Using *tert*-butyl groups in a ligand to identify its binding site on a protein

Wan-Na Chen[§,‡], and Gottfried Otting[§,*]

[§]Research School of Chemistry, Australian National University, Canberra, ACT 2601, Australia

[‡]College of Pharmacy, Jinan University, Guangzhou, Guangdong 510632, China

KEYWORDS Boc group; nuclear Overhauser effects; ligand binding site

ABSTRACT: Few methods allow determining the binding site of tightly binding ligands. We show that ligands containing a *tert*-butyl (e.g., Boc) group produce easily observable nuclear Overhauser effects (NOE) with the target protein even when the *tert*-butyl group is not highly solvent exposed. NOEs with methyl groups of the target protein are readily assigned by selectively isotope labeling, presenting a practical and quick way to pinpoint the location of the ligand without any prior specific NMR assignments of the protein. The approach works for non-exchanging ligands as well as for weakly binding ligands.

Nuclear magnetic resonance (NMR) spectroscopy is a widely used tool for drug discovery in the pharmaceutical industries.[1,2] A frequent scenario is that the crystal structure of the target protein is available but not in complex with the ligand of interest. In this situation, further development of any ligand hinges on determining its binding site on the target and establishing its specificity of binding. In principle, NMR spectroscopy is uniquely positioned to yield structural information on the protein–ligand interaction at atomic resolution, but the sensitivity of chemical shifts with regard to changing chemical environments means that NMR resonance assignments of the protein change in response to different ligands, compromising throughput. Here we present a new strategy, which harnesses intermolecular NOEs with *tert*-butyl groups in the ligand molecule to determine its binding site on the protein. It is applicable also to protein–ligand complexes with tightly bound ligands. Our strategy relies on the crystal structure of the free protein but requires no prior resonance assignments of the protein NMR spectrum.

Tert-butyl groups produce intense signals in the ^{1}H-NMR spectrum that can readily be detected against the background from the numerous NMR signals of a protein. The signals are narrow because splittings due to J_{HH} couplings remain unresolved and the effective correlation times are short due to rapid rotation of the methyl groups and of the *tert*-butyl group itself. The signal of a *tert*-butyl group is also strong, as it corresponds to nine equivalent protons, greatly facilitating its detection. We have previously shown that site-specific incorporation of *O*-*tert*-butyl-tyrosine (Tby) into proteins allows the observation of the *tert*-butyl group in molecular systems as large as 320 kDa without dilution of the proton spin density by perdeuteration, provided the *tert*-butyl group is highly solvent exposed to allow rapid reorientation of the C–H bonds of the *tert*-butyl group.[3] In the case of Tby, the reorientation is promoted further by rotation about the C–O bonds linking the *tert*-butyl group to the rest of the amino acid. Broader ^{1}H-NMR signals are observed, however, when the *tert*-butyl group is only partially solvent exposed. For example, the ^{1}H-NMR signal of the *tert*-butyl group of an inhibitor bound to the 27 kDa dengue virus NS2B-NS3 protease were more difficult to identify in 1D ^{1}H-NMR spectra of paramagnetically labeled samples, ultimately requiring NOE cross-peaks with the *tert*-butyl group for their unambiguous identification.[4] NOEs also proved necessary for spectral identification of highly solvent exposed *tert*-butyl and trimethylsilyl groups in 95 kDa complexes between single-stranded DNA binding protein (SSB) and single-stranded DNA.[5]

The present work demonstrates the feasibility of detecting intermolecular protein-ligand NOEs, when the ligand contains *tert*-butyl groups located in binding pockets of the protein. These NOEs allow analysis of the binding site of the ligand without any prior assignments of the NMR resonances of the protein. While there are many NMR methods for studying protein-ligand complexes of weakly binding ligands, it remains difficult to identify binding sites on the target protein, if the bound ligand does not exchange with its free form on the time scale of the experiment.[6] The present strategy opens a way for binding site determination that is applicable to tightly binding ligands.

The concept is illustrated using the complex between the hepatitis C virus (HCV) inhibitor asunaprevir (ASV; also referred to as BMS-650032) and the full-length NS3 protein (69 kDa) and its protease domain (NS3pro; 23 kDa), each in a construct with an 11-residue polypeptide segment from the co-factor NS4A. ASV in combination with daclatasvir has been shown to be effective against HCV genotype 1b infection in phase 3 clinical trials.[7] It is a tightly binding ligand, inhibiting the proteolytic activity of HCV NS3 with IC_{50} values in the low-nanomolar range for a wide range of HCV genotypes.[8] It contains two *tert*-butyl groups, one of them as part of a Boc group (Figure 1A).

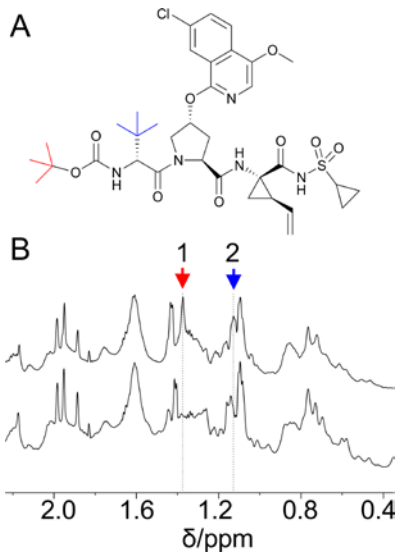


Figure 1. ^{1}H-NMR signals of the *tert*-butyl groups of the inhibitor ASV in complex with HCV NS4A-NS3pro. (A) Chemical structure of the inhibitor ASV. The *tert*-butyl groups 1 and 2 of the inhibitor are highlighted in red and blue, respectively. (B) 800 MHz 1D ^{1}H-NMR spectra of 50 µM solutions of HCV NS4A-NS3pro in NMR buffer (20 mM sodium phosphate, pH 6.5, 150 mM NaCl, 5 mM DTT, 10% D_2O) at 25 °C. The spectra were recorded in the presence of an equimolar amount of inhibitor (top spectrum) or without inhibitor (bottom spectrum). The red and blue arrows identify the chemical shifts of the *tert*-butyl groups 1 and 2 of the inhibitor.

