National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Canada

UNIVERSITY OF ALBERTA

Expression of an Active Hepatitis C Virus Serine Proteinase in *E. coli*

by

COLIN GUANGSHUN LUO

A THESIS SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF MEDICAL MICROBIOLOGY AND INFECTIOUS DISEASES

EDMONTON, ALBERTA

Spring, 1995

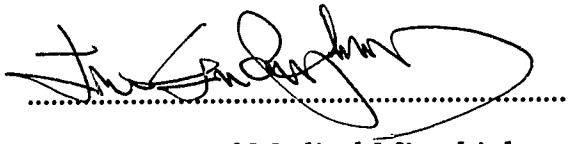
ISBN 0-612-01626-9

Canada

UNIVERSITY OF ALBERTA

LIBRARY RELEASE FORM

NAME OF AUTHOR:	Colin Guangshun Luo
TITLE OF THESIS:	Expression of an Active Hepatitis C Virus Serine Proteinase in *E. coli*
DEGREE:	Master of Science
YEAR THIS DEGREE GRANTED:	1995



Department of Medical Microbiology and Infectious Diseases, University of Alberta, Edmonton, Alberta, Canada

DATED Jan. 25/95

UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled Expression of an Active Hepatitis C Virus Serine Proteinase in *E. coli* submitted by Colin Guangshun Luo in partial fulfillment of the requirements for the degree of Master of Science.

.....
Dr. John F. Elliott (Supervisor)

.....
Dr. David L. J. Tyrrell

.....
Dr. Randy J. Read

.....
Dr. Bruce Malcolm (External Examiner)

Date Jan 18/95

ABSTRACT

Hepatitis C virus (HCV) is a significant cause of human hepatitis, for which there is at present only limited therapies. The virus expresses a serine proteinase which is itself first cleaved from the viral polyprotein; this enzyme then acts to complete processing of the HCV polyprotein. Enzymatic and structural characterization of the HCV serine proteinase is a key first step in the development of specific proteinase inhibitors which may have therapeutic application. Such characterization requires significant quantities of HCV serine proteinase in an enzymatically active form. To obtain a cDNA fragment encoding the HCV serine proteinase, viral RNA was extracted from the serum of infected patients, reverse transcribed, and amplified using the polymerase chain reaction (PCR). The resulting fragments were cloned into a plasmid vector, characterized by DNA sequencing, and after further engineering used to express recombinant protein in *E. coli*. Two slightly different proteins, each fused to an N-terminal Met-His 6 segment, were expressed and purified using a nickel chelating column. Both proteins were deposited as insoluble inclusion bodies in *E. coli*, necessitating the use of a denaturing agent during the extraction and purification process. To obtain an enzymatically active product, attempts were made to refold both of these purified, denatured proteins. Since the precise N-terminus of the HCV serine proteinase was not known at the time this work began, the first protein was engineered to give an N-terminus which was reasonable by comparison with other known serine proteinase enzymes. This protein was used to generate rat

immune sera, but was not tested for catalytic activity, since soon after it was made new data published by another group showed that our first protein did not contain the N-terminal 10 residues of the mature serine proteinase. The second protein was constructed so as to contain the necessary additional N-terminal residues, and in such a way that the mature N-terminus could be generated by enterokinase cleavage. This second recombinant protein was shown to be catalytically active after enterokinase cleavage and refolding, and generated the expected peptide fragment using a model synthetic peptide substrate.

ACKNOWLEDGEMENTS

I would like to express my warmest appreciation to my supervisor Dr. John F. Elliott, for his guidance, encouragement, and endless patience, and to my supervisory committee including Dr. David L. J. Tyrrell and Dr. Randy J. Read. Dr. Bruce Malcolm of the Biochemistry Department, and Dr. David Wishart of the PENCE have been especially friendly and helpful, and I am grateful for their time and interest.

My sincere gratitude extends to a number of individuals who have been helpful in teaching me many of the techniques and offering helpful discussions. They are Dean Smith, Brian Taylor, Li Li, Joud Shafiq, Jia Li, Chris Lowe, and Rajan George.

I am grateful for expert technical assistance from Rick Fawcett, who synthesized oligonucleotides, did most of the animal injections, test bleeds, and part of the peptide purification.

I would like to gratefully thank the following people: Linda Luo, who performed DNA sequencing for some of the sequences described in this thesis; Xiao-Bo Tang, who collaborated in the early stages of this work; and Anne Hudson, who read through my thesis draft. Photographic assistance from San Vinh is also gratefully appreciated.

I would like to thank my family, and many of my friends, for their continuous encouragement and spiritual support through-out the course of the project.

Last, but not the least, thank you, Canada.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION 1

1. Hepatitis C Virus - History, Epidemiology and Clinical Features 1

2. Genome Structure and Proteins of the Hepatitis C Virus 4

3. Proteinases Which Act on the Hepatitis C Polyprotein 5

4. General Approaches to Expressing Recombinant Proteinases in E. coli 8

a) Systems for Expressing Recombinant Proteins in E. coli 8

b) (His)6 Affinity Purification System 12

c) Active Viral Proteinases Which Have Been Expressed in E. coli 14

5. Assays for Proteinases Using Peptide Substrates 15

6. Research Goals and Rationale 16

CHAPTER II

MATERIALS AND METHODS 18

(I) REAGENTS 18

(II) GENERAL METHODS 20

(1) Gel electrophoresis 20

(a) SDS PAGE 20

(b) Agarose gel electrophoresis 21

(c) DNA sequencing gel electrophoresis 21

(2) Preparation of Plasmid DNA and single stranded phagmid DNA 21

(a) Mini-preparation of plasmid DNA by the alkaline lysis method 21
(b) Maxi-preparation of plasmid DNA by QIAGEN Maxi Kit 22
(c) Maxi-preparation of plasmid DNA by CsCl method 22
(d) Preparation of single stranded DNA template 24
(3) DNA sequencing reactions 25
(4) Isolation of viral RNA 25
(5) Reverse transcription-polymerase chain reaction (RT-PCR) and DNA cloning 26
(6) Construction of expression plasmids 27
(7) Expression of protein in Escherichia coli 29
(a) Preliminary assays for expression 29
(b) Large-scale protein expression 30
(8) Nickel-chelate affinity chromatography 30
(9) Dialysis and refolding of the purified proteins 31
(a) Dialysis 31
(b) Refolding of proteins 31
(c) Protein quantitation 32
(d) Enterokinase cleavage 32
(10) HCV serine proteinase activity assay 33
(11) Calculation of Molecular weight for peptides/polypeptides 34
(12) Construction of peptide substrate 34
(a) Synthesis 34
(b) Chemical modification and cleavage 35
(c) HPLC purification 35
(13) Immunization of laboratory animals 36

(a) Rats 36
(b) Rabbits 36
(c) Mice 37
(14) Enzyme-linked immunosorbent assay (ELISA) 37
CHAPTER III
RESULTS AND DISCUSSION 39
Cloning of a DNA fragment encoding the HCV serine proteinase 39
Characterization of HCV proteinase cDNA clones 41
Expression of HCV proteinase using pT7-H6/HCVP 43
HCV proteinase polypeptide expressed at high levels 43
Recombinant HCV serine proteinase present in inclusion bodies 44
Nickel-chelate affinity chromatography offers one step purification 45
Recombinant proteinase has the expected N-terminus 46
Generation of antiserum against the H6/HCVP protein 46
Construction of pT7-H6/HCVPfix and production of a second recombinant HCV proteinase 47
Production of E. coli thioredoxin fusion proteins 50
Generation of synthetic peptide substrate 51
Enterokinase cleaved H6/HCVPfix is catalytically active 53
1) Evidence from the colorimetric assay using TNBS 53
2) HPLC purification and mass spectroscopy of the major cleavage product derived from the peptide substrate 55
CHAPTER FOUR
FINAL DISCUSSION AND CONCLUSIONS 57
BIBLIOGRAPHY 107

LIST OF TABLES

Table 1. HCV-encoded proteins and proposed functions 69

Table 2. Proteinase recognition sites in HCV polyprotein known to be cleaved 70

Table 3. Other consensus sites in the HCV polyprotein which are not cleaved 71

Table 4. Rate at which primary amino groups are released from substrate peptide under two different buffer conditions 72

LIST OF FIGURES

Figure 1. HCV polyprotein cleavage products and relationship to the recombinant proteins expressed 73

Figure 2. The probability of recovery of an intact protein after cleavage of affinity tag by various proteinases 74

Figure 3. Viral cDNA sequences flanking the putative HCV serine proteinase domain 75

Figure 4. Non-denaturing agarose gel electrophoresis of PCR products 76

Figure 5. Nucleotide sequence and deduced amino acid sequence of the clone #6 cDNA insert 78

Figure 6. Comparison of the protein sequence of trypsin with that of the putative HCV serine proteinase 79

Figure 7. DNA sequence and the deduced protein sequence in the polylinker of expression vectors 80

Figure 8. Amino acid sequences of the various recombinant proteins 81

Figure 9. SDS-PAGE of recombinant HCV serine proteinase from construct pT7-H6/HCVP 83

Figure 10. Affinity purification of recombinant protein by nickel chelate chramatography 85

Figure 11. SDS-PAGE of protein purified by Ni2+-NTA affinity chromatography 86

Figure 12. ELISA assays of antisera raised in laboratory animals 88

Figure 13. SDS-PAGE of test expression from construct pT7-H6/HCVPfix 89

Figure 14. SDS-PAGE of cells after lysis by French press. 91

Figure 15. SDS-PAGE of purified recombinant protein from the construct pT7-H6/HCVPfix 93

Figure 16. SDS-PAGE of recombinant HCV serine proteinase processed by enterokinase 95

Figure 17. SDS-PAGE of total cellular proteins of cultures grown at various temperature 97

Figure 18. SDS-PAGE of proteins from cells lysed by osmotic shock in different solutions 99

Figure 19. HPLC profile of synthetic peptide substrate 101

Figure 20. Generation of primary amino groups by refolded HCV serine proteinase 102

Figure 21. Initial velocity at which primary amino groups are generated 103

Figure 22. HPLC profile of the proteolyzed substrate peptide 104

Figure 23. Plasma desorption mass spectra (PDMS) of the differential peak 106

ABBREVIATIONS AND DEFINITIONS

2XYT:	a bacterium culture medium composed of 1.6% (w/v) bacto-tryptone, 1.0% (w/v) bacto-yeast extract, and 0.5% (w/v) NaCl.
2XYT/MT:	2XYT containing 10 mM Tris-HCl (pH7.5), 2 mM $MgCl_2$.
ABI:	Applied Biosystems Inc., Foster City, CA.
ABTS:	2, 2'-azino-bis(3-ethylbenz-thiazoline-6-sulfonic acid).
BCA:	bicinchoninic acid.
bp:	base pair(s).
BRL:	Bethesda Research Laboratories, Inc., Gaithersberg, MD.
BSA:	bovine serum albumin.
cDNA: enzyme	complementary deoxyribonucleic acid; DNA which has been copied from messenger RNA using the reverse transcriptase.
CID:	chimpanzee infectious dose.
DEPC:	diethyl pyrocarbonate.
DMSO:	dimethyl sulfoxide.
DNA:	deoxyribonucleoic acid.
dNTP's:	the four common 2'-deoxynucleotide-5'-triphosphates (G,A,T,C).
DTT:	1,4-dithiothreitol.
EDTA:	ethylenediaminetetraacetic acid.
ELISA:	enzyme-linked immunosorbent assay.
FLAG:	a name for the tetrapeptide DYKD used as a affinity tag.
GST:	glutathione-S-transferase.

