

Non-Peptidic Small-Molecule Inhibitors of the Single-Chain Hepatitis C Virus NS3 Protease/NS4A Cofactor Complex Discovered by Structure-Based NMR Screening

Daniel F. Wyss,* Ashok Arasappan, Mary M. Senior, Yu-Sen Wang, Brian M. Beyer, F. George Njoroge, and Mark A. McCoy

Schering-Plough Research Institute, 2015 Galloping Hill Road, Kenilworth, New Jersey 07033

Received October 9, 2003

NMR-based screening of a customized fragment library identified 16 small-molecule hits that bind weakly ($K_D \approx 100 \mu\text{M}$ to 10 mM) to substrate binding sites of the NS4A-bound NS3 protease of the hepatitis C virus (HCV). Analogues for five classes of NMR hits were evaluated by a combination of NMR and biochemical data yielding SAR and, in most cases, optimized hits with improved potencies ($K_D \approx K_I \approx 40 \mu\text{M}$ to 1 mM). NMR chemical shift perturbation data were used to establish the binding location and orientation of the active site directed scaffolds in these five analogue series. Two of these scaffolds, which bind the enzyme at the proximal S1–S3 and S2' substrate binding sites, were linked together producing competitive inhibitors of the HCV NS3 protease with potencies in the micromolar range. This example illustrates that the low molecular weight scaffolds discovered from structure-based NMR screening can be optimized with focused structure-guided chemistry to produce potent nonpeptidic small-molecule inhibitors of the HCV NS3 protease.

Introduction

The RNA virus hepatitis C (HCV) has chronically infected an estimated 170 million people worldwide, of whom 10–20% will ultimately develop cirrhosis and 1–5% will develop liver cancer.¹ HCV contains a positive single-stranded RNA genome that encodes a polyprotein of about 3010 amino acids that is processed into structural (capsid, E1, and E2), p7, and nonstructural (NS2, NS3, NS4A, NS4B, NS5A, and NS5B) proteins by cellular and virus-encoded proteases NS2 and NS3.² The NS3 protein is a 631 amino acid residue bifunctional enzyme with a serine protease localized at the N-terminal 180 residues and an RNA helicase located at the C-terminal 451 residues³ and represents an attractive molecular target for drug development.⁴ Both protease and helicase domains retain their *in vitro* activity when expressed separately.⁵ Although the protease domain of NS3 (NS3p) has intrinsic proteolytic activity, complex formation with the viral NS4A polypeptide is essential for efficient processing of the NS3–NS4A and NS4B–NS5A sites and for improved cleavage at the NS4A–NS5B and NS5A–NS5B junctions.⁶ Three-dimensional (3D) structures have been determined for the HCV NS3 protease domain by itself,⁷ for its complex with NS4A-derived peptides,⁸ and for engineered single-chain HCV NS4A–NS3 proteins containing either the protease domain of NS3 (ns4a–ns3p)⁹ or both the protease and helicase domains of NS3 (ns4a–ns3)¹⁰ C-terminally linked to an NS4A-derived peptide. Collectively, these studies revealed that the N-terminal β -barrel of NS3p undergoes large structural rearrangements upon NS4A binding, resulting in proper alignment of the catalytic triad and stabilization of this subdomain.

While conventional bioassay-based high-throughput screening (HTS) of large chemical libraries remains a mainstream approach for lead discovery, its limitations have driven the development of alternative and complementary tools. In this regard, novel NMR-based approaches that have emerged over the past few years show great promise.^{11–13} We have used structure-based NMR screening approaches (SbN) for a variety of drug targets to identify druglike small-molecule hits from customized libraries, which subsequently could be optimized into potent leads through focused, structure-guided chemistry. Focus was placed on targets for which HTS of large compound libraries failed to identify suitable leads. SbN was carried out on a diverse set of targets to identify and optimize novel scaffolds, which subsequently were linked or further optimized by the parallel synthesis of a large number of analogues or used as templates for structure-based inhibitor design. A critical step in these fragment-based methods, which take a modular approach to drug design, is the optimization of weak NMR hits (K_D approximately in the micromolar to millimolar range) into potent leads (K_D less than micromolar quantities) using chemistry guided by 3D structural data. However, such information cannot always be obtained easily. We have therefore developed novel tools that utilize protein chemical shift perturbation (ΔCS) data, generated in heteronuclear single quantum coherence (HSQC) based NMR screens, to rapidly provide an accurate structural representation of protein/ligand complexes under conditions unfavorable for traditional structural work using X-ray crystallography and/or solution NMR.^{14,15} These tools use the ΔCS data in a more quantitative manner than the traditional “chemical shift perturbation mapping” widely used to qualitatively map molecular interfaces, and hence, they allow rapid determination of the ligand

* To whom correspondence should be addressed. Phone: (908) 740-3299. Fax: (908) 740-7305. E-mail: daniel.wyss@spcorp.com.

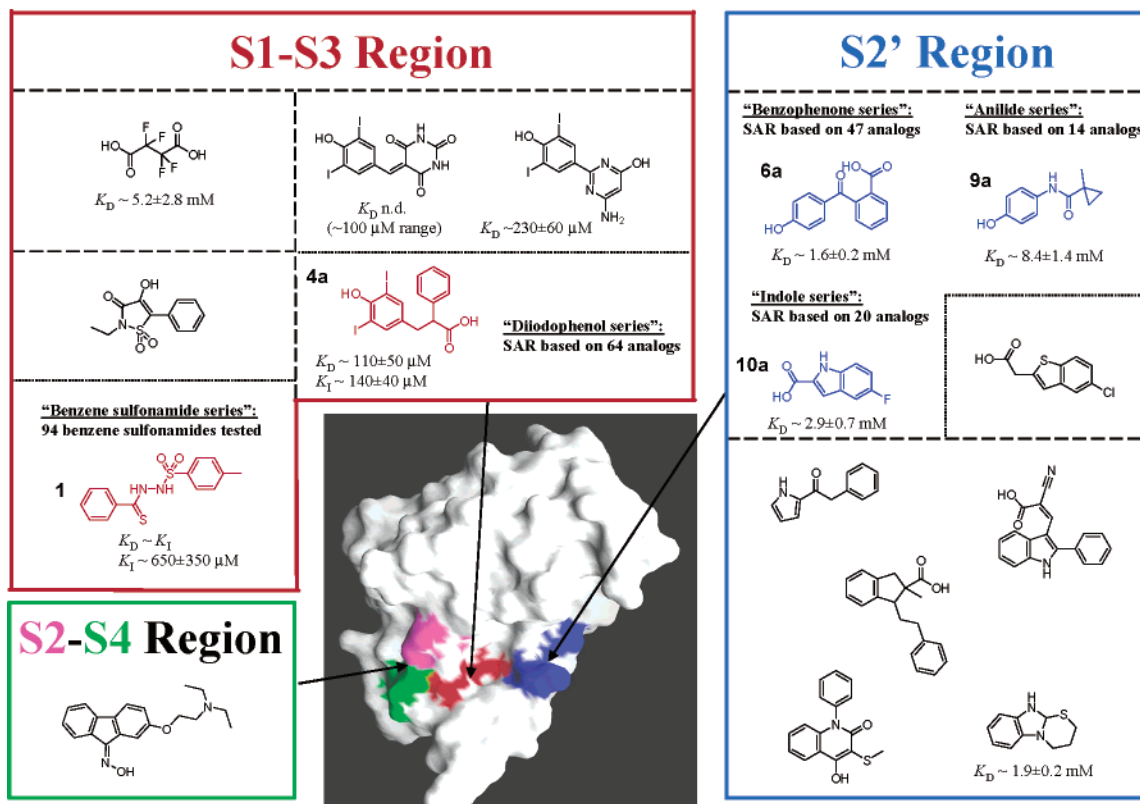


Figure 1. Validated active-site directed "NMR hits" that bind to substrate binding sites of the ^{15}N -labeled ns4a–ns3p. They were identified by screening 3639 compounds from a customized small-molecule library by 2D ^{15}N – ^1H HSQC NMR. The NMR hits are grouped according to their approximate binding sites on the HCV NS3 protease. The respective substrate binding sites are colored on the surface of the protease domain of the ns4a–ns3 crystal structure (PDB accession code 1CU1). SAR, K_D and/or K_I values, and detailed binding orientations could be derived from the analogue series for the five compounds that are colored according to their binding sites.

binding site with much higher precision.¹³ By careful comparison of ΔCS data of closely related analogues, these methods also often enable an unambiguous determination of the binding mode of a small-molecule ligand. Herein, we describe the application of SbN to the discovery of small-molecule inhibitors of the HCV NS3 protease/NS4A cofactor complex.

Results and Discussion

Identification of Small-Molecule Hits to HCV NS3 Protease Substrate Binding Sites by NMR-Based Screening. Because of the importance of the NS4A cofactor for efficient catalytic activity⁶ and structural integrity^{7–10} of the NS3p domain, we used a single-chain HCV NS3 protease/NS4A complex (ns4a–ns3p)⁹ for our NMR-based screens. Furthermore, since NS4A binding stabilizes S' residues in the N-terminal subdomain of NS3p, NS4A-bound NS3p will be more relevant for the design of nonpeptidic inhibitors that bind to S' pockets of the HCV NS3 protease. Ns4a–ns3p was optimized for SbN with regard to NMR spectral quality, long-term stability, large-scale production, and recycling efficiency.

From a customized small-molecule library, 3639 compounds were screened against the uniformly ^{15}N -labeled ns4a–ns3p protein using two-dimensional (2D) ^{15}N heteronuclear single quantum coherence (^{15}N – ^1H HSQC) NMR. From over 50 hits in our NMR screen, 16 compounds were validated as active-site hits that were pure and specifically bound the HCV NS3 protease at

substrate-binding sites as judged by the NMR chemical shift perturbation data (Figures 1 and 2). Six and nine of these fragments bind the ns4a–ns3p in the S1–S3 region and S2' region, respectively; one additional fragment binds the enzyme in the S2–S4 region. These hits and their active analogues (see below) bind the ns4a–ns3p with K_D values between $100 \mu\text{M}$ and 10 mM as determined by NMR titration experiments. The compounds that bind the ns4a–ns3p in the S1–S3 region were also active in an HCV protease biochemical assay (K_I values) in which conditions similar to those of the NMR experiments were used. In most cases the K_D values measured by NMR agreed well with the K_I values. In contrast, compounds that bind the ns4a–ns3p in the S2' region were inactive in the biochemical assay. This result is expected because the biochemical assay substrate spans only the S6–S1' region. Thus, S2' binding fragments are not expected to interfere with substrate binding.