The non-structural protein NS3 of HCV encodes a helicase preceded by a serine protease, NS3pro, that functions together with the N-terminal part of NS4 (NS4A) as cofactor.[9] Crystal structures have been solved for the complex between ASV and NS4A-NS3pro,[10,11] and for NS4A-NS3 in complex with different inhibitors.[12,13] Structural information for ASV bound to full-length NS3 is still lacking. The crystal structure of NS4A-NS3pro shows that ASV binds in a pocket of the protease facing the helicase domain. The *tert*-butyl group of the Boc group is about 60% solvent exposed, whereas the other *tert*-butyl group is more buried. The ^{1}H nuclei of the two *tert*-butyl groups of ASV are as close as 5.5 Å of each other and within 5 Å of methyl groups of the protein. NMR spectroscopy of NS4A-NS3pro and full-length NS4A-NS3 is made difficult by poor stability of the protein samples against precipitation. In the past, stability was enhanced by the addition of 4% glycerol and 0.1 - 1.5 mM CHAPS.[14,15] To demonstrate the performance of our method in difficult cases, we measured in aqueous solution without co-solvents.

The free inhibitor is only poorly soluble in water. In DMSO, the chemical shifts of the *tert*-butyl groups were reported to be 1.10 ppm and about 0.96 ppm.[10] In the complex with NS4A-NS3pro, the ^{1}H-NMR signal of the Boc group (*tert*-butyl group 1, at about 1.35 ppm) is readily detected by comparison with the 1D ^{1}H-NMR spectrum of the free protein (Figure 1). The second *tert*-butyl group (*tert*-butyl group 2, at about 1.1 ppm) is obscured by overlap with similarly intense signals of the protein (Figure 1B; ultimately, the specific assignments were confirmed by NOEs, see below). The chemical shifts of the *tert*-butyl groups became clear in a NOESY spectrum, where the cross-peaks with the *tert*-butyl groups stood out for their narrow line shape, making them taller than most of the intra-protein cross-peaks. A notably intense cross-peak was observed between the resonances of the *tert*-butyl groups (Figure 2A). The NOEs with methyl groups of the protein yielded the specific assignment of the two *tert*-butyl groups and identification of the ligand binding site on the protein.

Figure 2. NOESY spectra of the complex between HCV NS4A-NS3pro (150 µM) and ASV. The spectra were recorded at 25 °C, using a mixing time of 200 ms. The chemical shifts of the *tert*-butyl groups 1 and 2 are identified by dotted red and blue lines, respectively. NOEs with valine and isoleucine residues and an amide proton are identified. Cross-peaks between the *tert*-butyl groups appear at 1.11 and 1.37 ppm. (A) NOESY spectrum of the complex with unlabeled NS4A-NS3pro. (B) NOESY spectrum recorded of a complex, where NS4A-NS3pro was prepared with perdeuterated valine and uniformly ^{13}C-labeled isoleucine. Comparison with the spectrum in (A) confirms that the *tert*-butyl group 1 makes NOEs with a valine residue. (C) Same as (B), except without ^{13}C-decoupling in the δ_1 dimension. Comparison with the spectrum in (B) confirms

that the *tert*-butyl group 2 makes NOEs with an isoleucine residue.

Initially, no resonance assignments were available for NS4A-NS3pro or ASV in the complex. Clearly, however, some of the NOESY cross-peaks displayed by the *tert*-butyl groups in the methyl region must arise from intermolecular NOEs with the protein, as ASV contains only a single methyl group beyond the *tert*-butyl groups (Figure 1A). NOEs with methyl groups in the spectral range between 0 and 1 ppm are very likely with isoleucine, leucine, and valine. To identify the residue types, we prepared a protein sample selectively labeled with deuterated valine and ^{13}C-labeled isoleucine. In the NOESY spectrum recorded of this sample, three cross-peaks disappeared which were thus assigned to valine (Figure 2B). Furthermore, two cross-peaks became undetectable, when ^{13}C-decoupling was omitted, assigning them to isoleucine (Figure 2C). In this way, a single selectively isotope-labeled sample afforded residue-type specific assignments of the methyl resonances of isoleucine, leucine, and valine (any remaining cross-peaks with methyl groups would have been assigned to leucine).

