	>6	ND[a]	Buti *et al.* (C-SALVAGE) [9]
I170	V	ND[a]	ND[a]	

Time to loss of NS3-4A protease RAVs reported in the literature as detected by population and/or clonal sequencing.
[a] ND, no data available.
[b] No RAVs observed.

rapidly in HCV subtype 1b- than 1a-infected patients (Table 1) [3]. Moreover, some key resistance-associated AA substitutions are specifically observed in subtype 1a but not in 1b (Table 1). Such observations likely depend on the subtype-specific genomic background of the resistance mutations and are potentially explained by a genotype- and subtype-specific suite of resistant variants.

The NS3-4A protease plays an essential role in the HCV replication cycle and is a prime target site for direct antiviral therapy. By proteolytically processing nonstructural proteins from the viral polyprotein, the protease is involved in ribonucleic acid (RNA) synthesis [10]. Moreover, it modulates virus assembly by interference with the C-terminal RNA helicase domain of NS3 [10]. An allosteric site in the domain interface regulates NS3 function *via* stabilizing conformational states [11]. In addition, the protease blocks innate immune signaling of infected cells by cleavage of adaptor molecules, in particular Toll/interleukin-1 receptor domain containing adapter inducing interferon-β TRIF [12] and mitochondrial antiviral signaling protein MAVS (also called Cardif, VISA, and IPS-1) [13,14] (Fig. 1). Binding of PIs to the protease active site and the related resistance-associated AA substitutions can interfere with these processes [15,16]. Clinical data suggest low fitness costs due to drug resistance mutations in a subtype 1a genomic background [17], yet there is little understanding of the underlying molecular mechanisms.

We apply consensus residue networks to characterize drug resistance mutations in their subtype-specific protein structure environment to deduce biological phenotypes of drug-resistant viral variants dependent on their genomic background in different HCV subtypes.

Results

Consensus residue network approach

We made use of residue-interaction networks (RINs) for a subtype-specific characterization of PI resistance-associated sites and their potential impact on protein fold and function in NS3-4A (Fig. 2). To compare available multiple structure models from the Protein Data Bank RCSB PDB, we constructed consensus residue networks from subtype 1a and 1b NS3-4A protease structure ensembles, enriched with information on molecular properties of AA residues, sequence conservation, and inter-subtype variability (see Materials and Methods). The networks provide information on the location and nature of resistance-associated sites and their neighboring residues in the NS3-4A protease structure to residues previously identified to pertain to different NS3 functional sites. Based on the network topology and the bonding patterns of resistance-associated sites and functional sites, we predict the subtype-specific impact of resistance-associated AA substitutions on the replicative fitness of the PI-resistant variants.

Global network topology and sequence analysis

To compare the overall network topology between subtype 1a and 1b, we computed global topology measures for the subtype-specific consensus networks (Table 2). Each network comprises a total of 181 nodes, referring to residues of the NS3-4A protease. The network diameter is 12 edges in subtype 1a *versus* 11 edges in subtype 1b. The average number of neighbors, connected nodes, and network density
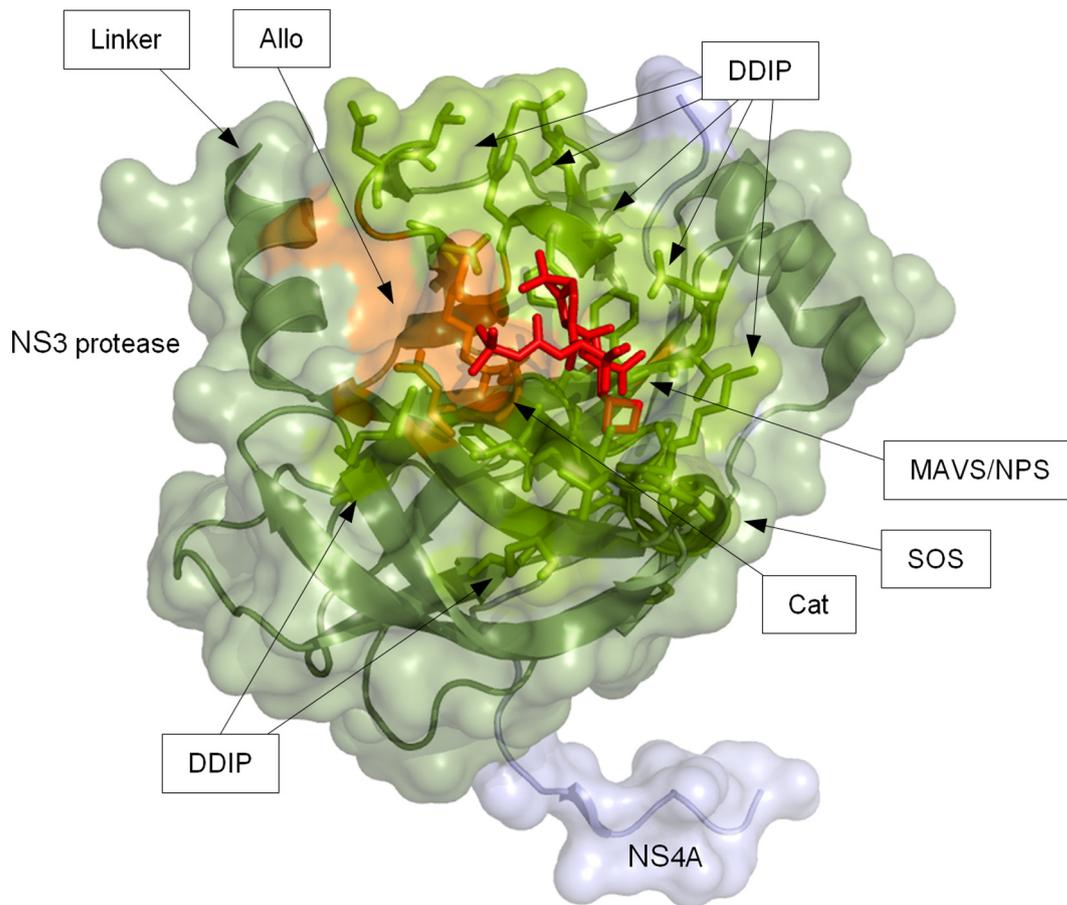
Fig. 1. 3D protein structure of the NS3-4A protease domain. The protein structure shows a subtype 1a NS3 protease domain from PDB entry 2OC8 co-crystallized with NS4A (light blue) and a linear PI (purple stick model). The protein backbone is given as ribbon model with transparent surface depiction. Residues where PI resistance-associated AA substitutions occur are highlighted as orange stick models; functionally important residues are shown as green stick models. Functionally important sites are indicated by arrows: allosteric site (Allo); protease active/catalytic site (Cat); domain–domain interaction sites (DDIP); linker region (Linker); MAVS binding site (MAVS); binding site for the natural protease substrate (NPS); SOS-binding site for TRIF (SOS).

are similar between both subtypes (Table 2). Minor differences are observed in overall residue interactions between subtypes. Of the interaction edges, 86% in the subtype 1a consensus network can be mapped to 94% of the edges in the subtype 1b network. Hence, the residue interaction difference is only 14% (77 edges) for subtype 1a and 6% (29 edges) for subtype 1b. Despite this overall similarity, the consensus networks show subtype-specific differences in clustering coefficient, degree, and betweenness centrality; measures that reflect residue connectivity and control over neighboring residues (Table 2) [18–21]. Although the subtype 1a network shows a higher clustering coefficient (3.4%) and higher average number of neighbors (6.5%), the average distance from resistance-associated sites to NS3-4A protease functional sites in subtype 1a is larger than in subtype 1b (2.5%) (Table 2, Suppl. Tables 1 and 2).

We characterized the residue frequency distance (RFD) between subtypes to assess similarity between

AA frequencies at equivalent sequence positions in the NS3-4A protease (Suppl. Tables 3 and 4). Thereby, large RFD values (>0.1) indicate a tendency to have a different AA distribution between subtypes (see Materials and Methods). From the 181 NS3-4A protease residues, only 19 residues (10.5%) show RFD values >0.1, whereas the majority of residues (152 residues; 84%), comprising most of the PI resistance-associated sites, shows RFD values <0.001, thus have the same or similar amino-acid profile in subtype 1a and 1b (Suppl. Tables 1 and 2). A subtype-specific different AA profile at PI resistance-associated sites is observed at residue 80 and 170 (RFD > 0.1).