Our next goal was to evaluate analogues for some of these novel scaffolds to potentially improve their potency, obtain the SAR, and determine their binding mode in the complex with the ns4a–ns3p for focused, structure-guided chemistry. For five of these "NMR hits", analogues from our corporate database or commercial sources were evaluated by NMR methods and in the biochemical assay to identify "optimized hits" (Figure 1 and see below). Distinct SAR and higher-affinity hits were obtained in most analogue series, which in addition helped to better understand the

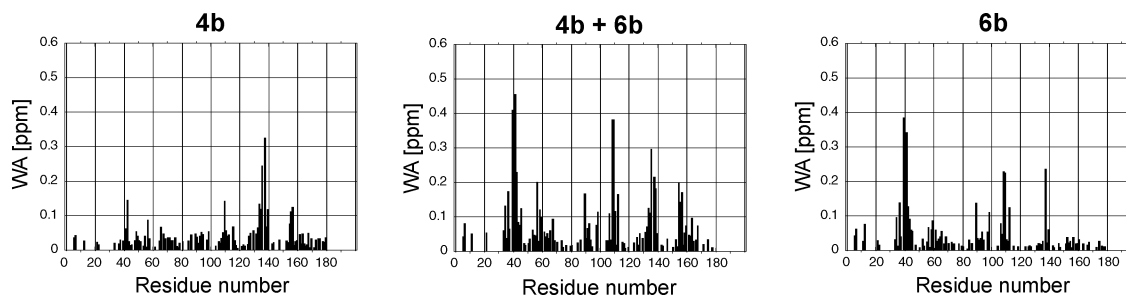


Figure 2. Per residue weighted average chemical shift perturbations ($WA = [(\Delta\delta\{^1H\})^2 + (0.2 \Delta\delta\{^{15}N\})^2]^{1/2}$) of 200 μ M ^{15}N -labeled ns4a–ns3p in the presence of 1 mM each of **4b** (left), **4b** and **6b** (middle), and **6b** (right), illustrating that no global protein conformational changes occur upon ligand binding (WA for most residues are below 0.04 ppm) and the simultaneous binding of ligands **4b** and **6b** to ns4a–ns3p is possible.

details of their binding mode in the complex with the ns4a–ns3p (see below).

Optimization of Ligands That Bind the S1–S3 Region. In an effort to optimize the “benzene sulfonamide hit” **1**, we evaluated a total of 94 related compounds in directed NMR screens. No distinct SAR could be derived from these studies, since direct analogues of ligand **1** were not available. However, two additional weak hits to the S1–S3 region, compounds **2** and **3** with potencies of ~ 560 and ~ 150 μ M, respectively, were identified in this process.

Sixty-four analogues of the “diiodophenol hit” **4a** were evaluated by NMR, and if found to bind to ns4a–ns3p, their biological activity was also determined in the biochemical assay. The SAR derived from this series is summarized in Table 1. To probe the importance of the 2-phenylpropionic acid moiety, several analogues of **4a** were tested in which the 4-hydroxy-3,5-diiodophenyl moiety was maintained (**4b–n**). These data collectively revealed that the carboxylic acid functionality is critical for activity, and its position with respect to the ring seems important. The phenyl substituent at the 2-position of the propionic acid moiety, on the other hand, does not appear to be critical for activity, since compounds **4b** and **4c** showed activities similar to that of the parent ligand **4a** and various substituents at the 2-position were well tolerated as exemplified by compounds **4d–f**. Bulky groups at the 2-position, however, precipitated the protein (e.g., compounds **4g–i**, which were tested for protein binding below their respective solubility limits). The spacing between the carboxylate group and the diiodophenol ring seems important, but not critical (compare compounds **4b**, **4j**, and **4k**), with the propionic acid derivative **4b** being the most potent analogue. In addition, removal of the negatively charged carboxylate lowered the solubility and gave inactive compounds (e.g., compounds **4l–n**), suggesting that the carboxylate is critical for solubility and possibly binding. Interestingly, several tyrosine derivatives with various N-protection groups were also active with K_I values in the 200–300 μ M range (e.g., compounds **4d–f**). The latter scaffolds can more easily be used in the construction of focused libraries of linked molecules by linking together optimized NMR hits.

Since the iodides make the hydroxyl group acidic (pK_a -[3,5-diiodotyrosine] = 5.32 (25 $^{\circ}$ C) [*CRC Handbook Chemistry and Physics*, 76th edition]) and therefore are not desired in the final drug molecule, analogues were tested in which substitution of the iodides and the acidic hydroxyl group was attempted (e.g., compounds **5a–g**).

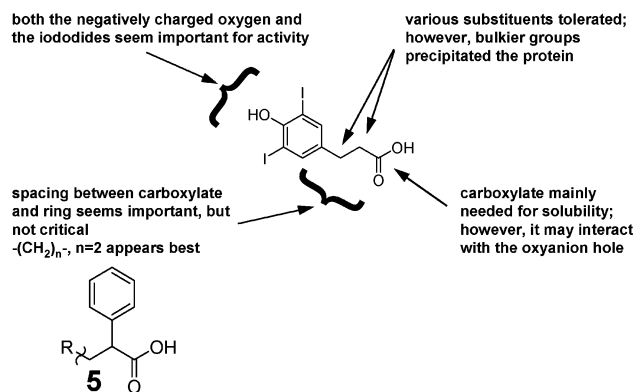
However, none of the limited set of analogues that were available in the corporate database showed any activity in the NMR assay. Collectively, the data suggest that at neutral pH both the negatively charged hydroxyl oxygen (e.g., compare **4a** with **5c**) and the iodides (e.g., compare **4a** with **5a** and **5b**) seem important for activity.

Optimization of Ligands That Bind the S2' Region. Forty-seven analogues of the “benzophenone hit” **6a** were evaluated by NMR. The SAR derived from this series is summarized in Table 2. Several substitutions for the *p*-hydroxyl of NMR hit **6a** were tolerated with the analogues showing about 4-fold differences in K_D (ligands **6a–f**) and **6b** being the most potent analogue in this series. A *p*-hydroxyl to carboxylate substitution, however, reduced the molecule’s specificity for the S2’ site, since **6g** exhibited a trend for nonspecific binding to additional exposed sites on the ns4a–ns3p. The *p*-butyric acid analogue of NMR hit **6a** (**6h**) still preferred the S2’ binding pocket but with an approximately 40-fold lower affinity. This latter compound **6h** also bound to additional exposed sites on the ns4a–ns3p. In contrast to the *p*-Cl analogue of ligand **6a** (**6d**), the *m*-Cl analogue (**6i**) was inactive in the NMR assay, suggesting that it is at least 4-fold less active. Substitution of the keto linker by an ether linker resulted in analogues that were about 2-fold less potent (compare **6j** and **6k** with **6d** and **6e**, respectively). All compounds with a methylene linker were inactive in the NMR assay at the concentrations tested, and hence, they were at least 3-fold less potent than their analogues with a keto linker (compare **6l** and **6m** with **6d** and **6a**, respectively).

A substitution of the *o*-COOH group of the benzoic acid moiety of ligand **6d** by an *o*-CH₂COOH group in **7a** reduced the potency only slightly. All other modifications of the benzoic acid moiety resulted in inactive analogues (**7b–h**), suggesting that the carboxylic acid functionality is critical for activity and solubility. The results, however, do not reveal if the COOH group is tolerated at other positions on the ring. We could not identify such analogues in the corporate database at quantities sufficient for the NMR assay. We tested several analogues in which the *p*-hydroxyl and the benzoic acid moieties of NMR hit **6a** were simultaneously replaced by various groups (**8a–n**), with four of these analogues containing a carboxamide linker in place of the keto linker (**8k–n**). However, all these analogues showed either no binding activity or bound the protease with a much-reduced affinity ($K_D > 20$ –100 mM).

Table 1. SAR Derived from an Evaluation of 64 Analogues in the "Diiodophenol" Series

| compd | R | Dissociation Constant K_D (μM) ^a | Inhibition Constant $K_I \pm \text{SEM}$ (μM) |
|-----------|---|---|---|
| 4a | | 110±50 | 140±40 |
| 4b | | 90±10 | 80±40 |
| 4c | | 40±10 | 100±60 |
| 4d | | active ^b | 210±100 |
| 4e | | active ^b | 260±70 |
| 4f | | active ^b | 310±60 |
| 4g | | ppt ^c | n.d. ^d |
| 4h | | ppt ^c | n.d. ^d |
| 4i | | ppt ^c | n.d. ^d |
| 4j | | 140±50 | 180±100 |
| 4k | | 420±300 | 600±500 |
| 4l | | inactive ^e | n.d. ^d |
| 4m | | inactive ^e | n.d. ^d |
| 4n | | > ~2,500 ^f | n.d. ^d |

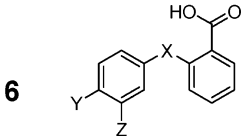


| compd | R | Dissociation Constant K_D (μM) ^a | Inhibition Constant $K_I \pm \text{SEM}$ (μM) |
|-----------|---|---|---|
| 5a | | > ~8000 ^g | n.d. ^d |
| 5b | | > ~8000 ^g | n.d. ^d |
| 5c | | > ~1600 ^f | n.d. ^d |
| 5d | | > ~8000 ^g | n.d. ^d |
| 5e | | > ~8000 ^g | n.d. ^d |
| 5f | | > ~8000 ^g | n.d. ^d |
| 5g | | > ~8000 ^g | n.d. ^d |

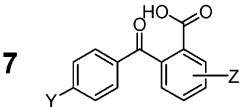
^a K_D derived from NMR titration experiments using ^{15}N -labeled ns4a–ns3p. ^b Significant chemical shift perturbations (ΔCS) at 1:2 compound [100 μM]/protein ratio. ^c No ΔCS , but protein precipitation at 1:2 compound [100 μM]/protein ratio. ^d Not determined. ^e No ΔCS at 5:1 compound [1 mM]/protein ratio, but compound insoluble in NMR buffer. ^f No ΔCS at maximal solubility of compound in NMR buffer; K_D estimated assuming greater than 10% of protein must be bound to observe ΔCS . ^g No ΔCS at 5:1 compound [1 mM]/protein ratio; K_D estimated assuming greater than ~10% of protein must be bound to observe significant ΔCS .