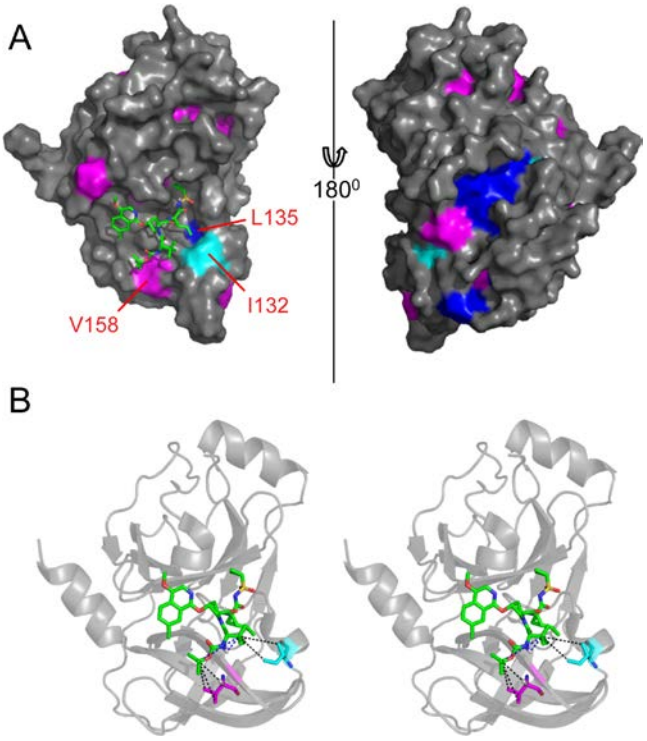


Figure 3. ASV binding site and NOEs with HCV NS4A-NS3pro. The figure shows the crystal structure of the complex between ASV and HCV NS4A-NS3pro (PDB ID: 4NWL; Scola2014). (A) Surface representation of the protein with the bonds of ASV shown as sticks. Isoleucine, valine and leucine residues are highlighted in cyan, magenta and blue, respectively, and assigned if they line the ligand binding site. (B) Stereo view of ASV bound to the protease. The protein is shown as grey ribbons. The inhibitor and the amino acid residues Ile132 and Val158 are shown in stick representations. Dashed lines highlight short distances between the *tert*-butyl groups and the methyl groups of Ile132 and Val158. Corresponding ^{1}H–^{1}H distances are below 5 Å.

To test the capability to identify plausible binding sites of ASV on the protein without reference to the crystal structure of the complex, we used the programs RaptorX[16] and COACH.[17] Starting from the amino acid sequence of HCV NS4A-NS3pro as the sole input data, the programs make use of available X-ray structures of homologous proteins to suggest binding sites. Both programs predicted ligand binding in proximity to Ile132, Leu135, and Val158 (Table 1). Indeed, based on the crystal structure of the NS4A-NS3pro complex with ASV (Figure 3), the ^{1}H–^{1}H distances between the *tert*-butyl group 1 and the methyl groups of Val158 are as short as 2.6 Å, and the ^{1}H–^{1}H distances between the *tert*-butyl group 2 and the methyl groups of Ile132 are as short as 3.3 Å. Notably, there are few sites on NS4A-NS3pro, where methyl groups of isoleucine, leucine, or valine (ILV) are solvent exposed, and there are very few sites, where the *tert*-butyl groups of ASV could simultaneously contact methyl groups of different ILV residue types (Figure 3A). With the crystal structure of the complex with ASV in hand, the intermolecular NOEs assign the ^{1}H-NMR signal of the *tert*-butyl group 1 to the Boc group of ASV.

While the characteristic chemical shifts of ILV methyl groups provide a straightforward way of NMR resonance assignment by residue type, NOESY cross-peaks in other spectral regions are much more difficult to attribute to specific residue types. For example, the NOE at about 8 ppm in Figure 2 can be assigned to an amide proton by virtue of its absence from a spectrum recorded in D_2O (Figure S2) but it is difficult to attribute to specific protein residues without residue-selective isotope labeling of many different amino acid types. We did not attempt to do so, as the most likely assignment is the carbamate proton of the ligand next to the *tert*-butyl group 2 (Figure 1A).

To probe the broader applicability of the assignment strategy described above, we recorded a NOESY spectrum of ASV in complex with a 69 kDa construct of full-length NS4A-NS3. The spectrum displayed only a single NOE that could be attributed to the *tert*-butyl groups, namely the NOE between the *tert*-butyl groups 1 and 2 (Figure 4A). The limited sensitivity of the experiment arose from the low concentration of the complex as well as its limited stability as, in our hands, all NMR samples of the protein-ligand complex were unstable at room temperature, preventing any NMR experiments lasting longer than 12 hours. Nonetheless, it is interesting to note that the chemical shifts of the *tert*-butyl groups changed compared with the complex with NS4A-NS3pro (Figure 2A). This is in agreement with the crystal structures of NS4A-NS3, which show that the binding site of ASV faces the helicase domain (Figure 4B).[12,13] As expected for the increased molecular weight, the cross-peaks between the NMR signals of the *tert*-butyl groups were broader in the complex with full-length NS4A-NS3. Their continued presence suggests that the helicase domain does not significantly affect the binding mode of ASV to the protease domain.

Table 1. Ligand binding site residues of HCV NS4A-NS3pro suggested by the programs RaptorX and COACH *

RaptorX[16]	COACH[17]
Q41, F43	Q41, T42, F43
H57, G58	H57, G58
L82	V78, D81
R123	R123
I132, L135, K136, G137, S139	I132, L135, K136, G137, S138, S139, G140
F154, R155, A156, A157, V158, C159	F154, R155, A156, A157, V158, D168

* Only the amino acid sequence was provided to the respective webservers. The residue numbering follows the convention used in the crystal structure. Underlining identifies the isoleucine, leucine, and valine residues suggested by both programs to be involved in ligand binding.

It is remarkable that the NOE between the *tert*-butyl groups 1 and 2 of ASV was readily observable even in the complex with full-length NS4A-NS3, considering that the crystal structure indicates that the shortest $^{1}H–^{1}H$ distance between these two groups is slightly longer than 5 Å (Figure 3B). Possibly, this result reflects the greater mobility of the Boc group (*tert*-butyl group 1), which may transiently shorten the distance to the *tert*-butyl group 2. The NOESY spectrum, however, showed no evidence for a narrower line shape of the *tert*-butyl group 1 than for the *tert*-butyl group 2 (Figure 2) and the latter is unlikely to have any greater degrees of freedom beyond rotations of the *tert*-butyl group and the methyl groups within (Figure 3B). Fundamentally, the shortening of the effective rotational correlation time effected by such rotations reduces the NMR line widths and raises the maxima of NOESY cross-peaks above the level of the much broader intra-protein cross-peaks. The present work shows that this holds true not only for highly solvent-exposed *tert*-butyl groups but also for *tert*-butyl groups embedded in a ligand binding pocket. NOESY cross-peaks between two *tert*-butyl groups can thus be expected to be especially tall.

