Self-association and conformational variation of NS5A domain 1 of hepatitis C virus

Serap Beldar,1 Malathy Sony Subramanian Manimekalai,1 Nam-Joon Cho,2 Kwanghee Baek,3 Gerhard Grüber1 and Ho Sup Yoon1,3,*

Abstract

Direct-acting antivirals (DAAs) targeting the non-structural 5A (NS5A) protein of the hepatitis C virus (HCV) are crucial drugs that have shown exceptional clinical success in patients. However, their mode of action (MoA) remains unclear, and drug-resistant HCV strains are rapidly emerging. It is critical to characterize the behaviour of the NS5A protein in solution, which can facilitate the development of new classes of inhibitors or improve the efficacy of the currently available DAAs. Using biophysical methods, including dynamic light scattering, size exclusion chromatography and chemical cross-linking experiments, we showed that the NS5A domain 1 from genotypes 1b and 1a of the HCV intrinsically self-associated and existed as a heterogeneous mixture in solution. Interestingly, the NS5A domain 1 from genotypes 1b and 1a exhibited different dynamic equilibria of monomers to higher-order structures. Using small-angle X-ray scattering, we studied the structural dynamics of the various states of the NS5A domain 1 in solution. We also tested the effect of daclatasvir (DCV), the most prominent DAA, on self-association of the wild and DCV-resistant mutant (Y93H) NS5A domain 1 proteins, and demonstrated that DCV induced the formation of large and irreversible protein aggregates that eventually precipitated out. This study highlights the conformational variability of the NS5A domain 1 of HCV, which may be an intrinsic structural behaviour of the HCV NS5A domain 1 in solution.

INTRODUCTION

The hepatitis C virus (HCV) is a positive-sense RNA virus, belongs to the Flaviviridae family and affects nearly 2–3% of the world’s population [1]. HCV infection causes chronic hepatitis C (CHC) in 75% of infected individuals, and CHC can develop into liver steatosis, hepatocellular carcinoma or cirrhosis over the long term [2]. Until 2014, the combination of interferon-α (IFN-α) and ribavirin was the most common treatment for HCV infections, but was associated with severe side effects and poorly sustained virologic response rates [3]. However, newly discovered drugs that directly target HCV proteins, henceforth called direct-acting antivirals (DAAs), have dramatically changed the landscape of treatment regimens over the last three years [4]. These DAAs are highly efficient and are classified by their therapeutic targets, such as protease inhibitors, nucleoside polymerase inhibitors and NS5A inhibitors [5, 6]. To date, the NS5A inhibitors, DCV- and DCV-like compounds, have become essential components of IFN-free treatments [7] and they have shown promise in the elimination of HCV infections worldwide [8].

NS5A protein is a membrane-associated multi-functional phosphoprotein [9, 10] that contains three domains separated by two low-complexity sequences [11]. Due to its potential function(s) in the IFN signalling pathway, NS5A protein became popular as a therapeutic target [12]. Recently, it has attracted broad research interest because the current HCV inhibitors have been found to be targeting NS5A [13], though the mode(s) of action (MoA) of NS5A inhibitors are unclear [14]. NS5A is an active player in the replication machinery of HCV, with no known enzymatic function [15, 16]. Moreover, it has other roles in cellular processes, from innate immunity to dysregulated cell

Received 6 July 2017; Accepted 11 December 2017

Author affiliations: 1School of Biological Sciences, Nanyang Technological University, Singapore 637551, Singapore; 2School of Materials Science and Engineering, Nanyang Technological University, Singapore 639798, Singapore; 3Department of Genetic Engineering, College of Life Sciences, Kyung Hee University, Yongin-si, Gyeonggi-do, 444-701, Republic of Korea.

*Correspondence: Ho Sup Yoon, hsyoon@ntu.edu.sg

Keywords: hepatitis C virus; NS5A protein; conformational variation; direct-acting antivirals; drug resistance; daclatasvir.

Abbreviations: AH, amphipathic helix; BMH, 1,6-Bis(maleimido)hexane; CHC, chronic hepatitis C; DAAs, direct-acting antivirals; DCV, daclatasvir; DLS, dynamic light scattering; Dmax, maximum particle dimension; DSP, dithiobis(succinimidyl propionate); EC50, half maximal effective concentration; HCV, hepatitis C virus; IFN, interferon; MoA, mode of action; NS5A, nonstructural 5A; P(max), distance distribution function; PDB, protein data bank; Rg, radius of gyration; SAXS, small-angle X-ray scattering; SEC, size-exclusion chromatography; SUMO, small ubiquitin-like modifier.

Seven supplementary figures are available with the online version of this article.
protein mediates its multiple functions through interactions with viral and host cell proteins (~130 interactors) [14, 20].

The structure of the N-terminus region of NS5A has been determined, by nuclear magnetic resonance spectroscopy, to be an amphipathic helix (AH) and has been shown to be essential for genome replication in addition to membrane association [21]. X-ray crystallography of the domain 1 of NS5A lacking an AH peptide revealed that NS5A domain 1 has an ordered structure and can form dimers and tetramers. The monomer structures of NS5A domain 1 are virtually the same in all structures and contain two subdomains. However, domains 2 [22] and 3 [23] have been reported to be largely unstructured.