Table 3 summarizes the SAR of the "anilide series", which is based on 14 analogues of NMR hit **9a**. Replacement of the 1-methylcyclopropyl of **9a** in the three analogues **9b–d** reduced the affinity by about 2- to 7-fold. In analogue **9b** methylation of the *p*-hydroxyl (**9e**) reduced the affinity by approximately 4-fold, whereas three additional methylated compounds (**9f–h**) were completely inactive by NMR. Interestingly, some bind-

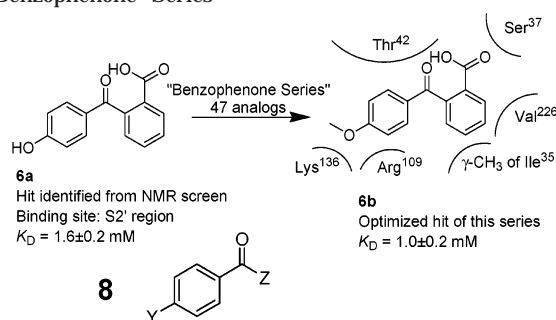
ing activity was restored in ligand **9i**, which contains an *o*-benzoic acid like the most active ligands in the "benzophenone series". However, a trifluoromethyl group next to the *p*-hydroxyl in the ring of **9a** increased the affinity about 12-fold (**9j**). The resulting "optimized NMR hit" **9j** binds the ns4a–ns3p with a K_D of 0.69 ± 0.07 mM as determined by NMR titration experiments. Three analogues of ligand **9j** could be tested in which





Table 2. SAR Derived from an Evaluation of 47 Analogues in the "Benzophenone" Series


| No. | X | Y | Z | K_D [mM] ^a |
|-----|-----------------|--------------------------------------|----|--------------------------------|
| 6a | C=O | OH | H | 1.6±0.2 |
| 6b | C=O | OCH ₃ | H | 1.0±0.2 |
| 6c | C=O | CH ₃ | H | 1.6±0.4 |
| 6d | C=O | Cl | H | ~3 |
| 6e | C=O | F | H | ~4 |
| 6f | C=O | CF ₃ | H | ~4 |
| 6g | C=O | COOH | H | different sites ^b |
| 6h | C=O | (CH ₂) ₃ COOH | H | ~40 less specific ^b |
| 6i | C=O | H | Cl | (> ~10) |
| 6j | O | Cl | H | ~7 |
| 6k | O | F | H | ~7 |
| 6l | CH ₂ | Cl | H | (>~7) |
| 6m | CH ₂ | OH | H | (>~7) |



| No. | Y | Z | K_D [mM] ^a |
|-----|------------------|---|-------------------------|
| 7a | Cl | <i>o</i> -CH ₂ -COOH | ~4 |
| 7b | Cl | <i>o</i> -CO-N(CH ₃) ₂ | (> ~10) |
| 7c | Cl | <i>o</i> -CO-OCH ₃ | (> ~10) |
| 7d | Cl | <i>o</i> -Cl | insoluble |
| 7e | OH | <i>p</i> -OH | (> ~10) |
| 7f | OH | <i>p</i> -CH ₃ | (> ~5) |
| 7g | OH | <i>m</i> -Br | (> ~0.6) |
| 7h | OCH ₃ | <i>p</i> -OCH ₃ | insoluble |



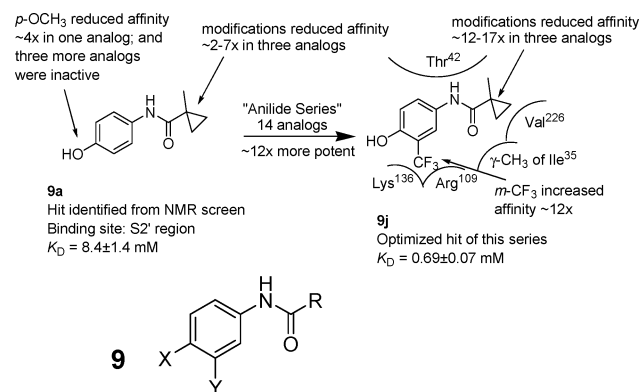
| No. | Y | Z | K_D [mM] ^a |
|-----|---|---|-----------------------------|
| 8a | COOH | CF ₃ | different site ^b |
| 8b | OCH ₃ | 2-pyrrol | (> ~20) |
| 8c | OCH ₃ | cyclopropyl | (> ~20) |
| 8d | O-CH ₂ -COOH | CH ₃ | (> ~20) |
| 8e | O-(CH ₂) ₂ -COOH | CH ₃ | (> ~20) |
| 8f | NH ₂ | benzoyl | ~20 |
| 8g | NH ₂ | 3-(2-amino)-pyridyl | ~70 |
| 8h | NH ₂ | <i>p</i> -anilin | (> ~20) |
| 8i | NH ₂ | phenyl | (> ~30) |
| 8j | NH ₂ | OH | (> ~40) |
| 8k | NH ₂ |  | ~20 |
| 8l | NH ₂ |  | ~100 |
| 8m | NH ₂ |  | (> ~20) |
| 8n | NH ₂ |  | (> ~20) |

^a K_D values were determined by NMR only for the three most potent analogues. If significant binding was observed in the NMR assay, the affinity was estimated on the basis of the extent of the perturbations relative to the parent compound. It was assumed that the chemical shifts of the ¹⁵N-labeled ns4a–ns3p in the bound form are identical for each analogue. If no binding was observed, a lower limit for the K_D is indicated in parentheses. It was assumed that less than 5% of the ns4a–ns3p was bound at the highest compound concentration that was used in the NMR assay. ^b Binds to additional exposed sites of ns4a–ns3p.

the 1-methylcyclopropyl was replaced by other groups (**9k–m**). Like in the parent compound **9a** none of the replacement groups were favored and reduced the affinity significantly (~12- to 17-fold). Two additional analogues (**9n** and **9o**) were inactive. These contained a *p*-carboxylate and different groups in place of the 1-methylcyclopropyl of **9j** and were inactive.

Table 4 summarizes the SAR of the "indole series", which is based on 20 analogues of NMR hit **10a**. 5-Bromine (**10b**) and 5-chlorine (**10c**) substitutions

increased the affinity by about an order of magnitude, whereas a 5-methyl substitution (**10d**) was only slightly better than the NMR hit **10a** and a 5-methoxy substituent (**10e**) appeared to be slightly worse. Removal of the fluorine at the 5-position reduced the affinity about 4-fold (**10f**). In contrast, the 6-chloro analogue **10g** showed a significant increase in potency, being similarly potent compared to its 5-chloro analogue **10c**. Also, a chlorine at either the 4- (**10h**) or the 7-position (**10i**) or a bromine at the 7-position (**10j**) increased the affinity

Table 3. SAR Derived from an Evaluation of 14 Analogues in the "Anilide" Series

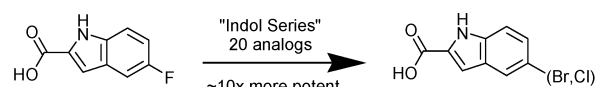
| no. | X | Y | R | K _D [mM] ^a |
|-----------|-------------------|-----------------|---|----------------------------------|
| 9a | OH | H | 1-methylcyclopropyl | 8.4 ± 1.4 |
| 9b | OH | H | CH(CH ₃) ₂ | ~17 ± 8 |
| 9c | OH | H | CH ₂ CH ₃ | ~34 ± 17 |
| 9d | OH | H | CH ₂ CH ₂ Cl | ~55 ± 14 |
| 9e | OCH ₃ | H | CH(CH ₃) ₂ | ~70 |
| 9f | OCH ₃ | H | 1-methylcyclopropyl | (≥ 60) |
| 9g | OCH ₃ | H | C(CH ₃) ₃ | (≥ 100) |
| 9h | OCH ₃ | H | CH ₂ CH ₂ CO ₂ H | (≥ 110) |
| 9i | OCH ₃ | H | <i>o</i> -benzoic acid | ~22 ± 4 |
| 9j | OH | CF ₃ | 1-methylcyclopropyl | 0.69 ± 0.07 |
| 9k | OH | CF ₃ | CH(CH ₃) ₂ | ~8 ± 3 |
| 9l | OH | CF ₃ | N(CH ₃) ₂ | ~12 ± 2 |
| 9m | OH | CF ₃ | N(CH ₃)(OCH ₃) | ~8 ± 2 |
| 9n | CO ₂ H | H | NH-CH ₃ | (≥ 210) |
| 9o | CO ₂ H | H | CH=NOH | (≥ 250) |

^a K_D values were determined by NMR only for the most potent analogues. If significant binding was observed in the NMR assay, the affinity was estimated on the basis of the extent of the perturbations relative to the parent compound. It was assumed that the chemical shifts of the ¹⁵N-labeled ns4a–ns3p in the bound form are identical for each analogue. If no binding was observed, a lower limit for the K_D is indicated in parentheses. It was assumed that less than 5% of the ns4a–ns3p was bound at the highest compound concentration that was used in the NMR assay.

by about an order of magnitude compared to **10f**. These results show that halogenation of the indole at the 4-, 5-, 6-, or 7-position is favorable for the binding activity of the 2-carboxyindole scaffold.

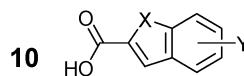
2-Carboxybenzothiophen **10k** and 2-carboxybenzofuran **10l** were about 7-fold and 4-fold more potent, respectively, than 2-carboxyindole **10f**. Similarly, N-methylation of 2-carboxyindole (**10m**) improved the affinity slightly compared to **10f**, suggesting that the HN is not forming an H-bond if the benzene ring is fully protonated. In contrast, an indole was about 7-fold more potent than a benzofuran when chlorinated at the 5-position (compare **10c** with **10n**) and about 2-fold more potent than a benzothiophen in the presence of a 5-methoxy group (compare **10e** with **10o**).