One of the *tert*-butyl groups of ASV resides in a Boc group. While Boc groups and tert-butyl groups in general are compatible with oral drug intake, they are not frequently found in commercial drugs. Boc groups are, however, very common protection groups in organic synthesis. Compared to synthesis of ligand molecules with ^{13}C or ^{15}N labels, which is difficult and costly, synthetic intermediates with a Boc protection group are much more readily available. The present work shows that a Boc group is suitable for the detection of intermolecular NOEs with a target protein. For detection of intermolecular protein–ligand NOEs in high-molecular weight systems, it may be possible to engineer proteins and ligand molecules in a way that the interaction results in intermolecular NOEs between two *tert*-butyl groups.

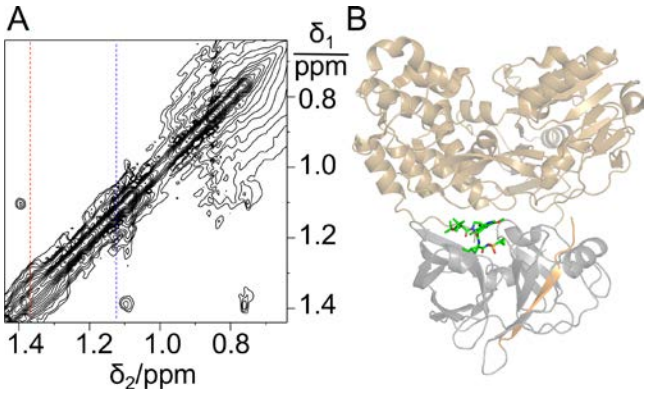


Figure 4. NOEs between the *tert*-butyl groups of ASV in the 69 kDa complex with HCV NS4A-NS3. (A) Region of the NOESY spectrum showing the NOEs between the *tert*-butyl groups. The spectrum was recorded of an approximately 150 μM sample with a mixing time of 200 ms. Red and blue dotted lines indicate the chemical shifts of the *tert*-butyl groups in the complex with NS4A-NS3pro (Figure 2). The change in chemical shifts indicates that the *tert*-butyl groups sense the presence of the helicase domain of NS3. (B) Structure and binding site of ASV in the complex with NS4A-NS3pro superimposed onto the crystal structure of NS4A-NS3 (PDB ID: 5FPS; Saalau2012), showing the inhibitor in a stick representation, NS4A in orange, and the protease and helicase domains in grey and gold, respectively. The helicase domain of NS3 and its C-terminus are close to the inhibitor binding site on NS3pro.

In summary, we have developed a practical strategy for assigning intermolecular protein–ligand NOEs that does not depend on prior sequence-specific resonance assignments of the protein and can be applied to tightly binding ligands characterised by slow off-rates. Importantly, *tert*-butyl groups provide the signal intensities required for their assignment in the NMR spectrum, if not from the 1D ^{1}H-NMR spectrum (as in Figure 1B), then from a string of intense NOESY cross-peaks (as in Figure 2). Our results show that, in a 27 kDa protein, the requisite cross-peaks can readily be observed even if the *tert*-butyl group is embedded in a binding pocket with only partial exposure to solvent, limiting its degree of freedom for reorientational motions. Attributing the ^{1}H-NMR signals of methyl groups from the protein to different residue types was achieved with a single sample produced with a deuterated and a ^{13}C-labeled amino acid. Intermolecular NOEs between *tert*-butyl groups of the ligand and methyl groups of the protein provide a powerful means to pinpoint the location of the ligand on the protein. Finally, the strategy presented here does not depend on perdeuterated protein samples, which is particularly important for unstable samples such as the HCV NS4A-NS3 protease and eukaryotic proteins, which are not amenable to perdeuteration.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website.

Experimental details, amino acid sequences used in the present work (Figure S1), assignment of an NOE with the *tert*-butyl group 2 to an exchangeable amide proton (Figure S2), ^{13}C-HSQC spectrum of NS4A-NS3pro selectively labeled with $^{2}H/^{15}N$-valine and $^{13}C/^{15}N$-isoleucine in complex with ASV (Figure S3) (PDF)

AUTHOR INFORMATION

Corresponding Author

* gottfried.otting@anu.edu.au. Phone: +61 (02) 6125 6507

Author Contributions

W.-N. Chen performed the experiments. W.-N. and G. Otting designed the work and prepared the manuscript.

ORCID

Gottfried Otting: 0000-0002-0563-0146

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

This work was supported by the Australian Research Council, Chinese National Natural Science Foundation (grant 31700670) and Guangdong Natural Science Foundation (grant 2017A030313202). G. O. thanks the Australian Research Council for a Laureate Fellowship.

ABBREVIATIONS

NMR, nuclear magnetic resonance; NOE, nuclear Overhauser enhancement; Tby, *O-tert*-butyl-tyrosine; HCV, hepatitis C virus; ASV, asunaprevir.

REFERENCES

(1) Erlanson, D. A.; Fesik, S. W.; Hubbard, R. E.; Jahnke, W.; Jhoti, H. Twenty Years on: the Impact of Fragments on Drug Discovery. *Nat. Rev. Drug Discov*. **2016**, 15, 605–619.

(2) Harner, M. J.; Mueller, L.; Robbins, K. J.; Reily, M. D. NMR in Drug Design. *Arch. Biochem. Biophys*. **2017,** 628, 132–147.

(3) Chen, W.-N.; Kuppan, K. V.; Lee, M.; Jaudzems, K.; Huber T.; Otting, G. O-tert Butyltyrosine, an NMR Tag for High-Molecular-Weight Systems and Measurements of Submicromolar Ligand Binding Affinities. *J. Am. Chem. Soc*. **2015**, 137, 4581–4586.

(4) Chen, W.-N.; Nitsche, C.; Pilla, K. B.; Graham, B.; Huber, T.; Klein, C. D.; Otting, G. Sensitive NMR Approach for Determining the Binding Mode of Tightly Binding Ligand Molecules to Protein Targets. *J. Am. Chem. Soc*. **2016**, 138, 4539–4546.

