*Proc. Natl. Acad. Sci. USA*
Vol. 91, pp. 888–892, February 1994
Biochemistry

# Molecular model of the specificity pocket of the hepatitis C virus protease: Implications for substrate recognition

**(model building/nonstructural protein NS3/N-terminal sequencing/sequence alignment)**

ELISABETTA PIZZI*, ANNA TRAMONTANO*, LICIA TOMEI*, NICOLA LA MONICA*, CRISTINA FAILLA*, MOHINDER SARDANA†, THERESA WOOD†, AND RAFFAELE DE FRANCESCO*‡

*Istituto di Ricerche di Biologia Molecolare, P. Angeletti Pomezia, 00040 Rome, Italy; and †Merck Research Laboratories, West Point, PA 19486

*Communicated by Edward M. Scolnick, October 12, 1993*

**ABSTRACT** **We have built a model of the specificity pocket of the protease of hepatitis C virus on the basis of the known structures of trypsin-like serine proteases and of the conservation pattern of the protease sequences among various hepatitis C strains. The model allowed us to predict that the substrate of this protease should have a cysteine residue in position P1. This hypothesis was subsequently proved by N-terminal sequencing of two products of the protease. The success of this "blind" test increases our confidence in the overall correctness of our proposed alignment of the enzyme sequence with those of other proteases of known structure and constitutes a first step in the construction of a complete model of the viral protease domain.**

Hepatitis C virus (HCV) is the major cause of parentally transmitted non-A non-B hepatitis and is also implicated as the cause in many sporadic cases of non-A non-B hepatitis (1, 2). Chronic infection with HCV has also been linked to the development of hepatocellular carcinoma (3–5).

HCV is an enveloped virus with a positive-strand RNA genome of ≈9.5 kb (6–13). This virus is probably related to flaviviruses and pestiviruses in view of the similarity of the deduced amino acid sequences and hydropathy profile of the viral proteins (14). As for pestiviruses and flaviviruses, the HCV genomic RNA includes a single open reading frame encoding a precursor polyprotein that is cleaved co- or posttranslationally into mature viral polypeptides. The putative viral structural proteins are encoded by the 5′ quarter of the genome, whereas the remaining part encodes the nonstructural proteins (NS2, NS3, NS4A, NS4B, NS5A, and NS5B) that are believed to be components of the viral replication apparatus (15).

The gene order in the HCV genome has been shown to be the following: 5′-C-E1-E2-NS2-NS3-NS4A-NS4B-NS5A-NS5B-3′ (16, 17). C is a basic protein that can bind RNA (E. Santolini and N.L.M., unpublished results) and it is postulated to be the nucleocapsid protein. E1 and E2 are putative envelope glycoproteins. NS2 through NS5B represent the viral nonstructural proteins and could be functional analogues of the corresponding proteins of pestiviruses and flaviviruses (18, 19).

Processing of the polyprotein requires both host and viral proteases. The signal peptidase of endoplasmic reticulum is responsible for generating the N termini of E1, E2, and possibly NS2 (20). Whereas it is not clear what protease is responsible for the cleavage between NS2 and NS3, we (16) and others (21) have identified NS3 as the protease responsible for cleavage at the NS3–NS4A, NS4A–NS4B, NS4B–NS5A, and NS5A–NS5B junctions. Analysis of the HCV NS3 amino acid sequence has suggested that this viral protein may contain a trypsin-like serine protease domain that could process the viral polyprotein as do such proteins in flaviviruses and pestiviruses (14, 22, 23).

Three-dimensional structures of serine proteases have been extensively studied by x-ray crystallography. Their active site is formed by three residues, histidine, aspartate, and serine. In all known structures of serine proteases, the relative spatial position of the catalytic residues in the three-dimensional structure is extremely well conserved. The peptide bond is cleaved by a nucleophilic attack of the serine hydroxyl group on the scissile carbonyl bond, forming an acyl enzyme intermediate. The histidine imidazole group abstracts the alcohol proton and conveys it in the amine leaving group, while the aspartate stabilizes the positive charge that develops on histidine in the intermediate complex (24).

The residues His-1083, Asp-1107, and Ser-1165 (numbered according to their location in the polyprotein of BK strain HCV; ref. 10), found in the N-terminal domain of the NS3 protein, are highly conserved among all HCV strains sequenced so far and have been proposed to constitute the catalytic triad of the HCV protease (14, 25). Consistent with this prediction, when the proposed catalytic Ser-1165 of the polyprotein was mutated into an Ala, processing at the NS3–NS4A, NS4A–NS4B, NS4B–NS5A, and NS5A–NS5B sites was abolished (16, 21). Interestingly, cleavage at three of the four sites of the mutated polyprotein could be restored by supplying a functional NS3 in trans, while processing at the NS3–NS4A only occurred in cis (16). This finding indicates that the cleavage of the latter site is an intramolecular event.

