High Resolution Footprinting of the Hepatitis C Virus Polymerase NS5B in Complex with RNA

Received for publication, March 7, 2007, and in revised form, April 13, 2007 Published, JBC Papers in Press, April 18, 2007, DOI 10.1074/jbc.M701973200

Jérôme Devai,1 Claudia M. D’Abramo,2 Zhuojun Zhao,3 Suzanne McCormick4 Dimitrios Coutsinos4 Sonja Hess5,6 Mamuka Kvaratskhelia4, and Matthias Götte4§‡

From the 1Departments of Microbiology and Immunology, 4Medicine, and 1Biochemistry, McGill University, Montreal, Quebec H3A 2B4, Canada, the 2College of Pharmacy, Center for Retrovirus Research and Comprehensive Cancer Center, Ohio State University, Columbus, Ohio 43210, and 5,6NIDDK, National Institutes of Health, Bethesda, Maryland 20892

The nucleic acid binding channel of the hepatitis C virus RNA polymerase remains to be defined. Here we employed complementary footprinting techniques and show that the enzyme binds to a newly synthesized duplex of approximately seven to eight base pairs. Comparative analysis of surface topologies of free enzyme versus the nucleoprotein complex revealed certain lysines and arginines that are protected from chemical modification upon RNA binding. The protection pattern helps to define the trajectory of the nucleic acid substrate. Lys81, Lys98, Lys100, Lys106, Arg158, Arg386, and Arg394 probably interact with the bound RNA. The selective protection of amino acids of the arginine-rich region in helix T points to RNA-induced conformational rearrangements. Together, these findings suggest that RNA-protein interaction through the entire substrate binding channel can modulate intradomain contacts at the C terminus.

Hepatitis C virus (HCV)3 infection is a serious public health concern that affects about 170 million people worldwide (1, 2). HCV belongs to the Flaviviridae family, which comprises other human pathogens, such as dengue virus, West Nile virus, as well as yellow fever virus. The single (plus)-stranded HCV RNA genome encodes a polyprotein, which is processed into several smaller mature structural proteins, including the capsid protein (C), the envelope proteins (E1 and E2), and the nonstructural (NS) proteins (NS2, NS3, NS4A, NS4B, NS5A, and NS5B). Initial cis cleavage through NS2–NS3 releases the NS3 protein, which, in turn, continues to process the precursor (for a review, see Ref. 3).

NS5B is a 65-kDa RNA-dependent RNA polymerase that is capable of initiating RNA synthesis de novo, in the absence of a primer (4–7). However, the detailed mechanism of de novo initiation remains elusive. Crystallographic snapshots of initial stages of the reaction that include ternary or quaternary complexes of NS5B with a bound RNA template and nucleotide substrates are not yet available. Modeling studies, on the basis of structures of the apoenzyme, suggest that the enzyme must undergo extensive conformational changes to accommodate the newly synthesized double-stranded RNA (4, 8). The structure of HCV NS5B is reminiscent of a human right hand, although, in contrast to many other DNA or RNA polymerases with a similar fold, extensive interaction between the “thumb” and the tips of the “fingers” subdomains encircle the active site of NS5B (8–10). The NTP substrates can enter through a well defined tunnel. A β-hairpin or flap that comprises residues Leu443 to Ile454 and the C-terminal tail of the protein are both located in the vicinity of the active site and may cause steric conflicts with the newly synthesized double-stranded primer-template (4, 8). At the same time, it should be noted that the precise positioning of the C terminus remains to be defined, because most of the structural and biochemical studies were performed with C-terminally truncated proteins to facilitate expression and purification procedures.

HCV NS5B has also been co-crystallized with GTP (11). In addition to the active site, the authors show that GTP can bind to the thumb domain in close proximity to the fingertip Δ1 loop that is involved in interdomain interactions between the thumb and fingers. Although this allosteric GTP site lies 30 Å away from the polymerase active site, it is implicated in the regulation of dynamic interactions between the fingers and thumb subdomains (12, 13). A complex of NS5B with a short RNA oligomer (rU5) provides structural information with regard to interaction between the 5’-end of the single-stranded primer-template at the entrance channel of the substrate (14). The 3’-end of the RNA substrate is seen in the vicinity of the active site, and it is evident that binding of the short oligomer does not affect the overall fold of the enzyme. However, it remains to be seen whether longer, biologically relevant substrates affect the conformation of the enzyme and how such substrates are accommodated. Structural information in this regard may provide the basis for a better understanding of both initiation and elongation processes.

Here we employed two complementary footprinting methods that were designed to characterize the RNA-protein inter-
RNA Binding to HCV Polymerase

face. RNA footprinting revealed that the NS5B enzyme is in contact with approximately seven to eight newly synthesized residues, and mass spectrometric nucleoprotein footprinting enabled us to identify specific lysines and arginines that are involved in RNA binding. Moreover, the protection pattern in the thumb subdomain points to structural rearrangements that affect intradomain interactions at the C-terminal region of the NS5B enzyme.