(5) Jabar, S.; Adams, L.; Wang, Y.; Aurelio, L.; Graham B.; Otting, G. Chemical Tagging with tert-Butyl and Trimethylsilyl Groups for Measuring Intermolecular Nuclear Overhauser Effects in a Large Protein–Ligand Complex. *Chem. Eur. J*. **2017**, 23, 13033–13036.

(6) Nitsche, C.; Otting, G. NMR Studies of Ligand Binding. *Curr. Opin. Struct. Biol*. **2018**, 48, 16–22.

(7) Manns, M.; Pol, S.; Jacobson, I. M.; Marcellin, P.; Gordon, S. C.; Peng, C.-Y.; Chang, T.-T.; Everson, G. T.; Heo, J.; Gerken, G.; Yoffe, B.; Towner, W. J.; Bourliere, M.; Metivier, S.; Chu, C.-J.; Sievert, W.; Bronowicki, J.-P.; Thabut, D.; Lee, Y.-J.; Kao, J.-H.; McPhee, F.; Kopit, J.; Mendez, P.; Linaberry, M.; Hughes E.; Noviello, S. All-Oral Daclatasvir Plus Asunaprevir for Hepatitis C Virus Genotype 1b: a Multinational, Phase 3, Multicohort Study. *Lancet* **2014**, 384, 1597–1605.

(8) McPhee, F.; Sheaffer, A. K.; Friborg, J.; Hernandez, D.; Falk, P.; Zhai, G.; Levine, S.; Chaniewski, S.; Yu, F.; Barry, D.; Chen, C.; Lee, M. S.; Mosure, K.; Sun, L. Q.; Sinz, M.; Meanwell, N. A.; Colonno, R. J.; Knipe J.; Scola, P. Preclinical Profile and Characterization of the Hepatitis C Virus NS3 Protease Inhibitor Asunaprevir (BMS-650032). *Antimicrob. Agents Chemother*. **2012**, 56, 5387–5396.

(9) Lin, C.; Thomson, J. A.; Rice, C. M. A Central Region in the Hepatitis C Virus NS4A Protein Allows Formation of an Active NS3-NS4A Serine Proteinase Complex in Vivo and in Vitro. *J. Virol*. **1995**, 69, 4373–4380.

(10) Scola, P. M.; Wang, A. X.; Good, A. C.; Sun, L.-Q.; Combrink, K. D.; Campbell, J. A.; Chen, J.; Tu, Y.; Sin, N.; Venables, B. L.; Sit, S.-Y.; Chen, Y.; Cocuzza, A.; Bilder, D. M.; D'Andrea, S.; Zheng, B.; Hewawasam, P.; Ding, M.; Thuring, J.; Lix, J.; Hernandez, D.; Yu, F.; Falk, P.; Zhai, G.; Sheaffer, A. K.; Chen, C.; Lee, M. S.; Barry, D.; Knipe, J. O.; Li, W.; Han, Y.-H.; Jenkins, S.; Gesenberg, C.; Gao, Q.; Sinz, M. W.; Santone, K. S.; Zvyaga, T.; Rajamani, R.; Klei, H. E.; Colonno, R. J.; Grasela, D. M.; Hughes, E.; Chien, C.; Adams, S.; Levesque, P. C.; Li, D.; Zhu, J.; Meanwell, N. A.; McPhee, F. Discovery and Early Clinical Evaluation of BMS-605339, a Potent and Orally Efficacious Tripeptidic Acylsulfonamide NS3 Protease Inhibitor for the Treatment of Hepatitis C Virus Infection. *J. Med. Chem*. **2014**, 57, 1708–1729.

(11) Soumana, D. I.; Ali, A.; Schiffer, C. A. Structural Analysis of Asunaprevir Resistance in HCV NS3/4A Protease. *ACS Chem. Biol*. **2014**, 9, 2485–2490.

(12) Schiering, N.; D'Arcy, A.; Villard, F.; Simić, O.; Kamke, M.; Monnet, G.; Hassiepen, U.; Svergun, D. I.; Pulfer, R.; Eder, J.; Raman P.; Bodendorf, U. A Macrocyclic HCV NS3/4A Protease Inhibitor Interacts with Protease and Helicase Residues in the Complex with its Full-Length Target. *Proc. Natl. Acad. Sci. USA* **2011**, 108, 21052–21056.

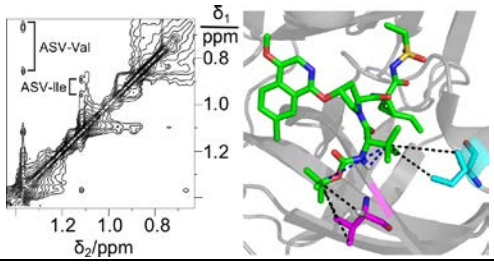
(13) Saalau-Bethell, S. M.; Woodhead, A. J.; Chessari, G.; Carr, M. G.; Coyle, J.; Graham, B.; Hiscock, S. D.; Murray, C. W.; Pathuri, P.; Rich, S. J.; Richardson, C. J.; Williams P. A.; Jhoti, H. Discovery of an Allosteric Mechanism for the Regulation of HCV NS3 Protein Function. *Nat. Chem. Biol*. **2012**, 8, 920–925.

(14) Barbato, G.; Cicero, D. O.; Nardi, M. C.; Steinkühler, C.; Cortese, R.; De Francesco R.; Bazzo, R. *J. Mol. Biol*. **1999**, 289, 371–384.

(15) Cicero, D. O.; Barbato, G.; Koch, U.; Ingallinella, P.; Bianchi, E.; Nardi, M. C.; Steinkühler, C.; Cortese, R.; Matassa, V.; De Francesco, R.; Pessi, A.; Bazzo, R. Structural Characterization of the Interactions of Optimized Product Inhibitors with the N-Terminal Proteinase Domain of the Hepatitis C Virus (HCV) NS3 Protein by NMR and Modelling Studies. *J. Mol. Biol*. **1999**, 290, 385–396.