Binding Mode of Optimized NMR Hits. The NMR-detected low molecular weight scaffolds are most useful if combined with structural information derived from the small ligands when complexed to the target. Attempts to determine the 3D structures of complexes between these NMR-detected ligands and ns4a–ns3p were unsuccessful by both X-ray crystallography and conventional NMR methods. Therefore, we used two recently developed tools that utilize the chemical shift perturbation (Δ CS) data, generated in the target-detected NMR screen, to provide a structural represen-

Table 4. SAR Derived from an Evaluation of 20 Analogues in the "Indole" Series

10a
 Hit identified from NMR screen
 Binding site: S2' region
 K_D = 2.9 ± 0.7 mM

10b, 10c
 Optimized hits of this series
 K_D ~ 0.1–0.5 mM



| no. | X | Y | K _D [mM] ^a |
|------------|------------------|--------------------|----------------------------------|
| 10a | N | 5-F | 2.9 ± 0.7 |
| 10b | N | 5-Br | ~0.2 ± 0.1 |
| 10c | N | 5-Cl | ~0.3 ± 0.2 |
| 10d | N | 5-CH ₃ | ~1.2 ± 0.6 |
| 10e | N | 5-OCH ₃ | ~4.0 ± 2.0 |
| 10f | N | 5-H | ~12.0 ± 0.4 |
| 10g | N | 5-H, 6-Cl | ~0.4 ± 0.3 |
| 10h | N | 5-H, 4-Cl | ~0.8 ± 0.1 |
| 10i | N | 5-H, 7-Cl | ~1.4 ± 0.4 |
| 10j | N | 5-H, 7-Br | ~1.0 ± 0.3 |
| 10k | S | 5-H | ~1.7 ± 0.2 |
| 10l | O | 5-H | ~6.4 ± 0.6 |
| 10m | NCH ₃ | 5-H | ~8.1 ± 1.1 |
| 10n | O | 5-Cl | ~2.2 ± 0.5 |
| 10o | S | 5-OCH ₃ | ~8.0 ± 1.6 |

^a A rigorous K_D determination by NMR was only done for the parent compound **10a**. The affinities of the analogues were estimated. Estimations were based on the extent of the perturbations relative to the parent compound when measured at four compound concentrations. It was assumed that the chemical shifts of the ¹⁵N-labeled ns4a–ns3p in the bound form are identical for each analogue.

tation of protein/ligand complexes under conditions that are not favorable to traditional structural work.^{13–15} In the first approach we use NMR Δ CS of the protein to construct a "j-surface", which provides a spatial representation of the most likely position for the aromatic rings of a ligand when it is bound to its protein target.^{13,14} In most cases, this procedure made it possible to rapidly, and with good precision, visualize the binding site of the weakly interacting small-molecule ligands with respect to the surface of ns4a–ns3p. In addition, the j-surface was useful for evaluating the consistency and quality of protein-detected chemical shift perturbations and provided a visual summary of the Δ CS data used to restrain the protein/ligand complex. If the quality of the Δ CS data was sufficient and compatible with a single binding mode of the ligand, as judged by its j-surface. We then used a second approach in which the structures of protein/ligand complexes were determined by minimizing the difference between experimental and simulated chemical shift perturbations (DCS: Δ CS_{exp} – Δ CS_{sim}) that are produced by the ligand when it binds weakly to a protein surface.¹⁵ Recently, this procedure has been implemented as an algorithm in the program Sybyl [Moyna, G.; personal communication], which allows energy minimization of the protein/ligand structure against van der Waals, DCS, and possibly electrostatics.¹³ Using these approaches, we were able to obtain experimentally derived models of the complexes between the optimized NMR hits and ns4a–ns3p for structure-guided chemistry (Figure 3).

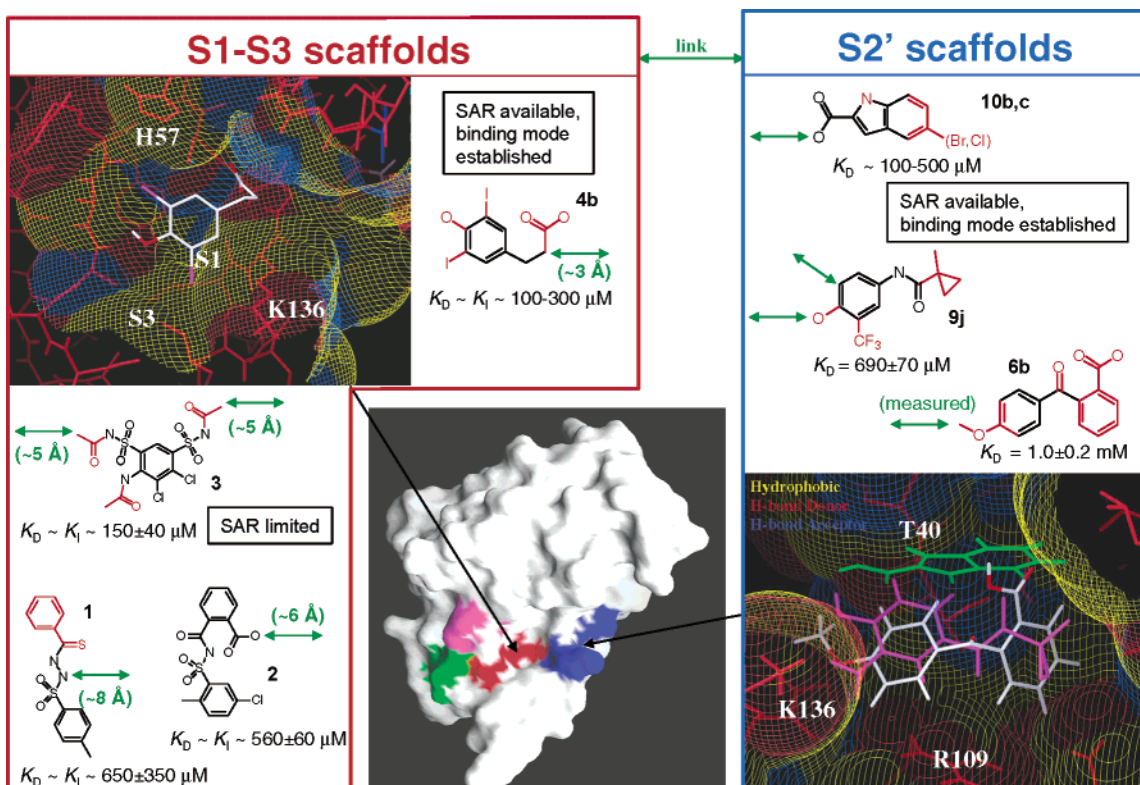


Figure 3. Binding orientation of optimized SbN hits to proximal substrate binding sites of the ns4a–ns3p (see text). Shown in green parentheses are the approximate distances measured. These distances would have to be spanned by an appropriate linker in order to connect the S1–S3 scaffolds with the S2' scaffolds at the respective sites (indicated with green arrows). Note that the binding mode of scaffold **3** is ambiguous and two different models rotated around a vertical axis were used to measure the two indicated distances. Red highlights indicate those parts of the chemical structures for which detailed SAR is available (see Tables 1–4).

Complexes to the Unprimed Side. NMR hits in the “diiodophenol series” (compounds **4a**, **4b**, and **4e**) all showed similar patterns of Δ CS, strongly affecting residues Ala 157, Lys 136, and Ser 138 in the S1–S3 region. Figure 4 shows the *j*-surface for three of the diiodophenol inhibitors. These surfaces clearly indicate the presence of an aromatic ring in S1 for all three compounds. **4a** contains two benzene rings; both rings could potentially generate large shifts when interacting with the protein surface, and either ring (or a combination of rings) could be responsible for the observed shifts. Analogues **4b** and **4e** contain only one ring, yet they show Δ CS values that are similar to those produced by **4a**. On the basis of relating the structural differences of closely related analogues to patterns in the Δ CS, the iodinated ring in all diiodophenol hits was placed in S1. The orientation of the iodo groups is fixed by the aromatic ring orientation in S1 and by the placement of the carboxylate group in the oxyanion hole (Figure 4). The Δ CS values for the diiodotyrosine analogues (such as **4e** in Figure 4) are similar to those of the diiodophenol carboxylic acid analogues; however, Thr 42 and Phe 43 are also perturbed. The model that emerges from these compounds places the *t*-BOC group toward the S2' pocket, which is consistent with our previous placement of the propionic acid group in the diiodophenol carboxylic acid analogues. We have observed some differences in the positioning of the aromatic ring between distinct members of the diiodophenol carboxylic acid analogues and diiodotyrosine analogues. These variations are evident by differences in the ratio of the shifts for residues Lys 136, Ser 138, and Ala 157,

which can vary by as much as a factor of 3. The signs of the shifts, however, do not change, which indicates that the binding mode is very similar for all compounds in this series. Note that the *j*-surfaces for these compounds show extensions of the ligand surface toward Ser 139 of the oxyanion hole (Figure 4) as well as extension of **4e** into the primed side.

Models for the best three hits of the “benzene sulfonamide series” **1–3** (Figure 3), which also bind to S1–S3, were created; however, their binding mode could not be derived from the Δ CS data with confidence. These compounds, when compared to those of the “diiodophenol series”, contained fewer ^1H perturbations that could be used for alignment. Hits **1** and **2** both contain two aromatic rings, which leads to ambiguity as to the correct placement of each ring. The correlation of the limited SAR with Δ CS was also not helpful. Hit **3**, which has only one ring, interacts strongly and nearly exclusively with Ser 139. Since there are no residues near Ser 139 that are strongly perturbed and **3** has a K_I in the 150 μM range, we assume that Ser 139 is the primary binding site and that the strong Δ CS is due to a hydrogen bond acceptor on **3** (there are seven carbonyls). The final placement of **3** is ambiguous, and the model should be used with caution.