Local network topology at PI resistance-associated sites

Next, we characterized the local residue neighborhood at PI resistance-associated sites to predict the
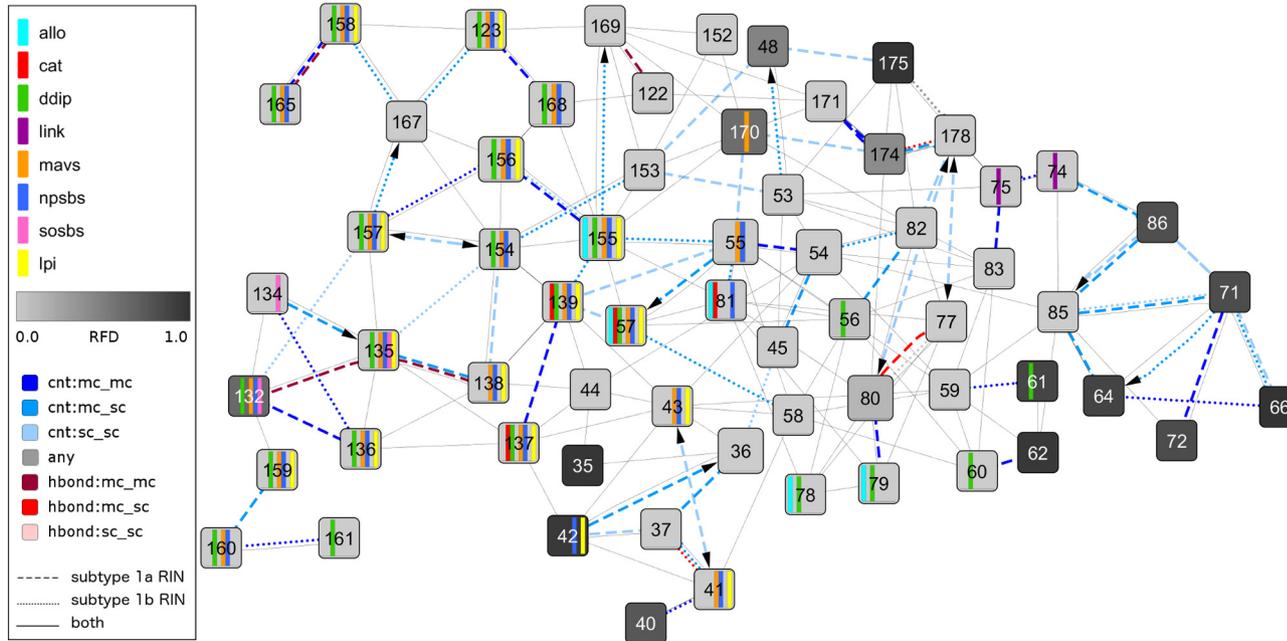
Fig. 2. Network comparison view between subtype 1a and 1b consensus RINs with focus on functional sites and PI resistance-associated sites. This comparison network view shows the noncovalent residue interactions between a subset of residues of the NS3-4A protease with focus on the differences between subtype 1a and 1b consensus RINs. The view focuses on the PI resistance-associated sites 36, 54, 55, 80, 155, 156, 168, and 170 and further includes their direct neighbors, all residues annotated as functionally important (see Materials and Methods), and residues that vary between subtype 1a and 1b. Functional residue annotations are shown as vertical bars (cyan, allosteric site; red, catalytic site; green, domain–domain interaction sites; violet, residues interacting with the linker region; orange, MAVS peptide; blue, natural peptide substrate; yellow, linear PI; purple, SOS-binding site of TRIF). The RFD is mapped to the node color using a white-to-gray gradient for small-to-large values. Solid edge lines represent noncovalent residue interactions present in both subtype consensus RINs, dashed edge lines are residue interactions from the subtype 1a consensus RIN, and dotted edge lines correspond to residue interactions in the subtype 1b consensus RIN. Edges that represent contacts are colored in blue, hydrogen bonds in red, and overlaps in gray; a darker color indicates interaction between main chain atoms, while lighter color and an arrow stand for sc atom interactions.

Table 2. Consensus RIN topology measures

Topology measure	Subtype	
	1a	1b
Radius	6	6
Diameter	12	11
Ave. path length	4.777	4.785
Ave. clustering coefficient	0.422	0.408
Ave. number of neighbors	6.157	5.78
Density	0.035	0.034
Connected nodes	178	173
Isolated nodes	3	7
Combi edges (unique[a])	548 (77)	500 (29)
Common[b] combi edges	85.9%	94.2%
Contact edges (unique[a])	726 (184)	676 (129)
H-bond edges (unique[a])	89 (30)	74 (15)

[a] Edges unique to either the subtype 1a or 1b RIN.
[b] Edges common to both subtype 1a and 1b RINs.

potential impact of AA substitutions on the residue-bonding patterns and protease functional sites.

Network topology at V36

V36 is connected to five neighboring residues, 35, 37, 42, 43, and 45; 2/5 are protease functional sites (natural protease substrate-binding site) (Fig. 3a). The clustering coefficient for V36 is higher in subtype 1a than 1b (0.333 *versus* 0.167) (Table 3, Suppl. Fig. 1). The subtype 1a-bonding pattern for V36 shows specific interactions with residue 42 (natural protease substrate-binding site) that also has a different AA distribution between subtypes (RFD 0.1). We identified an experimental protein structure comprising the AA substitution V36M in subtype 1a, PDB-2QV1, and constructed a comparison network from the subtype 1a-M36 mutant structure and the subtype 1a consensus network (Fig. 4a). In the comparison network, M36 retains all seven noncovalent interactions with first neighbor residues, connecting to five additional residues through seven new side chain (sc) interactions; van-der-Waals interactions (residues 34, 43, 45, 52, 64, and 85) and one H-bond interaction (residue 42). Although these changes amounts to a 100% gain of interactions for subtype 1a-M36 compared to wild type (WT), only few protease functional sites are affected; 7/10 interactions occur with nonfunctional sites. As expected from the residue-network analysis, V36M in subtype 1a shows no significant deficit in its replicative fitness as suggested by the long-term persistence of V36M-RAVs in subtype 1a patients (Table 1).

Network topology at T54

T54 is highly connected with neighboring residues with a node degree of 0.05 and a clustering coefficient of 0.25 in both subtypes (Table 3, Suppl. Fig. 1). It

shows direct interactions to nine neighboring residues, 43, 44, 45, 53, 55, 59, 82, 83, and 85 (Fig. 3b); 2/9 residues are protease functional sites (natural protease-substrate binding sites). Among those sites, residue 55 is previously reported with a key role in structural rearrangements at the protease–helicase interface [23]. The mechanism is known to involve a V55-neighboring surface loop in the NS3-4A protease structure [23]. By using the consensus residue networks, we identified a unique 54–55 interaction in subtype 1a and a related subtype 1a-specific subnetwork of residue interactions comprising the loop residues 56, 57, and 58 (Suppl. Fig. 2). Our network-based findings suggest a subtype-specific modulatory role of this loop structure on the protease–helicase interface in subtype 1a that likely explains the discrepancy in the persistence of T54S-RAVs between subtype 1a and 1b patients (Table 1).

Network topology at Q80

Q80 is highly connected with neighboring nodes, showing a similar clustering coefficient of 0.476 and 0.467 in subtype 1a and 1b, respectively (Table 3, Suppl. Fig. 1). In both subtypes, Q80 forms direct interactions with seven neighboring residues; 3/7 are protease functional sites [natural protease substrate-binding site, allosteric site, domain–domain interaction sites (DDIPs)] (Fig. 3c). Q80 has a distinct AA distribution between subtypes (RFD > 0.1) and a subtype 1a-specific bonding pattern with residues in the protease–helicase interface: residue 79 and 178 (DDIP, allosteric site, inter-domain linker). Moreover, Q80 interact indirectly *via* residue 178 with residues that neighbor the protease–helicase linker: residue 74 and 75 (Fig. 2). Based on the network topology, we hypothesized a subtype-specific modulatory role for Q80 in the protease–helicase interaction and hence on RNA replication and assembly of infectious virus particles [24,25]. Molecular rearrangements in the protease–helicase interface due to Q80K are also supported from the local residue-bonding pattern at Q80. Different from the WT residue, the lysine sc in the mutant K80 is positively charged and provides for salt-bridge interactions with negatively charged neighboring residues. A potential K80 sc-interaction partner is D79, a DDIP that pertain to the NS3 allosteric site (Fig. 3c).