The first crystal structure of NS5A domain 1 from genotype 1b, which was determined more than 10 years ago [24], shows that this domain has an asymmetric dimer structure, also known as a 'clam-like dimer', that is maintained by disulfide bonds. This structure accommodates a wide groove, which has been proposed to be the RNA-binding region. The second crystal structure of NS5A domain 1 from genotype 1b exhibits an alternative dimer structure of two parallel monomers [25]. A recent study has found two alternative dimers of NS5A domain 1 from the genotype 1a of HCV that can form a tetrameric structure [26]. Over the past decade, protein structural studies of NS5A domain 1 lacking an AH peptide region have revealed different dimeric forms of the protein, and this has provided structural insights that resulted in several models of the structure of membrane/DAA/RNA-associated NS5A domain 1 dimers [27–29]. Although these crystal structures show the existence of these dimers, there is limited biochemical evidence regarding their formation in solution [30, 31]. It is also still unknown how, and in which conformation(s), NS5A protein carries out its function(s). Currently, the biological relevance of the dimerization of NS5A protein is under debate [14].

We performed a comprehensive study of NS5A domain 1 and focused on comparing its differences between genotypes 1b and 1a. By combining a variety of biophysical characterization tools, such as size-exclusion chromatography (SEC), dynamic light scattering (DLS) and cross-linking assays, in combination with small-angle X-ray scattering (SAXS), we carried out a systematic study of the behaviour of NS5A domain 1 in solution. The use of these complementary techniques provided valuable insights regarding the structural variability of the protein and also allowed us to investigate the interactions of NS5A with daclatasvir (DCV).

RESULTS

NS5A domain 1 from genotype 1b of HCV self-associated in solution and formed reversible, higher-order oligomers

We first expressed the small ubiquitin-like modifier (SUMO)-tagged NS5A domain 1 protein lacking AH, His-SUMO-NS5A (38–198) (genotype 1b, Con1). We subsequently cleaved the His-SUMO-tag with SUMO protease and obtained NS5A (36–198) from the genotype 1b of HCV (referred to as NS5A-D1-g1b) for further experiments. Our gel filtration and SDS-PAGE results indicated that the protein was mostly in a monomeric state (~19 kDa) and a highly pure form (Fig. 1a).

Previous structural studies on NS5A domain 1 of genotypes 1a and 1b showed that the protein might form different dimers [24–26]. To better understand its structural characteristics, we analysed NS5A-D1-g1b by DLS to assess the size distribution profile of the purified protein. The monomer fractions of NS5A-D1-g1b (the shaded area in Fig. 1a) were collected and concentrated to ~2.5 mg ml⁻¹ and then followed by DLS analysis. Our data revealed a bimodal distribution of two peaks with maxima of ~150 and ~2550 Å (Fig. 1b). The DLS data shown in Fig. 1(b) are based on intensity-averaged size distribution. The intensity of scattered light was proportional to the sixth power of the particle diameter [32], and the intensity of smaller particles (peak 1: ~150 Å) was more than twofold higher than that of larger particles (peak 2: ~2550 Å), indicating that the smaller particles were more populated than the larger. The smaller particles should have corresponded to monomers and dimers. However, we also noted that peak 1 was quite broad and covered a range of up to 500 Å, which might have also included higher-order structures or small aggregates.

The DLS results indicated that the purified NS5A-D1-g1b was heterogeneous and self-associated in solution. Thus, we further investigated the self-association behaviour of the protein using SEC. We pooled and concentrated the fractions corresponding to monomer sizes at varying concentrations and reloaded the same column. We observed a clear shift toward smaller elution volumes at different concentrations, ranging from 58 to 268 µM (Fig. 1c), which suggested that the monomers associated to form higher-order structures. In self-associating systems, single peaks can appear if the dissociation constant, k_d, is <1 µM (single oligomer peak) or >1 mM (single monomer peak) [33]. However, an intermediate k_d value would cause a broad tail toward smaller elution volumes due to the diffusion and dispersion of the sample peak, which was similar to what we observed (Fig. 1c). After demonstrating that NS5A-D1-g1b formed oligomers at higher concentrations, we next examined whether these oligomers were reversible. To address this question, the size corresponding to the higher-order forms of the proteins was eluted after 10.5 ml (indicated by a red box in Fig. 1c), and the volumes were collected and injected into the SEC again. The elution profile of the higher-order molecules (Fig. 1d) demonstrated that they dissociated to smaller sizes, which confirmed that higher-order structures were not irreversible aggregates. We also examined the effect of concentration on the self-association behaviour of the proteins. Proteins (~500 µg) at concentrations in the range 51–164 µM were reloaded into the same column. The intensities of the absorbance spectra of the monomer...
Fig. 1. (a) SEC column chromatogram of NS5A-D1-g1b. Monomer eluted at round 17 ml. The inset shows the purity of the eluted fraction (shaded area) checked by SDS-PAGE. (b) Size distribution analysis of purified NS5A-D1-g1b by DLS. (c) SEC analysis of NS5A-D1-g1b. (d) Higher orders were collected from Fig. 1c (red box) and reloaded onto the same column. (e) Shown are the normalized SEC elution profiles of NS5A-D1-g1b at different concentrations. (f) Purified NS5A-D1-g1b was cross-linked by BMH at different concentrations and subsequently analysed on 10% SDS-PAGE.
significantly decreased, while those of higher-order oligomers slightly increased as protein concentration increased (Fig. S1a, available in the online version of this article). To better analyze the data, we normalized all absorbance spectra by the peak maxima of the monomer region for each concentration (Fig. 1e). The normalized SEC profiles of the proteins suggested that as the concentration of the protein increased, the equilibrium shifted toward the higher-order states of the protein. The SEC experiment provided an additional insight regarding the self-association behaviour of the protein: the dimeric forms were almost undetectable, even though all the available crystal structures were dimeric.