(16) Källberg, M.; Wang, H.; Wang, S.; Peng, J.; Wang, Z.; Lu, H.; Xu, J. Template-Based Protein Structure Modeling Using the RaptorX web server. *Nat. Protoc*. **2012**, 7, 1511–1522.

(17) Yang, J.; Roy A.; Zhang, Y. Protein–Ligand Binding Site Recognition Using Complementary Binding-Specific Substructure Comparison and Sequence Profile Alignment. *Bioinformatics* **2013**, 29, 2588–2595.


ASV-Val
ASV-Ile
δ1
ppm
0.8
1.0
1.2
1.2 1.0 0.8
δ2/ppm

Supporting Information

Using *tert*-butyl groups in a ligand to identify its binding site on a protein

Wan-Na Chen and Gottfried Otting

Expression constructs

A single-chain construct with NS4A fused to the N-terminus of NS3pro via a tetrapeptide linker is an established target for drug discovery,[1] where the cofactor NS4A stabilizes the active conformation of the HCV NS3 protease.[2] A phase 3 clinical trial determined that the combination of asunaprevir and daclatasvir is effective for treating HCV genotype 1b infection[3] and crystal structures of NS3 from HCV genotype 1b have been determined.[4] Therefore, we based the amino acid sequence of our constructs on this genotype. Specifically, our construct of HCV NS4A-NS3pro (Figure S1A) comprised an N-terminal tag translating into MASMTG (coded by the 5' nucleotide sequence of the bacteriophage T7 gene 10), a His_6 tag, the non-native tetrapeptide MKKK, the sequence of HCV $NS4A_{21-32}$-SGDT-$NS3_{5-180}$ derived from HCV genotype 1[5,6] and a C-terminal ASKKK tag.[7] The amino acid sequences of NS4A and NS3pro are the same as the sequences used in the crystal structure (PDB ID: 4NWL).[5]

The construct of full-length HCV NS4A-NS3 was based on a previously published single-chain construct comprising an N-terminal His_6-tag followed by $NS4A_{21-32}$, the tetrapeptide linker GSGS and $NS3_{3-631}$.[8] Our construct of full-length NS4A-NS3 was based on the amino acid sequence of our NS4A-NS3pro construct, inserting a TEV cleavage site after the N-terminal His_6-tag, removing the C-terminal ASKKK tag and extending NS3 by the linker region and helicase domain to make the 665-residue HCV $NS4A_{21-32}$-SGDT-$NS3_{5-631}$ construct (shown in Figure S1B) .

The nucleotide sequences of both constructs were codon-optimized for expression in *E. coli* (IDT DNA Technologies, Iowa) and inserted into the T7 expression plasmid pETMCSI.[9] *E. coli* TOP10 and BL21(DE3) cells were used for plasmid progression and *in vivo* protein expression, respectively.

(A)

MASMTGHHHHHHMKKKGSVVIVGRINLSGDTAYAQQTRGEEGCQETSQTGRDKNQVEGEVQ
IVSTATQTFLATSINGVLWTVYHGAGTRTIASPKGPVTQMYTNVDKDLVGWQAPQGSRSLT
PCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRA
AVCTRGVAKAVDFIPVESLETTMRASKKK

(B)

MASMTGHHHHHHENLYFQGMKKKGSVVIVGRINLSGDTAYAQQTRGEEGCQETSQTGRDKN
QVEGEVQIVSTATQTFLATSINGVLWTVYHGAGTRTIASPKGPVTQMYTNVDKDLVGWQAP
QGSRSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPISYLKGSSGGPLLCPAGH
AVGIFRAAVCTRGVAKAVDFIPVESLETTMRSPVFTDNSSPPAVPQSFQVAHLHAPTGSGK
STKVPAAYAAQGYKVLVLNPSVAATLGFGAYMSKAHGIDPNIRTGVRTITTGAPVTYSTYG
KFLADGGCSGGAYDIIICDECHSTDSTTILGIGTVLDQAETAGARLVVLATATPPGSVTVP
HPNIEEVALSNTGEIPFYGKAIPIEAIRGGRHLIFCHSKKKCDELAAKLSGLGINAVAYYR
GLDVSVIPTIGDVVVVATDALMTGYTGDFDSVIDCNTCVTQTVDFSLDPTFTIETTTVPQD
AVSRSQRRGRTGRGRRGIYRFVTPGERPSGMFDSSVLCECYDAGCAWYELTPAETSVRLRA
YLNTPGLPVCQDHLEFWESVFTGLTHIDAHFLSQTKQAGDNFPYLVAYQATVCARAQAPPP
SWDQMWKCLIRLKPTLHGPTPLLYRLGAVQNEVTLTHPITKYIMACMSADLEVIT

Figure S1. Amino acid sequences used in the present work. Non-native sequences are underlined by solid lines. The NS4A segment is underlined by a wavy line. (A) Construct of HCV NS4A-NS3pro. The molecular weight is 22.7 kDa. (B) Construct of HCV NS4A-NS3. Dashed underlining identifies the linker and helicase domain. The molecular weight is 69.0 kDa after cleavage with TEV protease.

Preparation of complex between HCV NS4A/NS3pro and ASV

NS4A-NS3pro samples were produced by continuous cell-free protein synthesis at 30 ^{0}C for about 14 h from PCR-amplified DNA as template.[10-12] Selectively isotope-labeled samples were prepared with ^{2}H/^{15}N-valine (ISOTEC, St. Louis, MO, USA) and ^{13}C/^{15}N-isoleucine (Cambridge Isotope Laboratories, Andover, MA, USA). Typical volumes were 4 mL reaction mixture in 40 mL outside buffer. Purification of the protein was performed by loading and re-loading the supernatant onto a spin column filled with 2 mL Ni-NTA agarose (Qiagen, Hilden, Germany) 3 times, washing twice with 6 mL buffer A (50 mM Tris-HCl, pH 7.5, 300 mM NaCl, 5% glycerol), 5 times with 6 mL buffer A plus 10 mM imidazole and finally eluting with 2.5 mL buffer A containing 300 mM imidazole. The eluted fraction was then applied onto a PD-10 column (GE Healthcare, Chicago, IL, USA) pre-equilibrated with 30 mL NMR buffer (20 mM sodium phosphate, pH 6.5, 150 mM NaCl, 5 mM DTT). After the solution had completely entered the packed bed, 3.5 mL of fresh NMR buffer was added to elute the protein. The average yield was about 0.2 mg purified protein per mL cell-free reaction mixture.