HBTU:	2-(1H-Benzotriazol-1-yl)-1,1,3,3,-tetramethyl uronium hexafluorophosphate
HEPES:	N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid.
HCC:	hepatocellular carcinoma.
HCV:	hepatitis C virus.
HPLC:	high-performance liquid chromatography.
IPTG	isopropyl-β-thiogalactopyranoside
MBP:	maltose-binding protein.
NANB:	non-A, non-B hepatitis; a former name for hepatitis C.
Ni^{2+}-NTA:	nickel-nitrilotriacetic acid adsorbent column
NS:	non-structural.
OD:	optical density.
PBS:	phosphate buffered saline, pH7.2.
PBST:	PBS containing 0.5% (v/v) Tween-20.
PCR:	polymerase chain reaction.
pfu:	plaque forming unit.
PMSF:	phenylmethylsulfonyl fluoride.
PVDF:	polyvinylidene difluoride
RNA:	ribonucleic acid.
RNase:	ribonuclease.
RT:	the enzyme reverse transcriptase; reverse transcription.
SDS:	sodium dodecyl sulphate.
TE:	Tris-EDTA buffer (10 mM Tris-HCl, 1 mM EDTA).
TNBS:	2,4,6-trinitrobenzenesulfonic acid
TNP-peptide:	2,4,6-trinitrophenyl-peptide.

Tris-HCl: tris (hydroxymethyl) aminomethane hydrochloride.

CHAPTER I

INTRODUCTION

1. Hepatitis C Virus - History, Epidemiology and Clinical Features

Hepatitis C virus (HCV) was discovered by Choo et al. (19) using a purely molecular cloning approach. These investigators constructed a prokaryotic cDNA expression library from RNA extracted from the plasma of a chimpanzee infected with non-A, non-B (NANB) hepatitis. By screening their library with immune serum from patients suffering from NANB hepatitis, a single positive cDNA clone was initially isolated. The cloned nucleic acid fragment was not derived from human DNA (i.e. it gave a negative result when used to probe Southern blots of human DNA), and it hybridized specifically to RNA extracted from the plasma or liver of patients suffering from NANB hepatitis. Thus the original clone appeared to be expressing a virus-specific antigen and to contain a virus-specific DNA sequence. The DNA sequence of the cloned fragment was similar to a DNA sequence found in flaviviruses and pestiviruses, and it was concluded that a fragment from the genome of the causative agent for NANB hepatitis had been cloned. This agent was renamed hepatitis C virus (HCV), and subsequently the entire genome of this virus was cloned and sequenced.

It is now clear that HCV is the predominant, but probably not the sole etiological agent of a 'catch all' group of diseases which were formerly described as transfusion associated non-A, non-B (NANB) hepatitis. For the purposes of this introduction, NANB hepatitis and HCV hepatitis will be used interchangeably, depending on whether the reference was written before or after the introduction of the term HCV, and the more refined definition of the

disease. In practical terms, older references which refer to transfusion associated NANB hepatitis are for the most part describing HCV hepatitis. HCV hepatitis is a parenterally transmitted or in some cases sporadically acquired disease which is found throughout the world. In 1985 in the United States, NANB hepatitis (i.e. hepatitis C) accounted for 20 to 40% of cases of acute viral hepatitis (4). Today HCV appears to be the single most common cause of acute hepatitis in most western countries. The disease is a well known complication in the treatment of hemophiliacs with Factor VIII concentrates and other blood products (23). Other risk factors for infection by HCV include intravenous drug use, hemodialysis, or receipt of an organ transplant (27). HCV now accounts for at least 85% of the cases of transfusion-associated hepatitis worldwide (1).

In addition to the parenteral route, hepatitis C virus appears to be transmitted by other routes as well. Data from the CDC's Sentinel Countries Study of Acute Hepatitis showed that in 1987 in the United States approximately 62% of patients with NANB hepatitis (i.e. HCV) had a recognized risk factor for the disease (3,5). Predominant risk factors included: 1) a history of intravenous drug abuse (42% of cases), 2) a history of heterosexual or household contact with a person who had hepatitis, or a history of heterosexual activity with multiple partners (10%), 3) a history of recent blood transfusion (5%) and 4) a history of medical or dental employment (5%). This study showed that approximately 40% of patients presenting with acute NANB hepatitis had no evident source of infection, and were considered to be sporadic cases. This suggests that HCV hepatitis can be transmitted by other, as yet unknown mechanisms.

Infection by HCV frequently (if not universally) leads to persistent disease, and results in a range of clinical conditions, from an inapparent or asymptomatic carrier state, which may persist for life, to severe acute hepatitis, chronic active hepatitis, cirrhosis, or even hepatocellular carcinoma (HCC; (45)). After transmission by parenteral or non-parenteral routes, the hepatitis C virus appears to be specifically located within the liver, since if tissue damage does occur, it is limited to this organ (13). In parenterally transmitted NANB hepatitis (i.e. Hepatitis C), the mean incubation period from transfusion to the first elevated liver enzymes (e.g. ALT) was found to range from 2 to 26 weeks, with a peak of onset between 6 and 12 weeks (24). Interestingly, infections often remain asymptomatic. When acute illness does develop, it generally consists of a range of symptoms including fever, chills, headache, weakness and jaundice. Compared with hepatitis B, although the acute course of Hepatitis C is more likely to be clinically mild (peak ALT levels of 200 to 600 UI/liter) and less commonly icteric (25% of cases) (2,75), individual cases may be severe and can be clinically indistinguishable from severe cases of hepatitis A or B.

The hallmark of HCV infection is persistent, chronic hepatitis. This in turn can lead to cirrhosis and hepatocellular carcinoma (HCC). Approximately 50% of patients with HCV hepatitis eventually develop biochemical and histological evidence of chronic hepatitis. After only 5 to 10 years of follow up, 25% of patients with chronic NANB hepatitis had developed cirrhosis which appeared to be more indolent than alcohol-related cirrhosis (27). Seventy percent of Japanese patients with HCC have been found to be HCV seropositive (68). More recent studies have revealed that the high prevalence (70%) of anti-HCV antibodies in chronic, non-alcoholic, and persistently anti-

hepatitis B negative patients with hepatocellular carcinoma indicates a strong association between HCV infection and tumor development (17,21,29,49,81,93). The time interval from transfusion to the clinical presentation of chronic hepatitis, cirrhosis and hepatocellular carcinoma may be extremely long. Historically, for NANB hepatitis the mean interval to recognition of chronic hepatitis was 14 years; to the development of cirrhosis, 18 years; and to the development of hepatocellular carcinoma, 23 years (49).

Until recently, the only accepted animal model for hepatitis C virus infection was the chimpanzee. Unfortunately, to date the virus has not been satisfactorily propagated in any *in vitro* cell culture system, despite a few reports of limited successes (38,86). However, recent unpublished reports suggest that propagation of the virus *in vitro* in eukaryotic cell lines may be feasible in the near future (Personal communication, Jonathan Coates, Glaxo Inc.).

The only available treatment for HCV infection is the administration of Interferon α (IFN-α), which can cure about 30% of patients without signs of recurrence. Unfortunately this therapy is very expensive, and it is not possible to predict which patients will respond.

2. Genome Structure and Proteins of the Hepatitis C Virus

The hepatitis C virus genome consists of a 9.4 kilobase (kb) positive-sense, single-stranded, linear RNA. The viral RNA is complexed with nucleoproteins and packaged in a lipid envelope. Recently hepatitis C virus has been assigned to a new, independent genus in the *Flaviviridae* family. The life cycle of HCV is not known thus far, partly because no acceptable *in vitro* culture system has been established for the virus.

cDNA sequence analysis has indicated that the viral genome contains a 5' untranslated region of 341 nucleotides, followed by a single long open reading frame (ORF) encoding a precursor polyprotein of 3010 or 3011 amino acids, and then a 3' untranslated region of variable length (20,47,92). The 5' end of the viral RNA may contain a 5Me-dG cap structure, and a poly (U) (16,47,69,92,94) or a poly (rA) (20,36) tract occurs at the 3' end. The genomic organization of HCV resembles that of the flaviviruses and pestiviruses, with structural proteins located in the N-terminal region and a variety of nonstructural (NS) proteins in the C-terminal region of the polyprotein. The deduced amino acid sequence of some of the HCV-NS proteins suggests that they have significant sequence similarities to those of other pestiviruses and flaviviruses. The gene order obtained by analogy with those related viruses, and by *in vitro* transcript-ion/translation studies or transient expression assays is 5'-C-E1-E2-NS2-NS3-NS4A-NS4B-NS5A-NS5B-3' (33,41,96,97) (Figure 1 A).

The functions of all the various HCV proteins are not yet known with certainty (Table 1). C is a basic protein that binds RNA (84) and is postulated to be the nucleo-capsid protein. E1 and E2 are envelope glycoproteins. NS2 through NS5B are the viral nonstructural proteins that are proposed to be responsible for polyprotein processing and viral genome replication. For the purpose of the present thesis, cleavage of the polyprotein is of particular significance, since this process generates the mature functional proteins that replicate the viral genome, and mediate invasion of hepatocytes by the mature virion.