Network topology at R155

R155 is a functionally important protease residue (natural protease substrate-binding site, allosteric site, DDIP). The subtype-specific network topology at R155 differs in betweenness centrality and node degree that are both higher in subtype 1b than 1a (Table 3, Suppl. Fig. 1). Both measures indicate that R155 in the subtype 1b protease structure is more important for information spread and control over neighboring
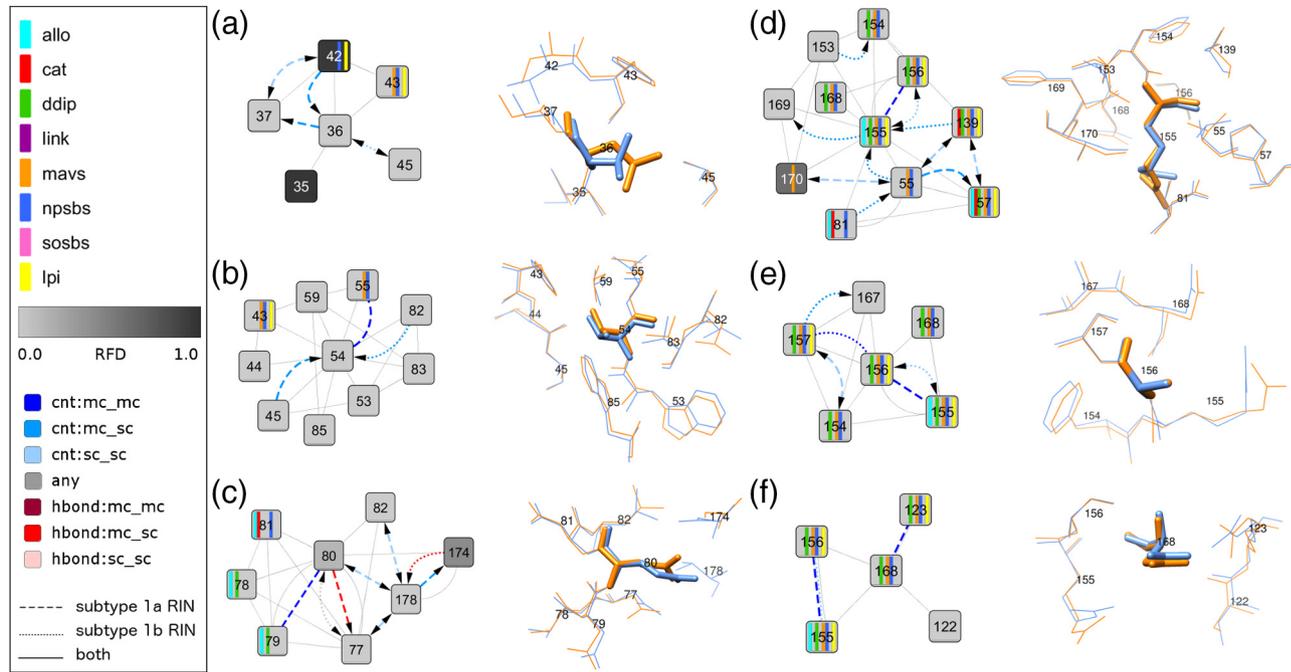
Fig. 3. Network comparison view and corresponding 3D structure close-up for the subtype-specific PI resistance-associated sites: (a) 36, (b) 54, (c) 80, (d) 155, (e) 156, and (f) 168. Each comparison network view shows the noncovalent residue interactions between the specified residue and its direct neighbors from the subtype 1a and 1b consensus RINs. Node and edge colors as well as functional annotations are described in the graphical legend above and in the Fig. 2 legend. The corresponding residues from the subtype 1a reference structure from PDB entry 2OBQ and the subtype 1b reference structure from PDB entry 1DXP are aligned and shown as blue and orange sticks, respectively.

Table 3. Network topology of PI resistance-associated sites

Class	Res	Subtype	Topology measures			
			Degree	Clust	Clos	Betw
L	36	1a	0.022	0.333	0.211	0.001
		1b	0.022	0.167	0.224	0.003
	54	1a	0.050	0.250	0.282	0.125
		1b	0.050	0.250	0.269	0.100
	55	1a	0.050	0.278	0.262	0.052
		1b	0.039	0.429	0.249	0.034
	170	1a	0.044	0.286	0.257	0.037
		1b	0.034	0.400	0.245	0.022
M[a]	80	1a	0.039	0.476	0.222	0.006
		1b	0.034	0.467	0.214	0.007
B	155	1a	0.050	0.278	0.262	0.043
		1b	0.056	0.222	0.262	0.072
	156	1a	0.028	0.500	0.235	0.007
		1b	0.028	0.500	0.229	0.008
	168	1a	0.022	0.333	0.222	0.004
		1b	0.022	0.167	0.217	0.006

Class, class of compound (L, linear; M, macrocyclic; B, both) [22]; Res, residue; Degree, degree centrality; Clust, clustering coefficient; Clos, closeness centrality; Betw, betweenness centrality. All values are rounded for better readability.
   [a] Exception: Sovaprevir [22].

residues than in subtype 1a. R155 in both subtypes is connected to 10 neighboring residues; 8/10 interacting sites are functionally important (Fig. 3d). In subtype 1b, a unique R155 sc-interaction with S139 of the protease catalytic triad suggests a critical impact of AA substitutions at R155 on the proteolytic activity. We identified an experimental protein structure comprising the AA substitution R155K in subtype 1a, PDB-2OIN. To characterize the structure impact of subtype 1a-R155K, we constructed a comparison network from the mutant structure and the subtype 1a-specific consensus network (Fig. 4b) and quantified the fraction of changes in residue interactions. Interestingly, K155 in subtype 1a retains all 12 noncovalent interactions with first neighbors and only gains two additional interactions (17%). We find no direct impact on protease functional sites in subtype 1a-R155K.

Other resistance sites

Only few clinical data are available on RAV persistence for AA substitutions at V55, A156, D168, and I170. These residues form dense bonding patterns with functionally important neighboring residues. For A156 and D168, we find only minor subtype-specific differences in the local residue-bonding patterns (Fig 3EF). In contrast, V55 and I170 form a unique sc interaction in subtype 1a (Suppl. Fig. 3) and show subtype-specific differences in node degree, clustering coefficient and betweenness centrality (Table 3, Suppl. Fig. 1).

Replicative fitness of Q80K in different genomic backgrounds

Q80K is clinically most important and shows particularly strong subtype specificity. Based on our

network analysis, we predicted subtype-specific structural rearrangements that impact RNA replication and virus assembly differentially in subtype 1a and 1b. To prove our *in silico* predictions and explore the mechanisms underlying the Q80K prevalence for subtype 1a, we applied reverse genetics in subtype-specific infectious HCV cell culture models. The AA substitution Q80K was created within the subtype 1a genomic background H77S.3 and the subtype 1b genomic background N.2, and their impact on replication of the viral RNA and production of infectious virus determined in RNA-transfected cells. Detailed information on experimental procedures is given in Materials and Methods. Both viruses are adapted to replicate in cell culture and as reported previously, the kinetics of RNA replication and virus production differ between H77S.3 and N.2 viruses [26].

(i) Replicative fitness: In both genomic backgrounds, Q80K showed only modest impairment of RNA replication capacity as measured by GLuc activity lower than the parental (WT) RNA. Compared to the parental H77S.3/GLuc2A and N.2/Gluc2A RNA, we observed a 20% reduction of replication in H77S.3 *versus* 6% reduction in N.2 at 72 h after transfection (Fig. 5a).