We further examined the molecular characteristics of the protein by conducting a chemical cross-linking experiment. The fractions corresponding to the monomer size were again collected and concentrated. The fractions were then cross-linked by their cysteine residues using 1,6-bis(maleimido)hexane (BMH). BMH has an arm length of 13.0 Å; Fig. S2a shows the cysteine residues on the dimer, and distances less than 13.0 Å between two cysteines are indicated. We aimed to load the same amount of (−15 µg) cross-linked samples at different concentrations (ranging from 40 to 164 µM) on a 10 % SDS-PAGE followed by Coomassie blue staining (Fig. 1f). Chemical cross-linking of the NS5A-D1-g1b resulted in monomers, dimers and higher-order oligomers. Interestingly, the dimers were an abundant species on the chemically cross-linked proteins whereas the monomer and higher-order bands were very faint.

**Distinct oligomeric structures of NS5A-D1-g1b appeared in the form of repeating units of dimers**

To investigate the structural features of NS5A-D1-g1b in solution, we performed SAXS measurements of the protein at two different concentrations, 2.5 and 3.7 mg ml\(^{-1}\) (Fig. 2a). SAXS analysis of a protein in solution provides information about the size [radius of gyration (\(R_g\)) and the maximum particle dimension (\(D_{\text{max}}\)], low-resolution shape, conformation and assembly state. In the scattering curve, the size and shape information are at low angles (the Guinier region) and the surface-to-volume ratio is at higher angles (Porod region). The Guinier plots of NS5A-D1-g1b at low angles appeared linear and confirmed good data quality with no indication of protein aggregation (Fig. 2a, inset). The \(R_g\)-values derived from the Guinier region were determined to be 37.55±1.35 and 46.10±1.41 Å at 2.5 and 3.7 mg ml\(^{-1}\), respectively (Table 1). The extended scattering curve is converted using indirect Fourier transform to provide the distance distribution function \([P(r)]\), which is a histogram of distances between all possible pairs of atoms within a particle. The \(P(r)\) of NS5A-D1-g1b showed a single peak and had a right-skewed distribution (Fig. 2b) with a \(D_{\text{max}}\) of 127±10 and 145±10 Å at 2.5 and 3.7 mg ml\(^{-1}\), respectively, indicating that NS5A-D1-g1b has an elongated structure in solution. The \(R_g\)-values extracted from the \(P(r)\) function were 39.38±0.48 and 46.80±0.34 Å at 2.5 and 3.7 mg ml\(^{-1}\), respectively, and were in agreement with the \(R_g\)-values extracted from the Guinier region within the error range (Table 1). The scattering curve is transformed to a normalized Kratky plot [(\(qR_g^2[I(q)/I(0)]\) vs \(qR_g\)], which provides information about the potential flexible behaviour and/or extended shape of the examined particle [34]. For a well-folded globular protein such as a lysozyme, the plot will show a defined bell curve profile as seen in Fig. 2(c). By comparison, the normalized Kratky plot of NS5A-D1-g1b exhibited a very broad bell-shaped profile, and its peak was shifted well toward the right with respect to standard globular proteins, indicating the presence of motion and/or having been extended. Based on the volumes extracted from the higher angle of the scattering data, the Porod, the excluded volumes and the volume of correlation, the molecular mass of the NS5A-D1-g1b was calculated to be around 65 kDa at a concentration of 2.5 mg ml\(^{-1}\) and around 100 kDa at a concentration of 3.7 mg ml\(^{-1}\) in solution. The increase in \(R_g\) and \(D_{\text{max}}\), values, as well as the molecular mass, indicated that as the solute concentration increased the particle size increased in solution, suggesting the formation of higher-order oligomers. A theoretical scattering curve computed using CRYSO [35] for the two different crystal structure dimers from the genotype 1b, the ‘clam-like’ dimer (Protein Data Bank (PDB) ID: 1zh1, Fig. S3a) and ‘back-to-back’ dimer (PDB ID: 3fqg, Fig. S3b), had a poor fit to the experimental scattering data with a discrepancy (\(\chi^2\)) of 4 (the similarity of experimental data and the dimer model was rejected based on the correlation map test [36] for \(n=957, C=136, P=10^{-6}\) at a concentration of 2.5 mg ml\(^{-1}\). Neither the monomer nor the tetramer (generated based on symmetry-related molecules in the crystal structure dimer) fit the data \(\chi^2 = 8.7 (P<10^{-6} \text{ with } C=218)\) for monomer and 1.5 \((P<10^{-6} \text{ with } C=96)\) for tetramer for the 2.5 mg ml\(^{-1}\) concentration data). The higher-order oligomers generated using the ‘clam-like’ dimer (PDB ID: 1zh1) had a better fit to the experimental scattering data than those generated using the ‘back-to-back’ dimer (PDB ID: 3fqg). Because no single oligomer population fitted the experimental data, a mixture of oligomers was considered for further analysis of the scattering data for NS5A-D1-g1b. The results from the SEC studies also exhibited a polydisperse population in solution. The OLIGOMER program [37] was used; this program fits the experimental scattering intensities from a multi-component mixture of proteins and estimates the volume fractions of each independent scattering component in the solution. The analysis revealed that NS5A-D1-g1b had a population mixture composed of monomers, hexamers and decamers, with a volume fraction of 61:18:21 at 2.5 mg ml\(^{-1}\) of protein and 41:17:42 at 3.7 mg ml\(^{-1}\) of protein. It also agreed with the experimental scattering data (\(\chi^2\) value of 0.5; Fig. 2d). A correlation map assessment test [36] revealed that the similarity of experimental data and the oligomer population mixture model could not be rejected \((n=571, C=11, P=0.2411 \text{ for } 2.5 \text{ mg ml}^{-1} \text{ and } C=12, P=0.1283 \text{ for } 3.7 \text{ mg ml}^{-1})\). The OLIGOMER program [37] analysis also suggested that the higher-order oligomer ratio increased as the protein concentration increased, which agreed with the SEC data discussed above. Two possible explanations can be derived from these
findings. First, the larger particles observed in the DLS and SEC analyses were indeed higher-order structures rather than irregular aggregates. These higher-order structures were formed by dimeric subunits of NS5A-D1-g1b. Second, the dimeric form was virtually absent in the free form in solution, suggesting that the NS5A-D1-g1b monomers formed dimers. These dimers did not remain free in solution for very long and were not converted to higher-order structures. One possible reason for the abundance of dimers in the BMH cross-linking assay is due to the looseness of the higher-order structures. Another explanation is that BMH, the cross-linking agent we used, might be failing to covalently link the higher-order structures although it was successful in doing so for dimers. Therefore, it is thought that the higher-order forms are unstable and dissociated during the crosslinking experiments while their subunits, dimers, became highly stable through successful cross-linking (Fig. 1f). We have employed other cross-linking...

