

## In Vivo Selection of Protease Cleavage Sites by Using Chimeric Sindbis Virus Libraries

LAURA PACINI, ALESSANDRA VITELLI, GESSICA FILOCAMO, LINDA BARTHOLOMEW,  
MIRKO BRUNETTI, ANNA TRAMONTANO, CHRISTIAN STEINKÜHLER,  
AND GIOVANNI MIGLIACCIO\*

*Istituto di Ricerche di Biologia Molecolare P. Angeletti, 00040 Pomezia (Rome), Italy*

Received 15 May 2000/Accepted 15 August 2000

**Identifying protease cleavage sites contributes to our understanding of their specificity and biochemical properties and can help in designing specific inhibitors. One route to this end is the generation and screening of random libraries of cleavage sites. Both synthetic and phage-displayed libraries have been extensively used in vitro. We describe a novel system based on recombinant Sindbis virus which can be used to identify cleavage sites in vivo, thus eliminating the need for a purified enzyme and overcoming the problem of choosing the correct in vitro conditions. As a model we used the serine protease of the hepatitis C virus (HCV). We engineered the gene coding for this enzyme and two specific cleavage sites in the Sindbis virus structural gene and constructed libraries of viral genomes with a random sequence at either of the cleavage sites. The system was designed so that only viral genomes coding for sequences cleaved by the protease would produce viable viruses. With this system we selected viruses containing sequences mirroring those of the natural HCV protease substrates which were cleaved with comparable efficiencies.**

In the past 10 years, random peptide libraries, either chemically synthesized or displayed on phage, have been widely used as a source of protein binding molecules and applied to the identification of peptides able to bind specifically to several types of proteins, including antibodies, receptors, cytokines, and chaperones (21). Peptide libraries have also been used to isolate proteases substrates (16, 25). For this application, libraries can only be used in vitro with purified enzymes, and consequently the selection of substrates may be biased by the experimental conditions. This aspect could be particularly relevant for intracellular proteases of eukaryotic cells, whose activity often depends on the particular microenvironment in which each enzyme is located.

The development of genetic libraries in eukaryotic viruses represents a strategy to overcome the limitations of the in vitro approach. Recently, Buchholz and colleagues described the first example of in vivo selection of protease cleavage sites in mammalian cells (4). Using a retroviral display library, these authors identified a number of cleaved sequences with a common motif characteristic for proteases belonging to the family of proprotein convertase. The experimental strategy employed did not allow a specific protease to be chosen, and individual sequences were selected based on their ability to be cleaved by unknown cellular proteases.

We report a novel type of peptide library based on recombinant Sindbis virus (SBV) and its application to the identification of substrates of the serine protease of the hepatitis C virus (HCV) (10). SBV, the prototype of the alphavirus genus, is a positive-stranded RNA virus which propagates lytically in mammalian cells (20). With the advent of readily available infectious cDNA, it became possible to use SBV to express foreign genes in cultured cells and experimental animals (11, 19). The SBV genomic RNA contains two genes: one is located at the 5' end, spans two-thirds of the genome, and codes for a

precursor protein which is processed to form the replication machinery; the other occupies the rest of the genome, is translated from the 26S subgenomic RNA, and encodes the precursor of all structural proteins. The 26S region is dispensable for replication and can be manipulated to express foreign genes (11, 19, 20).

HCV is a human pathogen which affects about 2% of the world population and for which no effective therapy is available (10). Its genomic RNA encodes a precursor polyprotein of about 3,000 amino acids which is processed by cellular and viral proteases in at least 10 individual proteins, in the order C, E1, E2, p7, NS2, NS3, NS4A, NS4B, NS5A, and NS5B. The HCV serine protease (NS3-4Ap) is a heterodimeric enzyme comprising the N-terminal domain of the NS3 protein and the central domain of the NS4A protein and cleaves the nonstructural portion of the viral polyprotein at four junctions: NS3/NS4A cleavage is the first event and occurs only intramolecularly (*cis*), followed by cleavages at the NS5A/NS5B, NS4A/NS4B, and NS4B/NS5A sites, which can also be intermolecular (*trans*) (1, 5, 9, 26, 28). Being responsible for the release of the components of the virus replication machinery, this enzyme is considered a pivotal player in the virus life cycle and is thus a candidate target for developing HCV therapies (3). This protease has been extensively characterized at both the biochemical and structural level (12, 14, 18, 23, 24) and its specificity has been defined by identification and mutagenesis of its natural cleavage sites (2, 9, 13, 18, 27, 29). Thus, it is an ideal enzyme for validating the use of SBV-based libraries as a tool for identifying protease substrates.

We recently described the generation of stable SBV-HCV chimeric viruses whose propagation depends on the activity of NS3-4Ap (6, 7). In these viruses, the release of the SBV capsid-autoprotease protein (C) from the structural polyprotein precursor was transformed from an event dependent on the autoproteolytic activity of C into an event dependent on the activity of NS3-4Ap (Fig. 1A). NS3-4Ap was fused at the N terminus of the inactivated SBV capsid-autoprotease (C), and specific cleavage sites were engineered between these two proteins (NS4A/C) and between C and the SBV PE2 glycoprotein

\* Corresponding author. Mailing address: Istituto di Ricerche di Biologia Molecolare P. Angeletti, Via Pontina km 30.600, 00040 Pomezia (Rome), Italy. Phone: 39 06 91093239. Fax: 39 06 91093225. E-mail: migliaccio@irbm.it.