3. Proteinases Which Act on the Hepatitis C Polyprotein

Both host and viral proteinases are required for processing of the HCV

polyprotein, and these result in at least nine distinct cleavage products (Figure 1 A and Table 1). The signal peptidase present in the endoplasmic reticulum is responsible for generating the N termini of E1, E2 and possibly NS2 (41). A novel virus-encoded *cis*-acting zinc-dependent metalloproteinase is responsible for the cleavage between NS2 and NS3 (31,42). The viral serine proteinase derived from the NS3 N-terminal region (Figure 1 B) is responsible for cleavage at the NS3-NS4A, NS4A-NS4B, NS4B-NS5A and NS5A-NS5B junctions (7,8,25,32,96).

Analysis of the amino acid sequence of the NS3 polypeptide of HCV suggests that this viral protein contains a trypsin-like serine proteinase domain in the N-terminal region as well as a helicase domain in the C-terminal region (Figure 1 B) (9). It is not known exactly where the proteinase domain ends and the helicase domain begins, nor has the isolated proteinase domain been shown to be active independent of the helicase. The serine proteinase domain functions in the processing of viral polyprotein as described above.

The active site of any serine proteinase invariably contains three residues: histidine, aspartate, and serine. These residues maintain the same relative spatial position in all known structures of these enzymes, and they constitute the catalytic triad. The histidine imidazole group abstracts the alcohol proton of the serine and transfers it to the amine-leaving group, while the aspartate stabilizes the positive charge developed on the histidine in the intermediate complex (72). The histidine 1083, aspartate 1107 and serine 1165 residues are highly conserved among all HCV strains sequenced so far (Figure 6) and have been proposed to constitute the catalytic triad of the HCV proteinase. Using a transient eucaryotic expression assay (the BHK-21 cell line or the A16 subclone of the human hepatoma HepG2 cell line co-infected with vaccinia virus-HCV

recombinants plus vTF7-3 --- a vaccinia virus recombinant expressing the T7 DNA-dependent RNA polymerase), Gracoui *et al.* (32) recovered NS4A, NS4B, NS5A, and NS5B proteins by immunoprecipitation with HCV region specific polyclonal rabbit antiserum and human patient serum. These radio-labeled HCV-specific proteins were separated by SDS gel electrophoresis, transferred to Immobilon polyvinylidene difluoride membranes, localized by autoradiography, and sequenced by partial N-terminal amino acid sequence analysis. From this analysis the NS3 proteinase-dependent cleavage sites were localized. Sequence comparison of the residues flanking these cleavage sites for all sequenced HCV strains suggested that the HCV proteinase encoded in NS3 region may have substrate specificity for an acidic residue (Asp or Glu) at the P6 position, a Cys or Thr residue at the P1 position, and a Ser or Ala residue at the P1' position (Table 2). Nev rtheless, five additional sites conforming to this same motif have also been identified from the polyprotein of several HCV strains (Table 3). So far, no product consistent with cleavage at these sites has been observed, suggesting that additional factors, such as substrate conformation, may be involved in determining the cleavage preferences of the HCV serine proteinase. It may be that the sites which are not cleaved are not accessible (i.e. they are in the interior of the folded proteins).

Failla and colleagues have suggested that, in addition to NS3, a 33-amino-acid fragment from the C-terminus of the NS4A protein is required for cleavage at the NS3-NS4A and NS4B-NS5A sites. Furthermore, this fragment apparently accelerates the rate of cleavage at the NS5A-NS5B junction. Thus, NS4A supplied *in trans* apparently activates the NS3 serine proteinase (26). Most recently Lin *et al.* reported that NS4A appears to be absolutely required for

trans-cleavage at the NS4B-NS5A sites, but it is not an essential cofactor for HCV serine proteinase activity (57).

4. General Approaches to Expressing Recombinant Proteinases in *E. coli*

a) Systems for Expressing Recombinant Proteins in *E. coli*

Since 1976, biotechnology has become a reality as the methodologies for DNA cloning, oligonucleotide synthesis, and gene expression have been perfected so that virtually any protein can now be expressed from recombinant DNA. The popular expression systems include the bacteria *E. coli* and *Bacillus subtilis*, yeast, and cultured insect and mammalian cells. Bacterial cells offer simplicity, short generation times, and large yields of product with low costs. These characteristics, coupled with a long history of use in genetics and 18 years of application in recombinant gene expression, have established *E. coli* as the leading host for expression of genetically engineered proteins.

A number of *E. coli* expression systems have been developed since 1976. The most popular of these are all based on inducible expression, and use a variety of promoters including: 1) the T7 RNA polymerase/T7 promoter system, 2) phage λ regulatory sequences, 3) the *lac Z* or *trp E* promoters, or a hybrid of these, (i.e. the *tac* promoter, which is also regulated by the *lac* repressor) (82,91). Several vector systems, such as the pUR series of plasmids (79) and pMR100 (35), use the *lac* promoter to drive expression of *lac Z* fusion proteins. Both pUR and pMR100 vectors are thus IPTG inducible.

The pSKF and pEX series of vectors utilize bacteriophage λ regulatory signals, for example the powerful p_L promoter in pSKF vectors and the p_R promoter in pEX vectors. In these systems, gene expression is induced by

rapidly increasing the temperature to 42°C. The major advantage of using phage regulatory signals is that they are more tightly regulated (but see below) and the induced expression is highly efficient. The major disadvantage of this type of vector is that the increase in temperature induces a variety of endogenous heat shock proteins including several with strong proteolytic activities, in addition to inducing expression of the recombinant protein. The phage λ, *lac Z*, *trp E* and *tac* promoters all rely on *E. coli* RNA polymerase for their transcription. Unfortunately, these promoters are always "leaky". This means that a small amount of fusion protein is constantly being made, which if it is toxic will kill all expressing cells. For example, cells that are making recombinant protein due to basal expression will be lost during an overnight growth, and the culture (or colony) will be overtaken by cells containing either no plasmid or mutations that prevent the recombinant protein from being expressed. This can happen even at the DNA subcloning stage, during the plating of transformations, and mapping of the plasmid minipreps, and well before any attempts to actually express the protein of interest.

To overcome the problems of promoter "leakiness" during the initial DNA cloning steps, the T7 RNA polymerase system was developed (90). It puts expression of the recombinant gene under the control of a T7 promoter, which is only active when bacteriophage T7 RNA polymerase is present. This approach has a number of advantages: 1) T7 RNA polymerase is a very active enzyme-- it synthesizes RNA at a rate several times that of *E. coli* RNA polymerase, and in a very processive way (For example, transcription can 'circumnavigate' a plasmid more than once, resulting in RNA transcripts several times the unit length of the plasmid, although this is not an advantage in terms of protein expression). 2) The T7 promoter is not recognized by *E. coli*

RNA polymerase, therefore no basal expression is possible so long as no T7 RNA polymerase is present (i.e. there is no promoter 'leakiness' in the ordinary *E. coli* host). 3) T7 RNA polymerase is highly selective for initiation at its own promoter sequence, and does not initiate at any site on the *E. coli* DNA. 4) T7 RNA polymerase is resistant to antibiotics such as rifampicin which inhibit *E. coli* RNA polymerase, so that the addition of rifampicin to cells that are producing T7 RNA polymerase results in exclusive expression of only those genes which are under control of the T7 promoter.

Expression of recombinant *lac Z* or *trp E* fusion proteins in *E. coli* has been widely used for antigen production. For example, the pATH vectors (51) express *trp E* fusion proteins under the *tac* promoter. An advantage of producing fusion proteins is that the β-gal or trp E moiety can be used as a means of identifying the recombinant protein (e.g. on Western blot). Antibodies directed against the β-gal or trp E protein domain can also be used for affinity purification of the fusion protein. However, a potential problem in using fusion proteins for biochemical analysis is that the properties of the native protein may be significantly altered due to the inclusion of the foreign residues encoded by the *E. coli* protein.

A number of more modern fusion protein expression systems have also been developed to allow affinity purification of recombinant proteins. Four examples are included here for illustration. 1) The maltose-binding protein (MBP) vectors, such as the pMAL vectors (New England Biolabs, Beverly, MA), which allow for the expression of a cloned gene fused to the C-terminus of MBP encoded by the *mal E* gene of *E coli*. This method uses a strong, inducible *tac* promoter and the *mal E* translation initiation signal to give high level expression. Affinity purification of non-denatured protein is accomplished by

passing the fusion protein over an amylose resin (binds the MBP) and eluting with maltose (48). 2) The glutathione-S-transferase (GST) vectors (e.g. pGEX vectors (88)), use a *tac* promoter to express recombinant proteins as fusions with the C-terminus of glutathione-S-transferase (GST) from *Schistosoma japonicum*. Fusion with GST may help to make the recombinant protein more soluble, and allows it to be purified under nondenaturing conditions by absorption onto glutathione-agarose beads. Recombinant protein is eluted with glutathione, which can then be dialyzed away. 3) The flag vectors, such as pFLAG (Kodak, IBI, New Haven, CT), allow the expression, detection, and purification of a protein encoded by a cloned gene by fusing it to the C-terminus of the FLAG octapeptide DYKDDDDK. This system uses an inducible *tac* promoter and an *omp A* transcription initiation signal to drive gene expression. Ideally the *omp A* signal sequence (at the extreme N-terminus) produces a periplasmic fusion protein which can be purified by affinity chromatography over a column which has the murine anti-FLAG monoclonal antibody covalently attached to agarose. Binding of FLAG to the monoclonal antibody via the DYKD residues is critically dependent on the presence of Ca^{2+}, and recombinant protein can be eluted with a Ca^{2+} chelator such as EDTA. 4) The polyhistidine fusion system, such as the pET-14b to pET-23a vectors (Novagen, Inc. Madeson, WI) and the pBlueBacHis vector (Invitrogen Corporation, San Diego, CA). These systems express the recombinant protein as a fusion with six consecutive histidines, either at the C- or N-terminus. The poly-His 'tag' has a high affinity for immobilized metal ions, typically Ni^{2+} or Cu^{2+} ions, on a solid phase chelating resin. Host cell proteins do not bind tightly to the nickel resin and can be easily washed away, and the purified recombinant protein can then be eluted either with a low pH buffer under denaturing conditions or with an imidozole gradient under native conditions

(44). More recently, *E. coli* thioredoxin fusion proteins (55) and *E. coli* Gro EL (heat-shock protein 60) (54) fusion proteins have been made. These fusions may provide some advantage in efficient re-folding of the overexpressed recombinant protein domains.