Fig. 2. Solution X-ray scattering studies of NS5A-D1-g1b. (a) SAXS pattern (○) of NS5A-D1-g1b at concentrations of 2.5 mg ml⁻¹ (green) and 3.7 mg ml⁻¹ (forest green). Inset: the Guinier plots show linearity at all concentrations used, indicating no aggregation. The scattering profiles were offset for clarity by applying arbitrary scale factors. (b) Overlapping of the pair-distance distribution function P(r) at concentrations of 2.5 mg ml⁻¹ (—; green) and 3.7 mg ml⁻¹ (—; forest green) show differences in the profile. (c) The normalized Kratky plot of NS5A-D1-g1b at concentrations of 2.5 mg ml⁻¹ (●; green) and 3.7 mg ml⁻¹ (●; forest green) was compared to the compact globular lysozyme (●; grey) with a peak (—-; grey), representing the theoretical peak and assuming an ideal Guinier region of a globular particle. The scattering pattern of NS5A-D1-g1b exhibited a broad, bell-shaped profile shifted toward the right with respect to the standard globular proteins, indicating the presence of motion and/or an extended nature of the protein. (d) Theoretical scattering curve for the mixture of monomer-M, hexamer-H and decamer-D of NS5A-D1-g1b (—; red) was calculated using the OLIGOMER program and was fitted to the experimental scattering pattern at concentrations of 2.5 mg ml⁻¹ (○; green) and 3.7 mg ml⁻¹ (○; forest green), with a χ²-value of 0.5.
agents, DSP [dithiobis(succinimidyl propionate)], a lysine cross-linker, and glutaraldehyde. However, these agents also failed to cross-link the higher-order forms (Fig. S4a, b). Taken together, our results suggest that the higher-order oligomers are highly dynamic and transient in nature.

**Self-association behaviour of NS5A domain 1 from genotype 1a of HCV was different from that from genotype 1b**

We characterized NS5A domain 1 from genotype 1a in the same way as NS5A-D1-g1b by using SEC, DLS, chemical cross-linking and SAXS methods. NS5A-D1-g1a was purified using the same method as NS5A-D1-g1b. The SEC elution profile and its corresponding SDS-PAGE analysis for NS5A-D1-g1a are shown in Fig. 3(a). The fractions corresponding to the monomer state of the protein were collected and concentrated to 2.5 mg ml\(^{-1}\) for DLS analysis in SEC buffer. DLS measurements of purified NS5A-D1-g1a revealed a bimodal size distribution with peak maxima at around 100 and 1650 Å, as shown in Fig. 3(b). Although the particle sizes of NS5A-D1-g1b and NS5A-D1-g1a were not same, the smaller particles again dominated the NS5A-D1-g1a samples. To further investigate the self-association behaviour of NS5A-D1-g1a, we reloaded the monomeric protein into the SEC after collecting and concentrating it to different concentrations: 52, 109 and 176 µM. In this experiment, we again aimed to load the same amount of protein, \(~ 500 \mu g\). Fig. 3(c) shows the normalized SEC profiles of the protein at different concentrations (see Fig. S5a for the representation of raw data.).