EXPERIMENTAL PROCEDURES

Chemicals and Nucleic Acid—The RNA template sequences used in this study were 5'-AACAGUUUCUUU-CUCUC-3' (T20-RNA), and 5'-AACTTTTAGTCTTTTT-7TTTTTTTCCTCC-3' (T27-DNA), with the underlined base as the unique site of chain termination. The dinucleotide 5'-GG served as a primer. Both RNA and DNA oligonucleotides were chemically synthesized and purified on 12% polyacrylamide, 7 M urea gels, with the underlined base as the unique site of chain termination. The dinucleotide 5'-GG served as a primer. Both RNA and DNA oligonucleotides were chemically synthesized and purified on 12% polyacrylamide, 7 M urea gels, 5'-End labeling of the GG primer was conducted with [γ-32P]ATP and T4 poly nucleotide kinase according to the manufacturer's recommendations (Invitrogen). NTPs were purchased from Roche Applied Science, and 3'-dCTP was purchased from TriLink Biotechnologies. N-Hydroxysuccinimidobiotin (NHS-biotin) and p-hydroxyphenylglyoxal (HPG) were purchased from Pierce and freshly resuspended in solution prior to use.

Expression and Purification of HCV NS5B—The HCV NS5B sequence (BK) inserted into the expression vector pET-22 (Novagen) was expressed as a C-terminal truncated enzyme (NS5BΔ21) in Escherichia coli BL21(DE3) and purified utilizing a combination of metal ion affinity and ion exchange chromatography (4). All mutant enzymes were generated through site-directed mutagenesis using the Stratagene QuikChange kit according to the manufacturer's protocol. All sequences were confirmed by sequencing at the McGill University and Genome Quebec Innovation Centre.

RNA Synthesis and Rescue Assay—Standard chain termination reaction mixtures consisted of 1 μM RNA template, 1 μM HCV NS5B, and 0.25 μM radiolabeled GG primer in 40 mM HEPES, pH 8, buffer containing 10 mM NaCl, 1 mM dithiothreitol, and 0.2 mM MnCl2. Reactions were started by adding 10 μM GTP plus ATP and 3 μM 3'-dCTP with T20-RNA template and typically incubated for 30 min at room temperature. For the rescue reaction, 125 μM PpN, 100 μM CTP, and 5 μM 3'-dUTP were added to the chain termination mix for 30 min. The reactions were stopped and precipitated with isopropyl alcohol, heat-denatured for 5 min at 95 °C, and resolved on 12% polyacrylamide, 7 M urea gels.

Ribonuclease H (RNase H) Protection Assay—RNA/DNA duplexes were generated by mixing together 1 μM HCV NS5B and 0.25 μM radiolabeled GG primer in 40 mM HEPES, pH 8, buffer containing 10 mM NaCl, 1 mM dithiothreitol, and 0.2 mM MnCl2. Reactions were started by adding 10 μM GTP plus ATP and 3 μM 3'-dCTP with T27-DNA template. After chain termination, the footprinting reaction was initiated by adding 0.015 units of E. coli RNase H (Fermentas) with 8 mM MgCl2, incubated at 37 °C for 2 and 5 min.

![Image](313x333 to 563x733)

**FIGURE 1.** RNase H protection assay. **A,** principle of the reaction. In the presence of a 27-mer DNA template (solid line), NS5B (1 μM) extends a radiolabeled GG dinucleotide primer to the site of chain termination at position 19. The subsequent addition of E. coli RNase H to the product of the primer extension causes degradation of the RNA strand (dotted line) of free RNA/DNA duplex. **B,** autoradiogram of the RNase H footprinting experiment. The control lane (C) is the product of chain termination prior to nuclease reaction. Other lanes represent the same chain termination products incubated directly with RNase H (lane —), or after protein denaturation at 70 °C for 10 min (lane +). The region protected by NS5B is indicated.

Surface Topology Analysis of the Nucleoprotein Complex—The surface topology analyses of lysine and arginine residues were conducted as described previously (15, 16). Briefly, in parallel experiments, the preformed nucleoprotein complex comprising 1 μM purified NS5B and equimolar 20-mer RNA T20-RNA was compared with the free protein. Surface accessibility of lysine residues was monitored using 100 μM NHS-biotin in 40 mM HEPES, pH 8, buffer containing 10 mM NaCl, 1 mM dithiothreitol, and 0.2 mM MnCl2. NHS-biotin reacts specifically with primary amines, resulting in the covalent addition of a biotin molecule (226 Da) to lysine residues with the concomitant release of N-hydroxysuccinimide. The biotinylation reactions were carried out at 25 °C for 40 min and were terminated by the addition of 10 mM (final concentration) lysine in its free amino acid form. To monitor surface accessibility of arginine...
residues, NS5B-RNA complexes were preformed and subjected to HPG modification in 50 mM HEPES, 50 mM boric acid, pH 8.0, buffer containing 10 mM NaCl, 1 mM dithiothreitol, and 0.2 mM MnCl2. HPG reacts with guanidino groups on arginines, resulting in the 132-Da mass increment. The reactions were carried out at 30 °C for 40 min in the dark and were quenched with 10 mM (final concentration) arginine in its free amino acid form.

Following treatments, the reaction mixtures were subjected to SDS-PAGE, and NS5B was visualized by Coomassie Blue staining. The band of interest was excised and extensively destained in 50% methanol, 10% acetic acid. SDS was removed by washing the gel pieces with ammonium bicarbonate, dehydrated with 100% acetonitrile, and vacuum-desiccated. Samples were digested with 1 μg of trypsin (Roche Applied Science) in 50 mM ammonium bicarbonate overnight at room temperature. The supernatant was then recovered for MS and MS/MS analysis.