The comparative analysis of the three-dimensional structure of serine proteases obtained so far, with or without bound inhibitors, has shed light on the way diverse cleavage specificities are generated by the substrate binding sites. The structure of various serine proteases reveals substrate binding pockets that are tailored to bind the side chain of the residue preceding the scissile bond (P1). For protein substrates, the binding interaction can extend over several residues on the acyl-group side of the scissile bond; an array of secondary binding sites can facilitate the cleavage of the scissile bond by enhancing the binding energy (ref. 26 and references therein).

Structural information on the protease domain of HCV NS3, and in particular on the architecture of its substrate binding pocket, would be extremely helpful in making predictions on the specificity of the enzyme, as well as in designing structure-based inhibitors that could be used as therapeutic agents for treatment of HCV infections. To this end, we decided to undertake the modeling of the three-dimensional structure of the enzyme by taking advantage of the wealth of structural information available on serine

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: HCV, hepatitis C virus.

‡To whom reprint requests should be addressed.

proteases. In this paper, we describe the results we obtained in modeling the enzyme specificity pocket and compare our prediction of the P1 substrate side chain with experimental data. The agreement between the prediction and the experimental determination of the cleavage sites increased our confidence in our initial assumption on the protein folding as well as in the strategy used to obtain the molecular model and encouraged us to exploit the model of the enzyme substrate-binding pocket and the sequence alignments used to construct it to model the whole protease domain of HCV NS3.

## MATERIALS AND METHODS

**Sequence Alignment and Modeling.** Sequences used for the alignment were taken from Releases 24.0 and 36.0 of Swiss-Prot and EMBL sequence data bases, respectively. The program PILEUP of the Genetics Computer Group package (27) with default parameters was used to align the serotype sequences. Coordinates of the protease structures are from the Brookhaven data bank (28). The model was built using the program INSIGHT (29).

**Immunoaffinity Purification of Region-Specific Anti-HCV Antibodies.** The fusion proteins glutathione *S*-transferase (GST)–NS5A and TrpE–NS5B (16) were exploited to affinity-purify region-specific antibodies directed against NS5A and NS5B, respectively. Rabbit antibodies against NS5A and NS5B were affinity-purified from rabbit antisera as follows: TrpE–NS5B was prepared from *Escherichia coli* inclusion bodies as outlined in ref. 16. The protein was then solubilized in SDS/PAGE loading buffer and electrophoresed on a 10% polyacrylamide/SDS gel. The protein was transferred from the gel onto a nitrocellulose filter by electroblotting in a buffer containing 50 mM Tris, 380 mM glycine, and 10% (vol/vol) methanol and visualized by staining with Ponceau red. The protein of interest was then excised from the filter, washed extensively in phosphate-buffered saline (PBS), and blocked in PBS containing 5% (wt/vol) nonfat dry milk. This filter strip was subsequently incubated for 1 h at room temperature with rabbit antiserum that contained antibodies directed against the TrpE–NS5B protein. The filter was washed with PBS, the bound antibodies were eluted with 100 mM glycine (pH 2.5), and immediately neutralized with 0.1 vol of 1 M Tris·HCl (pH 8.0). Affinity-purified GST–NS5A fusion protein was prepared and cross-linked to activated Affi-Gel 10 chromatography matrix (Bio-Rad) essentially as described (16). The affinity matrix thus obtained was then incubated for 1 h at 4°C with rabbit antisera that contained antibodies directed against the GST–NS5A protein, poured in a small disposable chromatography column, and washed extensively with PBS. The bound antibodies were eluted with 100 mM glycine (pH 2.5) and neutralized with 0.1 vol of 1 M Tris·HCl (pH 8.0).

**Preparation of Labeled Cell Extracts.** HeLa cells seeded at $1 \times 10^6$ cells per 10-cm plate were infected with vaccinia virus vTF7-3 (30) at a multiplicity of five plaque-forming units per cell. vTF7-vaccinia-infected cells were transfected with 40 $\mu$g of recombinant plasmid pCITE-SX as described (16). pCITE-SX contains the cDNA encoding aa 991 to aa 3010 of the polyprotein from HCV BK (10, 16). Four hours after transfection, the medium was replaced with minimum essential medium (MEM) lacking the appropriate amino acid (GIBCO), and the cells were starved for 1 h at 37°C. Cells were then radiolabeled for 3 h with 1 mCi of [$^{35}$S]methionine, [$^3$H]leucine, or [$^3$H]valine (Amersham; 1 Ci = 37 GBq) in 5 ml of MEM lacking the appropriate amino acid and supplemented with 2% (vol/vol) dialyzed fetal calf serum. Cells were harvested and lysed for immunoprecipitation in $IPB_{150}$ (20 mM Tris·HCl, pH 8.0/150 mM NaCl/1% Triton) supplemented with 1 mM phenylmethylsulfonyl fluoride/1 mM EDTA/1 mM dithiothreitol.