Complexes to the Primed Side. Models were also obtained for the optimized NMR hits **6b**, **9j**, and **10b**, all of which interact with the S' side of the enzyme (Figure 3). **6b**, the best hit of the “benzophenone series”, perturbs HNs of residues Thr 40, Thr 42, Ser 138, Arg 109, His 110, Leu 35, Val 226 (from NS4A) and the side chain HNs of Gln 9. **6b** is largely aromatic so the

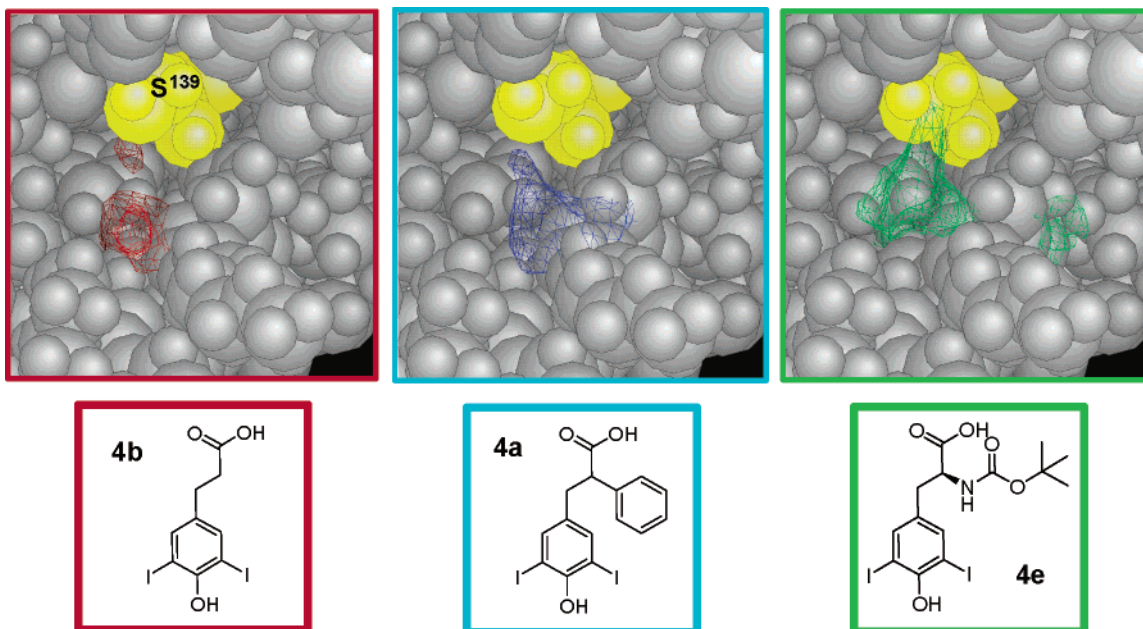


Figure 4. Spatial localization of ligand binding sites from electron current density surfaces¹⁴ calculated from NMR chemical shift perturbations. Ligand *j*-surfaces of closely related analogues in the “diiodophenol” series enable the localization of chemical extensions on the diiodophenol scaffold (see text). Three ligands are shown below their respective *j*-surfaces. Gray atoms are from the HCV NS3 protease domain. The active site Ser139 is shown in yellow. The ligand potency ranges from 80 to 260 μ M. The comparison of the three *j*-surfaces correlates with the diiodophenol binding into the S1–S3 subsites, the propionic acid moiety extending toward the oxyanion hole, and the phenyl ring (blue) and the *t*-BOC group (green) extending toward the S' area of the ns4a–ns3p.

assumption that ring current is the primary source of Δ CS is a good one, and the observed pattern of Δ CS is very sensitive to the relative ring orientation. While the orientation of the benzene rings was easily determined, the rings could not be distinguished. This ambiguity was resolved once the hits of the “anilide series” were characterized. The latter set of compounds have the phenol ring in common with the benzophenone analogues, and they showed nearly the same pattern of shifts around Thr 42 but did not perturb residues Leu 35, Val 226, and Gln 9. The final model of **6b** places the methoxy group of this ring in a hydrophobic environment that is largely produced by the Lys 136 side chain. The two benzene rings form a clamp around the side chain of Thr 42, and the carboxylate extends into solution. The final model of **9j**, the optimized hit in the “anilide series”, has the hydroxyl group pointed toward the S side, the trifluoromethyl group is oriented toward the backbone of Lys 136, the ring interacts with the side chain of Thr 42, and the methylcyclopropyl group extends into the hydrophobic region near the NS4A interface. A model for **10b**, the optimized hit in the “indole series”, and for similar compounds was produced. The face of the ring fits against the 38–42 loop, and the halogen (F, Cl, Br) is modeled as making hydrophobic contact with Leu 13.

Design, Synthesis, and Biological Activity of Linked SbN Hits. Since several classes of compounds had been identified by SbN that bind weakly (millimolar to micromolar range) to proximal substrate binding sites of the ns4a–ns3p, our next goal was to produce linked leads of these weakly binding compounds with significantly improved potency based on the information available from the NMR data.¹¹ In our first attempt to produce nonpeptidic, potent HCV protease inhibitors, we designed a limited set of linked molecules based on

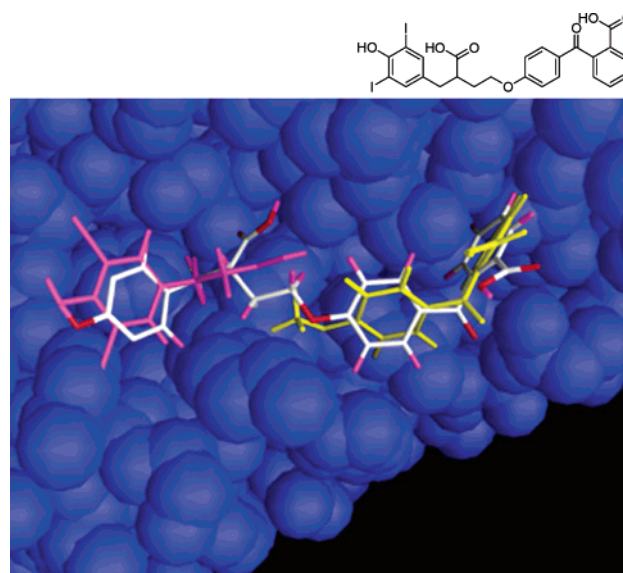
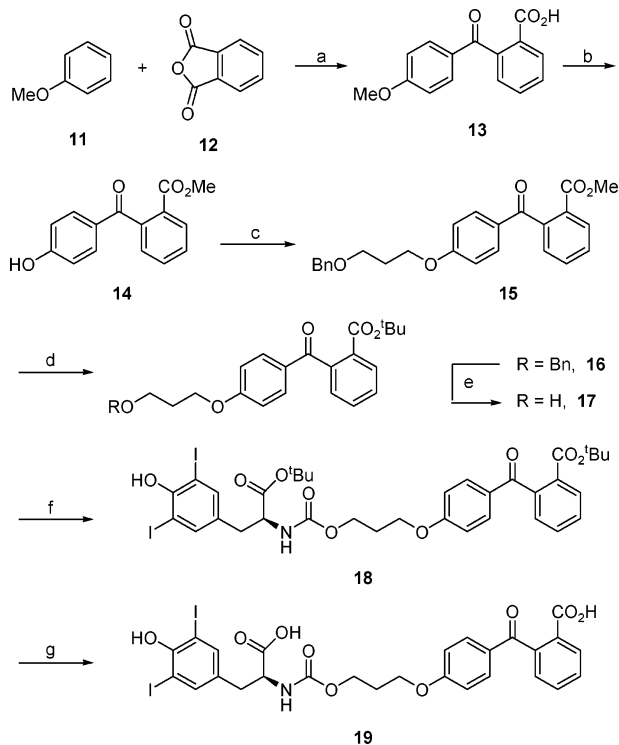


Figure 5. Shown in white is a model of a compound in which a two-methylene bridge was used to link the two “optimized SbN hits”, **4b** and **6b**, whose binding orientations are depicted in magenta and yellow, respectively. The binding orientations of the individual “building blocks”, which were derived from the NMR data (see Figure 3), are similar in the energy-minimized conformation of the linked molecule. This model suggests that, in principle, the individual “SbN hits” could be positioned in their preferred orientations in this bridged molecule, and hence, their intrinsic binding energies may be unchanged upon linking.

the binding orientation of SbN hits in the “diiodophenol series” (e.g., **4b**, **4a**, and **4e** in Figure 4) and **6b**, the optimized SbN hit in the “benzophenone series” (Figure 5), since we could demonstrate the simultaneous binding of these ligands to ns4a–ns3p (Figure 2). The synthesis

Scheme 1. Synthesis of Linked Compound **19**^a

of one of the designed linked molecules, compound **19**, is shown in Scheme 1.

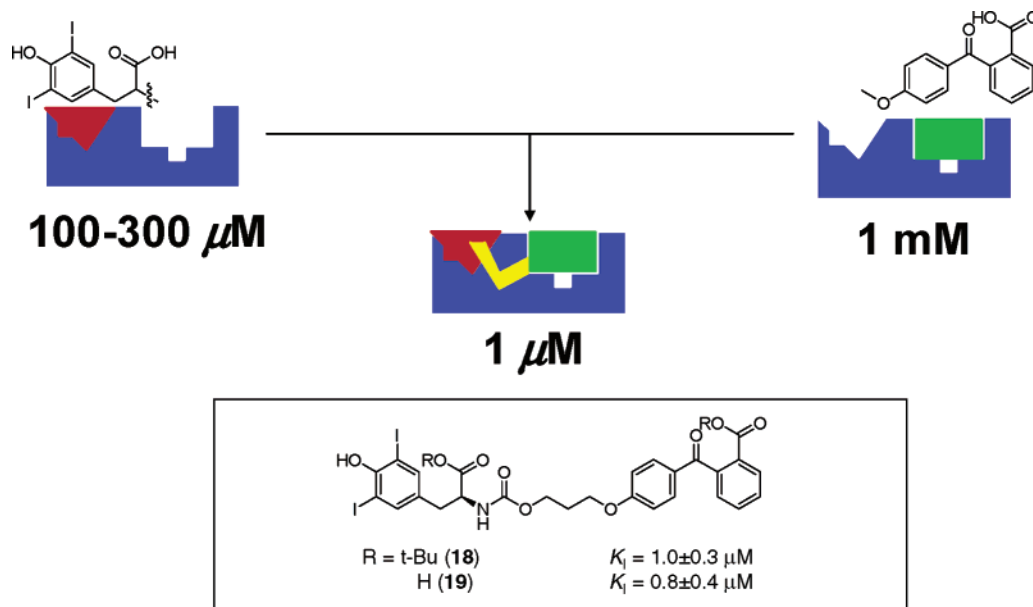
Lewis acid catalyzed Friedel–Craft's acylation of anisole, **11**, with phthalic anhydride, **12**, resulted in the benzophenone derivative **13**. Protecting group adjustments provided the *p*-hydroxybenzophenone derivative **14**. The linker was installed by treatment of **14** with

3-benzyloxypropanol under Mitsunobu protocol. Removal of the benzyl ether protecting group of **16** using transfer hydrogenation conditions afforded the free alcohol **17**. Treatment of the alcohol **17** with 4-nitrophenyl chloroformate provided the activated carbamate (not shown), which was subsequently reacted with diiodotyrosine *tert*-butyl ester resulting in the linked compound **18** as the diester derivative. Deprotection of both the ester functionalities under acidic conditions afforded the linked compound **19** in good yields.