(ii) Infectious virus yield: Yields of infectious virus generally correlate well with the RNA replication capacity [16]. In a subtype 1a genomic background, the infectious virus production for H77S.3/Q80K was not reduced compared to the parental genome. Here, Q80K even produced slightly more infectious virus particles than WT at all times measured. In contrast, Q80K in the subtype 1b genetic background N.2 showed greater impairment in its ability to produce infectious virus than expected from
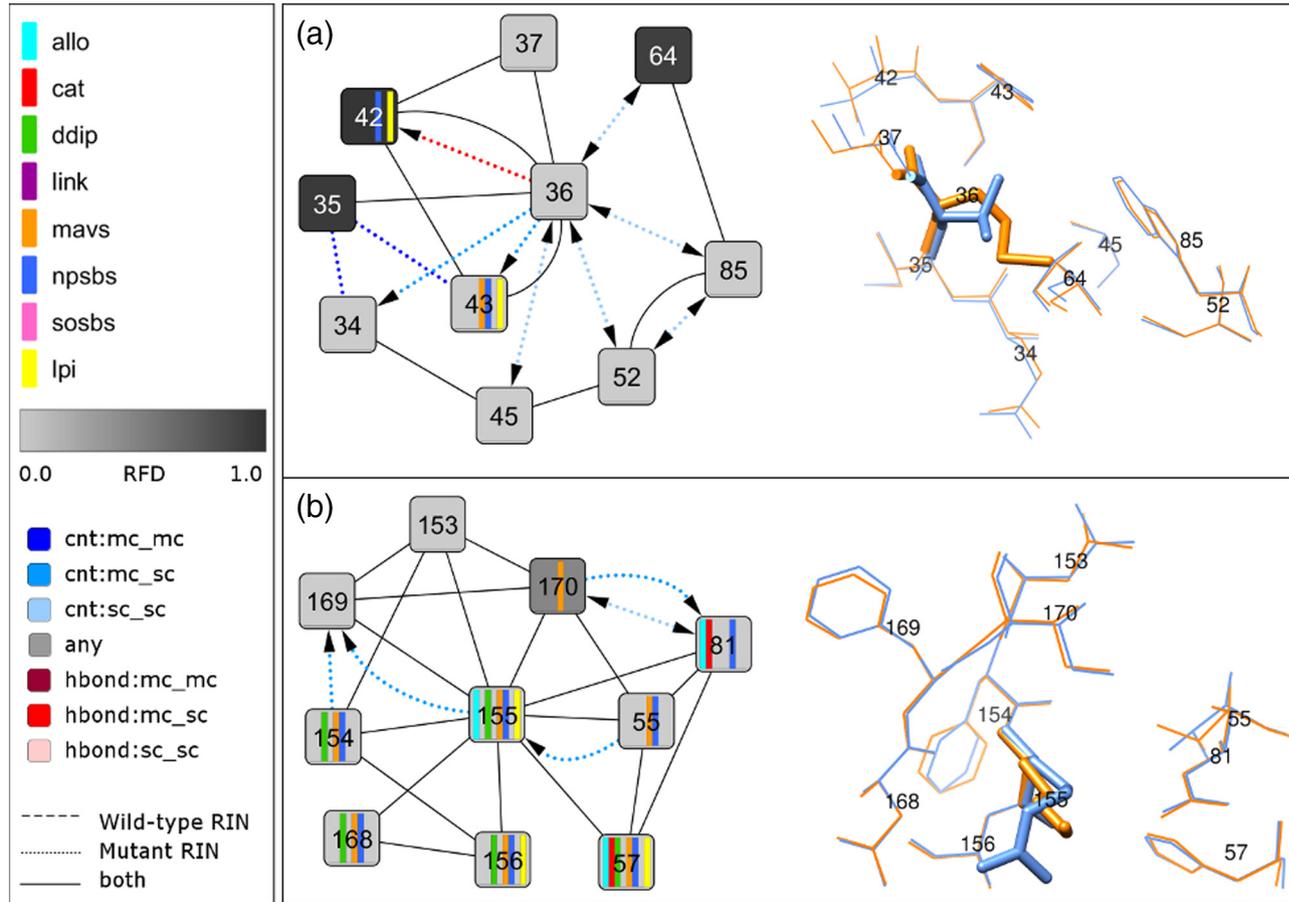
Fig. 4. Network comparison view between WT and mutant structure for V36M and R155K and corresponding 3D structure close-up. (a) The comparison network view between the subtype 1a consensus RIN and the RIN of the V36M mutant structure from PDB entry 2QV1 focuses on the direct noncovalent interactions of residue 36 with its neighbors. The corresponding residues in the structure are shown as blue and orange sticks, respectively. (b) The comparison network view between the subtype 1a consensus RIN and the RIN of the R155K mutant structure from PDB entry 2OIN focuses on the direct noncovalent interactions of residue 155 with its neighbors. The corresponding residues in the structure are shown as blue and orange sticks, respectively.
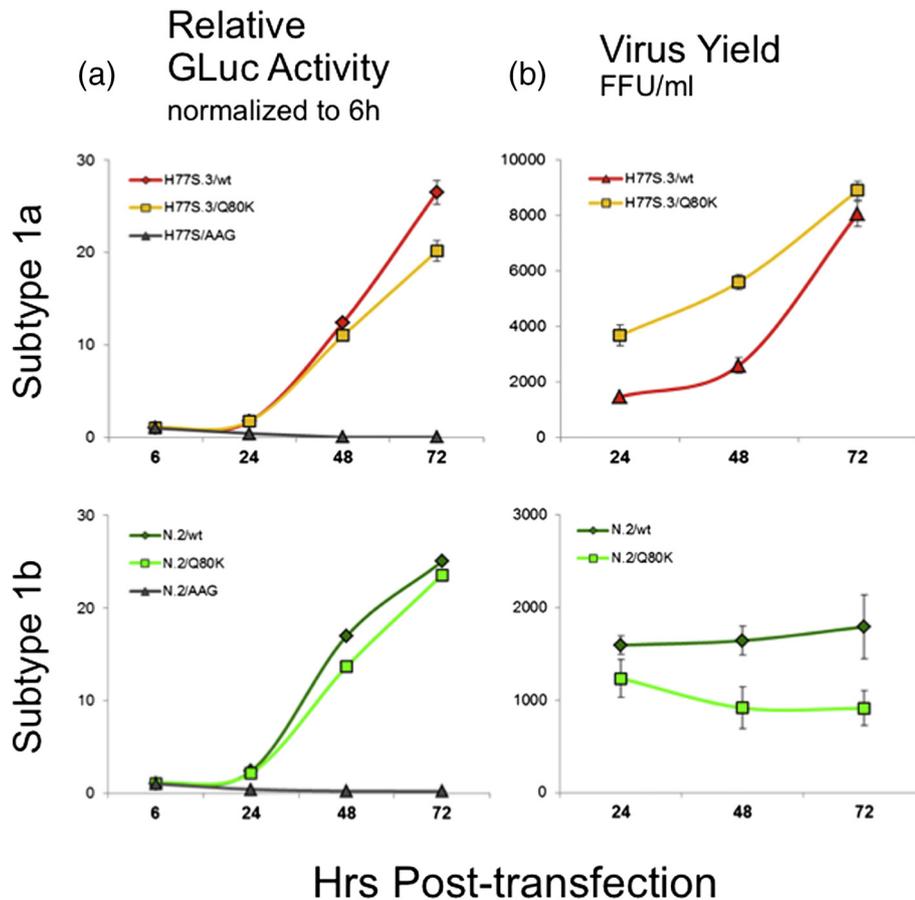
Fig. 5. Comparison of replication capacity and infectious virus yield of H77S.3 *versus* N.2 RNAs with the Q80K substitution in the NS3-4A protease. (a) The effect of NS3–Q80K on the RNA replication capacity was measured by GLuc assay in the H77S.3/GLuc2A (left, upper panel) *or* the N.2/GLuc2A genetic background (left, lower panel). Results are normalized to GLuc activity at 6 h post-transfection and represent the mean ± SD of triplicate samples. (b) The effect of NS3–Q80K on the infectious virus production from the H77S.3 (right, upper panel) *or* the N.2 background (right, lower panel) was measured by an FFU assay from cell culture supernatant fluids collected every 24 h post-transfection. Data represent the mean ± SD from at least three independent experiments.

reductions in RNA replication capacity. Virus production by N.2/Q80K was already lower at 24 h after transfection (23% compared to WT). By 72 h, virus production was almost 50% lower compared to parental N.2 (Fig. 5b). As predicted from the residue-network analysis detailed above, the production of infectious virus showed distinct subtype-specific differences. The discordance between RNA replication capacity and yields of infectious virus for N.2/Q80K suggests a negative impact particularly on infectious virus production for Q80K in a subtype 1b genomic background.

Discussion

We made use of publicly available sequence and structure data and present a graph-based approach to address subtype-specific differences in the replicative fitness of drug-resistant viral variants in HCV. Com-

parison views of subtype-specific consensus networks of the NS3-4A protease of HCV were constructed from protein structure ensembles and described as graphs that use common measures of node centrality. The consensus networks were characterized for their global network topology and local connectivity at PI resistance-associated sites. Different from coevolutionary analysis techniques that infer structural details of proteins based on primary sequence information [27,28], we characterize bonding patterns in RINs based on available experimental structure information from the PDB. The most novel aspect is the idea of creating consensus residue networks that make multiple experimental structure information accessible and comparable to gain additional insights into the structural and functional roles of interacting residues.

In the NS3-4A protease of HCV, the overall network topology between subtype 1a and 1b is similar. In both consensus networks, the drug resistance-associated sites are highly connected with protease functional

sites; however, we find that resistance-associated sites in the subtype 1a network are more distant from protease functional sites than in subtype 1b. This can explain lower replicative fitness costs of drug resistance mutations in a subtype 1a protease genomic background (as suggested by RAV persistence; Table 1). Although our approach is capable of detecting allosteric interactions as well as capturing effects of residues distant in the protein structure (Suppl. Fig. 2), we find the local analysis of residue interactions most suitable to explain clinical observations in this particular application scenario. To focus on the specific PI resistance-associated sites, we characterized their local residue-interaction patterns with near neighbor residues. Here, the bonding patterns at major resistance-associated sites suggest a potential for better variant replicative fitness in a subtype 1a structure-specific context.