### Table 1. Data collection and X-ray scattering-derived parameters for HCV NS5A domain1 from genotypes 1b and 1a

<table>
<thead>
<tr>
<th>Data collection parameters</th>
<th>NS5A-D1-g1b</th>
<th>NS5A-D1-g1a</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Instrument (source and detector)</strong></td>
<td>Bruker NanoStar equipped with MetalJet eXcillum X-ray source and VÀNTEC-2000 detector</td>
<td></td>
</tr>
<tr>
<td>Beam geometry</td>
<td>100 µm slit</td>
<td></td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.34</td>
<td></td>
</tr>
<tr>
<td>(q) range (Å(^{-1}))</td>
<td>0.016–0.4</td>
<td></td>
</tr>
<tr>
<td>Exposure time (min)</td>
<td>10 (10 frames)</td>
<td></td>
</tr>
<tr>
<td>Temperature (K)</td>
<td>288.15</td>
<td></td>
</tr>
<tr>
<td><strong>Protein sample</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Concentration (mg ml(^{-1}))</td>
<td>2.5</td>
<td>3.7</td>
</tr>
<tr>
<td><strong>Structural parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(I(0)) (arbitrary units) [from (P(r))]</td>
<td>18.98±0.26</td>
<td>27.20±0.25</td>
</tr>
<tr>
<td>(R_g) (Å) [from (P(r))]</td>
<td>39.38±0.48</td>
<td>46.80±0.34</td>
</tr>
<tr>
<td>(I(0)) (arbitrary units) (from Guinier)</td>
<td>18.96±0.49</td>
<td>27.76±0.66</td>
</tr>
<tr>
<td>(R_g) (Å) (from Guinier)</td>
<td>37.55±1.35</td>
<td>46.10±1.41</td>
</tr>
<tr>
<td>(D_{max}) (Å)</td>
<td>127±10</td>
<td>145±10</td>
</tr>
<tr>
<td>Porod volume estimate ((V_p)) (Å(^3))</td>
<td>–102931</td>
<td>–157357</td>
</tr>
<tr>
<td>DAMMIF excluded volume ((V_{ex})) (Å(^3))</td>
<td>–137400</td>
<td>–208100</td>
</tr>
<tr>
<td>Dry volume from sequence (Å(^3))†</td>
<td>–23003</td>
<td>–23003</td>
</tr>
<tr>
<td><strong>Molecular mass determination</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calculated monomeric MM (kDa) (from sequence*)</td>
<td>–19</td>
<td>–19</td>
</tr>
<tr>
<td>MM from Porod invariant ((V_p/1.6)) (kDa)</td>
<td>64±6</td>
<td>98±10</td>
</tr>
<tr>
<td>MM from excluded volume ((V_{ex}/2)) (kDa)</td>
<td>68±7</td>
<td>104±10</td>
</tr>
<tr>
<td>MM from volume of correlation ((V_c)) (kDa)</td>
<td>60±6</td>
<td>97±10</td>
</tr>
<tr>
<td><strong>Oligomer Mixture</strong></td>
<td>Mixure‡ (61 : 18 : 21)</td>
<td>Mixure‡ (41 : 17 : 42)</td>
</tr>
<tr>
<td><strong>Software used</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary data reduction</td>
<td>BRUKER SAS</td>
<td></td>
</tr>
<tr>
<td>Data processing</td>
<td>PRIMUS</td>
<td></td>
</tr>
<tr>
<td>(Ab) initio analysis</td>
<td>DAMMIF</td>
<td></td>
</tr>
<tr>
<td>Validation and averaging</td>
<td>DAMAVER</td>
<td></td>
</tr>
<tr>
<td>Computation of model intensities</td>
<td>CRYSOL</td>
<td></td>
</tr>
<tr>
<td>Computation of volume fractions of mixtures</td>
<td>OLIGOMER</td>
<td></td>
</tr>
<tr>
<td>3D graphics representations</td>
<td>PyMOL</td>
<td></td>
</tr>
</tbody>
</table>

†http://www.basic.northwestern.edu/biotools/proteincalc.html.
‡Monomer : hexamer : decamer.
suggested that the exchange rates between different conformations were significantly different for NS5A-D1-g1a and NS5A-D1-g1b. While higher-order structures were clearly observed in the SEC profile of NS5A-D1-g1b, such forms were almost undetectable for NS5A-D1-g1a. The overlapping SEC profiles shown in Fig. S5b highlight these differences between NS5A-D1-g1b and NS5A-D1-g1a. This crucial finding was surprising because the protein sequences were highly similar, with a similarity ratio of around 91% (Fig. S6).

We also performed BMH cross-linking experiments for NS5A-D1-g1a, as for NS5A-D1-g1b (Fig. 1f), which showed that the ratio of bands (monomer/dimer) was higher than that observed with cross-linking SDS-PAGE for NS5A-D1-g1b (Fig. 3d).

**Low-resolution solution structure of NS5A-D1-g1a revealed a dimeric state**

SAXS studies of the NS5A-D1-g1a of HCV were performed at two concentrations, 2.5 and 3.0 mg ml\(^{-1}\) (Fig. 4a). The SAXS patterns were of good quality and the Guinier plots at low angles appeared linear (Fig. 4a, inset), with derived \(R_g\) values of 29.51±1.11 and 29.55±0.98 Å for concentrations of 2.5 and 3.0 mg ml\(^{-1}\), respectively (Table 1). The \(P(r)\) function was similar for both concentrations and showed a single peak that was right-skewed (Fig. 4b), with a \(D_{\text{max}}\) of 106±6 Å (Table 1), demonstrating that NS5A-D1-g1a was elongated. The normalized Kratky plot of NS5A-D1-g1a had a broad, bell-shaped profile, demonstrating that it was extended and less rigid in solution (Fig. 4c). The calculated monomeric molecular mass from sequence is about 19 kDa. The molecular mass was determined based on the Porod
Fig. 4. Solution X-ray scattering studies of NS5A-D1-g1a. (a) SAXS pattern (○) of NS5A-D1-g1a at concentrations of 2.5 mg ml$^{-1}$ (orange) and 3.0 mg ml$^{-1}$ (red). Inset: the Guinier plots show linearity at all concentrations used, indicating no aggregation. The scattering profiles were offset for clarity by applying arbitrary scale factors. (b) Overlapping of the pair-distance distribution function $P(r)$ at concentrations of 2.5 mg ml$^{-1}$ (---; orange) and 3.0 mg ml$^{-1}$ (---; red) showed a similar profile. (c) Normalized Kratky plot of NS5A-D1-g1a at a concentration of 3.0 mg ml$^{-1}$ (●; red) compared to the compact globular lysozyme (●; grey) with a peak (---; grey), representing the theoretical peak and assuming an ideal Guinier region of a globular particle. The scattering pattern of NS5A-D1-g1a exhibited a broad, bell-shaped profile shifted toward the right with respect to standard globular proteins, indicating the presence of motion and/or an extended nature of the protein. (d) Theoretical scattering curve fit (---) of dimer 1 (cyan) and dimer 2 (blue) to the experimental scattering pattern (○) of protein concentrations of 2.5 mg ml$^{-1}$ (orange) and 3.0 mg ml$^{-1}$ (red). Dimer 2 fitted better to the experimental data, with $\chi^2$-values of 0.46 and 0.55 for concentrations of 2.5 mg ml$^{-1}$ (orange) and 3.0 mg ml$^{-1}$ (red), respectively. (e) Averaged and
and excluded volumes, and volume of correlation confirmed the existence of NS5A-D1-g1a as a dimer in solution at the concentration used (Table 1).