The linked molecule, **19** (Scheme 1), and its diester derivatives **18** showed competitive inhibition with K_I values in the micromolar range when tested in the biochemical assay (Figure 6). Thus, the linked compounds are approximately 200-fold and 1000-fold more potent than the individual unlinked "SbN building blocks" **4e** and **6b**, respectively. This shows that much more potent inhibitors against the HCV protease can be obtained by linking weaker "SbN hits" based on the NMR data, which was first demonstrated by Fesik and co-workers in their pioneering work on stromelysin.¹⁶ Moreover, **19** showed sufficient solubility in the aqueous NMR buffer for NMR binding studies. The pattern of Δ CS values observed for the linked compound was consistent with the proposed model in Figure 5. However, the complex between **19** and ns4a–ns3p was not very stable over time, slowly precipitating out of solution, and attempts to determine its 3D structure by NMR and/or X-ray crystallography were unsuccessful.

Conclusions

NMR-based screening was utilized to identify 16 initial weak ($K_D \approx 100 \mu\text{M}$ to 10 mM) small-molecule hits to substrate binding sites of the NS4A-bound NS3 protease of HCV. Five classes of NMR hits were further explored by a combination of NMR and a biochemical assay yielding, in most cases, optimized hits with improved potency, SAR, and binding mode information. As demonstrated by the potency of the linked compound



Binding of **19** confirmed by NMR

Figure 6. Linking of two SbN fragments that bind to proximal subsites of ns4a–ns3p yields compounds with improved potencies.

19, which incorporates two low-affinity SbN hits to the S1–S3 and S2' substrate binding regions of ns4a–ns3p, respectively, the information presented herein can be utilized by structure-guided chemistry to produce potent nonpeptidic small-molecule inhibitors of the NS4A-bound HCV NS3 protease.

Experimental Section

Production, Purification, and Recycling of ¹⁵N-Labeled ns4a–ns3p. A standard operating procedure was established for large-scale (100 L) routine expression of ¹⁵N-labeled ns4a–ns3p yielding soluble expression levels of approximately 10 mg/L protein. A 2-day protocol was set up for the purification of roughly 150 mg of ¹⁵N-labeled ns4a–ns3p per 20 L of culture. A recycling procedure was developed and implemented to produce active protein for NMR-based screening with a recycling yield of ~42 ± 10%.

BL21(DE3)pLysS cells were transformed with NS4A-tethered NS3 protease (NS4A_{21–32}-GSGS-NS3_{3–181}/I17K, strain 1a),⁹ a single colony was selected, grown in 100 mL (250 mL baffled flask, LB or Special B media containing 25 μg/mL kanamycin and 34 μg/mL chloramphenicol, 37 °C, 300 rpm) for approximately 4–6 h to an OD₆₀₀ of 1.5, and was stored at 4 °C overnight. Cells were then pelleted utilizing a benchtop centrifuge and resuspended in complete minimal media containing ¹⁵NH₄Cl as the nitrogen source. One liter of complete minimal media (2 L baffled flask) was inoculated with the resuspended pellet so that the starting OD₆₀₀ was ~ 0.1. The sample in the flask was subsequently grown at 37 °C (300 rpm) and expanded until 1/10 of the final tank volume was obtained for inoculation.

The sample in the 100 L tank (ABEC-150) was then inoculated, grown, induced, and checked for soluble expression using the above protocol (fermentation parameters rpm = 200, air flow = 32). At harvest, the cells were pelleted, resuspended at 1/10 original volume of lysis buffer (25 mM Hepes, pH 7.4, 1 M NaCl, 10% glycerol, 0.1% *n*-octylglucoside) with 0.2 mM DTT, 5 mL protease inhibitor cocktail III (Calbiochem), and Benzoase (Sigma; final concentration of 14 300 units/L) and frozen in an ethanol/dry ice bath. Pellets were stored at –20 °C until harvest.

Resuspended cells were subsequently thawed in a 30 °C water bath and gently poured into ultracentrifuge tubes and spun at 100000g for 30 min at 4 °C. The supernatant was brought to 25 mM imidazole (pH will shift to ~7.8) and batch-absorbed for 1 h at 4 °C onto 20 mL of Qiagen Ni–NTA agarose resin that had been equilibrated in lysis buffer (with 25 mM imidazole). Resin was packed in a column (Pharmacia XK-26) and washed with 10 bed volumes of lysis buffer (with 25 mM imidazole, pH 7.4, and 0.2 mM DTT). His-tagged ¹⁵N-labeled ns4a–ns3p was eluted with two to three column volumes of elution buffer (lysis buffer with 250 mM imidazole, pH 7.4). Eluted fractions were pooled, and the His tag was removed by addition of human thrombin (15 U/mg based on Biorad protein assay) followed by dialysis against gel filtration buffer (75 mM K₂PO₄, pH 6.5, 0.015% sodium azide containing 1 mM DTT) for 18 h at 4 °C. Dialyzed thrombin cleaved protease was applied to a HiLoad 26/60 Superdex 200 column (Pharmacia) that was pre-equilibrated in gel filtration buffer. Protein-containing fractions were pooled, clarified by centrifugation, and concentrated (final concentration of 4–5 mg/mL) by ultrafiltration on an Amicon stirred cell using a YM-10 membrane. Final protein was determined by A₂₈₀ (extinction coefficient is 17 700 L mol⁻¹ cm⁻¹; MW = 21 173), and samples were submitted for MS analysis and N-terminal sequencing to confirm the protein sequence identity and extent of ¹⁵N-labeling.

Recycling of ¹⁵N-labeled ns4a–ns3p was accomplished by dialyzing 15 mL (~50 mg) of protein/compound mixtures, which were inactive in the NMR assay, against 4 L of gel

filtration buffer. This was accomplished using four 1 L exchanges, and the final material was subjected to high-speed (100000g) ultracentrifugation. The sample was then loaded onto the column (Pharmacia HiLoad 26/60 Superdex 200), protein-containing fractions were pooled (based on BioRad assay), and the final material was concentrated as described above. The yield of this recycling procedure was 42 ± 10%, providing active protein that could be reused for NMR-based screening.

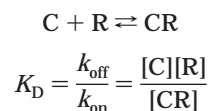
Detection of Ligand Binding Using NMR. Samples for NMR-based screening were composed of 500 μL of uniformly ¹⁵N-labeled ns4a–ns3p at a concentration of 0.2 mM in 75 mM K₂HPO₄/KH₂PO₄ buffer (pH 6.5), 20 mM DTT, 95/5% (v/v) H₂O/D₂O, and 0.015% sodium azide by weight of the sample (w/v). Ligand binding was detected at 25 °C by acquiring 1–2 h of ¹⁵N HSQC spectra in the presence and absence of added compound using a sample changer on either a Bruker DRX500 or a Varian INOVA600 spectrometer. Compounds from a custom-built library were added to the protein NMR sample as DMSO-*d*₆ stock solutions in groups of either 5 or 10 compounds, or individually during deconvolution of active clusters, each depending on their solubility both in DMSO-*d*₆ and aqueous buffer. The final concentration of each compound was either 1 or 0.1 mM, and the final DMSO-*d*₆ concentration of each sample was adjusted to 1 vol %. The pH and appearance of these clusters were measured and recorded. The pH of the samples was within ±0.1 pH unit. For every set of SbN samples prepared, a control reference protein sample containing 1% DMSO-*d*₆ was also prepared.

NMR data processing, hit detection, and chemical shift perturbation analysis were carried out in an automated manner on SGI workstations using custom-built C-shell scripts and macros inside the program FELIX (Accelrys). Chemical shift changes were considered significant, suggesting a hit if the weighted chemical shift difference,

$$WA = \sqrt{(\Delta\delta\{^1H\})^2 + (0.2 \Delta\delta\{^{15}N\})^2}$$

was greater than 0.04 ppm. If the HSQC spectrum of a protein cluster showed chemical shift changes that indicated ligand binding at a desired site of the target protein, the cluster was subjected to deconvolution. Deconvolution involved testing each compound of an active cluster individually. Once a compound was determined to be active, it was analyzed by one-dimensional proton NMR in both DMSO-*d*₆ and aqueous buffer containing a quantitative internal standard. The purpose of these evaluations was to determine (1) if the compound had the correct chemical structure, (2) if it was pure, and (3) the quantitative amount that dissolved into aqueous solution. If these tests revealed an incorrect structure or significant levels of impurities, the compound was subjected to further testing that included mass spectrometry and carbon-13 NMR. Next, *K*_D and/or *K*_I and binding site information were obtained for selected compounds to decide which of the active compounds should be further optimized for chemistry.

Determination of Dissociation Constants by NMR Titration Experiments. The dissociation constants of selected NMR hits were derived from an analysis of the changes in amide chemical shifts of residues in the binding site as a function of the concentration of the compound. Known amounts of the compound were added incrementally to a ¹⁵N-labeled protein sample of known concentration. ¹⁵N HSQC spectra were collected after each addition of compound. The data were then fit using a single binding site model for an interaction of a compound (C) with a protein (R):



This equation can be directly correlated to chemical shifts as

follows:

$$\frac{|\delta - \delta_f|}{|\delta_f - \delta_b|} = \frac{([C]_0 + [R]_0 + K_D) - \sqrt{([C]_0 + [R]_0 + K_D)^2 - 4[C]_0[R]_0}}{2[R]_0} \quad (1)$$

where $[C]_0$ and $[R]_0$ are the total concentrations of compound and protein, respectively, $[CR]$ is the concentration of the complex, δ is the chemical shift of the protein measured at each concentration $[CR]$, δ_f is the chemical shift of the protein in the absence of the compound $[C]_0 = 0$, and δ_b is the chemical shift of the protein at saturation with compound.

Nonlinear regression methods were used to estimate K_D and δ_b in each titration experiment. Data from an experiment consisted of chemical shift (δ) values measured at a number of different compound concentrations. The values of $[C]_0$, $[R]_0$, and δ_f were known. Estimates of K_D and δ_b were computed by fitting the data to eq 1 using nonlinear least squares in the statistical package SAS (SAS Institute Inc, Cary, NC). From a nonlinear fit, estimates of the standard errors were obtained for K_D and δ_b .