**Immunopurification and N-Terminal Sequence Analysis.** Prior to immunoprecipitation, SDS and dithiothreitol were added to the cell lysates to a final concentration of 2% (wt/vol) and 10 mM, respectively. The radiolabeled lysates were then heated at 95°C for 10 min. Affinity-purified antibody (100 $\mu$g) was bound to 100 $\mu$l of protein A-Sepharose. The cell lysate (200 $\mu$l) was diluted in 4 ml of $IPB_{150}$ and incubated for 1 h at 4°C with the protein A-Sepharose-bound antibody. The suspension was then layered on 0.9 ml of 0.5× NDET (0.5% Nonidet P-40/0.2% sodium deoxycholate/33 mM EDTA/10 mM Tris·HCl, pH 7.4) containing 30% (wt/vol) sucrose and pelleted by centrifugation. The pellet was resuspended in 1 ml of 10 mM ammonium carbonate (pH 7) and washed with water. The radiolabeled protein was then released from the protein A-Sepharose-bound antibody with 100 $\mu$l of 0.1% trifluoroacetic acid and neutralized with ammonia.

The radiolabeled HCV peptides NS5A and NS5B were subjected to Edman degradation on an Applied Biosystems model 470A gas-phase sequencer. The fractions at each cycle were collected and monitored for release of radioactivity in a liquid scintillation counter (Beckman LS3901).

## RESULTS AND DISCUSSION

**Analysis of the Structure of Known Serine Proteases and Model Building of the Specificity Pocket.** Serine proteases have been classified into two major families: trypsin-like and subtilisin-like according to their overall folding and to the order in which the residues of the catalytic triad occur in the sequence (31). Trypsin-like proteases in turn have been further classified according to their size into large cellular proteases (e.g., trypsin and elastase) and small cellular proteases (e.g., $\alpha$ lytic protease and *Streptomyces griseus* protease A and B) (20). Enzymes from both these classes have been crystallized, in some cases as complexes with their inhibitors (32–41). We have analyzed these structures to derive common features that could be helpful in understanding the folding and predict the substrate specificity of the HCV enzyme. The sequence of the protease domain of the HCV NS3 protein and the relative position in the sequence of the amino acids of the catalytic triad indicate that this protease belongs to the trypsin-like family (25). Unfortunately, the poor sequence homology of the HCV protease domain with the other serine proteases has not allowed a clear overall alignment to be derived. As a consequence, additional information is needed to construct a structural model of the viral enzyme.

Several HCV strains have been isolated and sequenced. A multiple alignment of these sequences is instrumental to the purpose of analyzing the folding and the substrate specificity of the protease in that (*i*) within the protease sequence, important residues involved in the specificity of the enzyme are likely to be conserved and (*ii*) the predicted molecular weight of the products derived from the HCV polyprotein by its protease processing allows rough definition of the putative cleavage sites of the protease: in the regions proposed to be cleavage sites for the protease, the putative specificity residues have to be conserved, at least in their physicochemical character.

Table 1 lists the trypsin-like proteases of known structure used in our analysis. We superimposed the catalytic triad of all these structures by best least squared fit of their main-chain atoms and analyzed the resulting superimposed structures to locate, within each sequence, the regions that are structurally conserved (within 1.5 Å). These regions (according to the chymotrypsin numbering system) span residues 41–43, 52–59, 101–107, 138–140, 189–199, 210–216, and 225–229. Obviously, these portions of the proteins include the catalytic triad and the regions surrounding it as well as the

Table 1. Trypsin-like serine proteases of known structure used for the analysis

| Protease | Code | Type | Position P1 of the substrate | Ref. | Resolution, Å |
|---|---|---|---|---|---|
| *Bos taurus* β trypsin | 1ntp | L | Arg, Lys | 32 | 1.8 |
| *Rattus rattus* tonin | 1ton | L | Arg, Trp | 33 | 1.8 |
| *Rattus rattus* cell protease | 3rp2 | L | Large hydrophobic | 34 | 1.9 |
| *Sus scrofa* elastase | 3est | L | Small hydrophobic | 35 | 1.65 |
| *Homo sapiens* elastase | 1hne | L | Small hydrophobic | 36 | 1.84 |
| *Streptomyces griseus* trypsin | 1sgt | L | Basic | 37 | 1.7 |
| *Bos taurus* α-chimotrypsin A | 5cha | L | Large hydrophobic | 38 | 1.7 |
| α-Lytic protease | 2alp | S | Small hydrophobic | 39 | 1.7 |
| *Streptomyces griseus* protease B | 3sgb | S | Aromatic | 40 | 1.8 |
| *Streptomyces griseus* protease A | 2sga | S | Aromatic | 41 | 1.5 |