When a recombinant protein is to be used in a biochemical or biological experiment, in addition to its purity the authenticity of the protein's function (e.g. enzymatic activity, biological activity) is very important. To achieve this, proteins expressed as fusions may require cleavage away from the non-native affinity 'tag', using a highly specific proteinase. The proteinases commonly used for this purpose include Factor Xa, thrombin, and enterokinase. Each of them recognizes and cleaves one particular peptide sequence: IEGR↓ for Factor Xa, LVPR↓GS for thrombin, and DDDDK↓ for enterokinase. Factor Xa and enterokinase cleave at the C-terminus of the recognition sequence, so that the authentic N-terminus of the protein is released, providing that the protease cleavage site is fused at the proper point in the recombinant construct. The cleavage of thrombin is within the recognition sequence, and it leaves two extra residues attached to the target protein when fused at the N-terminus. Efficient recovery of intact, authentic proteins is obtained using cleavage with enterokinase (FLAG™ Biosystem catalogue, IBI, New Haven, CT, and Figure 2).

b) $(His)_6$ Affinity Purification System

In 1975 Porath and co-workers introduced immobilized metal ion affinity chromatography for the purification of peptides and proteins (73). The principle of this technique is based on the coordination between the electron donor groups on a peptide surface and immobilized transition metal ions such as Ni^{2+}, Cu^{2+}, or Zn^{2+}. The predominant electron donor groups are histidine

residues which are present on a protein surface (39). The apparent dissociation constant pK_2 (imidazole) of histidine is 5.97. At neutral pH (pH7) more than 91% of the imidozole residues are unprotonated (44). Therefore the histidine residues in a protein are most likely to form stable coordination bonds with metal chelates at neutral or slightly alkaline pH. In 1990, Hochuli used genetic engineering to add a $(His)_6$ affinity tag to the C-terminus of mouse dihydrofolate reductase (DHFR). The fusion protein was expressed in *E. coli* and successfully purified from inclusion bodies using a Ni^{2+}-nitrilotriacetic acid adsorbent (Ni^{2+}-NTA adsorbent) column under denaturing conditions (6 M guanidine hydrochloride). In the same series of experiments, a $(His)_2$ tag/DHFR fusion protein was purified by passage over the same column under native conditions (44). This method has now become a widely accepted and routine technique for purifying recombinant proteins. After loading the cell extract onto a Ni^{2+}-NTA column, proteins binding nonspecifically to the column are washed away with extraction buffer and wash buffer. Elution can then be carried out at a constant pH by ligand exchange (using a gradient of 0 to 200 mM imidazole), or by changing the pH in a linear or stepwise fashion. In the case of pH gradients, the imidazole residues of the histidines become protonated as the pH decreases, and this reverses their capacity to bind to the immobilized metal ion.

The $(His)_6$ tag/Ni^{2+}-NTA affinity column approach offers a simple, quick, and effective method for purifying recombinant proteins. This system is especially well suited for the purification of those proteins which aggregate in the form of inclusion bodies when overexpressed in bacteria. Such proteins require denaturing conditions to dissolve, but with Ni^{2+}-NTA, purification is still possible in the presence of the denaturant (e.g. 8 M urea or 6 M guanidine

hydrochloride). Another advantage of the $(His)_6$ tag over maltose-binding protein (MBP) and glutathione-S-transferase (GST) protein fusions is that the affinity tag is relatively small in size. If each of the fusion proteins accumulates to the same level, say 5% of the total cellular protein, then the total quantity of authentic protein (i.e. after removal of affinity 'tag') obtained from the $(His)_6$ system will be significantly more than that from the MBP or GST systems. Although the FLAG fusion protein system also has this advantage, it requires an expensive and relatively delicate antibody column constructed from a murine anti-FLAG monoclonal antibody covalently attached to agarose. Unfortunately, this antibody column also binds some unrelated *E. coli* proteins (author's unpublished data). The Ni^{2+}-NTA adsorbent columns are, on the other hand, relatively inexpensive and more robust, and they can be regenerated and reused multiple times (44).

c) Active Viral Proteinases Which Have Been Expressed in *E. coli*

Many viruses encode proteinases for the post-translational processing of virally encoded proteins or polyproteins. The expression and purification of active recombinant viral proteinases allows for biochemical characterization of those enzymes and can lead to a better understanding of viral replication. Biochemically pure proteinases can be used to screen for specific inhibitors of those enzymes which might be effective as antiviral compounds. Several virally encoded proteinases have been successfully expressed in an active form in *Escherichia coli.* Active 3C encoded hepatitis A virus (HAV) proteinase was over-expressed in *E coli* and biochemically characterized, and the active site cysteine residue identified through site-directed mutagenesis (60). Poliovirus 3C proteinase has also been expressed in *E. coli* , purified, and its specific cleavage activity on natural and synthetic peptide substrates characterized

(67,71). Human rhinovirus 3C proteinase was similarly cloned and expressed in an active form (50,56).

Active recombinant human immunodeficiency virus 1 (HIV-1) proteinase (an aspartylproteinase) from *E. coli* has contributed to our understanding of the enzyme and its role in the viral life-cycle (11,12,58,63,64,76). Recently, efficient processing of the NS2-NS3 junction of the HCV polyprotein has been observed using the *cis*-acting NS2-NS3 encoded Zn^{2+} dependent HCV metalloproteinase expressed in *E coli* (31). By the time this thesis was completed, self-cleavage of HCV NS3 in *E. coli* using a construct which contained adjacent parts of NS2 and NS4 had also been reported (100).

5. Assays for Proteinases Using Peptide Substrates

The use of synthetic peptides as substrates to assay for proteinase activity is a widely accepted method. This approach has been used to assay viral proteinase activity, including recombinant proteinases expressed in *E. coli.* For example, it has been applied to studies of the HAV 3C proteinase (60), the poliovirus 3C proteinase (67,71), the human rhinovirus 3C proteinase (50,56), and the HIV proteinase (11,12,64,76).

Several methods for quantitative determination of peptide cleavage by proteinases have been described, including: 1) colorimetric assays using 2,4,6-trinitrobenzenesulfonic acid (TNBS) (11,60,70,85), 2)HPLC analysis (12,64), 3) thin-layer electrophoresis (TLE) (52), 4) continuous spectrophotometric monitoring (66,78,95), 5) radiometric assays (10,99), and 6) fluorometric methods (28,43,62). The colorimetric assay uses 2,4,6-trinitrobenzenesulfonic acid (TNBS) to detect the primary amino groups generated by cleavage of terminally blocked peptide substrates. The reaction of an exposed primary

amino group with TNBS yields a yellow-colored product which has a peak absorbance near 340 nm or 420 nm (70,85). This method provides a high capacity assay for enzymes and can be performed on microtiter plates in a fast, simple and discontinuous fashion.

6. Research Goals and Rationale

Clearly, more work needs to be done in order to characterize the HCV trypsin-like serine proteinase encoded in the NS3 region. Thus far, analysis of this proteinase has been carried out mainly in cell-free transcription/translation systems, and in transient expression systems in eucaryotic cells (e.g. using recombinant vaccinia virus). In addition to biochemical and enzymatic characterization, it would be useful to overexpress catalytically active HCV serine proteinase in *E. coli* in large quantities for a number of other reasons. (1) It would allow the generation of immune serum to enable the establishment of convenient, sensitive, and reproducible assays such as immunostaining, immunoprecipitation, and ELISA to detect the protein *in vitro* and in tissue samples, (2) Should sufficiently high expression levels be obtained, crystallization and X-ray diffraction analysis could be attempted. A structure solution would shed light on the properties of the catalytic site and the structural basis for the substrate specificity of the enzyme. This structural information could also be used to help design optimal peptide-based or other inhibitors. (3) It would facilitate the screening of non-peptide-based and peptide-based inhibitors. Once initial non-peptide or peptide-based inhibitors have been identified, medicinal chemistry, peptide modification strategies, etc. can be applied to enhance the activity and specificity of the inhibitors as well as optimizing their pharmacological properties. Such compounds could potentially be used in antiproteinase therapy for hepatitis C virus infections.

This project was designed at the outset to focus on the cDNA cloning and overexpression in *E. coli* of catalytically active HCV serine proteinase and on the generation of polyclonal, mono specific immune serum against the protein.

CHAPTER II

MATERIALS AND METHODS

(I) REAGENTS

Restriction enzymes and DNA modifying enzymes were purchased from New England Biolabs (Mississauga, ON), Boehringer Mannheim (Laval, PQ), GIBCO BRL (Burlington, ON) or Pharmacia (Baie d'urfé, PQ) and were used as recommended by the manufacturers; AMV super reverse transcriptase (SRT) was from Molecular Genetic Resources, Inc. (Tampa, FL); RNasin® ribonuclease inhibitor was from Promega/Fisher (Edmonton, AB); Sequenase™ version I and version II were from USB/Amersham (Oakville, ON); *Thermus aquaticus* DNA polymerase was produced in house (Engelke *et al.* 1990); *Pyrococcus furiosus* DNA polymerase was from Stratagene (PDI, Mississauga, ON); Deoxynucleotides were from Pharmacia (Baie d'urfé, PQ). Oligonucleotides used in the amplification of the target DNA / site directed mutagenesis by polymerase chain reaction (PCR) were synthesized in our laboratory on a 391 DNA synthesizer (Applied Biosystems, Mississauga ON) using chemicals from Applied Biosystems or Glen Research (BIO/CAN Scientific, Mississauga ON). Peptide synthesis regent (Fmoc-L-amino acids) were from BACHEM Bioscience (Philadelphia, PA).

Radioactive chemicals were purchased from Du Pont (Markham, ON) or Amersham (Oakville, ON).

ProBond™ resin was purchased from Invitrogen (San Diego, CA); Immobilon™ PVDF transfer memberane was from Millipore (Bedford, MA); Spectra/Por® 1 Molecularporous Dialysis Membrane (MWCO: 6,000-8,000) was

from Spectrum® (Houston, TX); Plasmid Maxi Kit was from QIAGEN (Chatsworth, CA); BCA Protein Assay Reagent was from PIERCE (Rockford, IL); Complete and incomplete Freund's adjuvant was from GIBCO® Laboratories Life Technologies, Inc. (Burlington, ON).