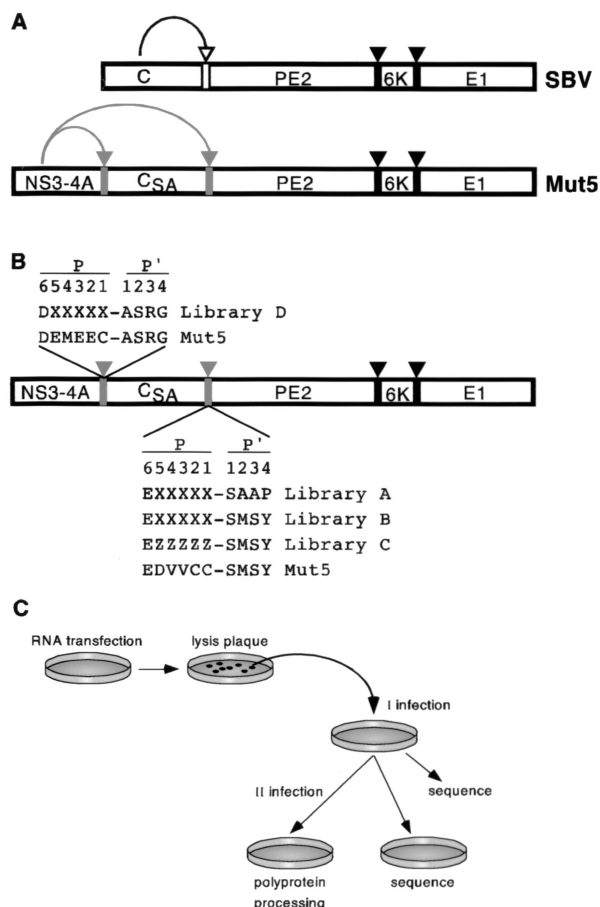


FIG. 1. Experimental outline. (A) Schematic representation of SBV and Mut5 virus polyprotein processing. SBV structural polyprotein is processed co- and posttranslationally to yield the nucleocapsid protein (C) and three membrane proteins (PE2, 6K, and E1), which form the viral envelope. C is an autoprotease and cleaves itself from the nascent precursor, which is then directed to the endoplasmic reticulum by a signal sequence located at the N terminus of PE2. PE2, 6K, and E1 are released from the precursor by signal peptidase. Processing of the Mut5 structural precursor depends on the activity of a fusion protein comprising the protease domain of NS3 and the entire NS4A (NS3-4A). NS3-4A is fused at the N terminus of C, which is inactivated by substitution of its catalytic serine ( $C_{SA}$ ), and two specific cleavage sites are engineered at the N and C termini of C. Empty, gray, and black arrows and bars indicate the cleavage sites of C, NS3-4A, and signal peptidase, respectively. (B) Schematic diagram of the structural polyprotein encoded by libraries of chimeric SBV viruses. The amino acid sequences of the NS4A/C and C/PE2 junctions are indicated in the single-letter code. X, any amino acid or stop codon; Z, any amino acid but cysteine. The position (P6 to P4') of each residue with respect to the scissile bond (—) is indicated. (C) Selection scheme. See Results for details.

pendent, Mut5 forms small but clearly detectable lysis plaques at 30°C (6).

Libraries were constructed by randomizing the nucleotide sequence coding for amino acids P5 to P1 of either the NS4A/C (library D) or the C/PE2 (libraries A, B, and C) cleavage site (Fig. 1B). In libraries A, B, and D the nucleotide sequence was entirely degenerated, thus allowing all codons at each position. Library C was assembled with a codon-based mutagenic strategy (17), such that each degenerated nucleotide triplet coded for all amino acids but cysteine, which is the P1 residue of the natural *trans*-cleavage sites of NS3-4A<sub>p</sub>. Libraries A and B differed only in the P2'-P4' sequence of the C/PE2 cleavage site: in library A the P' side was identical to that of SBV (SAAP); in library B it was derived from the NS5A/B site of HCV and was the same as Mut5 (SMSY). Libraries A, B, C,

and D comprised, respectively,  $4 \times 10^6$ ,  $5 \times 10^6$ ,  $10 \times 10^6$ , and  $1 \times 10^6$  independent bacterial clones, 70% or more of which contained the desired insert. Sequencing of 12 random clones from each library revealed that each clone had a distinct sequence, and no sequence consensus could be derived (not shown). With the exception of library C, which lacked cysteine and stop codons, all amino acids were represented in at least one position and 27% of the clones contained a stop codon in one of the degenerated positions.

Processing of the polyproteins encoded by the different libraries was analyzed by immunoblot. Libraries A, B, and C showed a correctly processed NS3-4A protein and aberrantly processed C and PE2 proteins (data not shown), suggesting that most of the clones were not processed at the C/PE2 junction. Library D showed not only a PE2 protein of the predicted size and aberrantly processed NS3-4A and C proteins but also a small amount of NS3-4A and C proteins of the correct size (data not shown), indicating that a small but significant fraction of the clones were processed at the NS4A/C junction.

**Library screening.** Libraries were screened by transfecting the in vitro-transcribed RNAs in BHK cells and isolating viruses able to form lysis plaques (Fig. 1C). In each experiment,  $1.6 \times 10^7$  BHK cells were transfected by electroporation with each RNA. Transfection efficiency varied between 40 and 80%. Three days after transfection, lysis plaques were picked and isolated viruses were amplified by passaging on BHK cells. RNA was extracted from infected cells, and the cDNA coding for the degenerated cleavage site was amplified by PCR and sequenced. Media containing infectious viruses were used for a second round of amplification on BHK cells to monitor processing of viral proteins and to confirm sequence.

Libraries A and B were screened three times: infectious viruses were recovered with higher frequency from library B ( $1 \times 10^{-4}$  to  $3 \times 10^{-4}$ ) than from library A ( $0.5 \times 10^{-5}$  to  $1.5 \times 10^{-5}$ ). Forty plaque-forming viruses were isolated from library A, and 70 were isolated from library B. Three viruses from library A and 34 from library B were completely characterized, while the remaining isolates yielded ambiguous sequences or could not be propagated.

The C/PE2 junctions of the three viruses isolated from library A all displayed a different sequence, while viruses from library B yielded 30 different sequences, 4 of which occurred twice (Table 1). In all selected viruses but A14, the fifth degenerated residue was cysteine. Sequences were aligned based on the cysteine-serine motif occurring either at the junction between the randomized and the constant region or, in the case of clone A14, at the fourth and fifth degenerated positions. Based on this alignment and similarly to the *trans*-cleavage sites of HCV, all clones had a cysteine in the P1 position. The P3 position accepted only valine (22 of 33), isoleucine (7 of 33), glutamic acid (3 of 33), and threonine (1 of 33). Notably, valine, glutamic acid, and threonine are the residues most frequently found in the P3 position of the natural cleavage sites of HCV, while isoleucine is present only in few HCV genotypes. Positions P5, P4, and P2 tolerated most residues, although position P2 displayed a preference for leucine (12 of 33), position P4 accepted mostly residues with a nonpolar aliphatic side chain (21 of 33), and position P5 preferred charged residues (13 of 33). Clones B14 and B46 encoded the amino acid sequence of the HCV NS5A/NS5B cleavage site engineered in the Mut5 virus, but the corresponding nucleotide sequences differed from that of Mut5.