Brookhaven data bank code is shown. Type: L, large cellular protease; S, small cellular protease.

backbone structures forming most of the specificity pockets (residues 189–194, 213–216, and 225–227). At the bottom of the specificity pocket the structures diverge, but this part of the molecule is only involved in the recognition of the P1 amino acid if it has a long side chain. The structure-derived alignment of the sequences of the aforementioned regions allowed us to derive a consensus that was used to include the HCV sequence in the alignment (Fig. 1). In this alignment, residues involved in specificity and catalysis are conserved among the HCV strains.

**Specificity Pocket.** Examples of large and small trypsin-like serine proteases complexed with their inhibitors are *Bos taurus* trypsin (42) and *Streptomyces griseus* protease B, respectively (41). The specificity pocket of trypsin, whose substrate has either an arginine or a lysine in the P1 position, is deep enough to allow the long side chains of these residues to reach its bottom, where Asp-189 makes hydrogen bonds with the substrate, either directly or through a water molecule. Backbone atoms of the specificity pocket of trypsin in the regions of aa 190–195, 214–216, 219–220, and 226 are in contact with the substrate. The trypsin–trypsin inhibitor complex structure suggests that the presence of glycines in positions 216 and 226 is essential in that they allow the pocket to be accessed by long side chain residues. This model has been elegantly proved by site-directed mutagenesis experiments (43). Side chains contacting the trypsin substrate are those of residues Asp-189, Ser-190, Glu-192, and Val-213.

The P1 residue of the inhibitor cocrystallized with *Streptomyces griseus* protease B is a leucine. Compared with the corresponding site of trypsin, the specificity pocket of this protease is smaller, due to the presence of a threonine at position 226. Ala-190 and Thr-213 are the only two side chains contacting the inhibitor.

We used our alignment of the HCV protease domain sequence with the conserved core regions of proteases listed in Table 1 to build manually the specificity pocket of the enzyme shown in Fig. 2. The model building was based on the alignment shown in Fig. 1, and the most commonly observed conformation for each amino acid was chosen (44).

Positions 213 and 190 in the HCV protease domain correspond to phenylalanine and tyrosine, respectively. The side chain of phenylalanine is so bulky that all of the residues that are located deep within the pocket are prevented from contacting the substrate: in fact, these residues, including Tyr-190, are not always conserved among HCV strains.

In all known proteases, residues corresponding to those contacting the substrate in our model have their side chains pointing away from the substrate. Thus the shape of the HCV protease pocket is essentially determined by main-chain atoms and by the phenylalanine ring.

The analysis of the shape of the modeled pocket almost immediately suggests that the P1 site of the HCV protease substrate has to be an amino acid with a short side chain, ideally a cysteine or a serine (Fig. 2). In addition, the pocket is very hydrophobic and is closed by the aromatic ring of the phenylalanine. The sulfhydryl group of cysteine has been shown to interact favorably with the aromatic ring system of phenylalanine (42). We analyzed the alignment of the regions containing the putative cleavage sites of all HCV serotypes, based only on the apparent molecular weights of the mature viral polypeptides. Only two amino acids are present and conserved in all these regions, namely, a cysteine and a leucine. This helped us to exclude that smaller residues, although theoretically able to fit in the modeled pocket, would be the preferred substrates. For the reasons given above, our model strongly favors cysteine, but of course it cannot be excluded that the protease has multiple specificities and/or that the initial alignment on which our model is based suffers from the uncertainty derived from the poor sequence homology. For this reason, we set out to substantiate our hypothesis with experimental data.