**RESULTS**

**NS5B Protects RNA from E. coli RNase H Cleavage**—In order to determine the size of the newly synthesized RNA that interacts with NS5B during RNA synthesis, we used E. coli RNase H as a tool to obtain footprints of the enzyme on its nucleic acid substrate. RNase H cleaves the RNA moiety of RNA/DNA hybrids (17), and the NS5B enzyme was shown to be able to synthesize RNA on a single-stranded DNA template, which is a prerequisite for this approach (18). In contrast to most nucleases that cleave double-stranded RNA, E. coli RNase H cleaves randomly, almost after each residue in the RNA/DNA duplex region. To initiate RNA synthesis, we employed a short radio-
labeled dinucleotide RNA primer (Fig. 1A). The reaction can be specifically blocked through incorporation of a 3′-dCMP chain terminator at position 19, which generates a stable, arrested elongation complex (Fig. 1B, control lane (C)). The nucleoprotein interaction can be destabilized at higher temperatures, which renders the RNA/DNA duplex fully accessible to RNase H cleavage (Fig. 1B, lane +). In contrast, non-denaturing conditions limit the RNase H-mediated cleavage to the 5′-end of the template and reveal a protected region of 11 residues (Fig. 1B, lane −). Thus, the NS5B enzyme appears to interact with approximately seven to eight base pairs of a newly synthesized RNA/DNA duplex, considering that *E. coli* RNase H interacts with at least four RNA residues downstream from the scissile bond (19). The crystal structure of NS5B in complex with the short rU₃ fragment suggests that the enzyme is also in contact with approximately five nucleotides of the template 5′-overhang (14). These measurements are consistent with recent modeling studies (4, 8); however, the trajectory of the bound nucleic acid and putative conformational changes upon RNA binding remain to be characterized.

Protein Footprinting—We employed a mass spectrometric protein footprinting approach to identify specific amino acids that interact with the RNA substrate. This method was proven successful in the characterization of a variety of protein nucleic acid interactions and consistently revealed functionally important amino acid contacts (15, 16, 20, 21). In our studies, lysines and arginines were chemically modified with NHS-biotin and hydroxyphenylglyoxal (HPG), respectively, to compare surface topologies of NS5B and the NS5B-RNA complex. For these experiments, we utilized the natural RNA substrate, since the specific contacts between DNA and NS5B may be different.

The first important step to successfully adopt this method to characterize the NS5B-RNA interface was to determine optimal concentrations of the modifying reagents under which the integrity of the preassembled complex would be preserved. For this purpose, we examined RNA synthesis in the presence of increasing concentrations of NHS-biotin and HPG, respectively. The specific set up allowed us to assess the enzymatic activity at two stages (Fig. 2A). NS5B initiates RNA synthesis from a GG primer at the 3′-end of the RNA template. The elongation of the primer stops when the enzyme incorporates a chain terminator at position 16. It is then possible to rescue RNA synthesis to generate a full-length product (20-mer) through the addition of pyrophosphate (PP₃), which promotes the excision of 3′-dCMP (22). Fig. 2 indicates that treatment of NS5B (Apo−E) with increasing concentrations of NHS-biotin or HPG, prior to the addition of the RNA template, results in the progressive decrease in the generation of the chain-terminated product (Fig. 2, B and C). In contrast, the preformed elongation complex with the incorporated chain terminator is significantly more resistant to inactivation by the modifying agents. The rescue of chain-terminated RNA synthesis is blocked at higher concentrations of NHS-biotin or HPG. Thus, functionally important lysines and arginines appear to be modified, which, in turn, impedes RNA binding to the isolated enzyme. In contrast, in the context of the nucleoprotein complex, the same amino acids are shielded from chemical modification through preformed protein-nucleic acid contacts. Fig. 2 indicates that 100 μM NHS-biotin and 200 μM HPG did not significantly disrupt the preformed NS5B-RNA complex, whereas the same treatment of the apoenzyme completely blocked the initiation of RNA synthesis. Therefore, these concentrations of modifying reagents were chosen to compare surface topologies of the isolated enzyme and NS5B prebound to a 20-mer RNA template.

### TABLE 1

| Conjugating agent | Fragment | Modified amino acid(s) | Protection* |
|-------------------|----------|------------------------|-------------|
| NHS-biotin⁵       | 51–56    | Lys⁵¹                 | +           |
|                   | 70–74    | Lys⁵²                 | −           |
|                   | 80–90    | Lys⁵¹                 | +           |
|                   | 91–100   | Lys⁵⁹/Lys⁶⁰         | +           |
|                   | 101–109  | Lys⁵⁰⁵¹             | +           |
|                   | 142–154  | Lys⁵¹                 | −           |
|                   | 155–158  | Lys⁵                                                   |
|                   | 201–211  | Lys⁴¹                 | +           |
|                   | 212–222  | Lys⁴²                 | −           |
|                   | 251–259  | Lys⁴³                  | +           |
|                   | 260–277  | Lys⁴⁰                  | −           |
|                   | 491–498  | Lys⁴¹                 | +           |
|                   | 509–517  | Arg⁴⁰                 | −           |
|                   | 565–570  | Lys⁴⁵                  | −           |
| HPG⁶             | 33–48    | Arg⁴³                  | −           |
|                   | 52–65    | Arg⁴⁶                  | −           |
|                   | 152–158  | Arg⁴⁴                 | +           |
|                   | 155–168  | Arg⁴⁶                 | −           |
|                   | 255–270  | Arg⁴⁴                 | −           |
|                   | 278–280  | Arg⁴⁷                 | −           |
|                   | 380–386  | Arg⁴⁰                  | −           |
|                   | 381–394  | Arg⁴⁰                  | −           |
|                   | 387–401  | Arg⁴⁰                  | −           |
|                   | 499–503  | Arg⁴⁰                  | −           |
|                   | 502–505  | Arg⁴⁰                  | −           |
|                   | 504–508  | Arg⁴⁰                  | −           |
|                   | 509–517  | Arg⁴⁰                  | −           |