To investigate if the presence of cysteine in the P1 position was necessary, we screened library C, in which each degenerated position encoded all amino acids but cysteine. In several experiments, plaque-forming viruses were found with a very

TABLE 1. Comparison of sequences of C/PE2 junctions of viruses selected from libraries A and B with those of the *trans*-cleavage sites present in the polyprotein of the majority of HCV isolates

Clone(s)	P <sup>a</sup>							P <sup>a</sup>					
	8	7	6	5	4	3	2	1	1	2	3	4	5
Library A													
A35	T	P	E	<b>K</b>	<b>V</b>	<b>V</b>	<b>L</b>	<b>C</b> <sup>b</sup>	S	A	A	P	
A1	T	P	E	<b>T</b>	<b>I</b>	<b>I</b>	<b>K</b>	<b>C</b>	S	A	A	P	
A14		T	P	E	<b>V</b>	<b>V</b>	<b>Y</b>	<b>C</b>	<b>S</b>	S	A	A	P
Library B													
B6	T	P	E	<b>D</b>	<b>Y</b>	<b>I</b>	<b>S</b>	<b>C</b>	S	M	S	Y	
B20	T	P	E	<b>E</b>	<b>Y</b>	<b>V</b>	<b>N</b>	<b>C</b>	S	M	S	Y	
B8	T	P	E	<b>Q</b>	<b>V</b>	<b>V</b>	<b>R</b>	<b>C</b>	S	M	S	Y	
B14, B46	T	P	E	<b>D</b>	<b>V</b>	<b>V</b>	<b>C</b>	<b>C</b>	S	M	S	Y	
B80	T	P	E	<b>D</b>	<b>V</b>	<b>V</b>	<b>Y</b>	<b>C</b>	S	M	S	Y	
B62	T	P	E	<b>D</b>	<b>V</b>	<b>I</b>	<b>Q</b>	<b>C</b>	S	M	S	Y	
B82	T	P	E	<b>E</b>	<b>I</b>	<b>V</b>	<b>A</b>	<b>C</b>	S	M	S	Y	
B10, B27	T	P	E	<b>L</b>	<b>V</b>	<b>V</b>	<b>P</b>	<b>C</b>	S	M	S	Y	
B3	T	P	E	<b>E</b>	<b>V</b>	<b>I</b>	<b>P</b>	<b>C</b>	S	M	S	Y	
B84	T	P	E	<b>S</b>	<b>L</b>	<b>V</b>	<b>P</b>	<b>C</b>	S	M	S	Y	
B28	T	P	E	<b>N</b>	<b>S</b>	<b>V</b>	<b>P</b>	<b>C</b>	S	M	S	Y	
B40	T	P	E	<b>N</b>	<b>M</b>	<b>V</b>	<b>K</b>	<b>C</b>	S	M	S	Y	
B78	T	P	E	<b>N</b>	<b>I</b>	<b>V</b>	<b>K</b>	<b>C</b>	S	M	S	Y	
B44	T	P	E	<b>M</b>	<b>T</b>	<b>V</b>	<b>K</b>	<b>C</b>	S	M	S	Y	
B56, B47	T	P	E	<b>T</b>	<b>T</b>	<b>V</b>	<b>N</b>	<b>C</b>	S	M	S	Y	
B9	T	P	E	<b>C</b>	<b>T</b>	<b>V</b>	<b>N</b>	<b>C</b>	S	M	S	Y	
B74	T	P	E	<b>S</b>	<b>V</b>	<b>I</b>	<b>N</b>	<b>C</b>	S	M	S	Y	
B2	T	P	E	<b>A</b>	<b>L</b>	<b>V</b>	<b>H</b>	<b>C</b>	S	M	S	Y	
B7	T	P	E	<b>E</b>	<b>T</b>	<b>V</b>	<b>L</b>	<b>C</b>	S	M	S	Y	
B1	T	P	E	<b>E</b>	<b>T</b>	<b>I</b>	<b>L</b>	<b>C</b>	S	M	S	Y	
B11	T	P	E	<b>E</b>	<b>A</b>	<b>V</b>	<b>L</b>	<b>C</b>	S	M	S	Y	
B19, B89	T	P	E	<b>R</b>	<b>L</b>	<b>V</b>	<b>L</b>	<b>C</b>	S	M	S	Y	
B81	T	P	E	<b>R</b>	<b>L</b>	<b>I</b>	<b>L</b>	<b>C</b>	S	M	S	Y	
B15	T	P	E	<b>R</b>	<b>V</b>	<b>V</b>	<b>L</b>	<b>C</b>	S	M	S	Y	
B41	T	P	E	<b>G</b>	<b>I</b>	<b>V</b>	<b>L</b>	<b>C</b>	S	M	S	Y	
B42	T	P	E	<b>V</b>	<b>E</b>	<b>V</b>	<b>L</b>	<b>C</b>	S	M	S	Y	
B4	T	P	E	<b>T</b>	<b>V</b>	<b>E</b>	<b>L</b>	<b>C</b>	S	M	S	Y	
B48	T	P	E	<b>D</b>	<b>V</b>	<b>E</b>	<b>L</b>	<b>C</b>	S	M	S	Y	
B43	T	P	E	<b>D</b>	<b>I</b>	<b>T</b>	<b>L</b>	<b>C</b>	S	M	S	Y	
B5	T	P	E	<b>T</b>	<b>L</b>	<b>E</b>	<b>F</b>	<b>C</b>	S	M	S	Y	
HCV													
NS4A/4B	E	F	D	<b>E</b>	<b>M</b>	<b>E</b>	<b>E</b>	<b>C</b>	A	S	H	L	
NS4B/5A	N	E	D	<b>C</b>	<b>S</b>	<b>T</b>	<b>P</b>	<b>C</b>	S	G	S	W	
NS5A/5B	A	S	E	<b>D</b>	<b>V</b>	<b>V</b>	<b>C</b>	<b>C</b>	S	M	S	Y	

<sup>a</sup> Presumed positions are given with respect to the cleavage site.  
<sup>b</sup> Selected sequences are shown in boldface type, with 3 and 4 adjacent N- and C-terminal amino acids, respectively.

low frequency ( $0.5 \times 10^{-6}$  to  $1.5 \times 10^{-6}$ ). Forty viruses were isolated but none could be propagated, thus suggesting that the P1 cysteine residue was required for efficient cleavage of the C/PE2 junction.

Library D was screened once and yielded infectious viruses with high frequency ( $5 \times 10^{-2}$ ), suggesting that NS3-4Ap could cleave a large number of different sequences at the NS4A/C site. Twenty plaques were isolated and used to infect BHK cells, and nine viruses were propagated. The nine different sequences displayed by these viruses did not show an obvious consensus (Table 2) and were aligned assuming that the junction between the fifth degenerated residue and the serine residue of the constant sequence represented the scissile bond. Based on this alignment, the P1 position was occupied only by cysteine (7 of 9) or threonine (2 of 9), while the other degenerated positions did not show a preference for particular res-

idues, except for the presence of charged residues in P2 (6 of 9). Remarkably, a similar lack of consensus has been reported for the *cis*-cleaved NS3/NS4A junction of HCV, which is the only cleavage site with a threonine residue in P1 and tolerates mutations in almost all positions (2).