**Determination of NS3-Dependent Polyprotein Cleavage Sites.** To evaluate the predictive value of the specificity pocket model, we determined the sequence of the N termini of the NS5A and NS5B proteins. These two proteins are released from the polyprotein precursor by NS3 (16, 21). When cultured cells are transfected with a cDNA coding for the HCV polyprotein, NS5A and NS5B are generated at a faster rate than NS4B and NS4A, which are often present as an uncleaved NS4AB precursor. We excluded the NS3–NS4A cleavage site from our study, because this site can only

```
code 41-43     52-59      101-107  138-140    189-199       210-216   225-229

1ntp .FCG....VVSAAHCY....NDIMLIK....ISG....DSCQGDSGGPV....QGIVSWG....PGVYT.
1ton .LCG....VITAAHCY....NDLMLLH....ASG....DTCAGDSGGPL....QGITSGG....PAIYA.
3rp2 .ICG....VLTAAHCK....HDIMLLK....AAG....AAFMGDSGGPL....HGIVSYG....PAIFT.
3est .TCG....VMTAAHCV....YDIALLR....ITG....SGCQGDSGGPL....HGVTSFV....PTVFT.
1hne .FCG....VMSAAHCV....NDIVILQ....AMG....GVCFGDSGSPL....HGIASFV....PDAFA.
1sgt .GCG....VLTAAHCV....KDWALIK....VAG....DTCQGDSGGPM....VGIVSWG....PGVYT.
5cha .FCG....VVTAAHCG....NDITLLK....TTG....SSCMGDSGGPL....VGIVSWG....PGVYA.
2alp .LCS....FVTAGHCT....NDRAWVS....RSG....CMGRGDSGGSW....QGVMSGG....SSLFE.
3sgb .RCS....FLTAGHCT....NDYGIVR....RRG....CAEPGDSGGPL....IGLTSGG....TTFFQ.
2sga .RCS....ALTAGHCT....NDYGIIR....RSG....CAQPGDSGGSL....LGLTSGG....TTFYQ.
      ..     . .       ..        ...     .   . .....     ...         ..
HCV  .TCI....CWTVYHGA....QDLVGWP....RRG....SYLKGSSGGPL....VGIFRAA....TTMRS.
                                             ******          ****     *
```

FIG. 1. Alignment of structurally similar regions of trypsin-like serine proteases with regions of the HCV NS3 protein sequence. Aligned regions of known structure differ by <1.5 Å after best superposition of the catalytic triad. HCV NS3 residues matching the consensus are indicated by solid circles. Stars indicate residues forming the specificity pocket. The catalytic triad residues are underlined. Brookhaven data bank codes are from Table 1.

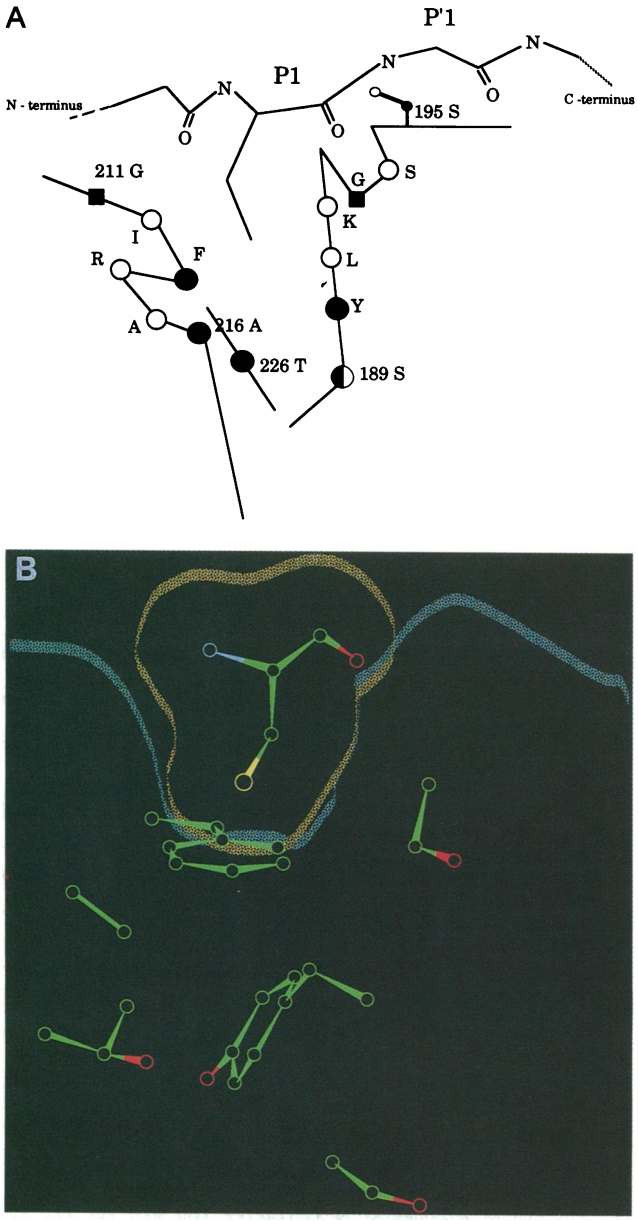


FIG. 2. (*A*) Schematic view of the model of the specificity pocket of HCV NS3. Amino acid residues whose side chains point either into or out of the pocket in all analyzed proteases are represented by solid and open circles, respectively; the residue identified by a partially solid circle is found in both conformations in different proteases; solid squares are conserved glycines; the catalytic serine (Ser-195) is also shown. (*B*) Predicted structure of the specificity pocket of HCV NS3 with a cysteine docked into the specificity pocket. Solvent-accessible surfaces of the protein (cyan) and of the putative cysteine residue in the pocket (yellow) are also shown. Atoms are indicated as follows: green, carbon; red, oxygen; blue, nitrogen; yellow, sulfur.