* a, residues that are shown to be protected from modification in the presence of RNA; −, residues readily modified in the presence and absence of RNA.

⁵ Total of 13 peptide fragments and eight protected ones.

⁶ Total of 14 peptides and six protected ones.

---

**FIGURE 3.** Mass spectrometric analysis of biotinylated NS5B peptide fragments. A, a representative segment of MALDI-TOF spectrum showing tryptic fragments of NS5B. Monoisotopic resolution was observed for all of the peptide peaks aiding assignment of unmodified and modified peaks to the NS5B sequence. Conventional amino acid sequence numbering for NS5B is used to indicate the peptide fragments. B, a representative segment of Q-TOF spectra depicting a doubly charged biotinylated NS5B peptide fragment (left). The molecular mass of this ion corresponds to peptide 101–109 plus biotin. MS/MS analysis of the parent biotinylated peptide ion shown confirms that K106 is biotinylated. The y- and b-ion peaks derived from internal fragmentation of peptide bonds provide amino acid sequence information read from peptide C and N terminus, respectively. The mass increment between y3 and y4 ions corresponds to a biotinylated lysine, whereas the masses of preceding and following y ions assign to 1⁰⁰°FYGYAKDVR¹⁰⁹ of the NS5B sequence. C, modifications of lysines in the presence of NHS-biotin. Representative segments of the MALDI-TOF data are depicted. Treatment of free NS5B with NHS-biotin results in modification of Lys⁵¹, yielding the peak corresponding to the 80–90 peptide containing biotinylated Lys⁵¹. In NS5B-RNA complex, this peak is significantly diminished due probably to direct interaction with the nucleic acid. In contrast, the modified 142–154 (Lys⁵¹) biotinylated species in the nucleoprotein complex, indicating that Lys⁵¹ remains surface-exposed in the complex. Unmodified NS5B peptide peaks C1–C4 serve as internal controls. D, modifications of arginines in the presence of HPG. The peak containing modified Arg⁵² is significantly diminished in the NS5B-RNA complex, whereas Arg⁵⁰⁵ remains susceptible to HPG modification in the nucleoprotein complex. Unmodified NS5B peptide peaks C5–C9 serve as internal controls.
MS and MS/MS Analysis of Modified Peptides within the NS5B-RNA Complex

To generate small peptide fragments amenable to MS and MS/MS analyses, treated free NS5B and NS5B-RNA complex were subjected to SDS-PAGE, and the purified protein bands were excised and exposed to in-gel proteolysis. Representative MS and MS/MS profiles are illustrated in Fig. 3. A and B show typical MALDI-TOF data and a Q-TOF spectrum of biotin-modified peptides, respectively. Internal fragmentation of the precursor ion primarily yielded y-ions, providing sequence information read from the C to N terminus. Our data indicate that 13 lysines and 14 arginines are readily susceptible to modification in the free protein (Table 1). These observations are consistent with the surface accessibility data for these basic residues, available from the crystal structures of NS5B (8–10, 14).

Comparative surface topology studies of free protein versus the nucleoprotein complex revealed lysines and arginines shielded from modification in the presence of RNA. Representative segments of the mass spectra showing protection of specific basic residues in the presence of bound RNA are depicted in Fig. 3C. Lys81, Lys100, and Lys106 are protected from biotinylation in the presence of RNA, whereas Lys151 remains solvent-accessible in the presence and absence of RNA. Fig. 3D shows two typical HPG-modified peptides containing Arg505 (protected) and Arg503 (unprotected). A summary of all modified residues is given in Table 1. Of the 13 lysines and 14 arginines readily susceptible to modification in the free protein, eight lysines and six arginines were protected in the NS5B-RNA complex.

Positioning of Modified Lysines and Arginines on the Surface of NS5B

The peptides detected by MALDI-TOF analysis cover about 90% of the sequence of NS5B (Fig. 4A). The modified residues were found in each subdomain of the protein exposed to solvent. The majority of protected residues are located within the fingers and thumb subdomains that are known to shape the nucleic acid binding channel of RNA and DNA polymerases. The structure of NS5B reveals that the protected residues, with the exception of Arg501 and Arg505, are all clustered at the side of the protein that is facing the active site within the palm. Thus, this region appears to define the surface area for RNA binding (Fig. 4B). In contrast, the basic residues that were not protected in our experiments are randomly spread on the outer surface of the protein.