In particular, we identified specific molecular details that might explain the prevalence of Q80K and R155K for subtype 1a and higher fitness costs in subtype 1b. From pooled samples from 1797 patients in phase 3 clinical trials with a PI, Sullivan *et al.* [3] observed no R155K-RAVs in subtype 1b but frequently observed R155K-RAVs in subtype 1a with long-term persistence (Table 1). Whereas a dense R155 residue-interaction pattern comprising multiple functional residues of the protease active site is observed in both subtypes, additional interactions to the protease catalytic triad are specific for subtype 1b. The unique noncovalent interactions suggest a larger subtype-specific structural impact on the protease catalytic site for R155K in subtype 1b. This notion is indirectly supported by experimental mutant structure data from R155K in subtype 1a. Here, the mutant structure revealed no direct impact on protease functional sites. Q80K is highly prevalent in subtype 1a (up to 47%) but rarely observed in subtype 1b (0.5%) [22]. As a natural polymorphism in treatment-naïve patients, Q80K has been associated with a reduced treatment response to simeprevir [5,29] and a median time until loss of 36 months in subtype 1a *versus* 24 months in subtype 1b [5]. By exploratory visual analysis of the RINs, we identified subtype-specific bonding patterns of Q80 with residues of the protease–helicase interface and indirect residue interactions with the protease–helicase linker that suggest a subtype-specific role of this residue in viral replicative fitness. Because the two enzymatic domains of NS3 are highly interdependent and sensitive to mutations in the domain interface [24,25], we predicted subtype-specific molecular rearrangements due to Q80K that, by modulating NS3 helicase activity, can impair RNA replication and in addition can interfere with virus assembly. We tested our hypothesis using infectious cell culture models of subtype 1a and 1b. Whereas Q80K showed only modest impairment of RNA replication in both subtypes, we observed a pronounced drop of infectious virus yield that was specific for subtype 1b, whereas Q80K in subtype 1a even produced slightly higher amounts of infectious virus particles than the parental WT virus. Given that high replicative fitness confers a general determinant of multidrug resistance [30], our data also potentially explain the lower sustained-virological response rates for voxilaprevir in subtype 1a patients [31], although Q80K is reported to not confer changes to voxilaprevir susceptibility *in vitro* [32].

Collectively, we depicted differences in the network topology and local residue-interaction patterns of PI resistance-associated sites in the NS3-4A protease of HCV that are subtype specific and affect functionally important sites in the protease. Our results suggest that the subtype 1a NS3-4A genomic background provides for lower replicative fitness costs due to PI resistance mutations than the subtype 1b genomic background.

Materials and Methods

Sequence data

Sequences of subtype 1a and 1b proteases were downloaded from the euHCV database [33] on April 23, 2014 after filtering based on the following criteria: single protein, standard name ns3, confirmed subtype and published before 2010. Identical sequences were removed. This resulted in 163 sequences for subtype 1a and 178 sequences for subtype 1b. As suggested by Berger *et al.* [34], we also downloaded sequences with identifiers between KC123434 and KC127656 from the GenBank database [35] to have a better representation of the natural genetic diversity. After filtering out NS5 sequences, we retrieved 2111 subtype 1a and 1335 subtype 1b NS3 sequences (as of December 12, 2014) and generated multiple sequence alignments for each subtype using MUSCLE [36] with default settings as provided by the EMBL-EBI bioinformatics web framework [37]. We computed the conservation as defined by Valdar [38] and the frequency of AA occurrence at each sequence position using the RINerator package [39]. The RFD between subtypes at each alignment position was computed using the cosine distance between the AA frequency vectors of each subtype. The cosine distance between two vectors is equal to one minus the cosine similarity, which is defined as the normalized dot product of the two vectors. Euclidean distance, cosine, and correlation similarity [40] correlated highly for all measures. We chose the cosine distance because they were more distinctive and easier to interpret than those with a similarity measure. Since RFD values agree very well between the two sequence data sources, we only discuss the GenBank RFD. We chose a threshold of 0.1 to distinguish between scores that represent conservation (RFD < 0.1) and substitution tendency (RFD > 0.1) based on the distribution of scores for

sequences from the euHCV and GenBank databases (Suppl. Fig. 4). There is a clear separation of scores for the euHCV sequences between RFD of 0.45 and 0.016 with only one residue with a score between these two. In addition, the 19 residues with an RFD > 0.1 overlap between both sets of sequences, although they do not have exactly the same values (see Suppl. Table 3 for RFD scores of each residue).

Structure data

The Protein Data Bank RCSB PDB [41] was used for a comprehensive analysis of available NS3-4A protease structure information using subtype 1a and 1b structure ensembles and RINs. We identified all database entries of the NS3-4A protease (102 as of June 5, 2014) and selected representative protein structures based on the following criteria: unbound, without helicase domain, high resolution and R-free, no mutations, that is 3KF2 [42], 1A1R [43], 2OBQ [44] for subtype 1a and 1DXP [45], and 1NS3 [46] for subtype 1b (Suppl. Table 5). Since the quality of 1NS3 in comparison to the other selected structures is poor, we made sure the consensus makes sense by comparing the number of edges between the consensus networks. For additional analysis, we selected 1CU1 [47], a subtype 1b 3D structure of the whole nonstructural protein NS3, as well as the subtype 1a structures 2QV1 [48] and 2OIN [49], which contain a V36M and a R155K AA substitution, respectively. For each structure, we used the longest resolved chain (chain A for 2OBQ, 1DXP, 1NS3, and 1CU1 and chain B for 3KF2 and 1A1R). We aligned the selected chains to create a residue mapping for each subtype using residue indices of 2OBQ and 1DXP for subtype 1a and 1b, respectively.

Residue interaction data

For selected PDB structures, we created RINs using RINerator [39]. Hydrogens were added to the 3D protein structures using the Reduce tool [50]. Contacts on the van-der-Waals surface of each atom were sampled by the Probe tool [51] and finally summarized into residue interactions by RINerator. Nodes in a RIN represent protein residues and edges indicate non-covalent residue interactions between residues; that is, van-der-Waals contacts, hydrogen bonds, overlaps of van-der-Waals radii, and a combination of the three (any of the previous three). We also distinguish between main chain and side chain (sc) interactions. The RINs and additional data associated with them are visualized in Cytoscape using the RINalyzer and structureViz2 apps [52] (Suppl. Fig. 5). We enriched the networks with sequence conservation, RFD, and functional site annotations. To define a representative RIN for each subtype, we combined noncovalent interactions from individual RINs from a subtype-specific structure ensemble into a single network and

weighed the edges by the fraction of structures, in which they are present. For subtype 1a, the representative RIN contains edges with weights 0.33, 0.66 and 1.0, which corresponds to interactions in one, two or all three structures, while for subtype 1b, there are only weights of 0.5 and 1.0 for interactions in only one or both structures (Suppl. Fig. 6). We created a consensus RIN for each subtype ensemble from all non-covalent interactions present in all individual RINs, which correspond to all edges with weight of 1.0.

Selection of functional sites

Residues in the RIN previously identified to pertain to different NS3 function were grouped as follows:

- Protease catalytic triad and oxyanion hole (residues 57, 81, 137, 139)
- Putative DDIPs (from PDB-1CU1, residues 56, 57, 60, 61, 78, 79, 123, 132, 135, 136, 137, 139, 154, 155, 156, 157, 158, 159, 160, 161, 165, 168, 186) [47]
- Allosteric site (residues 57, 78, 79, 81, 155) [11]
- Linker region (residues 182 to 195, but missing in the selected structures: residues 74, 75, and 181 interact with them in the RIN) [24]
- Binding site for the natural peptide (see, e.g., PDB-1CU1, residues 41, 42, 43, 55, 57, 58, 81, 123, 132, 135, 136, 137, 138, 139, 154, 155, 156, 157, 158, 159, 160, 165, 168) [47]
- MAVS binding specificity shows similar determinants to that of the protease natural substrate; MAVS binding sites (from PDB–3RC5, residues 41, 43, 55, 57, 123, 132, 135, 136, 137, 138, 139, 154, 155, 156, 157, 158, 159, 160, 161, 162, 168) [15,16]
- An eight-residue polyproline tract (SOS-binding motif) in TRIF is essential for binding to the protease [53]; SOS-binding site for TRIF (residues 132, 134, 135) [53]
- Linear PI binding site (from PDB-3SV6, residues 41, 42, 43, 57, 123, 135, 136, 137, 138, 139, 155, 156, 157, 158, 159) [54]

Binding sites for the natural peptide, MAVS and a linear PI were determined using PyMOL as all residues with an atom within 5 Å of the respective ligand or product, respectively, in the bound NS3 structure (1CU1 [47], 3RC5 [55], 3SV6 [54], respectively). DDIPs contain the residues from the NS3-4A protease domain that are directly interacting noncovalently with residues of the NS3 helicase domain in the RIN of the NS3 protease–helicase complex structure from PDB-1CU1 (Suppl. Fig. 7). According to Kohlway *et al.* [25], the NS3 inter-domain linker region comprises residues 182 to 195. Since these residues are missing from our selected subtype 1a and 1b structures, we identified all residues interacting with the linker region using the

noncovalent interactions in the RIN of PDB-1CU1 (Suppl. Fig. 8).