be processed intramolecularly and its recognition may, therefore, involve a different mechanism (16).

A portion of the HCV polyprotein containing most of the nonstructural region, from aa 991 to aa 3010, was transiently expressed in HeLa cells with the T7/vaccinia virus system (16, 30). After a pulse labeling of 4 h with [$^{35}$S]methionine, [$^{3}$H]leucine, or [$^{3}$H]valine, NS5A and NS5B were immunoprecipitated with region-specific rabbit polyclonal antibodies. The immunopurified proteins were directly subjected to automated solid-phase Edman degradation. The results obtained are shown in Fig. 3. For each of the sequenced proteins we could localize the N terminus within the HCV polyprotein

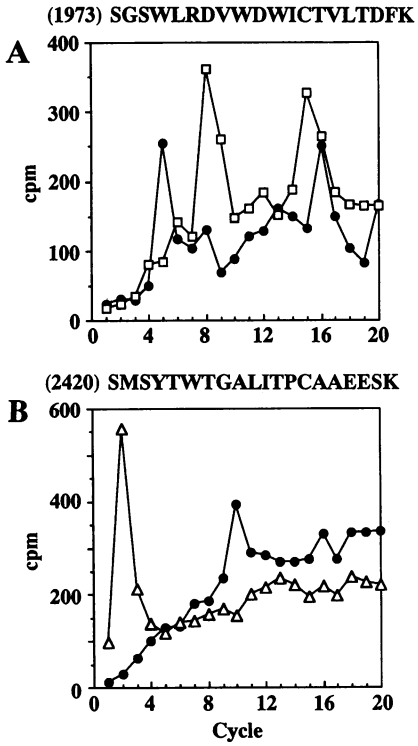


FIG. 3. N-terminal sequence analysis of HCV NS5A and NS5B proteins. The amount of radioactivity (cpm) released at each sequencing cycle is shown. The N-terminal sequences of NS5A (*A*) and NS5B (*B*) are shown at the top of *A* and *B*. The number in parentheses indicates the position of the N-terminal amino acid of each protein within the polyprotein of the BK strain of HCV (10). ●, [$^{3}$H]Leucine; □, [$^{3}$H]valine; △, [$^{35}$S]methionine.

in an unambiguous manner. The data position the N termini of NS5A and NS5B at residues 1973 and 2420 of the polyprotein, respectively. Both of these residues are serines and follow a cysteine residue in the polyprotein sequence. These results indicate that at least two of the scissile bonds cleaved by the HCV protease are found between a cysteine (P1) and a serine (P1′). This finding is in excellent agreement with the substrate specificity that was predicted by our model of the enzyme specificity pocket.

In Fig. 4, sequence a, we show the amino acid sequences surrounding these cleavage sites. A few common features are evident from the comparison of these sequences. Only the cysteine in position P1 and the serine in P1′ are strictly conserved among the two cleavage sites. In addition, an acidic residue is found at position P6 and an aromatic residue is present at position P4′.

It is interesting to observe that this pattern is not present elsewhere in the polyprotein, suggesting that it may not represent a strict consensus sequence. However, around the region predicted to be the N terminus of NS4B, only one cysteine residue is conserved among all the HCV strains. Alignment of this cysteine with the P1 cysteine of our

| | Sequence | Position |
|---|---|---|
| a) | DCSTP**C**SGSW | 1967 |
| | EDVVC**C**SMSY | 2414 |
| b) | DEMEE**C**ASHL | 1706 |
| c) | GNLPG**C**SFSI | 167 |

FIG. 4. Sequences: a, sequences of the NS4B–NS5A and NS5A–NS5B cleavage sites; b, consensus sequence at the putative NS4A–NS4B cleavage site; c, consensus sequence in the C-terminal part of core (see text). The boldface type indicates the P1 residue.

consensus sequence (Fig. 4, sequence b) shows that also in this case an acidic residue is present in P6. In some, but not all, HCV strains a serine is found in the P1′ position. We concluded that this sequence might be the NS4A–NS4B boundary. It is worth noting that no conserved cysteine is present in the region predicted to be at the NS3–NS4A boundary, confirming the unique character of this site.