**Characterization of selected cleavage sites.** In vivo processing of the polyproteins of selected viruses was initially investigated by immunoblotting with antibodies directed to the proteins flanking the degenerated junctions. With three exceptions described below, cells infected with viruses selected from libraries A and B produced PE2 and C proteins comigrating with those produced by the Mut5 virus (Fig. 2A and data not shown), demonstrating that the C/PE2 junctions of selected viruses were cleaved, presumably at the predicted cysteine-serine motif. Cells infected with clones B15, B19, and B81 produced correct-size PE2 proteins (not shown) and C proteins migrating slightly faster than that of Mut5 (Fig. 2A). Migration of these C proteins is probably due to the sequence of the degenerated region rather than to cleavage(s) occurring at an alternative or additional site(s) in the C protein. In fact, in vitro-synthesized ΔC proteins—bearing the sequences present in clones B15, B19, and B81 and truncated after the cysteine residue in position P1—also displayed anomalous migration (Fig. 2B). Similarly, cells infected with viruses selected from library D produced C proteins comigrating with those produced by the Mut5 virus (Fig. 2A and data not shown), indicating that also in this case the selected NS4A/C junctions were presumably cleaved at the predicted cysteine/threonine-alanine motif.

With the exception of the D1 virus, no accumulation of uncleaved NS3-4A/C (library D) or C/PE2 (libraries A and B) precursors was observed in cells infected with all selected viruses (data not shown), indicating that at steady state all but one selected sequences were cleaved to the same extent as the bona fide HCV sequences engineered in the Mut5 virus. This result demonstrated that the selection scheme yielded only sequences cleaved by NS3-4Ap well enough to ensure an almost-complete processing of the viral polyprotein. Nevertheless, it was possible that the cleavage efficiency of selected substrates could vary in a limited range and could affect the propagation ability of the corresponding viruses. Indeed, comparison of selected viruses showed that while the majority of them were indistinguishable from the Mut5 virus, four grew to

TABLE 2. Comparison of sequences of NS4A/C junctions of viruses selected from library D with that of the NS3/NS4A *cis*-cleavage site found in the majority of HCV isolates

Clone	P <sup>a</sup>							P <sup>a</sup>				
	8	7	6	5	4	3	2	1	1	2	3	4
Library D												
D1	E	F	D	<b>T</b>	<b>M</b>	<b>S</b>	<b>E</b>	<b>T</b> <sup>b</sup>	A	S	R	G
D3	E	F	D	<b>E</b>	<b>K</b>	<b>I</b>	<b>Q</b>	<b>T</b>	A	S	R	G
D43	E	F	D	<b>K</b>	<b>K</b>	<b>L</b>	<b>Q</b>	<b>C</b>	A	S	R	G
D32	E	F	D	<b>V</b>	<b>R</b>	<b>E</b>	<b>K</b>	<b>C</b>	A	S	R	G
D6	E	F	D	<b>V</b>	<b>A</b>	<b>T</b>	<b>K</b>	<b>C</b>	A	S	R	G
D48	E	F	D	<b>S</b>	<b>D</b>	<b>E</b>	<b>K</b>	<b>C</b>	A	S	R	G
D30	E	F	D	<b>A</b>	<b>T</b>	<b>P</b>	<b>T</b>	<b>C</b>	A	S	R	G
D45	E	F	D	<b>M</b>	<b>M</b>	<b>T</b>	<b>D</b>	<b>C</b>	A	S	R	G
D7	E	F	D	<b>S</b>	<b>Q</b>	<b>E</b>	<b>D</b>	<b>C</b>	A	S	R	G
HCV NS3/4A	S	A	D	<b>L</b>	<b>E</b>	<b>V</b>	<b>V</b>	<b>T</b>	S	T	W	V

<sup>a</sup> Presumed positions are given with respect to the cleavage site.  
<sup>b</sup> Selected sequences are shown in boldface type, with 3 and 4 adjacent N- and C-terminal amino acids, respectively.

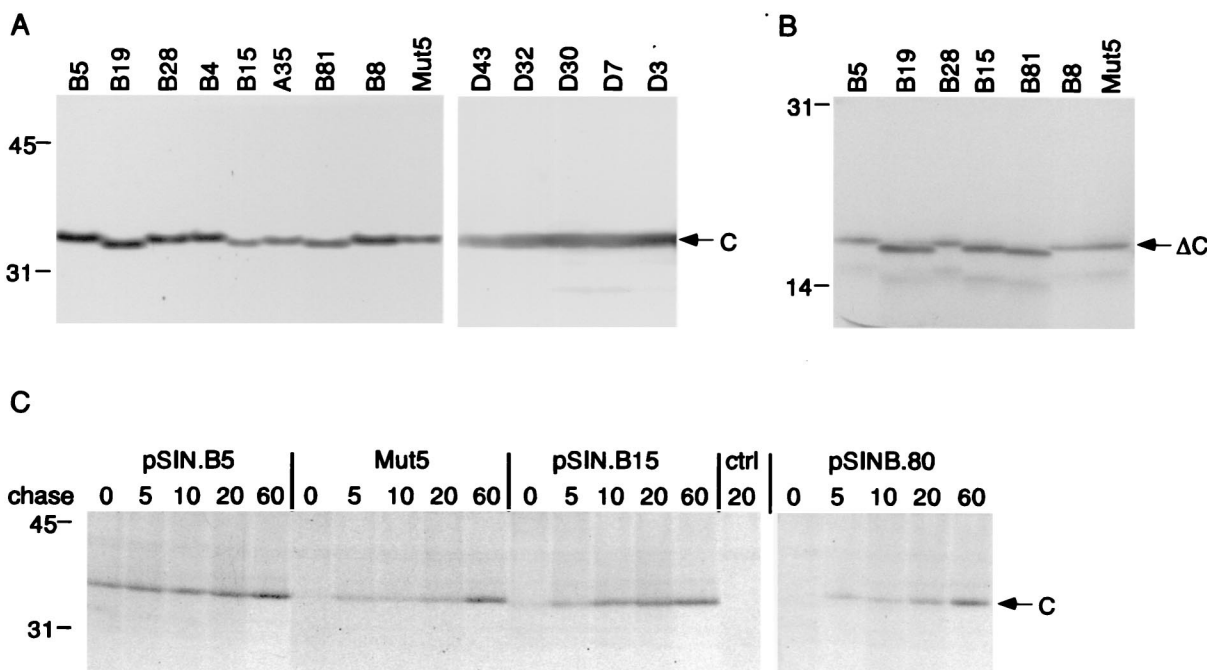


FIG. 2. In vivo processing of the polyprotein of selected viruses. (A) BHK cells were infected with the indicated viruses, and cell lysates were analyzed by immunoblot with the anti-SBV C antibody by SDS-10% PAGE. (B)  $\Delta$ C proteins, comprising amino acids 106 to 264 of SBV C and containing the P5-P1 sequence of the indicated viruses, were translated in vitro and analyzed by SDS-12% PAGE and autoradiography. (C) BHK cells were mock infected (ctrl) or infected with the indicated recloned viruses, pulse-labeled with  $^{35}$ S-labeled amino acids for 5 min at 37°C, and then chased for the time (in minutes) indicated above each lane. The C protein was immunoprecipitated and analyzed by SDS-12% PAGE and autoradiography. Positions of molecular mass standards (in kilodaltons), C, and  $\Delta$ C proteins are indicated. For clarity, only a representative subset of samples is shown.