A rigorous K_D determination by NMR was only carried out for the most potent analogues of each analogue series based on eight titration points. If significant binding of analogues was observed in the NMR assay, their affinity was estimated on the basis of the extent of the perturbations of three to four titration points relative to the parent compound. It was assumed that the chemical shift of the ^{15}N -labeled ns4a–ns3p in the bound form is identical for each analogue.

Structural Studies Based on Chemical Shift Perturbation NMR Data. Structural models of protein/ligand complexes were determined on the basis of the chemical shift perturbation NMR data using methods as previously described.^{13–15}

Determination of HCV NS3 Protease Inhibition in the Biochemical Assay. Inhibition was evaluated using a peptide-based chromogenic assay as described.¹⁷ The assays were performed at 30 °C in a 96-well microtiter plate. A sample of 100 μL of 20 nM ns4a–ns3p was added to 100 μL of assay buffer (25 mM MOPS, pH 6.5, 20% glycerol, 0.3 M NaCl, 0.05% lauryl maltoside, 5 μM EDTA, 5 μM DTT) containing chromogenic substrate Ac-DTEDVVP(Nva)-O-PAP and inhibitor. The reactions were monitored at an interval of 30 s for 1 h to detect a change in absorbance at 370 nm using a Spectromax Plus microtiter plate reader (Molecular Devices). Enzyme concentration was adjusted to achieve ~12% substrate depletion over the course of the assay. Potency of protease inhibitors was determined at fixed concentrations of enzyme and substrate (40 μM). The inhibition constant K_I was calculated once inhibition was demonstrated to be competitive with substrate, that is,

$$v_i = \frac{V_{\max}S}{K_M(1 + I/K_I) + S}$$

Chemical Synthesis of the Linked Molecule 19. General. All reagents/solvents were purchased from commercial sources and used without further purification. When required, reactions were performed under N_2 or Ar atmosphere. Flash chromatography was carried out using Selecto silica gel (230–400 mesh) or Biotage Flash 40i or FlashElute system with an appropriate size prepackaged silica cartridge. NMR spectra were obtained in CDCl_3 unless otherwise specified.

Compound 13. To a solution of phthalic anhydride (**12**, 7.4 g, 50 mmol, 1.0 equiv) in dichloromethane (250 mL) was added anisole (**11**, 27 mL, 250 mmol, 5.0 equiv). The solution was cooled in an ice bath. Aluminum chloride (13.4 g, 100 mmol, 2.0 equiv) was added portionwise. The ice bath was removed after 10 min and warmed to room temperature over 1 h. The reaction mixture was refluxed overnight (~16 h), cooled to room temperature, and poured carefully into a stirred solution

of ice/1 N HCl (500 mL). The organic layer was separated, and the aqueous layer was extracted with dichloromethane (3 \times 250 mL). The combined organic layer was extracted with cold aqueous 1 N NaOH (700 mL). The aqueous layer was extracted with dichloromethane (1 \times 300 mL). The organic layers were discarded and the cold aqueous basic layer was acidified with concentrated HCl (~70 mL), resulting in a milky white suspension. This was extracted with dichloromethane (4 \times 250 mL), dried (Na_2SO_4), and concentrated to provide compound **13** (9.03 g, 71%) as a white solid. ^1H NMR δ 3.85 (s, 3 H), 6.87–6.89 (m, 2 H), 7.33–7.35 (m, 1 H), 7.52–7.56 (m, 1 H), 7.62–7.70 (m, 3 H), 8.06–8.08 (m, 1 H); LCMS m/z 257.1 (M + H)⁺.

Compound 14. To **13** (5.0 g, 19.53 mmol) in glacial acetic acid (25 mL) was added 48% aqueous HBr (25 mL), and the mixture was refluxed (bath temp = 115–125 °C) overnight (~26 h). The reaction mixture was cooled to room temperature and poured carefully into an ice/water mixture (250 mL). The aqueous layer was extracted with EtOAc (4 \times 200 mL). The combined organic layer was washed with water (400 mL), dried (Na_2SO_4), and concentrated to afford an off-white solid in quantitative yield.

To a solution of the crude material obtained from the above (3.75 g, 15.5 mmol) in MeOH (150 mL) was added concentrated H_2SO_4 (2 mL), and the mixture was refluxed overnight. The reaction mixture was cooled to room temperature and concentrated. The residue was taken in cold water, solid NaHCO_3 was added until the solution became basic, and the layer was extracted with Et₂O (2 \times 200 mL). The combined ether layer was dried (Na_2SO_4) and concentrated. The crude product was recrystallized from hexanes/EtOAc (~1/1) to provide 2.94 g (74%) of compound **14**. ^1H NMR δ 3.67 (s, 3 H), 6.38 (s, 1 H), 6.79–6.82 (m, 2 H), 7.39–7.41 (m, 1 H), 7.54–7.68 (m, 4 H), 8.03–8.06 (m, 1 H); LCMS m/z 257.1 (M + H)⁺.

Compound 15. To a stirred 0 °C solution of triphenylphosphine (1.53 g, 5.85 mmol, 1.5 equiv) in dichloromethane (15 mL) was added DIAD (1.15 mL, 5.85 mmol, 1.5 equiv) dropwise. After a few minutes, 3-benzyloxypropanol (0.75 mL, 4.69 mmol, 1.2 equiv) was added. Finally a slurry of **14** (1.0 g, 3.9 mmol, 1.0 equiv) in dichloromethane (2 \times 5 mL) was added and the reaction mixture was warmed to room temperature over 3 h. The reaction mixture was concentrated and purified by column chromatography (Biotage flash chromatography system using 40M silica cartridge) with 3/97 EtOAc/dichloromethane to afford **15** (1.39 g, 88% yield). ^1H NMR δ 2.11 (m, 2 H), 3.64–3.68 (m, 5 H), 4.15 (t, J = 6.3 Hz, 2 H), 4.53 (s, 2 H), 6.89–6.92 (m, 2 H), 7.38–7.41 (m, 1 H), 7.56–7.64 (m, 2 H), 7.71–7.74 (m, 2 H), 8.04–8.07 (m, 1 H); LCMS m/z 405.1 (M + H)⁺.

Compound 16. To a solution of **15** (3.0 g, 7.43 mmol) in THF/MeOH (15 mL each) was added water (9 mL) followed by aqueous 1 M LiOH (22 mL). The reaction mixture was vigorously stirred for 4 h and concentrated. The residue was taken in aqueous 10% citric acid and extracted with EtOAc (2 \times). The combined organic layer was washed with brine, dried (Na_2SO_4), and concentrated to afford the crude acid.

To a slurry of MgSO_4 (3.57 g, 29.72 mmol) in chloroform (40 mL) was added concentrated H_2SO_4 (0.41 mL), and the mixture was stirred vigorously for 15 min. This slurry was added to the crude acid from the above in chloroform (40 mL). The flask was rinsed with chloroform (20 mL), and the rinse was added to the reaction mixture. Finally *tert*-butyl alcohol (3.55 mL, 37.15 mmol) was added, the reaction flask was tightly closed, and the mixture was stirred vigorously overnight. The reaction was then quenched by adding saturated NaHCO_3 solution, and the mixture was concentrated (to remove most of the organic solvent). The residue was extracted with EtOAc, and the combined EtOAc layer was washed with brine, dried (Na_2SO_4), and concentrated. The crude material was purified by flash chromatography (Biotage flash system using 40M silica cartridge) with 2/98 MeOH/chloroform to provide **16** (3.06 g, 92% yield). ^1H NMR δ 1.25 (s, 9 H), 2.10 (m, 2 H), 3.65 (t, J = 6.0 Hz, 2 H), 4.14 (t, J = 6.2 Hz, 2 H), 4.52 (s, 2 H), 6.88–6.90 (m, 2 H), 7.33–7.35 (m, 1 H), 7.53–

7.59 (m, 2 H), 7.71–7.74 (m, 2 H), 7.99–8.00 (m, 1 H); LCMS m/z 447.1 (M + H)⁺.

Compound 17. Argon was bubbled into a solution of **16** (3.0 g, 6.73 mmol) in EtOH (125 mL) for approximately 1 h. It was then cooled to 0 °C and 10% Pd/C was added portionwise (3 × 1 g) followed by dropwise addition of 1,4-cyclohexadiene (6.3 mL, 67.3 mmol). The reaction mixture was warmed to room temperature over 30 min and heated to 45 °C over 2 h. TLC indicated completion of reaction. The reaction mixture was cooled to room temperature, filtered through a pad of Celite, and concentrated. Purification by flash chromatography (Biotage flash system using 40M silica cartridge) using 5/95 to 20/80 EtOAc/dichloromethane provided **17** (1.2 g, 50% yield). LCMS m/z 357.1 (M + H)⁺.

Compound 18. To a cold (0 °C) solution of **17** (1.0 g, 2.81 mmol, 1.0 equiv) in THF/MeCN (9 mL/1.3 mL) were added 4-nitrophenyl chloroformate (623 mg, 3.09 mmol, 1.1 equiv) and pyridine (dropwise, 0.25 mL, 3.09 mmol, 1.1 equiv). The reaction temperature was maintained between 0 and 5 °C for 30 min, and the mixture was slowly warmed to room temperature over 2 h. At this time, the reaction mixture was diluted with EtOAc (30 mL), washed with water (25 mL) and brine (25 mL), dried (MgSO₄), and concentrated. Purification by flash chromatography (Biotage flash system using 40M silica cartridge) using 0/100 to 20/80 EtOAc/dichloromethane afforded 1.04 g of the 4-nitrophenyl carbonate derivative.

To a 0 °C solution of the activated carbonate (500 mg, 0.95 mmol, 1.0 equiv) in DMF (4 mL) was added a solution of HCl salt of diiodotyrosine *tert*-butyl ester (500 mg, 0.95 mmol, 1.0 equiv) and triethylamine (0.15 mL, 1.05 mmol, 1.1 equiv) in DMF (4 mL + 2 mL). Finally a few crystals of imidazole were added and the reaction mixture was warmed to room temperature overnight. The reaction mixture was diluted with EtOAc, washed with 10% citric acid and brine, dried (Na₂SO₄), and concentrated. The crude material was purified by flash chromatography (Biotage flash system using 40M silica cartridge) with 3/97 to 25/75 EtOAc/dichloromethane to provide compound **18** (165 mg, 20% yield). ¹H NMR δ 1.26 (s, 9 H), 1.43 (s, 9H), 2.13 (m, 2 H), 2.94 (d, *J* = 5.6 Hz, 2 H), 4.08–4.10 (m, 2 H), 4.24–4.30 (m, 2 H), 4.42 (q, 1 H), 5.22 (d, 1 H), 5.68 (s, 1 H), 6.89–6.91 (m, 2 H), 7.32–7.34 (m, 1 H), 7.46 (s, 2 H), 7.51–7.60 (m, 2 H), 7.72–7.74 (m, 2 H), 7.99–8.01 (m, 1 H); ¹³C NMR δ 28.0, 28.5, 29.3, 36.8, 55.4, 62.2, 65.1, 82.5, 83.0, 83.5, 114.6, 127.9, 129.7, 130.4, 131.0, 131.4, 132.3, 132.4, 132.7, 140.5, 141.7, 153.1, 156.0, 163.2, 165.8, 170.5, 196.0; HRMS (FAB) m/z calcd for C₃₅H₄₀I₂NO₉ 872.0793 (M + H)⁺, found 872.0794.