Buffer and other reagent chemicals were purchased from Sigma (ST. Louis, MO), BDH (Toronto, ON), Bio-Rad (Mississauga, ON), GIBCO BRL (Burlington, ON); Applied Biosystems (Mississauga, ON), Millipore (Toronto, ON) or ICN (Mississauga, ON) and were the best quality available.

pBluescript® plasmids are described by Short *et al.* (87). The expression plasmid pT7-7 was obtained from S. Tabor (Harvard University) (90). The DNA sequence of the pT7-7 polylinker cloning sites was determined to be 5'-ATATACATATGGCTAGAATTCGCGCCCGGGGATCCTCTAGAGTCGACCT GCAGCCCAAGCTTATCGATGATAAGCTGTCAAAC-3', with translation starting signal at 5' most ATG. Using standard DNA manipulation methods (82), we constructed the expression plasmid pT7-7His6 by inserting the sequence 5'-CATATGCACCACCACCACCACCACCTGGTTCCGCGTGGTTCC GGAATTC-3' in between the *Nde* I and *Eco* RI sites of the polylinker.

Bacterial *E. coli* strain DH5α is *endA1 hsdR17* (r_K^- m_K^+) *supE44 thi-1 recA1 gyrA* (Nal^r) *relA1 Δ(lacZYA-argF) U169 deoR* (ϕ80*dlacΔ(lacZ)M15);* BL21(DE3) is F^- *ompT* $r_B^- m_B^-$; As a B strain, it is also deficient in the *lon* proteinase.

Experimental animals were provided by University of Alberta Health Sciences Laboratory Animal Services.

(II) GENERAL METHODS

(1) Gel electrophoresis

(a) SDS PAGE

SDS PAGE is a well established method (53). Here described is the specific protocol used throughout this project.

For a final volume of 20 ml of 15% separating gel solution: 10 ml 30% acrylamide and 0.8% N,N'-methylenebisacrylamide (8 ml for 12% separating gel solution and 2 ml H_2O added to adjust the volume) was mixed with 10 ml 0.75 M Tris-HCl (pH8.8) 0.2% SDS, 0.1 ml 10% ammonium persulfate and 13 µl TEMED to initiate polymerization.

For the stacking gel: 4.25 ml 0.134 M Tris-HCl (pH6.8) mixed with 0.75 ml 30% acrylamide and 0.8% N,N'-methylenebisacrylamide stock solution, 0.1 ml 10% ammonium persulfate and 6 µl TEMED to initiate the polymerization

Ten times (10X) concentrated SDS polyacrylamide gel electrophoresis buffer was made by dissolving 120 g Trisma base, 570 g glycine and 40 g SDS in H_2O to a final volume of 4 L. This was diluted ten times to make the working solution.

The protein samples were prepared in 50 mM Tris-HCl (pH6.8), 100 mM dithiothreitol (DTT), 2% SDS (electrophoresis grade), 10% glycerol, and 0.1% bromophenol blue and boiled for 5 minutes or more, then spun for 15 to 20 minutes in a micro centrifuge at 14,000 Xg to bring down any undissolved matter before being applied to the gel. Visualization of the proteins in the gel was achieved by staining the gel in 0.25% Coomassie Brilliant Blue R250, 50%

methanol and 10% acetic acid, and then destaining the gel to reduce the background with 10% acetic acid and 10% methanol.

(b) Agarose gel electrophoresis

Non-denaturing agarose gels buffered by Tris-borate-EDTA (TBE) were prepared and used in electrophoresis according to standard methods (82).

(c) DNA sequencing gel electrophoresis

Denaturing polyacrylamide gels containing 8 M urea buffered by Tris-borate-EDTA (TBE) were prepared and used as follows: for 100 ml of 8% gel, 50 g urea dissolved in 30 ml of distilled water was well mixed with 20 ml of 38% acrylamide/2% N,N'-methylenebisacrylamide stock solution. The solution was adjusted to 99 ml with distilled water and filtered through a 0.22 μm filter. Ammonium persulfate (1 ml 10% w/v solution) was added, and then 40 μl TEMED to initiate polymerization. The gel was cast using 0.4 mm spacers and a shark's tooth comb, and allowed to polymerize for at least 2 hours before mounting on a home-made or GIBCO/BRL (Burlington, ON) S2 sequencing gel electrophoresis apparatus. After sample loading, the gel (dimensions 30 X 40 cm) was run for 2-4 hours at 60 watts constant power using a ECPS 3000/150 power supply (Pharmacia). The gel was soaked in 10% acetic acid/10% methanol solution for 15 minutes to remove urea, dried at 80°C under vacuum on a slab gel drier (Hoefer Scientific Instrument, San Francisco, CA) and exposed 17 hours to Kodak X-ray film (XAR-5) at room temperature before developing.

(2) Preparation of Plasmid DNA and single stranded phagmid DNA

(a) Mini-preparation of plasmid DNA by the alkaline lysis method

Glass culture tubes containing two milliliters of 2XYT medium (82), 10 mM Tris-HCl (pH7.5) and supplemented with the appropriate antibiotic (e.g. 100μg/ml ampicillin) were inoculated with a single isolated bacterial colony and grown at 37°C overnight on a roller wheel at top speed. About 1.5 ml of the overnight culture was transferred into a Eppendorf tube and centrifuged for one minute to collect the bacterial cells. The pellet of the bacterial cells was drained well and resuspended completely in 100 μl of 50 mM glucose, 10 mM EDTA, 25 mM Tris-HCl (pH8.0). It was gently mixed with 200 μl of freshly prepared 0.2 M NaOH and 1% SDS and left on ice for 5 minutes. The solution was neutralized by adding 150 μl of 3 M potassium acetate (pH4.8), left on ice for 5 min., and pelleted 15 min. at 4°C in a microfuge. The supernatant was transferred to a fresh tube and the plasmid DNA precipitated with an equal volume of isopropanol. The pellet was resuspended in 400 μl of 50 mM Tris-HCl (pH8.0), 150 mM NaCl and treated with RNase A (10μg/ml) at 37°C for one hour. The plasmid solution was extracted with an equal volume of phenol, followed by an equal volume of phenol/chloroform (1:1/v:v), and then chloroform alone before precipitation using one-tenth volume of 3.0 M sodium acetate (pH5.2), 2.5 volumes of 95% ethanol and chilling to -20°C. The plasmid DNA was collected by centrifugation (15 min., at 4°C in micro centrifuge), rinsed with cold 70% ethanol (stored at -20°C) and dried under vacuum before being redissolved in Tris EDTA (TE).

(b) Maxi-preparation of plasmid DNA by QIAGEN Maxi Kit

Maxi-plasmid preparation using QIAGEN Maxi Kits was performed according to the manufacturer's protocol.

(c) Maxi-preparation of plasmid DNA by CsCl method

For each plasmid, 500 ml of 2XYT medium, containing 10 mM Tris-HCl (pH7.5) and supplemented with the appropriate antibiotic was inoculated with a 2 ml overnight bacterial culture grown from a single isolated colony carrying the plasmid of interest. After overnight incubation at 37°C with vigorous shaking in an air shaker, the bacterial cells were pelleted by centrifugation in a 1 L centrifuging bottle (3,500 Xg for 30 min.). The cell pellet was drained well, and resuspended in 7.5 ml of 50 mM glucose, 10 mM EDTA, 25 mM Tris-HCl (pH8.0). This was transferred to a fresh 250 ml centrifuging bottle, followed by a 5 ml rinse of the 1L bottle with the same buffer. The rinse was added to the same 250 ml bottle, and resuspended cells were gently mixed with 25 ml of freshly prepared 0.2 M NaOH, 1% SDS and left on ice for 10 minutes. The mixture was neutralized by adding 19 ml of 3.0 M potassium acetate (pH4.8), shaken vigorously, and then left on ice for 10 minutes. The cell debris was spun down (10,000 Xg, 4°C, 30 min.), and the supernatant poured into a new bottle through 4 layers of cheesecloth. The crude plasmid DNA was precipitated by adding 35 ml isopropanol, mixing at room temperature, and centrifuging at 15,000 Xg, 4°C for 30 min. The drained DNA pellet was rinsed with 70% ethanol, drained well, and resuspended in 8 ml TE (pH8.0). 8.62 gram CsCl and 0.8 ml of 10 mg/ml ethidium bromide were added and the contents were mixed till the salt dissolved. The sample was centrifuged at 4,600 Xg for 10 minutes at room temperature and the supernatant was transferred into a

labeled quick-seal tube with a pasture pipette. The tube was filled, balanced, heat sealed, and spun overnight (> 16 hours) at 60,000 r.p.m. 20°C in an 80 Ti fixed angle rotor (Beckman). The plasmid DNA band was harvested with a syringe and a 16 gauge needle and extracted 6 times with water saturated n-butanol to remove ethidium bromide. The DNA solution was diluted with 2.5 volumes of TE, brought to 0.3 M sodium acetate, and precipitated with an equal volume of isopropanol overnight at -20°C followed by centrifugation (5,500 Xg, 4°C, 30 min.). The pellet was rinsed with 70% ethanol, redissolved in 400 µl TE and extracted twice with phenol, twice with phenol/chloroform (1:1 /v:v) and twice with chloroform containing 4% iso-amyl alcohol. The plasmid DNA was adjusted to 0.3 M sodium acetate, precipitated with 2.5 volumes 95% ethanol, rinsed with 70% ethanol and vacuum dried in a speedvac for 3 minutes before dissolving in TE.

(d) Preparation of single stranded DNA template

Two culture tubes each containing 2 ml 2XYT/MT medium were inoculated with 0.2 ml/tube of a fresh overnight culture of E. coli TG2 containing the plasmid to be sequenced. They were grown at 37°C on a roller wheel for one hour, and then 0.2 ml helper phage R408 (5 X 10^{11} pfu) was added to each tube and the cultures were allowed to grow at 37°C on a roller wheel for 6 to 8 hours with vigorous aeration. Cultures were distributed to three Eppendorf tubes and pelleted in a microfuge at 4°C for seven minutes. The supernatants were transferred to fresh tubes and centrifuged again at 4°C for 12 minutes. The supernatants were transferred to fresh tubes, and the phage precipitated by adding 0.3 ml of 20% polyetheleneglycol (PEG 8,000), 2.5 M NaCl, mixing well, and storing on ice for 15 minutes. The phage was collected by centrifugation at 4°C in a microfuge for 15 minutes. Phage pellets were

resuspended in 233 µl TE, pooled into a single tube, and extracted two times with phenol, two times with phenol/chloroform (1:1 /v:v) and two times with chloroform containing 4% iso-amyl alcohol. The single stranded DNA was precipitated by adding 60 µl 3 M sodium acetate, and 2.5 volumes of ethanol. Tubes were stored on powdered dry ice for 12 minutes before centrifuging at 4°C in a microfuge for 15 minutes. The single stranded DNA pellet was rinsed with 1 ml of 70% ethanol, dried in a speedvac, and resuspended in 50 µl sterile Milli Q water. The yield and purity of the single strand DNA template was determined by electrophoresis on a 0.7% nondenaturing agarose gel, with helper phage alone as a control. Single-stranded DNA templates were stored at -70°C prior to DNA sequencing.