Interestingly, a conserved C-S sequence is found at the C terminus of a stretch of hydrophobic residues in the HCV core protein (Fig. 4, sequence c). In the sequence of proteolytic events of the polyproteins of flaviviruses and pestiviruses, a similar hydrophobic region, postulated to be anchoring the core protein to the membrane, is eventually cleaved. For flaviviruses, the viral protease has been suggested to be responsible for this cleavage (19). Should this be the case also for HCV, the NS3 protease is a good candidate for the enzyme carrying out this cleavage.

While this report was in preparation, Grakoui *et al.* (21) published similar experiments on the cleavage sites of NS3–NS4A, NS4A–NS4B, NS4B–NS5A, and NS5A–NS5B. Their data are consistent with our finding of a cysteine in position P1. They also report that the P1 residue for the cleavage between NS3–NS4A is a threonine. When a threonine side chain is built into our modeled specificity pocket, its contacts are less favorable than those of cysteine. This is consistent with the observation that the NS3–NS4A site is poorly recognized by the HCV protease, since it can only be cleaved intramolecularly and not at all intermolecularly, as all the other sites (16).

The model presented here is limited to the substrate binding pocket of the enzyme and does not allow further speculation on its mechanism for substrate recognition. However, the success of the model in predicting the specificity of the enzyme before its experimental verification increases our confidence in the general correctness of our observations. We believe that the specificity pocket model can serve as an anchoring point for extension of the model to the whole enzyme.

We are grateful to Prof. Riccardo Cortese for continuous encouragement and support, to Drs. Armin Lahm, Josef Jiricny, and Andrew Wallace for useful suggestions, and to Yves Cully for the artwork.