higher titers and formed slightly larger (A35 and B15) or clearly larger (B5 and B19) lysis plaques (Fig. 3 and data not shown). Recloning the selected junctions in the Mut5 genome and analysis of the resulting viruses confirmed that the sequences of the C/PE2 junctions were responsible for the different plaque phenotypes of the selected viruses (Fig. 3 and data not shown). To ascertain whether the size of lysis plaques reflected differences in the processing kinetics overlooked by the immunoblot, we analyzed by pulse-chase labeling the biogenesis of the C proteins of recloned viruses (pSIN.A35, pSIN.B5, pSIN.B15, pSIN.B19, pSIN.B28, and pSIN.B80). The amount of C protein immunoprecipitated from infected cells increased over time and was similar for all tested viruses after 60 min of chase (Fig. 2C and data not shown), indicating that the C protein was released from the precursor in a time-dependent fashion. However, the kinetics of release for viruses which formed larger plaques was slightly (pSIN.A35 and pSIN.B15) or clearly (pSIN.B5 and pSIN.B19) faster than that of Mut5, pSIN.B28, and pSIN.B80 (Fig. 2C and data not shown). Since these viruses only differed in the sequence of the C/PE2 site, this result indicated that the C/PE2 junctions of clones A35, B15, B5, and B19 were cleaved in vivo with faster kinetics.

In vitro characterization of selected sequences was restricted to clones derived from libraries A and B, since we anticipated that, similarly to the natural *cis*-cleavage site of HCV (27, 28), sequences derived from library D would not be efficiently cleaved in *trans*. To demonstrate that the selected C/PE2 junctions were cleaved by NS3-4Ap, we tested whether this enzyme processed in vitro-translated  $\Delta$ C- $\Delta$ PE2 substrates containing the C/PE2 junctions present in selected viruses. In the absence of the protease, all  $\Delta$ C- $\Delta$ PE2 proteins migrated on gel as predicted by their mass (Fig. 4). Addition of increasing amounts of

recombinant NS3-4Ap (8) resulted in the decrease of the  $\Delta$ C- $\Delta$ PE2 substrate and in the appearance of the expected  $\Delta$ C and  $\Delta$ PE2 cleavage products, proving that all selected sequences were cleaved by NS3-4Ap (Fig. 4 and data not shown). All  $\Delta$ C- $\Delta$ PE2 proteins were cleaved with indistinguishable efficiency in several experiments (data not shown), indicating that although all selected C/PE2 junctions were good cleavage sites, this in vitro assay could not be used to appreciate kinetic differences.

To obtain quantitative data, we determined the kinetic parameters of decapeptide substrates containing the sequences present in a subset of selected viruses (B5, B8, B14, B15, B19, and B28). To overcome synthetic difficulties, in all peptides the methionine at position P2' was substituted by norleucine, and the cysteine present at position P2 of clone B14 was replaced by aminobutyric acid. In vitro, the stability of the NS3-NS4A complex and affinity for substrates are influenced by the buffer composition and in particular by the effect of ionic strength (8, 22). Thus, we determined  $K_m$  and  $k_{cat}$  values for selected peptide substrates both in the presence and absence of 150 mM NaCl. All selected peptides displayed an overall catalytic efficiency of the same order of magnitude of that measured for decapeptides containing the sequence present in the NS5A/NS5B and NS4A/NS4B sites of HCV (Table 3). Notably, while the overall catalytic efficiency ( $k_{cat}/K_m$ ) of the peptides B5s and NS4A/Bs decreased significantly in the presence of NaCl, other peptides were cleaved equally well under the two tested conditions (B8s, B19s, B28s, and B14s) or even more efficiently in the presence of salt (B15s). In almost all cases these differences were due to  $K_m$  changes, although there were also small  $k_{cat}$  variations. These results indicated that comparison between substrates was biased by the experimental conditions, justifying the apparent discordance with the processing kinetics observed

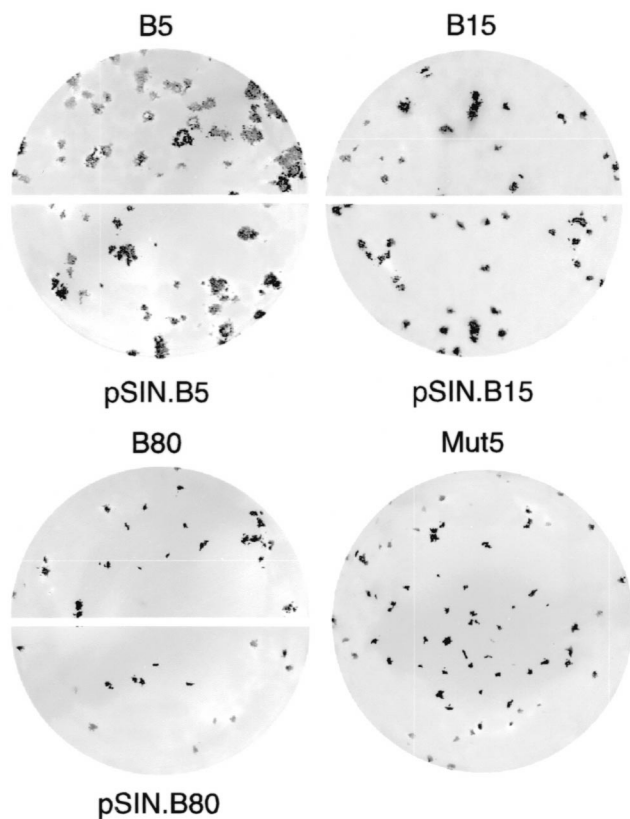


FIG. 3. Comparison of the plaque phenotype of selected and recloned viruses. Shown are plaques produced in BHK cells after infection with the selected (B5, B15, and B80) and recloned (pSIN.B5, pSIN.B15, pSIN.B80, and Mut5) viruses. Plaques were revealed by immunostaining with the rabbit anti-SBV E2 serum after 3 days of incubation at 30°C.

in vivo. Finally, HPLC analysis indicated that the C-terminal cleavage products of all substrates had the same retention time, further confirming that cleavage occurred at the predicted C-S junction.

Peptides corresponding to the N-terminal cleavage products of NS3-4Ap are competitive inhibitors of this enzyme (22). To determine whether selected sequences were also able to inhibit the enzyme, we tested hexapeptides corresponding to residues P6 to P1 of the cleavage sites present in viruses B5, B14, B15, and B19 (Table 4). These peptides showed inhibitory potencies

similar to or higher than those derived from the natural HCV sites, but also in this case the affinity for the enzyme was influenced by the experimental conditions. These data confirmed that selected sequences, independently of the P' region, had by themselves a good affinity for the enzyme, and suggested that they could be used as a scaffold for the generation of peptidomimetic inhibitors.