(3) DNA sequencing reactions

Single strand DNA sequencing reactions were performed according to the manufacturer's "Protocol For DNA Sequencing With Sequenase® Version 2.0" (USB, Cleveland, Ohio). Double strand DNA sequencing reactions were performed as described (Cullmann et al. 1993).

(4) Isolation of viral RNA

This protocol is adopted from Chomczynski and Sacchi (18). Fresh serum from patients infected with hepatitis C virus was mixed thoroughly with three volumes of total RNA extracting solution (2 M guanidium isothiocyanate, 13 mM sodium citrate pH7.0, 0.5% sarcosyl, 0.1 M β-mercaptoethanol, 0.1 M sodium acetate pH4.0, and 50% water saturated phenol). Chloroform (0.2 volumes) was added and thoroughly mixed for 30 seconds, the solution was cooled on ice for 15 minutes, and then centrifuged at 10,000 Xg, 4°C for 20 minutes. The aqueous phase containing RNA was transferred to a fresh tube, 5

µl carrier polyacrylamide (100 mg/ml non-cross-linked but polymerized acrylamide in TE) was added followed by one volume of isopropanol. The tube was placed at -20°C for 1 hour or overnight to precipitate the RNA, and this was harvested by centrifugation at 10,000 Xg, 4°C for 20 minutes. The resulting RNA pellet was rinsed in 75% ethanol, drained well, and dissolved in 0.3 ml 0.1% SDS, 2 mM EDTA. This was transferred into a 1.5-ml Eppendorf tube, and precipitated with 0.3 M sodium acetate pH5.2 and 3 volumes of 95% ethanol. After centrifugation in an Eppendorf centrifuge for 15 minutes at 4°C, the RNA pellet was rinsed twice with 75% ethanol, vacuum dried, and dissolved in desired buffer (0.1% SDS, 2 mM EDTA or DEPC treated autoclaved milli Q water containing 1 unit/ml RNasin).

(5) Reverse transcription-polymerase chain reaction (RT-PCR) and DNA cloning

One quarter of the total RNA prepared from 5 ml of hepatitis C patient serum was mixed with the oligonucleotide primer HCV-NS3-2A: 5'-TT(A/G)GT GCTCTTGCCGCTGCC-3' in water in a total volume of 15 µl. It was heated at 100°C for 5 minutes and then allowed to cool at room temperature for 20 minutes. The sample was combined with 0.05 units RNasin, reverse transcription buffer (25 mM Tris-HCl pH8.8 [pH8.2 at 42°C], 25 mM KCl, 3 mM $MgCl_2$), 0.1 mM each dNTPs, 20mM DTT and 0.02 unit AMV super reverse transcriptase. The reaction was incubated at 42°C for 90 min. and then terminated by heating the sample for 10 minutes at 68°C. The RT reaction (2 µl)was added to a 50 µl reaction and amplified through 36 cycles of polymerase chain reaction (PCR) (80). The PCR reaction contained 50 mM KCl, 10 mM Tris-HCl [pH8.3 at room temperature], 1.5 mM $MgCl_2$, 0.01% gelatin, 200 µM each of dNTPs and 0.5 µg each of the primers HCV-NS3-B: 5'-

ACCGCGGCGTGTGGGGATATCATC-3' (5' end primer) and HCV-NS3-2A: 5'-TT(A/G)GTGCTCTTGCCGCTGCC-3' (3' end primer). Denaturing was at 94°C for 1 minute, primer annealing at 50°C for 90 seconds, and elongation at 72°C for 2 minutes. The primers were designed to anneal to relatively conserved regions of the sequence which bracket the putative serine proteinase (see results, Figure 3). The PCR reactions were analyzed by non-denaturing agarose gel electrophoresis and ethidium bromide staining. The resultant bands of the expected size were amplified again through 25 cycles of PCR. The amplified DNA fragments were blunted using the Klenow fragment of *E. coli* DNA polymerase I, and cloned into the *Eco* RV site of the plasmid pBluescript® KS^+. Clones were verified by DNA sequencing using the dideoxy method (83).

(6) Construction of expression plasmids

Standard methods were used for recombinant DNA manipulations (82). Subclones containing DNA fragments from the PCR amplifications were in all cases verified by nucleotide sequence analysis. The PCR reactions (total volume 50 µl) were performed with 1 µg of each primer, 2.5 units of Pfu polymerase (Stratagene), 20 mM Tris-HCl (pH 8.2), 10 mM KCl, 6 mM $(NH4)_2SO_4$, 2 mM $MgCl_2$, 0.1% Triton X-100, 500 ng nuclease-free BSA, 100 µM dNTPs, 5% DMSO, and 10 ng of template DNA from plasmid HCV #6. Each PCR involved 25 cycles of amplification, with denaturing at 95°C for 90 seconds, elongation at 74°C for 2 minutes, and annealing for 2 minutes (48°C for the first 5 cycles, then 60°C for the remainder).

For the construct pT7-H_6/HCVP (Figure 7), a DNA fragment containing the putative HCV serine proteinase beginning at the HCV polyprotein residue 1037 and ending at 1221 was constructed using the following primers:

5'-GGAATTCGGGGCCTCTTGGGATGCATAAT-3' (5' primer) and 5'-TCTCTG TCGACTCACTGGGGCACTGCTGGTGG-3' (3' primer). These primers added an 'in frame' *Eco* RI site at the 5' end and a stop codon followed by *Sal* I site at the 3' end of the DNA. The PCR fragment was inserted into the *Eco* RI and *Sal* I sites of pT7-7His6 (Figure 8) after digestion of both the DNA fragment and the vector with the appropriate restriction enzymes.

The construct pT7-H6/HCVPfix was nearly identical, but included additional 5' nucleotides from HCV, so as to encode the actual N-terminal sequence of the HCV serine proteinase which was discovered during the course of this project (32). The pT7-H6/HCVPfix construct encodes polyprotein residues 1027 to 1221. It was constructed by PCR engineering in the same way as was pT7-H_6/HCVP, using HCV#6 plasmid DNA for the starting template and using the same 3' oligonucleotide primer but a different 5' oligonucleotide primer (sequence 5'-GGAATTCGCGACGACGATGACAAGGCACCCATTA CGGCGTATGCCCAGCAGACAAGGGGCCTCTT-3'). The 5' primer was designed to engineer an 'in frame' 5' *Eco* RI site followed by a short segment of DNA encoding an enterokinase cleavage sequence. This was positioned in such a manner that cleavage of the protein by enterokinase would free the actual N-terminus of the HCV serine proteinase (i.e. Alanine 1027). The oligonucleotide also corrected the single base (dC) deletion mutation at position 1033 which was present in our original clones HCV #5 and #6 (see results).

Constructs pT7trx/HCVP and pT7trx/HCVPfix were constructed by placing the *E. coli* thioredoxin A coding sequence 'in frame' between the *Nde* I and *Eco* RI sites of pT7-H6/HCVP and pT7-H6/HCVPfix. The thioredoxin A gene was obtained by PCR amplification of *E. coli* strain K 12 genomic DNA

using the primers 5'-TCTCTCATATGAGCGATAAAATTATTCAC-3' (5' primer) and 5'-GGAATTCCCTTGTCATCGTCATCACCAGAACCAGAACCG GCCAGGTTAGCGTCGAGGAA-3' (3' primer). The amplified fragment contained a *Nde* I site at the 5' end, and an enterokinase site follow by an 'in frame' *Eco* RI site at the C-terminal end of the thioredoxin A.

(7) Expression of protein in *Escherichia coli*

Overexpression of recombinant proteins in *E. coli* using the pT7 system is a well established method as described by Studier et al. (89). The protocol used in this project is described below.

(a) Preliminary assays for expression

A 13 X 100 mm culture tube containing 5 ml of 2XYT medium and 100 µg/ml ampicillin was inoculated with a single isolated BL-21 (DE3) colony carrying the pT7-HCV serine proteinase recombinant plasmid to be expressed. Cultures were grown at 37°C on a roller wheel until the OD_{600} reached 0.7. A 1 ml aliquot of the culture was mixed with 200 µl of autoclaved 50% glycerol, frozen on powdered dry ice, and stored at -70°C. A second 1 ml aliquot of the culture was pelleted for 1 minute in a micro centrifuge and the cell pellet stored on ice to provide the t=0 sample. 1.5 ml of the culture was transferred to a new tube and induced with IPTG at a final concentration of 0.4 mM. The remaining 1.5 ml of the culture continued to be grown but without induction. Both induced and non-induced cultures were grown on the roller wheel at 37°C for an additional 3 to 4 hours, and then equivalent numbers of cells (i.e. OD_{600} equivalents) as were obtained in the t=0 sample were spun down from both cultures. The drained cell pellets were resuspended in 50 µl of 10 mM Tris-HCl

(pH8.0), 1 mM EDTA containing lysozyme (0.7 mg/ml), and then subjected to three cycles of freeze-thawing before being processed for SDS-PAGE analysis.

(b) Large-scale protein expression

To express the protein, 2 ml overnight culture grown from a single isolated colony was used to innoculate a 1L culture flask containing 500 ml of 2XYT medium supplemented with 100 μg/ml ampicillin. This was grown at 37°C with vigorous shaking until the OD_{600} reached 0.7. Protein expression was induced by adding IPTG to a final concentration of 0.4 mM, and the incubation continued at 37°C for three and a half hours before the cells were harvested by centrifugation (3,500 Xg, 20°C, 30 min.). The cell pellet was drained well, re-suspended in 35 ml of 0.1M NaH_2PO_4 /10 mM Trisma base (pH8.0) and passed through a French-press three times (8,000 psi). The resulting suspension was centrifuged at 15,000 Xg, 4°C for 30 minutes to pellet the insoluble components. In a typical case preliminary experiments with small-scale expression had established that the recombinant protein was present as insoluble inclusion bodies. The pelleted inclusion bodies were drained, washed in the same buffer as above, and dissolved in 6 M guanidine-HCl, 0.1 M NaH_2PO_4, 0.01 M Tris (pH8.0), 1 mM β-mercaptoethanol. Remaining cell debris was removed from the guanidine solution by centrifugation at 15,000 Xg, 4°C for 30 minutes, and the solution was used in nickel-chelate affinity chromatography.