Topological network analysis

We performed topological analysis on consensus and comparison RINs by computing topology measures using the NetworkAnalyzer plugin for Cytoscape [56] and the NetworkX Python package [57]. Measures that describe the overall topology of a network are its radius and diameter, average shortest path length, and network clustering coefficient. The length of a path between two nodes is the number of edges forming it with the length of the shortest path referred to as distance. Eccentricity is defined as the maximum distance between a given node and any other node in the network; largest and smallest eccentricities are called network diameter and radius, respectively. The average shortest path length is the average distance between two connected nodes. The network clustering coefficient is the average of the clustering coefficients for all nodes in the network.

For each residue, we calculated the following topology measures; for detailed definitions, see Ref. [58]:

- Average distance to functional sites is the average shortest path length from a given node to any functional residue and gives a quantitative measure of how close a residue is to functionally important residues in NS3-4A.
- Degree centrality is the fraction of all network nodes connected to a given node. This measure distinguishes between nodes with many neighbors and those that only have a few neighbors (for instance, nodes that represent residues in the protein inside have higher degree centrality than residues on the protein surface).
- Clustering coefficient (density) measures the fraction of edges between direct neighbors of a given node relative to the total number of possible edges between all these direct neighbors and indicates how fully connected a group of nodes is.
- Closeness centrality is defined as the reciprocal of the sum of the shortest path distances from a given node to all other nodes and is a measure of how fast information spreads from a given node to other reachable nodes [18–21].
- Betweenness centrality is computed as the fraction of shortest paths that pass through a given node. This measure reflects the amount of control that a node exerts over interactions of other nodes in the network and favors nodes that join dense subnetworks, rather than nodes that lie inside such subnetworks or that lie in the periphery [18–21].
- Edge fraction is the fraction of edges with a specific property that connect a given node to its

direct neighbors, that is, in the subtype comparison network this property describes the original network (subtype 1a, subtype 1b or both) to which the edges belong.

Cells and viruses

Huh7.5 hepatoma cells (Apath LLC) [59] were maintained in DMEM with 10% FBS, 100 U/ml penicillin/streptomycin, 1× non-essential AAs, 1× sodium pyruvate, and 1× Glutamax (all Gibco/Life Technologies). For virus production and GLuc assays, medium was supplemented with 1 µM vitamin E (α-tocopherol) (Sigma) and 50 mM Hepes buffer (pH 7.4; Cellgro) [26]. The cell culture infectious subtype 1a HCV genome H77S.3, and its derivative H77S.3/GLuc2A, encoding a GLuc reporter gene, were both described in Ref. [16]. The cell culture infectious subtype 1b HCV genome N.2 and its derivative N.2/GLuc2A, encoding a GLuc reporter gene, were both described in Ref. [60]. The Q80K AA substitution was generated in H77S.3 and N.2 backgrounds using the Quikchange Site-Directed Mutagenesis Kit (Agilent).

Assays of viral fitness

5×10^6 cells were electroporated with 10 µg HCV genomic RNA as described previously [61] and plated to 25-cm^2 cell culture vessels for assays to assess intracellular viral RNA abundance (GLuc assay) or extracellular infectious virus production [focus-forming unit (FFU) assay]. For GLuc assays, cell culture medium was harvested and replaced at 6, 24, 48, and 72 h after electroporation. Cell culture medium was analyzed for GLuc activity using the BioLux Gaussia substrate (New England BioLabs) on a Synergy 2 Multimode Plate Reader (Biotek). For FFU assay, cell culture medium was harvested and replaced at 24, 48, and 72 h after electroporation. Infectious virus was measured as described previously [61].

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmb.2019.04.034.

Present address: N.T. Doncheva, Center for Non-Coding RNA in Technology and Health, Department of Veterinary and Animal Sciences & NNF Center for Protein Research, University of Copenhagen, Copenhagen, Denmark.

Present address: D.R. McGivern, Center for Biologics Evaluation and Research, U.S. Food and Drug Administration, Silver Spring, MD, USA.

Abbreviations used:
AA, amino acid; DDIP, domain–domain interaction sites; FFU, focus-forming unit; HCV, hepatitis C virus; MAVS, mitochondrial antiviral signaling protein; NS, nonstructural protein; PDB, Protein Data Bank; PI, protease inhibitor; RAV, resistance-associated amino acid variant; RFD, residue frequency distance; RIN, residue-interaction network; RNA, ribonucleic acid; sc, side chain; TRIF, Toll/interleukin-1 receptor domain containing adapter inducing interferon-β; WT, wild type.

References

[1] A. Krishnan, J.P. Zbilut, M. Tomita, A. Giuliani, Proteins as networks: usefulness of graph theory in protein science, Curr. Protein Pept. Sci. 9 (2008) 28–38.

[2] C. Welsch, Genetic barrier and variant fitness in hepatitis C as critical parameters for drug resistance development, Drug Discov. Today Technol. 11 (2014) 19–25.

[3] J.C. Sullivan, S. De Meyer, D.J. Bartels, I. Dierynck, E.Z. Zhang, J. Spanks, A.M. Tigges, A. Ghys, J. Dorrian, N. Adda, E.C. Martin, M. Beumont, I.M. Jacobson, K.E. Sherman, S. Zeuzem, G. Picchio, T.L. Kieffer, Evolution of treatment-emergent resistant variants in telaprevir phase 3 clinical trials, Clin. Infect. Dis. 57 (2013) 221–229.

[4] S. Susser, J. Vermehren, N. Forestier, M.W. Welker, N. Grigorian, C. Fuller, D. Perner, S. Zeuzem, C. Sarrazin, Analysis of long-term persistence of resistance mutations within the hepatitis C virus NS3 protease after treatment with telaprevir or boceprevir, J. Clin. Virol. 52 (2011) 321–327.

[5] O. Lenz, T. Verbinnen, B. Fevery, L. Tambuyzer, L. Vijgen, M. Peeters, A. Buelens, H. Ceulemans, M. Beumont, G. Picchio, S. De Meyer, Virology analyses of HCV isolates from genotype 1-infected patients treated with simeprevir plus peginterferon/ribavirin in phase IIb/III studies, J. Hepatol. 62 (2015) 1008–1014.

[6] M.P. Manns, J.M. Vierling, B.R. Bacon, S. Bruno, O. Shibolet, Y. Baruch, P. Marcellin, L. Caro, A.Y. Howe, C. Fandozzi, J. Gress, C.L. Gilbert, P.M. Shaw, M.P. Cooreman, M.N. Robertson, P. Hwang, F.J. Dutko, J. Wahl, N. Mobashery, The combination of MK-5172, peginterferon, and ribavirin is effective in treatment-naive patients with hepatitis C virus genotype 1 infection without cirrhosis, Gastroenterology 147 (2014) 366-76.e6.

[7] P. Krishnan, G. Schnell, R. Tripathi, J. Beyer, T. Reisch, X. Zhang, C. Setze, L. Rodrigues Jr., M. Burroughs, R. Redman, K. Chayama, H. Kumada, C. Collins, T. Pilot-Matias, Analysis of hepatitis C virus genotype 1b resistance variants in Japanese patients treated with paritaprevir–ritonavir and ombitasvir, Antimicrob. Agents Chemother. 60 (2015) 1106–1113.

[8] S. Zeuzem, C. Hezode, J.P. Bronowicki, V. Loustaud-Ratti, F. Gea, M. Buti, A. Olveira, T. Banyai, M.T. Al-Assi, J. Petersen, D. Thabut, A. Gadano, R. Pruitt, M. Makara, M. Bourliere, S. Pol, M. Beumont-Mauviel, S. Ouwerkerk-Mahadevan, G. Picchio, M. Bifano, F. McPhee, N. Boparai, K. Cheung, E.A. Hughes, S. Noviello, Daclatasvir plus simeprevir with or without ribavirin for the treatment of chronic hepatitis C virus genotype 1 infection, J. Hepatol. 64 (2016) 292–300.