1. Choo, Q.-L., Kuo, G., Weiner, A. J., Overby, L. R., Bradley, D. W. & Houghton, M. (1989) *Science* **244**, 359–362.
2. Kuo, G., Choo, Q.-L., Alter, H. J., Gitnick, G. L., Redecker, A. G., Purcell, R. H., Myamura, T., Dienstag, J. L., Alter, M. J., Stevens, C. E., Tagtmeier, G. E., Bonino, F., Colombo, M., Lee, W.-S., Kuo, C., Berger, K., Shister, J. R., Overby, L. R., Bradley, D. W. & Houghton, M. (1989) *Science* **244**, 362–364.
3. Bruix, J., Calvet, X., Costa, J., Ventura, M., Bruguera, M., Castillo, R., Barrera, J. M., Ercilla, G., Sanchez-Tapias, J. M., Vall, M., Bru, C. & Rodes, J. (1989) *Lancet* **ii**, 1404–1406.
4. Colombo, M., Kuo, G. L., Choo, Q.-L., Donato, M. F., Ninno, E. D., Tommasini, M. A., Dioguardi, N. & Houghton, M. (1989) *Lancet* **ii**, 1406–1408.
5. Saito, I., Miyamura, T., Obayashi, A., Harada, H., Katayama, T., Kikuchi, S., Watanabe, Y., Koi, S., Onji, M., Ohta, Y., Choo, Q.-L., Houghton, M. & Kuo, G. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 6547–6549.
6. Chen, P.-J., Lin, M.-H., Tai, E.-F., Liu, P.-C., Lin, C.-J. & Chen, D.-S. (1992) *Virology* **188**, 102–113.
7. Choo, Q.-L., Richman, K. H., Han, J. H., Berger, K., Lee, C., Dong, C., Gallegos, C., Colt, D., Medina-Selby, A., Barr, P. J., Weiner, A. J., Bradley, D. W., Kuo, G. & Houghton, M. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 2451–2455.
8. Okamoto, H., Kurai, K., Okada, S. I., Yamamoto, K., Iizuka, H., Tanaka, T., Fukuda, S., Tsuda, F. & Mishiro, S. (1992) *Virology* **188**, 331–341.
9. Okamoto, H., Okada, S., Sugiyama, Y., Kurai, K., Iizuka, H., Machida, A., Miyakawa, Y. & Mayumi, M. (1991) *J. Gen. Virol.* **72**, 2697–2704.
10. Takamizawa, A., Mori, C., Fuke, I., Manabe, S., Murakami, S., Fujita, J., Onoshi, E., Andoh, T., Yoshida, I. & Okayama, H. (1991) *J. Virol.* **65**, 1105–1113.
11. Tanaka, T., Kato, N., Nakagawa, M., Ootsuyama, Y., Cho, M.-J., Nakazawa, T., Hijikata, M., Ishimura, Y. & Shimotohno, K. (1992) *Virus Res.* **23**, 39–53.
12. Inchauspe, G., Zebedee, S., Lee, D.-H., Sugitani, M., Nasoff, M. & Prince, A. M. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 10292–10296.
13. Kato, M., Hijikata, M., Ootsuyama, Y., Nakagawa, M., Ohkoshi, S., Sugimura, T. & Shimotohno, K. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 9524–9528.
14. Miller, R. H. & Purcell, R. H. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 2057–2061.
15. Houghton, M., Weiner, A., Han, J., Kuo, G. & Choo, Q.-L. (1991) *Hepatology* **14**, 381–388.
16. Tomei, L., Failla, C., Santolini, E., De Francesco, R. & La Monica, N. (1993) *J. Virol.* **67**, 4017–4026.
17. Grakoui, A. C., Wychowsky, C., Lin, C., Feinstone, S. M. & Rice, C. (1993) *J. Virol.* **67**, 1385–1395.
18. Collett, M. S. (1992) *Infect. Dis.* **15**, 145–154.
19. Rice, C. M., Strauss, E. G. & Strauss, J. H. (1986) in *Togaviridae and Flaviviridae*, eds. Schlesinger, S. & Schlesinger, M. J. (Plenum, New York), pp. 279–326.
20. Hijikata, M., Kato, N., Ootsuyama, Y., Nakagawa, M. & Shimotohno, K. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 5547–5551.
21. Grakoui, A. C., McCourt, D. W., Wychowsky, C., Feinstone, S. M. & Rice, C. (1993) *J. Virol.* **67**, 2832–2843.
22. Chambers, T. J., Weir, R. C., Grakoui, A. C., McCourt, D. W., Bazan, F. F., Fletterick, R. J. & Rice, C. M. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 8898–8902.
23. Wiskerchen, M. & Collet, M. S. (1991) *Virology* **184**, 341–350.
24. Phillips, M. A. & Fletterick, R. J. (1992) *Curr. Opin. Struct. Biol.* **2**, 713–720.
25. Bazan, J. F. & Fletterick, R. J. (1989) *Virology* **171**, 637–639.
26. Bazan, J. F. & Fletterick, R. J. (1990) *Semin. Virol.* **1**, 311–322.
27. Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395.
28. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R. & Kennard, O. (1977) *J. Mol. Biol.* **112**, 535–542.
29. Dayringer, H. E., Tramontano, A., Sprang, S. R. & Fletterick, R. J. (1986) *J. Mol. Graphics* **4**, 82–87.
30. Fuerst, T. R., Niles, E. G., Studier, F. W. & Moss, B. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 8122–8126.
31. Neurath, H. (1984) *Science* **224**, 350–357.
32. Kossiakoff, A. A. (1984) *Basic Life Sci.* **27**, 281–304.
33. Fujinaga, M. & James, M. N. G. (1987) *J. Mol. Biol.* **195**, 373–396.
34. Remington, S. J., Woodbury, R. G., Reynolds, R. A., Matthews, B. W. & Neurath, H. (1988) *Biochemistry* **27**, 8097–8105.
35. Meyer, E., Cole, G., Radahakrishnan, R. & Epp, O. (1988) *Acta Crystallogr. B* **44**, 26–38.
36. Navia, M. A., McKeever, B. M., Springer, J. P., Lin, T. J., Williams, H. R., Fluder, E. M., Dorn, C. P. & Hoogsteen, K. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 7–11.
37. Read, R. J. & James, M. N. G. (1988) *J. Mol. Biol.* **200**, 523–551.
38. Blevins, R. A. & Tulinski, A. (1985) *J. Biol. Chem.* **260**, 4264–4275.
39. Fujinaga, M., Delabaere, L. T. J., Brayer, G. D. & James, M. N. G. (1985) *J. Mol. Biol.* **184**, 479–502.
40. Read, R. J., Fujinaga, M., Sielecki, A. R. & James, M. N. G. (1983) *Biochemistry* **22**, 4420–4433.
41. Moult, J., Sussman, F., James, M. N. G. (1985) *J. Mol. Biol.* **182**, 555–566.
42. Bode, W., Walter, J., Huber, R., Wenzel, H. R. & Tschesche, H. (1984) *Eur. J. Biochem.* **144**, 185–190.
43. Craik, C. S., Langman, C., Fletcher, T., Roczniak, S., Barr, P. J., Fletterick, R. & Rutter, W. (1985) *Science* **228**, 291–297.
44. Ponder, Y. W. & Richards, F. M. (1987) *J. Mol. Biol.* **193**, 775–791.