(8) Nickel-chelate affinity chromatography

A typical column had dimensions of 15 X 50 mm, and was prewashed with autoclaved milliQ water and then equilibrated with 6 M guanidine-HCl, 0.1 M NaH_2PO_4, 0.01 M Tris pH8.0, 1 mM β-mercaptoethanol (Gu-HCl buffer).

The guanidine solution containing the recombinant protein was loaded at a flow rate of 1 ml/min. This flow rate was maintained throughout the course of the chromatography. The column was washed first with 10 column volumes of Gu-HCl buffer, then with 6 column volumes of 8 M urea, 0.1 M NaH_2PO_4 , 0.01 M Tris (pH8.0), 1 mM β-mercaptoethanol buffer, followed by 6 column volumes of the same buffer but with the pH adjusted to 6.3. Finally, the recombinant protein was eluted stepwise with 3 different elution buffers, using 6 column volumes each time. Buffers contained 8 M urea, 0.1 M NaH_2PO_4, 0.01 M Tris, 1 mM β-mercaptoethanol at the following pHs: 5.9, 5.3 and 4.5. Eluted proteins were followed by continuous UV monitoring at 280 nm using an LKB 2138 uvcord S. Fractions were collected using a LKB Bromma 2111 Multirac fraction collector, and each fraction was analyzed by SDS-PAGE to determine the purity and the yield.

(9) Dialysis and refolding of the purified proteins

(a) Dialysis

Fractions containing the purified recombinant protein (from the NTA Ni-chelating column) were pooled (volume ~ 80 ml) and transferred into prepared dialysis tubing. This was clamped at each end with plastic clamps (Spectrum, Houston, TX). The protein solution was dialyzed against six exchanges (4L each) of 1 mM DTT, 10 μM sodium acetate in milli-Q water, (measured pH was 6.3). The first exchange was made after one hour, the second after six to eight hours of dialysis, and the subsequent four exchanges were at 12 hour intervals.

(b) Refolding of proteins

This protocol was used when the protein came out of solution during the

dialysis procedure or during the course of lyophilization when the sample was inadvertently thawed. 10 mg of precipitated protein was dissolved in 3 ml of reducing buffer (6 M guanidine HCl, 100 mM Tris-HCl (pH8.0), 1 mM EDTA and 60 mM DTT) by stirring for 2 hours at room temperature. The protein solution was added slowly dropwise to 100 ml pre-chilled (5°C) refolding buffer (0.5 M guanidine-HCl, 100 mM Tris-HCl (pH8.2), 2 mM EDTA and 2 mM oxidized glutathione) with rapid continuous mixing on a stir platform in the cold room. The solution was kept stirring at 5°C for 24 to 36 hours or until it became clear. The refolded protein was dialyzed against 0.2 mM DTT, 2 μM sodium acetate (at least two 4L exchanges over 24 hours in milliQ water with 0.2 mM DTT, 0.2 μM sodium acetate, measured pH was 6.3), and then lyophilized and stored frozen at -20°C.

(c) Protein quantitation

Protein quantitation was accomplished using the PIERCE BCA protein assay reagent. The assay was performed by following the manufacturer's 'microtiter plate protocol' without modification.

(d) Enterokinase cleavage

This procedure was used to prepare the enzyme prior to assay on the peptide substrate.

H6/HCVPfix protein which had been previously purified over a Ni^{2+}-NTA column, dialyzed against 1 mM DTT, 10 μM sodium acetate (measured pH: 6.3), lyophilized, and stored at -20°C, was dissolved (2 mg/10 ml) in 50 mM sodium acetate pH 5.0, 15 mM calcium chloride, 0.1% triton X-100, and 4 M urea. Enterokinase (100 units, Sigma Chemicals) was added and the reaction

mixture was incubated overnight at 37°C with gentle shaking. Urea (2.6 g) was added to bring the urea concentration to 8 M. The cleaved (His)6 tag, together with uncleaved H6/HCVPfix protein, was removed by running the reaction mixture through a small nickel-chelate affinity column (about 1 ml resin bed). The flow-through HCV serine proteinase solution was adjusted to 2 mM DTT with 22 µl 1 M DTT and dialyzed against 1 mM DTT, 10 µM sodium acetate (measured pH: 6.3) for 4 hours, 0.2 mM DTT, 2 µM sodium acetate (measured pH: 6.3) overnight, 0.05 X PBS for 10 hours, and 0.02 X PBS overnight. The enzyme was concentrated 2 fold and washed once with reaction buffer (100 mM potassium phosphate, 6 mM sodium citrate, 2 mM EDTA, pH 7.5) using a centricon-10 spin column. It was stored at 4°C overnight and used in the enzyme assay.

(10) HCV serine proteinase activity assay

Reactions were carried out in duplicate wells using a flat bottomed 96 well microtiter plate. Each reaction contained 2.5 mM of peptide substrate in reaction buffer (100 mM potassium phosphate, 6 mM sodium citrate, 2 mM EDTA, pH 7.5), 0.5 mg/ml BSA, and 0.19 µM of the enterokinase cleaved HCV serine proteinase described above. Reactions with buffer only served as blanks, and those with the enzyme plus BSA, or the peptide substrate plus BSA served as negative controls. Reactions of enterokinase alone with the substrate and the recombinant malaria protein PfsY-c1 with substrate (PsfY-c1 was purified in exactly the same way as was the H6/HCVPfix protein), were also included as negative controls. At various times aliquots (10 µl) of the reaction mixture were transfered to new microtiter wells containing 0.25 M borate/0.13 M sodium hydroxide pH9.5 (quenching buffer, 50 µl). Freshly prepared 0.14 M 2,4,6-trinitrobenzene-sulfonic acid (TNBS) in 0.25 M borate (pH9.5) (12.5 µl) were

added to each well and the plate was incubated at room temperature on a shaker for 10 minutes. Immediately thereafter, color stabilization solution (3.5 mM Na_2SO_3, 0.2 M KH_2PO_4, 200 µl) were added to each well and the absorbance at 405 nm was measured using a microtiter plate reader and Microplate Manager II software. To generate a standard curve, duplicate aliquots of glycine solution at various concentrations ranging from 5 µM to 5 mM (10 µl each well)were added to wells containing quenching buffer (50 µl) and assayed in the same manner and at the same time as the proteinase samples.

(11) Calculation of Molecular weight for peptides/polypeptides

Calculations of Molecular weight for peptides/polypeptides were done using MacProMass computer software program.

(12) Construction of peptide substrate

(a) Synthesis

The peptide substrate was synthesized using Fmoc-L-amino acids on a PS3 Automated Solid Phase Peptide Synthesizer (RAININ instrument Co. Inc. Woburn, MA) at the 0.05 mmole scale. For each coupling step, 0.2 mmole HBTU activator and 0.2 mmole Fmoc-L-amino acid was used (i.e. a 4 fold molar excess). Double coupling procedures were used for each amino acid (program 4). The peptide synthesized had the sequence:

Ac-Tyr-Gln-Glu-Phe-Asp-Glu-Nle*-Glu-Glu-Cys-Ser-Gln-His-Leu-amide

* Nle stands for norleucine, which is substituted for methionine.

The C-terminal amide was achieved by using Rink resin for the synthesis (77). The Fmoc group on the N-terminal amino acid was removed during the course of synthesis on the instrument, and the peptide was acetylated in a subsequent chemical modification step prior to deprotection and cleavage from the resin.

(b) Chemical modification and cleavage

The resin to which the synthetic peptide was attached was drained and washed in turn with N,N-dimethylformamide, 100% ethanol, and then, dichloromethane. The resin was dried on an aspirator and the peptide acetylated by resuspending the resin in a mixture of 100 µl acetic anhydride, 1 ml dichloromethane, 150 µl triethanolamine. The reaction was allowed to proceed for 30 minutes at room temperature with gentle shaking. The resin was drained, washed 6 times with dichloromethane, and then dried in a vacuum desiccator. Peptide (0.05 mmole) on dried resin was added to a solution containing 10 ml of trifluoroacetic acid, 0.5 ml anisole, 0.5 ml thioanisole, and 0.5 ml water. The cleavage and deprotection reaction was allowed to proceed with gentle stirring for 2 hours at room temperature in a parafilm covered 30 ml beaker. The cleaved resin was filtered out using a fritted glass disc funnel and washed 4 times with 1 ml of trifluoroacetic acid (TFA) each time. The TFA was removed from the filtrate using a rotary evaporator at 30-32°C (typically a half hour operation). The peptide was precipitated from the remaining solution by adding 50 ml diethyl ether, and then filtered using a 15 ml fritted disc buchner funnel. It was washed 6 times with diethyl ether before being dried in a vacuum desiccator.

(c) HPLC purification

Crude synthetic peptide was dissolved in 50% acetic acid/water to a final

concentration of about 1-4 mg/ml and then purified by reverse phase high performance liquid chromatography (HPLC). About 2 mg crude peptide was injected each time onto a semi-preparative C-18 reverse-phase HPLC column (1/2" X 250 mm, Macherey-Nagel, Nuc.) using a 0-70% acetonitrile gradient with 0.05% trifluoroacetic acid in both solvents. A flow rate of 2 ml/min, 0.5%/min. was used. The column eluate was monitored at 210 nm. Fractions containing the major product peak were collected and verified by mass spectrometry.

(13) Immunization of laboratory animals

Methods used to immunize laboratory animals to produce antibody are as described by Harlow and Lane (37). The specific protocols used in this project are described below.

(a) Rats

Two female Sprague-Dawley rats, each about two months of age (≈200 g), were bled twice to obtain preimmune serum, and then immunized with 100 µl purified H6/HCVP protein (1 mg/ml in sterile milli-Q water) homogenized with 100 µl of complete Freund's adjuvant. Each rat was boosted at 4 week intervals with the same dose of antigen homogenized with 100 µl incomplete Freund's adjuvant. A single intraperitoneal site was used for each injection. Test bleeds were performed 10 to 14 days after the second and successive boosts. The serum antibody titer was determined by enzyme-linked immunosorbent assay (ELISA). Once high titer antibodies were induced, the rats were euthanized and bled out. The antiserum was separated and stored in 100 µl aliquots at -70°C.