The crystal structure of NS5A domain 1 from genotype 1a (PDB ID: 4c11) contained four molecules in the asymmetric unit, forming two types of dimer. In ‘dimer 1’ the monomers were arranged in an anti-parallel head-to-tail configuration, and in ‘dimer 2’ they formed an N-terminal head-to-head configuration (Fig. S3c and d). A theoretical scattering curve computed using CRYSTAL [35] for both dimers had a fit to the experimental scattering curve of \( \chi^2 = 0.82 \) for dimer 1 and 0.46 for dimer 2 for data at a concentration of 2.5 mg ml\(^{-1}\) (\( \chi^2 = 1.13 \) for dimer 1 and 0.55 for dimer 2 at 3.0 mg ml\(^{-1}\); Fig. 4d). The correlation map assessment test [36] of the rigid model fit suggests that the experimental data and dimer 2 are similar (\( n = 957 \), \( C = 15 \), \( P = 0.1159 \) for 2.5 mg ml\(^{-1}\) and \( C = 19 \), \( P = 0.0545 \) for 3.0 mg ml\(^{-1}\)). However, for dimer 1 the hypothesis being similar is rejected with \( n = 957 \), \( C = 40 \), \( P < 10^{-6} \) for 2.5 mg ml\(^{-1}\) and \( C = 85 \), \( P < 10^{-6} \) for 3.0 mg ml\(^{-1}\). The experimental data were in good agreement with the dimer 2 curve, suggesting the presence of an N-terminal head-to-head dimers in solution. The low-resolution shapes of NS5A-D1-g1a were reconstructed \textit{ab initio} using DAMMIF [38]. The final averaged and filtered model with a normalized spatial discrepancy of 0.56±0.04 is shown in Fig. 4e. The NS5A-D1-g1a solution model showed an elongated conformation and superimposed well onto the crystallographic structure of dimer 2 (Fig. 4e). The unoccupied density in the bead model around one of the monomers (represented by the arrow in Fig. 4e) suggested that the monomers had a highly dynamic behaviour and were in rapid motion in solution due to the lower interface area (311 Å\(^2\)) of the dimer. SAXS analysis of NS5A-D1-g1a confirmed the presence of an N-terminal head-to-head dimer structure in solution.

**Effect of the DCV on the oligomeric state of NS5A-D1**

Neither the full-length NS5A protein nor NS5A domain 1 have an enzymatic function that can be tested \textit{in vitro} [39, 40]. However, domain 1 is known to be targeted by NS5A inhibitors [13, 31]. Bristol-Myers Squibb Company has identified the drug daclatasvir (BMS-790052, DCV) for the treatment of HCV replicons [38]. The final averaged and filtered model with a normalized spatial discrepancy of 0.56±0.04 is shown in Fig. 4e. The NS5A-D1-g1a solution model showed an elongated conformation and superimposed well onto the crystallographic structure of dimer 2 (Fig. 4e). The unoccupied density in the bead model around one of the monomers (represented by the arrow in Fig. 4e) suggested that the monomers had a highly dynamic behaviour and were in rapid motion in solution due to the lower interface area (311 Å\(^2\)) of the dimer. SAXS analysis of NS5A-D1-g1a confirmed the presence of an N-terminal head-to-head dimer structure in solution.

The NS5A-D1-g1a solution model showed an elongated conformation and superimposed well onto the crystallographic structure of dimer 2 (Fig. 4e). The unoccupied density in the bead model around one of the monomers (represented by the arrow in Fig. 4e) suggested that the monomers had a highly dynamic behaviour and were in rapid motion in solution due to the lower interface area (311 Å\(^2\)) of the dimer. SAXS analysis of NS5A-D1-g1a confirmed the presence of an N-terminal head-to-head dimer structure in solution.

The red arrow indicates unoccupied density in the solution shape.
and 1b, respectively [41, 42]. It is worthwhile to note that DCV still maintains its efficacy against Y93H replicons at higher dosages, e.g. 32 and ~0.05 nM for g1a and g1b, respectively. Similarly, Berger et al. recently showed that Y93H resistance mutation exhibited a slight reduction (~30%) in the binding of NS5A to the biotinylated derivative of DCV in Huh7 cells [46]. Thus, previous studies collectively indicate that DCV also affects mutated NS5A. So, to observe the significance of the resistance effect, multiple point mutations might be required. Moreover, we should mention that as we have studied the domain I of NS5A lacking AH and post-translational modifications, our observations may not fully reflect the behaviour of physiologically relevant NS5A. Furthermore, given that DCV affects both wild and resistant types, it is perhaps a challenge to observe the difference quantitatively with the biophysical characterizations e.g. DLS and SAXS.

**DISCUSSION**

Over the past decade, crystallization studies of NS5A domain I from genotypes 1a and 1b of HCV have resulted in different conformations of homodimers and revealed an identical monomer fold [24–26]. However, much remains unknown about how NS5A protein exists in solution, how it carries out its multiple functions and which conformation (s) of NS5A is/are targeted by NS5A inhibitors.

We aimed to investigate the behaviour of NS5A domain I in solution. We demonstrated that NS5A-D1-g1b had an intrinsic behaviour of oligomerization in solution and it existed in an equilibrium of monomer, transient dimer and higher-order structures. Our SAXS study confirmed that the larger particles observed in DLS and SEC analyses were elongated and ordered structures rather than irregular aggregates. Despite the fact that there was some inconsistency, our analyses agreed with previous work [25] that demonstrated a model of the multimeric state of NS5A domain I (genotype 1b). This theoretically assembled model contained both dimer conformations (PDB ID: 1zh1 and 3fqq) and formed a superhelical array. Our solution scattering data of NS5A-D1-g1b suggested a mixture of population for monomer, hexamer and decamer as illustrated in Fig. 7(a). Since the dimer population is not present, it was not possible to extract the dimer density from SAXS data. However, we generated the higher-order oligomers from symmetry-related molecules from dimers (PDB ID: 1zh1 and 3fqq). Since the higher-order oligomers of 'clam-like' dimer (1zh1) had a better fit to the SAXS data than the other 'back-to-back' dimer (PDB ID: 3fqq), our SAXS model suggested that the higher-order structures of NS5A-D1-g1b were formed of dimers of the crystal structure, 1zh1, and also had an elongated helical shape. Based on our findings, we propose that the monomers of NS5A-D1-g1b formed dimers in solution and that these dimers were transient and turned into higher-order structures.