FIGURE 4. Summary of the MS footprinting results in the context of NS5B structure. A, mapping of the modified residues (lysine and arginine) along the amino acid sequence of NS5B. The assigned MS peptide peaks cover 90% of the NS5B sequence (in black). The segments that could not be detected by MS are in gray. Unprotected modified residues in free NS5B are highlighted with a purple circle. The residues significantly protected from modification in the nucleoprotein complex are shown in red circles. Underlined sequences refer to NS5 subdomains: fingers in blue, palm in orange, thumb in green, and C terminus in yellow. Only the secondary structure elements (plain box) and loops (open box) that are discussed in this work are represented. B, NS5B crystal structure (Protein Data Bank code 1NB7), with the same color code for subdomains as in A. The side chains of protected and unprotected amino acids are displayed in red and purple, respectively.
These residues in RNA binding. Two selected residues, Lys212, Lys100, 106, 158, 386, 394, and 491. These residues are distantly involved in the wild-type enzyme. In particular, the effect on enzymatic activity in NS5B resulted in a 75–95% loss of activity, compared with the wild-type enzyme. In particular, the effect on enzymatic activity in NS5B resulted in a 75–95% loss of activity, compared with the wild-type enzyme. In particular, the effect on enzymatic activity in NS5B resulted in a 75–95% loss of activity, compared with the wild-type enzyme. In particular, the effect on enzymatic activity in NS5B resulted in a 75–95% loss of activity, compared with the wild-type enzyme. In particular, the effect on enzymatic activity in NS5B resulted in a 75–95% loss of activity, compared with the wild-type enzyme.

The protection pattern of the footprinting analyses helps to define the trajectory of the exiting RNA product. Lys81, Lys106, Arg386, and Arg491 are protected in the presence of RNA, and these residues are seen in close proximity, within 4 Å, to the 5’-end of the RNA. Lys270 is likewise protected; however, the distance of ~12 Å between the primary amino group and the bound RNA appears to be too long to support any direct interaction with the substrate. In this case, the protection from the modifying reagent is likely to be attributable to conformational changes upon RNA binding. Structures of HCV NS5B in the absence of RNA show Thr122 of the A1 loop within 6 Å of the amino group of Lys270, whereas the structure of the binary complex points to a shortened distance that shows the hydroxyl and amino groups of Thr122 and Lys270 within 4 Å. Thus, our footprinting results are in good agreement with the available crystallographic data and suggest that the intradomain interaction between Thr122 and Lys270 is stabilized in the presence of the nucleic acid substrate.

The protection pattern of the footprinting analyses helps to define the trajectory of the exiting RNA product. Lys81, Lys106, Arg386, and Arg491 are protected in the presence of RNA, and these residues are seen in close proximity, within 4 Å, to the 5’-end of the RNA. Lys270 is likewise protected; however, the distance of ~12 Å between the primary amino group and the bound RNA appears to be too long to support any direct interaction with the substrate. In this case, the protection from the modifying reagent is likely to be attributable to conformational changes upon RNA binding. Structures of HCV NS5B in the absence of RNA show Thr122 of the A1 loop within 6 Å of the amino group of Lys270, whereas the structure of the binary complex points to a shortened distance that shows the hydroxyl and amino groups of Thr122 and Lys270 within 4 Å. Thus, our footprinting results are in good agreement with the available crystallographic data and suggest that the intradomain interaction between Thr122 and Lys270 is stabilized in the presence of the nucleic acid substrate.

The protection pattern of the footprinting analyses helps to define the trajectory of the exiting RNA product. Lys81, Lys106, Arg386, and Arg491 are protected in the presence of RNA, and these residues are seen in close proximity, within 4 Å, to the 5’-end of the RNA. Lys270 is likewise protected; however, the distance of ~12 Å between the primary amino group and the bound RNA appears to be too long to support any direct interaction with the substrate. In this case, the protection from the modifying reagent is likely to be attributable to conformational changes upon RNA binding. Structures of HCV NS5B in the absence of RNA show Thr122 of the A1 loop within 6 Å of the amino group of Lys270, whereas the structure of the binary complex points to a shortened distance that shows the hydroxyl and amino groups of Thr122 and Lys270 within 4 Å. Thus, our footprinting results are in good agreement with the available crystallographic data and suggest that the intradomain interaction between Thr122 and Lys270 is stabilized in the presence of the nucleic acid substrate.

The protection pattern of the footprinting analyses helps to define the trajectory of the exiting RNA product. Lys81, Lys106, Arg386, and Arg491 are protected in the presence of RNA, and these residues are seen in close proximity, within 4 Å, to the 5’-end of the RNA. Lys270 is likewise protected; however, the distance of ~12 Å between the primary amino group and the bound RNA appears to be too long to support any direct interaction with the substrate. In this case, the protection from the modifying reagent is likely to be attributable to conformational changes upon RNA binding. Structures of HCV NS5B in the absence of RNA show Thr122 of the A1 loop within 6 Å of the amino group of Lys270, whereas the structure of the binary complex points to a shortened distance that shows the hydroxyl and amino groups of Thr122 and Lys270 within 4 Å. Thus, our footprinting results are in good agreement with the available crystallographic data and suggest that the intradomain interaction between Thr122 and Lys270 is stabilized in the presence of the nucleic acid substrate.