[9] M. Buti, S.C. Gordon, E. Zuckerman, E. Lawitz, J.L. Calleja, H. Hofer, C. Gilbert, J. Palcza, A.Y. Howe, M.J. DiNubile, M.N. Robertson, J. Wahl, E. Barr, X. Forns, Grazoprevir, elbasvir, and ribavirin for chronic hepatitis C virus genotype 1 infection after failure of pegylated interferon and ribavirin with an earlier-generation protease inhibitor: final 24-week results from C-SALVAGE, Clin. Infect. Dis. 62 (2016) 32–36.

[10] R. Bartenschlager, V. Lohmann, F. Penin, The molecular and structural basis of advanced antiviral therapy for hepatitis C virus infection, Nat. Rev. Microbiol. 11 (2013) 482–496.

[11] S.M. Saalau-Bethell, A.J. Woodhead, G. Chessari, M.G. Carr, J. Coyle, B. Graham, S.D. Hiscock, C.W. Murray, P. Pathuri, S.J. Rich, C.J. Richardson, P.A. Williams, H. Jhoti, Discovery of an allosteric mechanism for the regulation of HCV NS3 protein function, Nat. Chem. Biol. 8 (2012) 920–925.

[12] K. Li, E. Foy, J.C. Ferreon, M. Nakamura, A.C. Ferreon, M. Ikeda, S.C. Ray, M. Gale Jr., S.M. Lemon, Immune evasion by hepatitis C virus NS3/4A protease-mediated cleavage of the toll-like receptor 3 adaptor protein TRIF, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 2992–2997.

[13] E. Meylan, J. Curran, K. Hofmann, D. Moradpour, M. Binder, R. Bartenschlager, J. Tschopp, Cardif is an adaptor protein in

the RIG-I antiviral pathway and is targeted by hepatitis C virus, Nature 437 (2005) 1167–1172.

[14] C. Welsch, K. Haselow, J. Gouttenoire, M. Schneider, K. Morikawa, Y. Martinez, S. Susser, C. Sarrazin, S. Zeuzem, I. Antes, D. Moradpour, C.M. Lange, Hepatitis C virus variants resistant to macrocyclic NS3-4A inhibitors subvert IFN-beta induction by efficient MAVS cleavage, J. Hepatol. 62 (2015) 779–784.

[15] D.R. McGivern, T. Masaki, W. Lovell, C. Hamlett, S. Saalau-Bethell, B. Graham, Protease inhibitors block multiple functions of the NS3/4A protease–helicase during the hepatitis C virus life cycle, J. Virol. 89 (2015) 5362–5370.

[16] T. Shimakami, C. Welsch, D. Yamane, D.R. McGivern, M. Yi, S. Zeuzem, S.M. Lemon, Protease inhibitor-resistant hepatitis C virus mutants with reduced fitness from impaired production of infectious virus, Gastroenterology 140 (2011) 667–675.

[17] EASL recommendations on treatment of hepatitis C, J. Hepatol. 66 (2016) 153–194.

[18] L.C. Freeman, Centrality in social networks conceptual clarification, Soc. Networks 1 (1978) 215–239.

[19] D. Koschützki, K.A. Lehmann, L. Peeters, S. Richter, D. Tenfelde-Podehl, O. Zlotowski, Centrality indices, in: U. Brandes, T. Erlebach (Eds.), Network Analysis: Methodological Foundations, Springer, Berlin Heidelberg, Berlin, Heidelberg 2005, pp. 16–61.

[20] S.P. Borgatti, 1. Centrality and network flow. Soc. Networks 27 (2005) 55–71.

[21] A.L. Barabasi, Z.N. Oltvai, Network biology: understanding the cell's functional organization, Nat. Rev. Genet. 5 (2004) 101–113.

[22] J. Vermehren, C. Welsch, C. Sarrazin, Protease inhibitor resistance, in: M. Gotte, A. Berghuis, G. Matlashewski, M. Wainberg, D. Sheppard (Eds.), Handbook of Antimicrobial Resistance, Springer, New York, 2015.

[23] C. Welsch, F.S. Domingues, S. Susser, I. Antes, C. Hartmann, G. Mayr, A. Schlicker, C. Sarrazin, M. Albrecht, S. Zeuzem, T. Lengauer, Molecular basis of telaprevir resistance due to V36 and T54 mutations in the NS3-4A protease of the hepatitis C virus, Genome Biol. 9 (2008) R16.

[24] R.K. Beran, A.M. Pyle, Hepatitis C viral NS3-4A protease activity is enhanced by the NS3 helicase, J. Biol. Chem. 283 (44) (Oct 31 2008) 29929–29937.

[25] A. Kohlway, N. Pirakitikulr, S.C. Ding, F. Yang, D. Luo, B.D. Lindenbach, A.M. Pyle, The linker region of NS3 plays a critical role in the replication and infectivity of hepatitis C virus, J. Virol. 88 (18) (2014 Sep) 10970–10974.

[26] D. Yamane, D.R. McGivern, E. Wauthier, M. Yi, V.J. Madden, C. Welsch, I. Antes, Y. Wen, P.E. Chugh, C.E. McGee, D.G. Widman, I. Misumi, S. Bandyopadhyay, S. Kim, T. Shimakami, T. Oikawa, J.K. Whitmire, M.T. Heise, D.P. Dittmer, C.C. Kao, S.M. Pitson, A.H.J. Merrill, L.M. Reid, S.M. Lemon, Regulation of the hepatitis C virus RNA replicase by endogenous lipid peroxidation, Nat. Med. 20 (8) (Aug 2014) 927–935.

[27] N. Halabi, O. Rivoire, S. Leibler, R. Ranganathan, Protein sectors: evolutionary units of three-dimensional structure, Cell. 138 (4) (Aug 21 2009) 774–786.

[28] M. Weigt, R.A. White, H. Szurmant, J.A. Hoch, T. Hwa, Identification of direct residue contacts in protein–protein interaction by message passing, Proc. Natl. Acad. Sci. U. S. A. 106 (1) (2009 Jan 6) 67–72.

[29] C. Sarrazin, E. Lathouwers, M. Peeters, B. Daems, A. Buelens, J. Witek, Y. Wyckmans, B. Fevery, T. Verbinnen, A.

Ghys, M. Schlag, A. Baldini, S. De Meyer, O. Lenz, Prevalence of the hepatitis C virus NS3 polymorphism Q80K in genotype 1 patients in the European region, Antivir. Res. 116 (2015) 10–16.

[30] J. Sheldon, N.M. Beach, E. Moreno, I. Gallego, D. Pineiro, E. Martinez-Salas, J. Gregori, J. Quer, J.I. Esteban, C.M. Rice, E. Domingo, C. Perales, Increased replicative fitness can lead to decreased drug sensitivity of hepatitis C virus, J. Virol. 88 (2014) 12098–12111.

[31] I.M. Jacobson, E. Lawitz, E.J. Gane, B.E. Willems, P.J. Ruane, R.G. Nahass, S.M. Borgia, S.D. Shafran, K.A. Workowski, B. Pearlman, R.H. Hyland, L.M. Stamm, E. Svarovskaia, H. Dvory-Sobol, Y. Zhu, G.M. Subramanian, D.M. Brainard, J.G. McHutchison, N. Brau, T. Berg, K. Agarwal, B.R. Bhandari, M. Davis, J.J. Feld, G.J. Dore, C.A.M. Stedman, A.J. Thompson, T. Asselah, S.K. Roberts, G.R. Foster, Efficacy of 8 weeks of sofosbuvir, velpatasvir, and voxilaprevir in patients with chronic HCV infection: 2 phase 3 randomized trials, Gastroenterology 153 (2017) 113–122.

[32] J.G. Taylor, T. Appleby, O. Barauskas, X. Chen, H. Dvory-Sobol, R. Gong, J. Lee, E. Nejati, B. Schultz, Y. Wang, C. Yang, M. Yu, S. Zipfel, K. Chan, P0899: preclinical profile of the pangenotypic HCV NS3/4A protease inhibitor GS-9857, J. Hepatol. 62 (2015) S681.

[33] C. Combet, N. Garnier, C. Charavay, D. Grando, D. Crisan, J. Lopez, A. Dehne-Garcia, C. Geourjon, E. Bettler, C. Hulo, P. Le Mercier, R. Bartenschlager, H. Diepolder, D. Moradpour, J.M. Pawlotsky, C.M. Rice, C. Trepo, F. Penin, G. Deleage, euHCVdb: the European hepatitis C virus database, Nucleic Acids Res. 35 (2007) D363–D366.