We also showed that NS5A-D1-g1a existed as a heterogeneous mixture. However, using DLS and SEC, we observed that the size distribution of NS5A-D1-g1a was significantly different from that of NS5A-D1-g1b. The monomers of NS5A-D1-g1a remained in a different equilibrium of smaller molecules, as shown in Fig. 3c. In the case of NS5A-D1-g1b, the monomers can form higher-order structures (eluted at 12–14 ml), as seen in Fig. 1e. Our SAXS model of the NS5A-D1-g1a confirmed the existence of the N-terminal head-to-head dimeric conformation of the crystal structure, 4cl1, which we named dimer 2 (Fig. 4e). In our experiments, we did not observe any higher-order structures for the NS5A-D1-g1a protein in solution. Changing the buffer conditions or increasing the concentration of the protein might reveal some tetramers of NS5A-D1-g1a. However, under the same conditions (purification protocol, buffers, concentration, etc.), the proteins NS5A-D1-g1b and NS5A-D1-g1a exhibited different conformations as illustrated in Fig. 7(a, b).
Current knowledge regarding the MoA of NS5A inhibitors points to disruption of the membranous replication complex [46]. However, at the molecular level, our understanding remains limited. After the multimer model of NS5A had been proposed [25], the model became very popular in explaining the extraordinary MoA of the NS5A inhibitors. A recent study estimated that there was one DCV molecule for every 47,000 NS5A proteins in cells, and it suggested that a small molecule could have a profound effect on a large number of NS5A molecules [47]. By utilizing the multimeric model of NS5A, an allosteric regulation model for the MoA of the DCV was proposed in which a single DCV bound to NS5A could disturb communication between the NS5A molecules, inhibiting the formation of the replication complex. As we observed the self-association of the NS5A domain 1, we tested the effect of DCV on the self-association behaviour of the protein. However, our analyses from both DLS and SAXS studies revealed the formation of irreversible protein aggregates in response to the DCV for both wild- and DCV-resistant type protein with a 1:1 ratio of drug to protein concentrations. Moreover, given that DCV causes aggregation of both types, it is perhaps a challenge to observe the difference quantitatively with biophysical characterizations, such as DLS and SAXS. Therefore, the formation of irreversible protein aggregates should be interpreted with caution, and it is plausible that a number of limitations might have influenced our results. First, given that our findings are based on biophysical observations on the isolated part of NS5A, the effect of DCV on the protein may not completely reflect the physiologically relevant MoA of the drug. Thus, it cannot be excluded that irreversible aggregation of the protein could be an artefact of the high drug ratio, as it does not represent in vivo conditions. The multimers of NS5A in cell culture were detected by fluorescent

Fig. 6. Particle size distribution of (a) NS5A-D1-g1a, (b) Y93H NS5A-D1-g1b and (c) Y93H NS5A-D1-g1a were measured in the presence and absence of DCV. (d) Overlaid DLS profiles as indicated.
tags covalently attached to NS5A, and the formation of larger foci of NS5A in the presence of NS5A inhibitors has been reported [48]. It has also been reported that aggregate formation of NS5A is concurrently observed with replicase [49]. These earlier studies prompted us to test DCV using isolated NS5A domain 1 with conformational versatility, which could provide insightful information into the inhibitory mechanism of direct-acting antivirals.

In conclusion, this study provided insightful information about the intrinsic structural dynamics of HCV NS5A domain 1 in solution. Our data suggest the significance of the contributions of protein sequences to the alternative conformational states we observed in NS5A-D1-g1a and NS5A-D1-g1b. The dynamic behaviours and conformational variability could be considered as key features of how proteins regulate their biological processes [50]. Adaptability of the protein structure and its conformational motions have generated much interest because of their critical roles in determining molecular functions in response to external stimuli [51]. Conformational variations of NS5A might add another layer of structural dynamics and contribute to its functional versatility.

**METHODS**

**Cloning, expression and purification of recombinant proteins**

Two different NS5A domain 1 constructs were used in this study. They lacked Ah and comprised residues 36 to 198. One was from the H77 strain of HCV (genotype 1a) and another from the isolate Con1 (genotype 1b). Primers were designed to encode (Asp)₄-Lys residues, the enterokinase recognition sequence, which was reported to enhance the solubility of the NS5A domain 1 [24, 25]. The amplified cDNA fragments encoding NS5A domain 1 were cloned into pETSUMO vector (LifeSensors, USA). The resulting recombinant plasmids were transformed into *E. coli* Rosetta (DE3) cells (Novagen, USA). Cells were grown at 37 °C to an absorbance at 600 nm of ~0.6–0.7. The culture was cooled at 4 °C for 30 min before IPTG induction. Protein expression was induced with 1 mM IPTG for 14 h at 18 °C. The cells were harvested, suspended in a lysis buffer (50 mM Tris-HCl, pH 8.0, 300 mM NaCl and 10% glycerol supplemented with 1 mM PMSF and benzonase; Sigma-Aldrich, Singapore) and lysed by sonication. The clarified cell lysate was obtained by centrifugation for 30 min at 30,000 g at 4 °C. The protein was purified by Ni²⁺-NTA column chromatography. The column was pre-equilibrated with 10 columns of washing buffer (50 mM Tris-HCl, pH 8.0, 300 mM NaCl, 40 mM imidazole). The protein was eluted with elution buffer (50 mM Tris-HCl, pH 8.0, 300 mM NaCl, 10% glycerol, 500 mM imidazole). The collected fractions were concentrated, buffer-exchanged and cleaved in the presence of SUMO protease (Sigma-Aldrich, Singapore) at 4 °C overnight. The SUMO tag-cleaved proteins were loaded onto a second Ni²⁺-NTA affinity column to remove SUMO tag and SUMO protease. The eluted protein was further purified on a Superdex 200 Increase 10/300 GL column (GE Healthcare Life Sciences, Singapore).
BMH chemical cross-linking experiment