DISCUSSION

Like other viral RNA polymerases, HCV NS5B is capable of initiating RNA synthesis de novo in the absence of a primer (4, 6, 23, 24). The crystal structures of HCV NS5B and structures of the related BVDV and phage 6φ enzymes show an obstructed active site that cannot accommodate a double-stranded nucleic acid substrate (10, 25). For HCV NS5B, the primer exit is blocked by the C-terminal tail and the β-hairpin (4, 8). Therefore, NS5B must undergo major conformational changes in order to accommodate the double-stranded RNA product. The mechanism of this process remains elusive, largely because the interaction between NS5B and its nucleic acid substrate is underexplored.

In this study, we conducted RNA footprinting experiments with the double strand-specific E. coli RNase H, and found that the enzyme interacts with approximately seven to eight nucleotides of the primer. These findings are in good agreement with the predicted length of the bound duplex region on the basis of modeling studies with the apoenzyme (8). Moreover, RNase H-mediated cleavage of the released RNA/DNA duplex provides direct evidence to show that NS5B does not unwind the newly synthesized primer strand from its complementary template (9, 26). Intracellular strand displacement of the viral genome probably requires the contribution of other viral and/or cellular partners of the replication complex, including the helicase domain of NS3 (27, 28).

To identify residues of NS5B that play important roles in nucleoprotein interactions, we used lysine- and arginine-modifying reagents and examined surface topologies of free protein versus the complex through mass spectrometric analyses. Our detailed footprinting approach allowed us to identify several amino acids protected from modification in the nucleoprotein complex by direct contacts with RNA or protein conformational change induced upon RNA binding. Of 27 lysines and arginines (13 lysines plus 14 arginines) accessible to chemical modifications in the free protein, 14 were shielded in the presence of RNA. The importance of these basic residues for NS5B enzymatic functions was confirmed by mutational analysis.

Fig. 6A illustrates the interactions between a short 4-mer RNA fragment and the template entrance channel of NS5B on the basis of the crystal structure of the binary complex (14). Lys98 and Lys106 are protected from chemical modification, and these residues are seen in close proximity, within 4 Å, to the 5’-end of the RNA. Lys270 is likewise protected; however, the distance of ~12 Å between the primary amino group and the bound RNA appears to be too long to support any direct interaction with the substrate. In this case, the protection from the modifying reagent is likely to be attributable to conformational changes upon RNA binding. Structures of HCV NS5B in the absence of RNA show Thr122 of the A1 loop within 6 Å of the amino group of Lys270, whereas the structure of the binary complex points to a shortened distance that shows the hydroxyl and amino groups of Thr122 and Lys270 within 4 Å. Thus, our footprinting results are in good agreement with the available crystallographic data and suggest that the intradomain interaction between Thr122 and Lys270 is stabilized in the presence of the nucleic acid substrate.

The protection pattern of the footprinting analyses helps to define the trajectory of the exiting RNA product. Lys81, Lys106, Arg386, and Arg491 are protected in the presence of RNA, and each of these residues is located in the vicinity of a modeled, double-stranded nucleic acid substrate (Fig. 6B). Lys81 is part of a long loop between α-helices C and D, and Lys106 is in the putative RNA-contacting α-helix E that could interact with the primer (8). Peptides containing Lys81 or Lys106 can be cross-linked to the bound RNA, which is consistent with our data (29). Arg386 and Arg491 are located within a region that includes α-helix O, which appears to contact newly synthesized RNA in close proximity to the active site. Superimposing the structures of NS5B and HIV-1 RT in complex with primer-template provided support for this notion (4, 8); however, the trajectory of the double-stranded RNA, which is reasonably well defined through Lys81, is difficult to predict using only the molecular modeling approach.

The data also show protection of residues Arg51, Arg154, Lys155, Arg278, and Arg491, all of which are located in close proximity to the cavity forming the active site of NS5B (Fig. 6C).
FIGURE 6. Positioning of RNA in the tridimensional structure of NS5B. A, surface representation of the RNA entrance channel of NS5B (Protein Data Bank code 1NB7), in complex with a 4-mer oligonucleotide (yellow). Only the surface of residues within 20 Å of the 3’-end of the RNA is displayed. The three lysines Lys98, Lys100, and Lys270 that were detected by mass spectrometry footprinting are represented in red. The catalytic triad (Asp220, Asp318, and Asp319) is atom-colored, and the two Mn²⁺ are in brown. Inset, ribbon and stick representation of proximity between Lys270 (red) and Thr12 (colored by atom). B, proposed interaction between NS5B and long nucleic acid. This model was constructed by superimposing NS5B with HIV-1 reverse transcriptase (Protein Data Bank code 1HYS) (37). In order to fit appropriately the RNA and avoid steric clashes with NS5B, we excluded the C terminus tail and the β-hairpin from the thumb. NS5B is shown in a ribbon representation, and the RNA is displayed in yellow. The view was kept identical to the one seen in Fig. 4 B. The side chains of all residues detected by MS are represented in purple for unprotected and in red for protected. Inset, detailed view of helix T, including Arg503, Arg505, Arg507, and Arg510. To be consistent with Fig. 4, the fingertip D1 loop is displayed in blue, and the C terminus is shown in yellow. C, representation of the NTP channel. Modified residues detected by mass spectrometry are represented in red when protected (Lys51, Arg154, Lys155, Arg158, and Lys491) and purple when unprotected (Lys151).
Residues Arg^{154}, Lys^{155}, and Arg^{158} are part of the conserved motif F, which is located in the loop Δ2 (30). The crystal structure shows Arg^{158} within 4 Å of the 3' terminus of the bound template, whereas the side chains of Arg^{154} and Lys^{155} point away from the nucleic acid substrate. However, it is difficult to unambiguously ascertain whether these residues, together with Lys^{151} and Lys^{161}, are directly in contact with the bound RNA or whether their protection from chemical modification is a result of a conformational change upon substrate binding. Peptides 142–154 and 155–168 can be directly cross-linked to short oligonucleotides (29), which is indicative of direct interactions with the bound RNA. In contrast, Lys^{151}, which is unprotected in our experiments, is completely exposed to the solvent and cannot participate in NTP or RNA binding without significant rearrangements. These data are in agreement with mutational analyses, implying that no basic residue is required at this position (31).