## DISCUSSION

This study describes a new viral display system for in vivo use and its application to defining the substrate specificity of the HCV serine protease.

N-terminal sequencing of the NS4A-to-NS5B proteins and comparison of HCV strains initially delineated a consensus for the NS3-4Ap cleavage (9, 18), which was then refined by mutational analysis in transfected cells (2, 13) and in vitro studies with synthetic peptides (27, 29). Overall, these studies indicated different requirements for cleavage at the *cis* and *trans* sites. The *cis*-cleaved NS3/NS4A junction of HCV is the only natural cleavage site with threonine in P1 and tolerates mutations at almost all positions, indicating that processing at this site is primarily determined by polyprotein folding (2, 13). In contrast, a cysteine residue at the P1 position is absolutely required for *trans* cleavage, and additional residues are crucial for efficient processing (2, 13, 27, 29). An acidic residue is present at the P6 position of all cleavage sites, and its substitution is detrimental for cleavage both in vivo and in vitro. Also, P3 and P4' residues contribute to efficient substrate recognition, with preference shown for residues present in the natural cleavage sites. Efficient in vitro cleavage requires a peptide substrate of at least 10 residues spanning P6 to P4', suggesting that ground-state substrate binding to NS3-4Ap is mediated by multiple, weak interactions involving distal residues (P6, P3, P1', and P4') and main-chain interactions between P5 and P2, while the P1 residue dictates the efficiency with which the substrate proceeds to the transition state.

Our results demonstrated not only that the SBV system could be used to depict a comprehensive picture of the NS3-4Ap specificity but also that it was sophisticated enough to reproduce the different requisites for *cis* and *trans* cleavage observed for the HCV polyprotein. Specificity of the protease at the *cis*-cleaved NS4A/C junction was highly degenerated, as indicated by the fact that 5% of the clones present in library D yielded infectious viruses and that sequences found in selected viruses did not show an obvious consensus, except for the presence of cysteine or threonine residues in the P1 position.

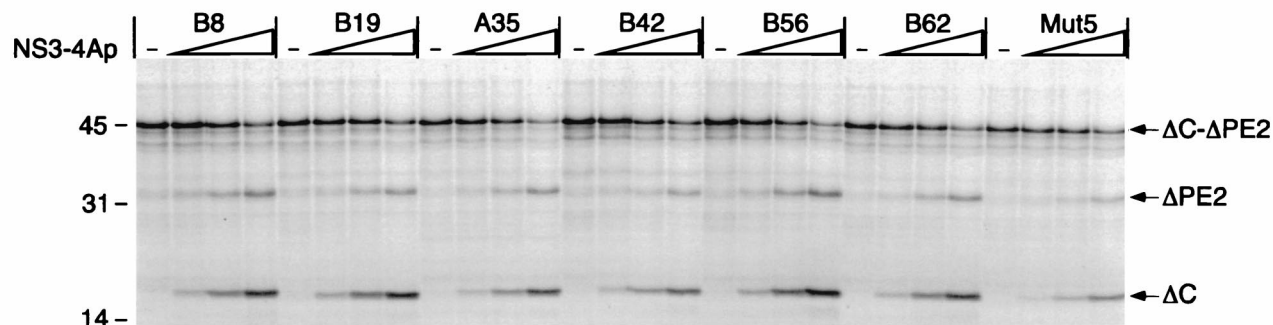


FIG. 4. In vitro processing of the polyprotein of viruses selected from libraries A and B. In vitro-translated  $\Delta$ C- $\Delta$ PE2 proteins, comprising amino acids 106 to 533 of SBV structural polyprotein and containing the C/PE2 cleavage sites of the indicated viruses, were incubated without (-) or with increasing amounts of full-length NS3-4Ap (final concentrations, 5, 20, and 80 nM, as symbolized by the triangles above the gel) and analyzed by SDS-12% PAGE and autoradiography. Positions of molecular mass standards (in kilodaltons),  $\Delta$ C- $\Delta$ PE2,  $\Delta$ C and  $\Delta$ PE2 proteins are indicated. For clarity, only a representative subset of samples is shown.

TABLE 3. Activity of substrate peptides based on the sequences of NS4A/B, NS5A/B, and selected cleavage sites<sup>a</sup>

Peptide	Sequence <sup>b</sup>	$K_m$ ( $\mu\text{M}$ )		$k_{\text{cat}}$ ( $\text{min}^{-1}$ )		$k_{\text{cat}}/K_m$ ( $\text{M}^{-1}\text{s}^{-1}$ )	
		+	-	+	-	+	-
B5 <sub>s</sub>	ETLEFCS $\bar{\text{A}}$ SY	27	3.5	36	23	22,200	109,000
B8 <sub>s</sub>	EQVVRCS $\bar{\text{A}}$ SY	21	7	10.5	3	8,300	7,100
B15 <sub>s</sub>	ERVVLC $\bar{\text{A}}$ SY	13	30	44.5	30	57,000	16,600
B19 <sub>s</sub>	ERLVLC $\bar{\text{A}}$ SY	12	11	49	30	68,000	45,000
B28 <sub>s</sub>	ENSVPCS $\bar{\text{A}}$ SY	56	15	14.5	4	4,300	4,400
B14 <sub>s</sub> (NS5A/B)	EDVV $\bar{\text{A}}$ CS $\bar{\text{A}}$ SY	10	4	47	34	78,000	141,000
NS4A/B <sub>s</sub>	DEMEECASHL	500	10	4	4	133	7,000

<sup>a</sup> Kinetic parameters were determined in the presence (+) or absence (-) of 150 mM NaCl.

<sup>b</sup>  $\bar{\text{a}}$ , norleucine.

These findings were in perfect agreement with the poor specificity observed for *cis* cleavage in the context of the HCV polyprotein (2, 13). Conversely, the amino acid sequences of the selected C/PE2 junctions delineated a *trans*-cleavage consensus entirely consistent with all previous data. The P1 residue of all selected viruses was cysteine, and we were not able to isolate and propagate viruses from the library lacking cysteine residues, indicating the absolute requirement for this residue in P1. The only four residues found at the P3 position (valine, isoleucine, glutamic acid, and threonine) were the same found at this position of HCV *trans*-cleavage sites. Although an acidic residue was present in the constant part of our libraries, the fact that it was in the putative P6 position in all but one selected virus underlined its importance. Comparison of libraries differing only for the P2'-P4' sequence emphasized the importance of the P' part of the substrate: the frequency of infectious viruses was higher in library B, which contained an HCV sequence, than in library A, which contained an SBV sequence. In addition, most viruses isolated from the latter library could not be propagated, suggesting that C/PE2 junctions of these viruses were cleaved inefficiently.