Varying concentrations of NS5A-D1-g1b and NS5A-D1-g1a recombinant proteins were cross-linked at cysteine residues by incubating with 1 mM BMH (Pierce) for 1 h at room temperature. The cross-linking reaction was quenched by 4 mM DTT. The cross-linked samples were analysed on 10% SDS-PAGE.

Dynamic light scattering

The average size and polydispersity of the recombinant proteins were analysed using a Zetasizer Nano ZS instrument (Malvern Instruments, UK). Each measurement was performed by modifying the protocol given in the literature [32], and all measurements were performed at 25°C. Approximately 50 µl of the protein samples were collected in a low-volume quartz batch cuvette (ZEN2112). The equilibration time was adjusted to 120 s in all measurements. At least 20 measurements with 15 s accumulations were recorded. The intensity-averaged hydrodynamic diameter results of a minimum of 15 data points were combined using the instrument software, DTS Version 6.7. The results are presented as diameter distributions. Daclatasvir dihydrochloride (DCV) was purchased from MedChem Express (USA).

Small-angle X-ray scattering

The SAXS data for HCV NS5A domain1 from genotypes 1b and 1a were collected with a BRUKER NANOSTAR SAXS instrument equipped with a Metal-Jet X-ray source (Excillum, Germany) and VANTEC 2000-system detector [52–54]. The scattering patterns were measured using a sample at a detector distance of 0.67 m and a wavelength of \( \lambda = 1.3414 \, \text{Å} \), which covered a range of momentum transfer of \( 0.016 < q < 0.4 \, \text{Å}^{-1} \) (\( q = 4 \pi \sin(\theta) / \lambda \), where \( \theta \) is the scattering angle). The SAXS experiments were carried out at 15°C at two different protein concentrations in 25 mM HEPES, pH 7.4, 150 mM NaCl, 10% glycerol) with a sample volume of 40 µl in a vacuum-tight quartz capillary. To monitor for radiation damage, 10 exposures (1 min) were collected for each protein sample; no radiation effect was observed. The data were normalized to the intensity of the transmitted beam. The scattering of the buffer was subtracted, and the data were normalized to the intensity of the transmitted beam. The scattering pattern was measured using the instrument software, DTS Version 6.7. The results are presented as diameter distributions. Daclatasvir dihydrochloride (DCV) was purchased from MedChem Express (USA).

All data processing steps were performed using program package PRIMUS [55] from the ATSAS package version 2.7.1. The experimental data obtained for all protein samples were analysed for aggregation using the Guinier approximation. These parameters were also computed from the extended scattering patterns using the indirect transform package GNOM [56], which provided the distance distribution function, \( P(r) \), the maximum particle dimension, \( D_{\text{max}} \) and \( R_g \). Qualitative particle motion was inferred by plotting the scattering patterns in the normalized Kratky plot \( [(qR_g)^2 (I(q)/I(0)) vs qR_g] \) [34]. Ab initio low-resolution models of the proteins were built by the program DAMMIF [38], which considers low-angle data (\( q < 2 \, \text{nm}^{-1} \)). Twenty independent ab initio reconstructions were performed for each protein and then averaged using DAMAVER [57]. The averaged and filtered ab initio model was superimposed with the atomic model using SUPCOMB [58]. Theoretical scattering curves from the atomic structures were generated and evaluated against the experimental scattering curves using CRYSO [35]. Utilizing the intensities from each component, the volume fraction of each component can be determined by OLIQOMER [37]. The oligomeric state of the protein was confirmed from the molecular mass calculation based on \( I(0) \), Porod volume \( (V_p) \), excluded volume \( (V_e) \) and the volume of correlation \( (V_c) \) [59, 60]. The goodness-of-fit test for the reduced \( \chi^2 \) was calculated using the correlation map (CorMap) [36] as implemented in ATSAS package that computes the probability \( (P) \) of similarity between data and model. The program reports the probability of obtaining an edge length larger than \( C \) within an \( n \)-by-\( n \) correlation matrix and if the \( P \)-value is less than a predetermined significance level (0.01), the observed difference is statistically significant [36].

Funding information

SB was supported Singapore International Graduate Award (SINGA) sponsored by NTU. This research was in part supported by Singapore Cancer Syndicate grant ZU44 to HSY, and Ministry of Education Singapore MOE Tier 3 grant (MOE2012-T3-1-008) to GG.

Conflicts of interest

The authors declare that there are no conflicts of interest.

References


---

**Five reasons to publish your next article with a Microbiology Society journal**

1. The Microbiology Society is a not-for-profit organization.
2. We offer fast and rigorous peer review – average time to first decision is 4–6 weeks.
3. Our journals have a global readership with subscriptions held in research institutions around the world.
4. 80% of our authors rate our submission process as ‘excellent’ or ‘very good’.
5. Your article will be published on an interactive journal platform with advanced metrics.

Find out more and submit your article at microbiologyresearch.org.