The two additional amino acids protected from the chemical probes are Arg^{501} and Arg^{505}. According to our model, these positions are too far away from the bound RNA to allow direct nucleoprotein contacts (Fig. 6B). We consider two different scenarios to explain the observed protection pattern in this region. One possibility is that these residues may be shielded from modification due to protein-protein interactions that are triggered upon RNA binding. Pull-down assays revealed that nucleolin recognizes α-helix T, which includes Arg^{501} and Arg^{505} (32), and the same authors have previously reported that this helix may be involved in intermolecular NS5B interaction (33). However, the specific protection pattern of arginines in helix T suggests that protein-protein interaction may not play a role in this regard. Arg^{508} is protected in the absence and in the presence of RNA, whereas Arg^{501} and Arg^{505} are solely protected in the presence of RNA, and both Arg^{503} and Arg^{510} are modified under both conditions. Protein-protein interactions in this region would probably diminish the accessibility of each of these adjacent residues.

We suggest that the nuances of the protection pattern are better explained through conformational changes upon RNA binding. We consider inter- and/or intradomain interactions in this regard. The side chains of Arg^{503} and Arg^{510} lie on the same side of helix T and point to the fingertip Δ1 loop, whereas Arg^{508}, Arg^{501}, and Arg^{505} lie on the other side of the helix, and their side chains point to the C-terminal loop region between residues Leu^{525} and Leu^{534}. Intradomain interaction between helix T and the fingertip Δ1 loop (residues 11–44) has been reported (34–36). Specifically, the enzyme can exist in “open” or “closed” conformations in which these intradomain interactions are modulated. Several nonnucleosidic inhibitors were shown to freeze the enzyme in an open, presumably inactive conformation. In contrast, extended intradomain interactions have been linked to a closed, active form of the enzyme. Together, these data suggested that the contacts between Helix T and the fingertip D1 loop are flexible. Our findings provide support for this notion, given that Arg^{503} and Arg^{510} are accessible. Both residues are also accessible in the presence of single-stranded RNA, which likewise points to an open conformation or flexible intradomain interaction in the nucleoprotein complex. However, the bound RNA template appears to affect the intradomain interaction between helix T and the adjacent C-terminal loop. Arg^{501} and Arg^{505} are the only arginines in helix T that become protected in the presence of RNA. These data suggest that the loop that comprises residues Leu^{525}–Leu^{534} moves closer to helix T under these conditions. This scenario can be interpreted as a consequence of a steric conflict between the C terminus of the NS5B enzyme and the bound RNA. As proposed for the double-stranded substrate, it appears that the C-terminal region moves from its position in the apoenzyme to accommodate the single-stranded RNA template.

Taken together, the comparison of surface topologies of free NS5B with the nucleoprotein complex suggests that single-stranded RNA can occupy the entire nucleic acid binding channel. The 3'-end of the template is not specifically positioned in close proximity to the active center to allow de novo initiation to occur. This result indicates that the single-stranded template binds preferentially to the entire channel to maximize contacts. Alternatively, a given population of nucleoprotein complexes can exist as a heterogeneous mixture, and the RNA can freely move in the nucleic acid binding groove. In either case, the 3'-end of the template is not specifically recognized by the NS5B enzyme. It is thus possible that the presence of nucleotide substrates is required to properly position the template for initiation, which remains to be elucidated. Our footprinting experiments suggest that this NS5B-RNA complex characterized in this study reasonably mimics the elongation state. The protection pattern around the nucleic acid binding groove is likely to be similar; however, the conformational change or rearrangement at the C-terminal region may be more pronounced or even different in nature when the enzyme needs to accommodate double-stranded RNA.