Compound 19. The di-*tert*-butyl ester **18** (50 mg, 0.057 mmol) was treated with dichloromethane/TFA (1 mL each) at room temperature for 2 h and then concentrated. To the residue was added AcOH (1 mL) and concentrated. This procedure was repeated with heptane (2×) to remove all the volatiles. The required product **19** was obtained in quantitative yield as a white solid. ¹H NMR (CD₃OD) δ 2.00–2.15 (m, 2 H), 2.77 (dd, 1H), 3.02–3.10 (m, 1H), 4.1–4.35 (m, 5H), 6.96 (d, 2H), 7.35 (d, 1H), 7.60–7.71 (m, 6H), 8.10 (d, 1H); ¹³C NMR (CD₃OD) δ 29.1, 35.5, 55.6, 61.6, 64.9, 84.1, 114.4, 127.6, 129.6, 130.3, 130.5, 132.0, 132.4, 133.7, 140.4, 154.6, 157.4, 163.6, 168.0, 173.9, 195.0; HRMS (FAB) m/z calcd for C₂₇H₂₄I₂NO₉ 759.9541 (M + H)⁺, found 759.9553.

Acknowledgment. The authors thank Dr. Hugh Eaton for assistance in automating the processing and analysis of NMR data, Mr. Ashwin Ranchod for small-molecule library construction, and Mr. Richard Ingram and Dr. Bruce Malcolm for the determination of the inhibition data. They also thank Dr. Lata Ramanathan, Mrs. Michelle Wendel, and Dr. Jennifer Gesell for assistance in the growth and purification of the protein. They thank the following people for providing analytical support: Dr. Charles McNemar for N-terminal sequencing; Dr. Yan-Hui Liu and Mrs. Shi-Hong Wang for mass

spectrometry data. They are grateful for the assistance of Dr. Bruce Belanger and Mrs. Ferdous Gheyas in the development of the *K_D* determination process, and of Dr. Bryan Marten in the modeling of linked compounds. They also thank Dr. Patricia Weber for her continued support of the project.

Supporting Information Available: ¹H NMR spectra of **18** and **19**. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References

- (1) Lauer, G. M.; Walker, B. D. Hepatitis C virus infection. *N. Engl. J. Med.* **2001**, *345*, 41–52.
- (2) Bartenschlager, R. Hepatitis C virus replicons: potential role for drug development. *Nat. Rev. Drug Discovery* **2002**, *1*, 911–916.
- (3) Reed, K. E.; Rice, C. M. Overview of hepatitis C virus genome structure, polyprotein processing, and protein properties. *Curr. Top. Microbiol. Immunol.* **2000**, *242*, 55–84.
- (4) Tan, S. L.; Pause, A.; Shi, Y.; Sonenberg, N. Hepatitis C therapeutics: current status and emerging strategies. *Nat. Rev. Drug Discovery* **2002**, *1*, 867–881.
- (5) Levin, M. K.; Patel, S. S. Helicase from hepatitis C virus, energetics of DNA binding. *J. Biol. Chem.* **2002**, *277*, 29377–29385.
- (6) (a) Bartenschlager, R.; Ahlborn-Laake, L.; Mous, J.; Jacobsen, H. Kinetic and structural analyses of hepatitis C virus polyprotein processing. *J. Virol.* **1994**, *68*, 5045–5055. (b) Failla, C.; Tomei, L.; De Francesco, R. Both NS3 and NS4A are required for proteolytic processing of hepatitis C virus nonstructural proteins. *J. Virol.* **1994**, *68*, 3753–3760. (c) Lin, C.; Prágai, B. M.; Grakoui, A.; Xu, J.; Rice, C. M. Hepatitis C virus NS3 serine proteinase: trans-cleavage requirements and processing kinetics. *J. Virol.* **1994**, *68*, 8147–8157. (d) Tanji, Y.; Hijikata, M.; Satoh, S.; Kaneko, T.; Shimotohno, K. Hepatitis C virus-encoded nonstructural protein NS4A has versatile functions in viral protein processing. *J. Virol.* **1995**, *69*, 1575–1581.
- (7) (a) Love, R. A.; Parge, H. E.; Wickersham, J. A.; Hostomsky, Z.; Habuka, N.; Moomaw, E. W.; Adachi, T.; Hostomska, Z. The crystal structure of hepatitis C virus NS3 proteinase reveals a trypsin-like fold and a structural zinc binding site. *Cell* **1996**, *87*, 331–342. (b) Barbato, G.; Cicero, D. O.; Nardi, M. C.; Steinkuehler, C.; Cortese, R.; De Francesco, R.; Bazzo, R. The solution structure of the N-terminal proteinase domain of the hepatitis C virus (HCV) NS3 protein provides new insights into its activation and catalytic mechanism. *J. Mol. Biol.* **1999**, *289*, 371–384.
- (8) (a) Kim, J. L.; Morgenstern, K. A.; Lin, C.; Fox, T.; Dwyer, M. D.; Landro, J. A.; Chambers, S. P.; Markland, W.; Lepre, C. A.; O'Malley, E. T.; Harbeson, S. L.; Rice, C. M.; Murcko, M. A.; Caron, P. R.; Thomson, J. A. Crystal structure of the hepatitis C virus NS3 protease domain complexed with a synthetic NS4A cofactor peptide. *Cell* **1996**, *87*, 343–355. (b) Yan, Y.; Li, Y.; Munshi, S.; Sardana, V.; Cole, J. L.; Sardana, M.; Steinkuehler, C.; Tomei, L.; De Francesco, R.; Kuo, L. C.; Chen, Z. 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.* **1998**, *7*, 837–847.
- (9) McCoy, M. A.; Senior, M. M.; Gesell, J. J.; Ramanathan, L.; Wyss, D. F. Solution structure and dynamics of the single-chain hepatitis C virus NS3 protease NS4A cofactor complex. *J. Mol. Biol.* **2001**, *305*, 1099–1110.
- (10) Yao, N.; Reichert, P.; Taremi, S. S.; Prosser, W. W.; Weber, P. C. Molecular views of viral polyprotein processing revealed by the crystal structure of the hepatitis C virus bifunctional protease-helicase. *Structure* **1999**, *7*, 1353–1363.
- (11) (a) Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. Discovering high-affinity ligands for proteins. *Science* **1997**, *278*, 497–499. (b) Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **1996**, *274*, 1531–1534.
- (12) (a) Fejzo, J.; Lepre, C. A.; Peng, J. W.; Bemis, G. W.; Ajay; Murcko, M. A.; Moore, J. M. The SHAPES strategy: an NMR-based approach for lead generation in drug discovery. *Chem. Biol.* **1999**, *6*, 755–769. (b) Pellecchia, M.; Meininger, D.; Dong, Q.; Chang, E.; Jack, R.; Sem, D. S. NMR-based structural characterization of large protein–ligand interactions. *J. Biomol. NMR* **2002**, *22*, 165–173. (c) van Dongen, M.; Weigelt, J.; Uppenberg, J.; Schultz, J.; Wikström, M. Structure-based screening and design in drug discovery. *Drug Discovery Today* **2002**, *7*, 471–478. (d) Moy, F. J.; Haraki, K.; Mobilio, D.; Walker, G.;

- Powers, R.; Tabei, K.; Tong, H.; Siegel, M. M. MS/NMR: a structure-based approach for discovering protein ligands and for drug design by coupling size exclusion chromatography, mass spectrometry, and nuclear magnetic resonance spectroscopy. *Anal. Chem.* **2001**, *73*, 571–581. (e) Stockman, B. J.; Dalvit, C. NMR screening techniques in drug discovery and drug design. *Prog. Nucl. Magn. Reson. Spectrosc.* **2002**, *41*, 187–231. (f) Meyer, B.; Peters, T. *Angew. Chem., Int. Ed.* **2003**, *42*, 864–890. (g) Jahnke, W.; Rüdiger, S.; Zurini, M. Spin label enhanced NMR screening. *J. Am. Chem. Soc.* **2001**, *123*, 3149–3150.
- (13) Wyss, D. F.; McCoy, M. A.; Senior, M. S. NMR-based approaches for lead discovery. *Curr. Opin. Drug Discovery Dev.* **2002**, *5*, 630–647.
- (14) McCoy, M. A.; Wyss, D. F. Spatial localization of ligand binding sites from electron current density surfaces calculated from NMR chemical shift perturbations. *J. Am. Chem. Soc.* **2002**, *124*, 11758–11763.
- (15) McCoy, M. A.; Wyss, D. F. Alignment of weakly interacting molecules to protein surfaces using simulations of chemical shift perturbations. *J. Biomol. NMR* **2000**, *18*, 189–198.
- (16) Hajduk, P. J.; Sheppard, G.; Nettlesheim, D. G.; Olejniczak, E. T.; Shuker, S. B.; Meadows, R. P.; Steinman, D. H.; Carrera, G. M., Jr.; Marcotte, P. A.; Severin, J.; Walter, K.; Smith, H.; Gubbins, E.; Simmer, R.; Holzman, T. F.; Morgan, D. W.; Davidsen, S. K.; Summers, J. B.; Fesik, S. W. Discovery of potent nonpeptide inhibitors of stromelysin using SAR by NMR. *J. Am. Chem. Soc.* **1997**, *119*, 5818–5827.
- (17) Zhang, R.; Beyer, B. M.; Durkin, J.; Ingram, R.; Njoroge, F. G.; Windsor, W. T.; Malcolm, B. A. A continuous spectrophotometric assay for the hepatitis C virus serine protease. *Anal. Biochem.* **1999**, *270*, 268–275.

JM0305117