Asunaprevir powder (MedChem Express, Monmouth Junction, NJ, USA) was dissolved in DMSO-d_6 to produce a 5 mM stock solution. 1 µL of this solution was added to the protein in NMR buffer and mixed well. This step was repeated 5 times for more complete complex formation. Since the ligand is poorly soluble in water, excess ligand separated by precipitation. The solution was then transferred to a centrifugal filter unit with molecular weight cut-off (MWCO) of 10 kDa (Amicon Ultra, Millipore, Billerica, MA, USA) and concentrated to a final volume of 200 µL containing 20 µL D_2O.

Preparation of full-length HCV NS4A-NS3

Full-length NS4A-NS3 was expressed in *E. coli* BL21(DE3) following a published expression protocol.[13] The protein was purified by a Ni-NTA column using buffer B (50 mM sodium phosphate, pH 6.5, 300 mM NaCl, 5% glycerol) and a gradient with increasing imidazole in buffer B up to 300 mM imidazole. The target protein co-eluted with other proteins as impurities. Therefore, fractions containing the target protein were pooled for cleavage by His_6-tagged TEV protease for 14 h at 4 ^{o}C.[14] The resulting mixture was loaded onto a Co-NTA gravity column and the flow-through was collected. Buffer exchange and complex formation were achieved as described in the protocol for NS4A-NS3pro preparation. The expression level of recombinant NS4A-NS3 was low and the purified product was unstable in our NMR buffer.

NMR spectroscopy

All NMR spectra were recorded on a Bruker 800 MHz NMR spectrometer equipped with a TCI cryoprobe. NOESY spectra were recorded with a pulse sequence using water flip-back by a water-selective 90° pulse prior to the last 90° pulse followed by a WATERGATE sequence. ^{13}C-decoupling in the δ_1 dimension was achieved by a 0.5 ms adiabatic ^{13}C-inversion pulse in the evolution time t_1, if t_1 was sufficiently long to accommodate the pulse. Other parameters were: mixing time = 200 ms, t_{1max} = 58 ms, t_{2max} = 116 ms, total recording time about 12 h. The ^{13}C-HSQC spectrum of Figure S3 was recorded with t_{1max} = 12 ms, t_{2max} = 116 ms, and a total recording time of 4 h.

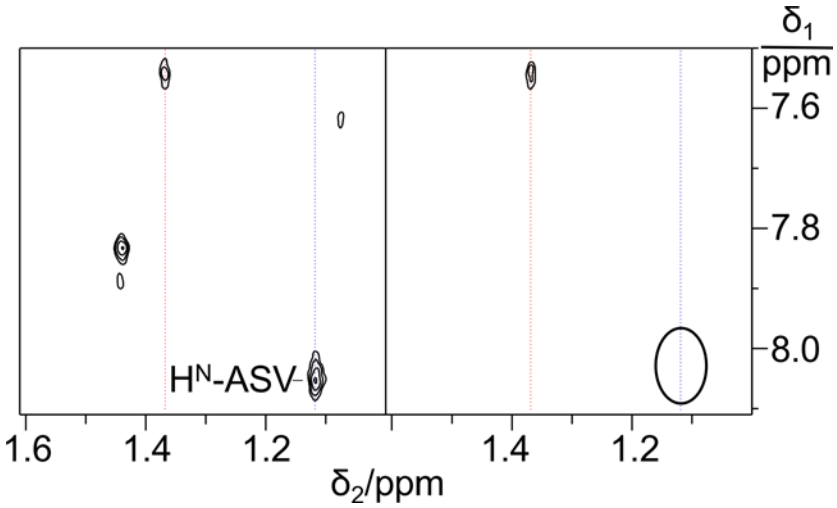


Figure S2. Assignment of an NOE with the *tert*-butyl group 2 to an exchangeable amide proton. The left panel shows a spectral region from the NOESY spectrum shown in Figure 2A. The chemical shifts of the *tert*-butyl groups 1 and 2 are indicated by red and blue dotted lines, respectively. The right panel shows the corresponding spectrum recorded after exchange of the sample into D_2O. The absence of the NOE at 8.04 ppm (circled) indicates that the NOE is with an exchangeable rather than aromatic proton.

Figure S3. ^{13}C-HSQC spectrum of NS4A-NS3pro selectively labeled with ^{2}H/^{15}N-valine and ^{13}C/^{15}N-isoleucine in complex with ASV (same sample as in Figure 2B). Dotted lines identify the ^{1}H-chemical shifts of the isoleucine signals involved in intermolecular NOEs with the *tert*-butyl group 2. The spectrum shows that, at these ^{1}H-chemical shifts, there are intense ^{13}C-HSQC cross-peaks that match $C^{\gamma 2}H_3$ and $C^{\delta 1}H_3$ groups of isoleucine, which supports the assignment of the intermolecular NOEs in Figure 2A to methyl groups from solvent-exposed isoleucine.

References

1 Taremi, S. S.; Beyer, B.; Maher, M.; Yao, N.; Prosise, W.; Weber, P. C.; Malcolm, B. A. Construction, Expression, and Characterization of a Novel Fully Activated Recombinant Single-Chain Hepatitis C Virus Protease. *Protein Sci.* **1998**, 7, 2143–2149.

2 Lin, C.; Thomson, J. A.; Rice, C. M. A Central Region in the Hepatitis C Virus NS4A Protein Allows Formation of an Active NS3-NS4A Serine Proteinase Complex in Vivo and in Vitro. *J. Virol*. **1995**, 69, 4373–4380.