REFERENCES

1. Wasley, A., and Alter, M. J. (2000) Semin. Liver Dis. 20, 1–16
2. Shepard, C. W., Finelli, L., and Alter, M. J. (2005) Lancet Infect. Dis. 5, 558–567
3. Lindenbach, B. D., and Rice, C. M. (2005) Nature 436, 933–938
4. Zhong, W., Ferrari, E., Lesburg, C. A., Maag, D., Ghosh, S. K., Cameron, C. E., Lau, J. Y., and Hong, Z. (2000) J. Virol. 74, 9134–9143
5. Ranjith-Kumar, C. T., Gutshall, L., Sarkisy, R. T., and Kao, C. C. (2003) J. Mol. Biol. 330, 675–685
6. Oh, J. W., Ito, T., and Lai, M. M. (1999) J. Virol. 73, 7694–7702
7. Hong, Z., Cameron, C. E., Walker, M. P., Castro, C., Yao, N., Lau, J. Y., and Zhong, W. (2001) Virology 285, 6–11
8. Bressanelli, S., Tomei, L., Roussel, A., Incitti, L., Vitale, R. L., Mathieu, M., De Francesco, R., and Rey, F. A. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 13039–13039
9. Ago, H., Adachi, T., Yoshida, A., Yamamoto, M., Habuka, N., Yatsunami, K., and Miyano, M. (1999) Structure 7, 1417–1426
10. Lesburg, C. A., Cable, M. B., Ferrari, E., Hong, Z., Mannarbon, A. F., and Weber, P. C. (1999) Nat. Struct. Biol. 6, 937–943
11. Bressanelli, S., Tomei, L., Rey, F. A., and De Francesco, R. (2002) J. Virol. 76, 3482–3492
12. Cai, Z., Yi, M., Zhang, C., and Luo, G. (2005) J. Virol. 79, 11607–11617
13. Ranjith-Kumar, C. T., Sarkisy, R. T., Gutshall, L., Thomson, M., and Kao, C. C. (2004) J. Virol. 78, 12207–12217
14. O’Farrell, D., Trowbridge, R., Rowlands, D., and Jager, J. (2003) J. Mol. Biol. 326, 1025–1035
15. Shkriabai, N., Datta, S. A., Zhao, Z., Hess, S., Rein, A., and Kvaratskhelia, M. (2006) Biochemistry 45, 4077–4083
16. Liu, Y., Kvaratskhelia, M., Hess, S., Qu, Y., and Zou, Y. (2005) J. Biol. Chem.
RNA Binding to HCV Polymerase

280, 32775–32783

17. Hausen, P., and Stein, H. (1970) Eur. J. Biochem. 14, 278–283
18. Ranjith-Kumar, C. T., Gutshall, L., Kim, M. J., Sarisky, R. T., and Kao, C. C. (2002) J. Virol. 76, 12526–12536
19. Nowotny, M., Gaidamakov, S. A., Crouch, R. J., and Yang, W. (2005) Cell 121, 1005–1016
20. Shell, S. M., Hess, S., Kvaratskhelia, M., and Zou, Y. (2005) Biochemistry 44, 971–978
21. Kvaratskhelia, M., Miller, J. T., Budihas, S. R., Pannell, L. K., and Le Grice, S. F. (2002) Proc. Natl. Acad. Sci. U. S. A. 99, 15988–15993
22. D’Abramo, C. M., Cellai, L., and Gotte, M. (2004) J. Mol. Biol. 337, 1–14
23. Luo, G., Hamatake, R. K., Mathis, D. M., Racela, J., Rigat, K. L., Lemm, J., and Colonna, R. J. (2000) J. Virol. 74, 851–863
24. Makeyev, E. V., and Bamford, D. H. (2000) EMBO J. 19, 6275–6284
25. Butcher, S. J., Grimes, J. M., Makeyev, E. V., Bamford, D. H., and Stuart, D. I. (2001) Nature 410, 235–240
26. Choi, H. B., Kim, Y. G., and Oh, J. W. (2003) Exp. Mol. Med. 35, 475–485
27. Lai, V. C., Dempsey, S., Lau, J. Y., Hong, Z., and Zhong, W. (2003) J. Virol. 77, 2295–2300
28. Piccininni, S., Varaklioti, A., Nardelli, M., Dave, B., Raney, K. D., and McCarthy, J. E. (2002) J. Biol. Chem. 277, 45670–45679
29. Kim, Y. C., Russell, W. K., Ranjith-Kumar, C. T., Thomson, M., Russell, D. H., and Kao, C. C. (2005) J. Biol. Chem. 280, 38011–38019
30. Bruenn, J. A. (2003) Nucleic Acids Res. 31, 1821–1829
31. Labonte, P., Axelrod, V., Agarwal, A., Aulabaugh, A., Amin, A., and Mak, P. (2002) J. Biol. Chem. 277, 38838–38846
32. Hirano, M., Kaneko, S., Yamashita, T., Luo, H., Qin, W., Shirotta, Y., Nomura, T., Kobayashi, K., and Murakami, S. (2003) J. Biol. Chem. 278, 5109–5115
33. Qin, W., Luo, H., Nomura, T., Hayashi, N., Yamashita, T., and Murakami, S. (2002) J. Biol. Chem. 277, 2123–2137
34. Biswal, B. K., Cherney, M. M., Wang, M., Chan, L., Yannopoulos, C. G., Bilimoria, D., Nicolas, O., Bedard, J., and James, M. N. (2005) J. Biol. Chem. 280, 18202–18210
35. Biswal, B. K., Wang, M., Cherney, M. M., Chan, L., Yannopoulos, C. G., Bilimoria, D., Bedard, J., and James, M. N. (2006) J. Mol. Biol. 361, 33–45
36. Di Marco, S., Volpari, C., Tomei, L., Altamura, S., Harper, S., Narjes, F., Koch, U., Rowley, M., De Francesco, R., Migliaccio, G., and Carfi, A. (2005) J. Biol. Chem. 280, 29765–29770
37. Sarafianos, S. G., Das, K., Tantillo, C., Clark, A. D., Jr., Ding, J., Whitcomb, J. M., Boyer, P. L., Hughes, S. H., and Arnold, E. (2001) EMBO J. 20, 1449–1461