All selected sequences were cleaved by NS3-4Ap *in vivo* and *in vitro* with an efficiency comparable to that of the natural *cis*- or *trans*-cleavage sites of HCV, demonstrating the reliability of the SBV library system. Interestingly, *in vivo* analysis by pulse-chase labeling indicated that four of the sequences selected from libraries A and B were processed faster than the natural NS5A/B substrate, albeit the corresponding peptides did not show enhanced kinetic parameters in all the *in vitro* assays. This apparent discrepancy could be rationalized considering that *in vitro* determinations of kinetic parameters (Table 3) are biased by the experimental conditions, which are unavoidably arbitrary and affect each substrate differently. In addition, the context in which the different substrates were analyzed, short peptides *in vitro* and larger precursor proteins *in vivo*, could also account for the different results obtained *in vitro* and *in vivo*. Therefore, we feel that comparison of substrates with *in vitro* assays based on peptide substrates, albeit fundamental for obtaining quantitative data, must be considered valid only for defined experimental conditions and should, whenever possible, be validated by *in vivo* experiments with full-length proteins.

Our results also showed a correlation between the propagation phenotype of selected viruses and the kinetics of *in vivo* processing of the corresponding cleavage sites, indicating that the plaque phenotype of the selected viruses can be utilized to identify sequences which are cleaved more efficiently *in vivo*. This correlation is probably not absolute, since it is possible that the sequence of the cleavage site affects the virus pheno-

TABLE 4. Activity of inhibitor peptides based on the sequences of NS4A/B, NS5A/B, and selected cleavage sites<sup>a</sup>

Peptide	Sequence <sup>b</sup>	IC <sub>50</sub> ( $\mu\text{M}$ )	
		+	-
B5 <sub>i</sub>	ETLEFC	37	1.3
B15 <sub>i</sub>	ERVVLC	60	60
B19 <sub>i</sub>	ERLVLC	42	40
B14 <sub>i</sub> (NS5A/B)	EDVV $\bar{\text{A}}$ C	100	4
NS4A/B <sub>i</sub>	DEMEEC	>100	1.5

<sup>a</sup> Kinetic parameters were determined in the presence (+) or absence (-) of 150 mM NaCl.

<sup>b</sup>  $\bar{\text{a}}$ , aminobutyric acid.

type also by influencing the structure and/or function of the cleaved C protein. Despite this limitation, we have shown that the possibility of ranking selected viruses on the basis of their plaque phenotype was very useful for identifying a manageable number of substrates whose processing efficiency could be estimated both *in vivo* and *in vitro* by biochemical and kinetic methods.

Similarly to other proteases, NS3-4Ap is competitively inhibited by peptides corresponding to the N-terminal cleavage products (22). We have exploited this feature to demonstrate that selected sequences inhibit the enzyme *in vitro* with potencies comparable to those of peptides derived from the natural cleavage sites. Because of their size and peptidic nature, these inhibitors do not enter efficiently into cells and cannot be tested *in vivo*. Nonetheless, the identification of new sequences efficiently recognized by the enzyme *in vivo* provides useful information for the development of synthetic compounds which effectively inhibit NS3-4Ap and eventually could be used to cure the disease caused by HCV.

SBV libraries represent a new tool for applying a combinatorial approach to the definition of protease specificity in the intracellular environment. Their most important property is that they are screened *in vivo*, thus avoiding all problems connected with defining the appropriate *in vitro* conditions, as is the case with phage or chemical libraries. This aspect is obviously a clear advantage when analyzing intracellular enzymes. In this regard, SBV libraries are similar to retroviral libraries (4), the main difference being that, in the present design, SBV libraries allow the target enzyme to be chosen while retroviral libraries do not. We believe that SBV libraries can be easily adapted to study other intracellular proteases. The HCV serine protease was engineered to cleave itself off a polyprotein precursor containing additional cleavage sites, which would not be feasible for other enzymes unable to cleave at their C terminus. However, the target enzyme does not need to be fused to the SBV structural polyprotein, and alternative expression strategies could be used, as SBV vectors are an extremely flexible experimental system. Duplication of the internal promoter of the viral replicase would allow expression of the protease and of the structural polyprotein, containing a randomized cleavage site at the C/PE2 junction, from two different mRNAs (11). Alternatively, the protease and the polyprotein could be expressed by a bicistronic mRNA obtained by the insertion of an internal ribosome entry site.

An important feature of SBV libraries is that they do not require a purified enzyme but only require the cognate cDNA. In theory, it is conceivable to study proteins which are predicted to be proteolytic enzymes on the basis of their sequence and for which no biochemical information is available. While phage-displayed and chemical peptide libraries can be used for

many different proteases, a new SBV library has to be constructed for each enzyme. However, since the SBV genome can be easily manipulated at the cDNA level, constructing libraries is not a limitation and in fact offers the opportunity to redesign the scaffold of the libraries to meet particular requirements of the target enzyme. An additional advantage of SBV libraries is that infectious viruses form lysis plaques and are thus easily isolated and characterized with a single round of selection.

The complexity of our SBV libraries ranged from  $1 \times 10^6$  to  $10 \times 10^6$  clones. Considering that 50% of the clones either did not contain the degenerated sequence or contained a stop codon, and assuming some redundancy, these libraries approached the theoretical diversity predicted for a library with five degenerated positions ( $3.2 \times 10^6$  possible protein sequences). These figures are similar to those reported for phage- or retrovirally displayed libraries (4, 16). Certainly there is still room for improvement: one useful expedient to increase diversity and reduce redundancy is to use a codon-based mutagenesis strategy (17), which also allows the assembly of libraries with selectively degenerated sequences, such as our library C.

#### ACKNOWLEDGMENTS

We acknowledge the contribution of Maria Scaturro in the initial part of this work. We thank Paolo Monaci for helpful discussion and critical review, J. Clench for editing the manuscript, M. Emili for graphics, P. Neuner for oligonucleotide synthesis, and A. Pessi for peptide synthesis. We are grateful to R. Cortese for continuous support and encouragement.