3 Manns, M.; Pol, S.; Jacobson, I. M.; Marcellin, P.; Gordon, S. C.; Peng, C.-Y.; Chang, T.-T.; Everson, G. T.; Heo, J.; Gerken, G.; Yoffe, B.; Towner, W. J.; Bourliere, M.; Metivier, S.; Chu, C.-J.; Sievert, W.; Bronowicki, J.-P.; Thabut, D.; Lee, Y.-J.; Kao, J.-H.; McPhee, F.; Kopit, J.; Mendez, P.; Linaberry, M.; Hughes E.; Noviello, S. All-Oral Daclatasvir Plus Asunaprevir for Hepatitis C Virus Genotype 1b: a Multinational, Phase 3, Multicohort Study. *Lancet* **2014**, 384, 1597–1605.

4 Saalau-Bethell, S. M.; Woodhead, A. J.; Chessari, G.; Carr, M. G.; Coyle, J.; Graham, B.; Hiscock, S. D.; Murray, C. W.; Pathuri, P.; Rich, S. J.; Richardson, C. J.; Williams P. A.; Jhoti, H. Discovery of an Allosteric Mechanism for the Regulation of HCV NS3 Protein Function. *Nat. Chem. Biol*. **2012**, 8, 920–925.

5 Scola, P. M.; Sun, L.-Q.; Wang, A. X.; Chen, J.; Sin, N.; Venables, B. L.; Sit, S.-Y.; Chen, Y.; Cocuzza, A.; Bilder, D. M.; D'Andrea, S. V.; Zheng, B.; Hewawasam, P.; Tu, Y.; Friborg, J.; Falk, P.; Hernandez, D.; Levine, S.; Chen, C.; Yu, F.; Sheaffer, A. K.; Zhai, G.; Barry, D.; Knipe, J. O.; Han, Y.-H.; Schartman, R.; Donoso, M.; Mosure, K.; Sinz, M. W.; Zvyaga, T.; Good, A. C.; Rajamani, R.; Kish, K.; Tredup, J.; Klei, H. E.; Gao, Q.; Mueller, L.; Colonno, R. J.; Grasela, D. M.; Adams, S. P.; Loy, J.; Levesque, P. C.; Sun, H.; Shi, H.; Sun, L.; Warner, W.; Li, D.; Zhu, J.; Meanwell, N. A.; McPhee, F. The Discovery of Asunaprevir (BMS-650032), an Orally Efficacious NS3 Protease Inhibitor for the Treatment of Hepatitis C Virus Infection. *J. Med. Chem.* **2014**, 57, 1730–1752.

6 (a) Wittekind, M.; Weinheimer, S.; Zhang, Y.; Goldfarb, V. Modified Forms of Hepatitis C NS3 Protease for Facilitating Inhibitor Screening and Structural Studies of Protease: Inhibitor Complexes. *US Patent 6333186*, **2001**; (b) Wittekind, M.; Weinheimer, S.; Zhang, Y.; Goldfarb, V. Modified Forms of Hepatitis C NS3 Protease for Facilitating Inhibitor Screening and Structural Studies of Protease: Inhibitor Complexes. *US Patent*

6800456, **2004**.

7 Bianchi, E.; Orrù, S.; Piaz, F. D.; Ingenito, R.; Casbarra, A.; Biasiol, G.; Koch, U.; Pucci, P.; Pessi, A. Conformational Changes in Human Hepatitis C Virus NS3 Protease upon Binding of Product-Based Inhibitors. *Biochemistry* **1999**, 38, 13844–13852.

8 Howe, A. Y.; Chase, R.; Taremi, S. S.; Risano, C.; Beyer, B.; Malcolm, B.; Lau, J. Y. A Novel Recombinant Single-Chain Hepatitis C Virus NS3A-NS4A Protein with Improved Helicase Activity. *Protein Sci.* **1999**, 8, 1332–41.

9 Neylon, C.; Brown, S. E.; Kralicek, A. V.; Miles, C. S.; Love, C. A.; Dixon, N. E. Interaction of the *Escherichia coli* Replication Terminator Protein (Tus) with DNA: a Model Derived from DNA-Binding Studies of Mutant Proteins by Surface Plasmon Resonance. *Biochemistry* **2000**, 39, 11989–11999.

10 Ozawa, K.; Headlam, M. J.; Schaeffer, P. M.; Henderson, B. R.; Dixon, N. E.; Otting, G. Optimization of an *Escherichia coli* system for Cell-Free Synthesis of Selectively ^{15}N-Labelled Proteins for Rapid Analysis by NMR Spectroscopy. *Eur. J. Biochem.* **2004**, 271, 4084–4093.

11 Wu, P. S. C.; Ozawa, K.; Lim, S. P.; Vasudevan, S. G.; Dixon, N. E.; Otting, G. Cell-Free Transcription/Translation from PCR-Amplified DNA for High-Throughput NMR Studies. *Angew. Chemie Int. Ed.* **2007**, 46, 3356–3358.

12 Apponyi, M. A.; Ozawa, K.; Dixon, N. E.; Otting, G. Cell-Free Protein Synthesis for Analysis by NMR Spectroscopy. *Methods Mol. Biol.* **2008**, 426, 257–268.

13 Poliakov, A.; Hubatsch, I.; Shuman, C. F.; Stenberg, G.; Danielson, U. H. Expression and Purification of Recombinant Full-Length NS3 Protease-Helicase from a New Variant of Hepatitis C Virus. *Protein Expr. Purif.* **2002**, 25, 363–71.

14 Cabrita, L. D.; Gilis, D.; Robertson, A. L.; Dehouck, Y.; Rooman, M.; Bottomley, S. P. Enhancing the Stability and Solubility of TEV Protease Using in Silico Design. *Protein Sci*. **2007**, 16, 2360–2367.