[34] K.L. Berger, I. Triki, M. Cartier, M. Marquis, M.J. Massariol, W.O. Bocher, Y. Datsenko, G. Steinmann, J. Scherer, J.O. Stern, G. Kukolj, Baseline hepatitis C virus (HCV) NS3 polymorphisms and their impact on treatment response in clinical studies of the HCV NS3 protease inhibitor faldaprevir, Antimicrob. Agents Chemother. 58 (2014) 698–705.

[35] D.A. Benson, M. Cavanaugh, K. Clark, I. Karsch-Mizrachi, D.J. Lipman, J. Ostell, E.W. Sayers, GenBank, Nucleic Acids Res. 41 (2013) D36–D42.

[36] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res. 32 (2004) 1792–1797.

[37] W. Li, A. Cowley, M. Uludag, T. Gur, H. McWilliam, S. Squizzato, Y.M. Park, N. Buso, R. Lopez, The EMBL-EBI bioinformatics web and programmatic tools framework, Nucleic Acids Res. 43 (2015) W580–W584.

[38] W.S.J. Valdar, Scoring residue conservation, Proteins 48 (2002) 227–241.

[39] N.T. Doncheva, K. Klein, F.S. Domingues, M. Albrecht, Analyzing and visualizing residue networks of protein structures, Trends Biochem. Sci. 36 (2011) 179–182.

[40] M.M. Deza, E. Deza, Encyclopedia of Distances, Springer-Verlag, Berlin Heidelberg, 2009.

[41] P.W. Rose, C. Bi, W.F. Bluhm, C.H. Christie, D. Dimitropoulos, S. Dutta, et al., The RCSB Protein Data Bank: new resources for research and education, Nucleic Acids Res. 41 (2013) D475–D482.

[42] M.D. Cummings, J. Lindberg, T.I. Lin, H. de Kock, O. Lenz, E. Lilja, S. Fellander, V. Baraznenok, S. Nystrom, M. Nilsson, L. Vrang, M. Edlund, A. Rosenquist, B. Samuelsson, P. Raboisson, K. Simmen, Induced-fit binding of the macrocyclic noncovalent inhibitor TMC435 to its HCV NS3/NS4A protease target, Angew. Chem. Int. Ed. Engl. 49 (2010) 1652–1655.

[43] J.L. Kim, K.A. Morgenstern, C. Lin, T. Fox, M.D. Dwyer, J.A. Landro, S.P. Chambers, W. Markland, C.A. Lepre, E.T. O'Malley, S.L. Harbeson, C.M. Rice, M.A. Murcko, P.R. Caron, J.A. Thomson, Crystal structure of the hepatitis C virus NS3 protease domain complexed with a synthetic NS4A cofactor peptide, Cell 87 (1996) 343–355.

[44] A.J. Prongay, Z. Guo, N. Yao, J. Pichardo, T. Fischmann, C. Strickland, J. Myers Jr., P.C. Weber, B.M. Beyer, R. Ingram, Z. Hong, W.W. Prosise, L. Ramanathan, S.S. Taremi, T. Yarosh-Tomaine, R. Zhang, M. Senior, R.S. Yang, B. Malcolm, A. Arasappan, F. Bennett, S.L. Bogen, K. Chen, E. Jao, Y.T. Liu, R.G. Lovey, A.K. Saksena, S. Venkatraman, V. Girijavallabhan, F.G. Njoroge, V. Madison, Discovery of the HCV NS3/4A protease inhibitor (1*R*,5*S*)-*N*-[3-amino-1-(cyclobutylmethyl)-2,3-dioxopropyl]-3- [2(*S*)-[[[(1,1-dimethy-lethyl)amino]carbonyl]amino]-3,3-dimethyl-1-oxobutyl]-6,6-dimethyl-3-azabicyclo[3.1.0]hexan-2(*S*)-carboxamide (Sch 503034) II. Key steps in structure-based optimization, J. Med. Chem. 50 (2007) 2310–2318.

[45] S. Di Marco, M. Rizzi, C. Volpari, M.A. Walsh, F. Narjes, S. Colarusso, R. De Francesco, V.G. Matassa, M. Sollazzo, Inhibition of the hepatitis C virus NS3/4A protease. The crystal structures of two protease-inhibitor complexes, J. Biol. Chem. 275 (2000) 7152–7157.

[46] Y. Yan, Y. Li, S. Munshi, V. Sardana, J.L. Cole, M. Sardana, C. Steinkuehler, L. Tomei, R. De Francesco, L.C. Kuo, Z. Chen, Complex of NS3 protease and NS4A peptide of BK strain hepatitis C virus: a 2.2 Å resolution structure in a hexagonal crystal form, Protein Sci. 7 (1998) 837–847.

[47] N. Yao, P. Reichert, S.S. Taremi, W.W. Prosise, P.C. Weber, Molecular views of viral polyprotein processing revealed by the crystal structure of the hepatitis C virus bifunctional protease–helicase, Structure 7 (1999) 1353–1363.

[48] Y. Zhou, D.J. Bartels, B.L. Hanzelka, U. Muh, Y. Wei, H.M. Chu, A.M. Tigges, D.L. Brennan, B.G. Rao, L. Swenson, A.D. Kwong, C. Lin, Phenotypic characterization of resistant Val36 variants of hepatitis C virus NS3-4A serine protease, Antimicrob. Agents Chemother. 52 (2008) 110–120.

[49] Y. Zhou, U. Muh, B.L. Hanzelka, D.J. Bartels, Y. Wei, B.G. Rao, D.L. Brennan, A.M. Tigges, L. Swenson, A.D. Kwong, C. Lin, Phenotypic and structural analyses of hepatitis C virus NS3 protease Arg155 variants: sensitivity to telaprevir (VX-950) and interferon alpha, J. Biol. Chem. 282 (2007) 22619–22628.

[50] J.M. Word, S.C. Lovell, J.S. Richardson, D.C. Richardson, Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation, J. Mol. Biol. 285 (1999) 1735–1747.

[51] J.M. Word, S.C. Lovell, T.H. LaBean, H.C. Taylor, M.E. Zalis, B.K. Presley, et al., Visualizing and quantifying molecular goodness-of-fit: small-probe contact dots with explicit hydrogen atoms, J. Mol. Biol. 285 (1999) 1711–1733.

[52] N.T. Doncheva, K. Klein, J.H. Morris, M. Wybrow, F.S. Domingues, M. Albrecht, Integrative visual analysis of protein sequence mutations, BMC Proc. 8 (2014) S2.

[53] J.C. Ferreon, A.C. Ferreon, K. Li, S.M. Lemon, Molecular determinants of TRIF proteolysis mediated by the hepatitis C virus NS3/4A protease, J. Biol. Chem. 280 (2005) 20483–20492.

[54] K.P. Romano, A. Ali, C. Aydin, D. Soumana, A. Ozen, L.M. Deveau, C. Silver, H. Cao, A. Newton, C.J. Petropoulos, W. Huang, C.A. Schiffer, The molecular basis of drug resistance against hepatitis C virus NS3/4A protease inhibitors, PLoS Pathog. 8 (2012), e1002832.

[55] K.P. Romano, J.M. Laine, L.M. Deveau, H. Cao, F. Massi, C.A. Schiffer, Molecular mechanisms of viral and host cell substrate recognition by hepatitis C virus NS3/4A protease, J. Virol. 85 (2011) 6106–6116.

[56] Y. Assenov, F. Ramirez, S.-E. Schelhorn, T. Lengauer, M. Albrecht, Computing topological parameters of biological networks, Bioinformatics 24 (2008) 282–284.

[57] A.A. Hagberg, D.A. Schult, P.J. Swart, Exploring network structure, dynamics, and function using NetworkX, Proceedings of the 7th Python in Science Conference (SciPy2008) 2008, pp. 11–15.

[58] N.T. Doncheva, Y. Assenov, F.S. Domingues, M. Albrecht, Topological analysis and interactive visualization of biological networks and protein structures, Nat. Protoc. 7 (2012) 670–685.

[59] K.J. Blight, J.A. McKeating, C.M. Rice, Highly permissive cell lines for subgenomic and genomic hepatitis C virus RNA replication, J. Virol. 76 (2002) 13001–13014.

[60] R.P. Kannan, L.L. Hensley, L.E. Evers, S.M. Lemon, D.R. McGivern, Hepatitis C virus infection causes cell cycle arrest at the level of initiation of mitosis, J. Virol. 85 (2011) 7989–8001.

[61] D.R. McGivern, T. Masaki, S. Williford, P. Ingravallo, Z. Feng, F. Lahser, E. Asante-Appiah, P. Neddermann, R. De Francesco, A.Y. Howe, S.M. Lemon, Kinetic analyses reveal potent and early blockade of hepatitis C virus assembly by NS5A inhibitors, Gastroenterology 147 (2014) 453–462.e7.