#### REFERENCES

- Bartenschlager, R., L. Ahlborn-Laake, J. Mous, and H. Jacobsen. 1994. Kinetic and structural analysis of hepatitis C virus polyprotein processing. *J. Virol.* **68**:5045–5055.
- Bartenschlager, R., L. Ahlborn-Laake, K. Yasargil, J. Mous, and H. Jacobsen. 1995. Substrate determinants for cleavage in *cis* and in *trans* by the hepatitis C virus NS3 proteinase. *J. Virol.* **69**:198–205.
- Bartenschlager, R. 1999. The NS3/4A proteinase of the hepatitis C virus: unravelling structure and function of an unusual enzyme and a prime target for antiviral therapy. *J. Vir. Hepatitis* **6**:165–181.
- Buchholz, C. J., K. W. Peng, F. J. Morling, J. Zhang, F. L. Cosset, and S. J. Russell. 1999. In vivo selection of protease cleavage sites from retrovirus display libraries. *Nat. Biotechnol.* **16**:951–954.
- Failla, C., L. Tomei, and R. De Francesco. 1994. Both NS3 and NS4A are required for proteolytic processing of hepatitis C virus nonstructural protein. *J. Virol.* **68**:3753–3760.
- Filocamo, G., L. Pacini, and G. Migliaccio. 1997. Chimeric Sindbis viruses dependent on the NS3 protease of the hepatitis C virus. *J. Virol.* **71**:1417–1427.
- Filocamo, G., L. Pacini, C. Nardi, L. Bartholomew, M. Scaturro, P. Delmastro, A. Tramontano, R. De Francesco, and G. Migliaccio. 1999. Selection of functional variants of the NS3-NS4A protease of hepatitis C virus by using chimeric Sindbis viruses. *J. Virol.* **73**:561–575.
- Gallinari, P., C. Paolini, D. Brennan, C. Nardi, C. Steinkühler, and R. De Francesco. 1999. Modulation of hepatitis C virus NS3 protease and helicase activities through the interaction with NS4A. *Biochemistry* **38**:5620–5632.
- Grakoui, A., C. Wychowski, C. Lin, S. M. Feinstone, and C. M. Rice. 1993. Expression and identification of hepatitis C virus polyprotein products. *J. Virol.* **67**:1385–1395.
- Houghton, M. 1996. Hepatitis C viruses, p. 1035–1058. *In* B. N. Fields, D. M. Knipe, and P. M. Howley (ed.), *Fields virology*. Lippincott-Raven Publishers, Philadelphia, Pa.
- Huang, H. V. 1996. Sindbis virus vectors for expression in animal cells. *Curr. Opin. Biotechnol.* **5**:531–535.
- Kim, J. L., K. A. Morgenstern, C. Lin, T. Fox, M. D. Dwyer, J. A. Landro, S. P. Chambers, W. Markland, C. A. Lepre, E. T. O'Malley, S. L. Harbeson, C. M. Rice, M. A. Murcko, P. R. Caron, and J. A. Thomson. 1996. Crystal structure of the hepatitis C virus NS3 protease domain complexed with a synthetic NS4A co-factor peptide. *Cell* **87**:343–355.
- Kolykhalov, A. A., E. V. Agapov, and C. M. Rice. 1994. Specificity of the hepatitis C virus NS3 serine protease: effects of substitution at the 3/4A, 4A/4B, 4B/5A, and 5A/5B cleavage sites on polyprotein processing. *J. Virol.* **68**:7525–7533.
- Love, R. A., H. E. Parge, J. A. Wickersham, Z. Hostomsky, N. Habuka, E. W. Moomaw, T. Adachi, and Z. Hostomska. 1996. The crystal structure of the hepatitis C virus NS3 proteinase reveals a trypsin fold and a structural zinc binding site. *Cell* **87**:331–342.
- Markland, W., B. L. Roberts, and R. C. Ladner. 1996. Selection for protease inhibitors using bacteriophage display. *Methods Enzymol.* **267**:28–51.
- Matthews, D. J., and J. A. Wells. 1993. Substrate phage: selection of protease substrates by monovalent phage display. *Science* **260**:1113–1117.
- Neuner, P., R. Cortese, and P. Monaci. 1998. Codon-based mutagenesis using dimer-phosphoramidites. *Nucleic Acids Res.* **26**:1223–1227.
- Pizzi, E., A. Tramontano, L. Tomei, N. La Monica, C. Failla, M. Sardana, T. Wood, and R. De Francesco. 1994. Molecular modelling of the specificity pocket of the hepatitis C virus protease: implications for substrate recognition. *Proc. Natl. Acad. Sci. USA* **91**:888–892.
- Rice, C. M., R. Levis, J. H. Strauss, and H. V. Huang. 1987. Production of infectious RNA transcripts from Sindbis virus cDNA clones: mapping of lethal mutations, rescue of temperature sensitive marker, and in vitro mutagenesis to generate defined mutants. *J. Virol.* **61**:3809–3819.
- Schlesinger, S., and M. J. Schlesinger. 1996. *Togaviridae: the viruses and their replication*, p. 825–841. *In* B. N. Fields, D. M. Knipe, and P. M. Howley (ed.), *Fields virology*. Lippincott-Raven Publishers, Philadelphia, Pa.
- Scott, J. K., and J. A. Wells. 1994. Random peptide libraries. *Curr. Opin. Biotechnol.* **5**:40–48.
- Steinkühler, C., G. Biasiol, M. Brunetti, A. Urbani, U. Koch, R. Cortese, A. Pessi, and R. De Francesco. 1998. Product inhibition of the hepatitis C virus NS3 protease. *Biochemistry* **37**:8899–8905.
- Steinkühler, C., L. Tomei, and R. De Francesco. 1996. In vitro activity of hepatitis C virus NS3 purified from recombinant baculovirus-infected Sf9 cells. *J. Biol. Chem.* **271**:6367–6373.
- Steinkühler, C., A. Urbani, L. Tomei, G. Biasiol, M. Sardana, E. Bianchi, A. Pessi, and R. De Francesco. 1996. Activity of purified hepatitis C virus protease NS3 on peptide substrates. *J. Virol.* **70**:6694–6700.
- Thornberry, N. A., T. A. Rano, E. P. Peterson, D. M. Rasper, T. Timkey, M. Garcia-Calvo, V. M. Houtzager, P. A. Nordstrom, S. Roy, J. P. Vaillancourt, K. T. Chapman, and D. W. Nicholson. 1997. A combinatorial approach defines specificities of members of the caspase family and granzyme B. Functional relationships established for key mediators of apoptosis. *J. Biol. Chem.* **272**:17907–17911.
- Tomei, L., C. Failla, E. Santolini, R. De Francesco, and N. La Monica. 1993. NS3 is a serine protease required for processing of hepatitis C virus polyprotein. *J. Virol.* **67**:4017–4026.
- Urbani, A., E. Bianchi, F. Narjes, A. Tramontano, R. De Francesco, C. Steinkühler, and A. Pessi. 1997. Substrate specificity of the hepatitis C virus serine protease. *J. Biol. Chem.* **272**:9204–9209.
- Urbani, A., R. De Francesco, and C. Steinkühler. 1999. Proteases of the hepatitis C virus, p. 61–91. *In* B. M. Dunn (ed.), *Proteases of infectious agents*. Academic Press Publishers, New York, N.Y.
- Zhang, R., J. Durkin, W. T. Windsor, C. McNemar, L. Ramanathan, and H. V. Le. 1997. Probing the substrate specificity of hepatitis C virus NS3 serine protease by using synthetic peptides. *J. Virol.* **71**:6208–6213.