(b) Rabbits

Two female New Zealand white rabbits, each 2 months of age, were initially bled twice to collect preimmune serum. They were injected subcutaneously at four different sites with a total of 900 µg purified H6/HCVP protein (dissolved in 500 µl sterile milli-Q water and homogenized with 500 µl complete Freund's adjuvant). Rabbits were boosted at six to nine week intervals with the same antigen dose, this time dissolved in 400 µl sterile water and homogenized with an equal volume of incomplete Freund's adjuvant. All booster injections were also at four different subcutaneous sites. Test bleeds were performed 10 to 14 days after the boosts, and serum antibody titers assayed as above.

(c) Mice

Four mice each 6 weeks of age, were bled for preimmune serum and then primed intraperitoneally with 50 µl antigen (1 mg/ml in sterile water) homogenized with 50 µl complete Freund's adjuvant. Booster injections were given intraperitoneally at three to six week intervals, using the same dose of antigen homogenized in 50 µl incomplete Freund's adjuvant. Terminal bleeds were performed 14 days after the third injection.

(14) Enzyme-linked immunosorbent assay (ELISA)

Ninety-six well microtiter plates (Falcon 3911 MicroTest III) were coated with H6/HCVP protein (2.0 mg/ml protein in phosphate buffered saline (PBS), 50 µl per well) by incubating at 37°C for 30 minutes. Each well was then blocked with 100 µl 20% fetal calf serum in PBS by incubating at 37°C for 10 minutes. Plates were washed with PBS containing 0.5% Tween-20 (PBST), and

the antiserum added to individual wells from serial dilution made in 1% bovine serum albumin (BSA)/PBST. Diluted preimmune serum (1:50, v:v) was also assayed. The plate was incubated at 37°C in a humidified CO_2 incubator for 30 minutes, and then washed three times with PBST. Secondary antibodies conjugated with horse radish peroxidase and against rat, mouse or rabbit Fc's were diluted 1: 5000 in 1% bovine serum albumin (BSA)/PBST and added in 50 µl aliquots to each well. This was incubated for 30 minutes at 37°C and the plate again washed with PBST. 0.1% 2, 2'-azino-bis(3-ethylbenz-thiazoline-6-sulfonic acid) (ABTS) in 0.2 M Na_2HPO_4, 0.1 M citric acid (pH7.5) and 0.003% H_2O_2 was added and the plate left at room temperature for an appropriate period of time to allow for color development. The absorbance was measured at 405 nm with a reference of 490 nm using a microtiter plate reader (Molecular Devices, Palo Alto, CA) and ΔSoft software (BioMetallics, Princeton, NJ).

CHAPTER III

RESULTS AND DISCUSSION

Cloning of a DNA fragment encoding the HCV serine proteinase

Previous analysis of the hepatitis C viral genome and the deduced polyprotein sequence indicated that the N-terminal domain of the NS3 region may represent a serine proteinase (9,20,30,65). These conclusions were based on the fact that the amino-terminal region of the HCV NS3 contains consensus sequences found in all serine proteinases, including the catalytic triad H, D, S and the consensus sequence GXSGXP surrounding the active-site serine residue. Furthermore, the spacing and order of the residues is similar to that found in other flavi- and pestiviruses for which NS3 has been demonstrated to be a viral serine proteinase involved in polyprotein processing (74).

The goal of this project was to clone and express in *E. coli* a catalytically active HCV serine proteinase which could be used to screen proteinase inhibitors. Such proteinase inhibitors might have therapeutic applications against HCV. Considering the relatively low titer of HCV in clinical samples (14), we chose to clone the proteinase gene by reverse transcription and polymerase chain reaction amplification (RT-PCR). However, despite the fact that several strains of HCV had been molecularly cloned and their genomes sequenced (16,19,46,47,69,92,94), little was known about the precise boundaries of either the serine proteinase or the NS3 region when we started this project in 1992. Initially, with the assistance of Dr. Bruce Malcolm, we made an informed 'guess' as to the boundaries of the HCV serine proteinase. These were placed between residues 1007 and 1221 in the polyprotein. At the time we were

designing the PCR primers, the NS3 proteinases of Dengue virus type 2 and Kunjin virus had been reported to cleave their respective polyproteins C-terminal to double basic amino acid residues to produce the mature N-termini of NS2B, NS3, and NS5 (15,74). We hypothesized that the HCV serine proteinase might also cleave at the N-terminus of NS3 (i.e. cleave its own N-terminus). Therefore our PCR primers included both the putative N-terminal region of the NS3 as well as the C-terminus of NS2 region. We hypothesized that if a HCV serine proteinase did cleave at the NS2/NS3 junction, then the proteinase activity might be assayed by self cleavage of the recombinant protein. Oligonucleotides primer pairs were derived from regions of relatively conserved DNA sequence for the known HCV sequences (Figure 3), and which flanked the putative HCV serine proteinase domain.

Since there appeared to be sequence variation among the HCV strains, and because of the fact that we did not know the DNA sequence or strain designation of the HCV patient samples we were using in RT-PCR, we actually generated a series of primer pairs as follows: 5'-ACCGCGGCGTGTG GGGATATCATC-3' (5' end primer) and 5'-AGGTCGGCCGACATGCATG CCATG-3' (3' end primer) and 5'-ACCGCGGCGTGTGGGGATATCATC-3' (5' end primer) and 5'-TT(A/G) GTGCTCTTGCCGCTGCC-3' (3' end primer). This trial-and-error approach to RT-PCR clone the required segment of the HCV genome also involved testing a variety of different PCR conditions. The following conditions were attempted: standard PCR buffers (50 mM KCl, 10 mM Tris-HCl pH8.3 at room temperature, 1.5 mM $MgCl_2$, 0.01% gelatin, 200 µM each of dNTPs, 0.1 nM each primer, 50 ng template cDNA from RT), with denaturing at 94°C for 1 minute, elongation at 72°C for 2 minutes, and primer annealing for 90 seconds at 1) 68°C for 10 cycles and then 65°C for 30 cycles; 2)

65°C for 40 cycles; 3) 62°C for 30 cycles; 4) 60°C for 36 cycles; 5) 55°C for 36 cycles; 6) 50°C for 10 cycles and then, 55°C for 30 cycles; 7) 49°C for 40 cycles; and 8) 50°C for 36 cycles. Eventually, using the PCR primers shown in Figure 3 and the RT and PCR conditions described in the Materials and Methods, we were able to generate the expected 749 base pair fragments (Figure 4) from three different patient samples. The cDNA fragments were cloned into the *Eco* RV site of pBluescript KS^+, giving rise to clones HCV#3, HCV#5, and HCV#6, all of which were verified by DNA sequence analysis.

Characterization of HCV proteinase cDNA clones

Each clone was sequenced across one strand using both single stranded and double stranded templates and the dideoxy method. A segment of DNA sequence from each clone was sent by e-mail to a blast search server using blastn@ncbi.nlm.nih.gov. The return message showed that the DNA segments from clones #3, #5, and #6 had relatively high sequence homology to HCV strains HPCPLYPRE and HPCJCG. The entire cDNAs were subsequently sequenced, and the data processed using DNA Strider 1.2 software. The cDNA sequences were translated in all three reading frames and compared to that of known strains of HCV. Clone HCV#3 (from one patient sample) contained two deletions (a dG deletion in the codon which normally codes for glycine 1007, and a dC deletion in the codon codes for alanine 1033), while clones HCV#5 and HCV#6 both had a single dC-deletion in the third position of the codon which normally codes for alanine 1033 of the HCV polyprotein (Figure. 5). Except for the single frame-shift mutation (circled dC missing), the derived amino acid sequence in Figure 5 corresponds to the HCV polyprotein residues 990 to 1238 inclusive. This includes 38 amino acid residues that belong to the C-terminal region of NS2 and the 211 N-terminal residues of the NS3 region (32).

Most importantly, the deduced protein sequence contains the proposed catalytic triad (His 1083, Asp 1107, and Ser 1165) and the consensus sequence GXSGXP surrounding the active-site serine residue found in all serine proteinases (Figure 5). The deduced amino acid sequences are all highly homologous to that of HCV strains HPCPLYPRE and HPCJCG.

The DNA sequence information suggested that the cloning of the cDNA segment of NS3 had been successful. However, all clones contained a dC deletion mutation at codon position 1033 of the HCV polyprotein. At the time this work was done, the real N- or C-terminus of NS3 had not yet been published. Although we had originally assumed that the serine proteinase might begin at residue 1007 (after the dibasic RR) and end at residue 1221, the occurrence of this point deletion in the cDNA led to a re-evaluation of the assumptions about the N-terminus of the HCV serine proteinase. Production of a recombinant protein starting at polyprotein residue 1007 would require site-specific mutagenesis of codon 1033 which would require some time and effort. The reasons for choosing residue 1007 as the N-terminus of the NS3 were relatively weak, and a comparison with rat pancreatic trypsin (Figure 6) suggested that starting at polyprotein residue 1037 might be equally successful, and not require the site-specific mutagenesis. Therefore the HCV serine proteinase was first expressed from residue 1037 to residue 1221 (construct pT7-H6/HCVP, Figure 8). This H6/HCVP protein was not ultimately used in the proteinase assay, since more accurate information about the boundaries of NS3 became available mid-way through this work (see below). However, the protein was used to generate antibodies in laboratory animals.

Expression of HCV proteinase using pT7-H6/HCVP

The pT7-H6/HCVP construct was created by re-amplifying the gene from clone #5 with a new pair of oligonucleotide primers (see Materials and Methods) to engineer an *Eco* RI site at 5' end and a *Sal* I site at 3' end for appropriate in-frame cloning into the expression plasmid pT7-7His6 (Figure 7 and 8B). The newly created copy of the gene encoding the HCV polyprotein residues 1037 to 1221 was cloned downstream of the T7 bacteriophage promoter in the vector. In addition to the T7 promoter, this expression vector contains a ribosome binding/Shine-Dalgarno sequence, a start codon (ATG) followed by six triplets (CAC) coding for $(histidine)_6$, a short DNA sequence encoding a thrombin cleavage site, and a polylinker segment beginning with an *Eco* RI site (Figure 7). The cloned gene is effectively transcribed by bacterophage T7 RNA polymerase and the transcripts translated to produce a recombinant protein fused with $(His)_6$ and a thrombin cleavage site at the N-terminus.

HCV proteinase polypeptide expressed at high levels

Various recombinant proteins were expressed at high levels throughout this project using the pT7 expression system in *E. coli* BL21(DE3). Figure 9 A shows the first expression trial which was conducted with the construct pT7-H6/HCVP. Total cell lysates of IPTG induced cultures were assayed side-by-side with those from non-induced cultures. An obvious band of recombinant protein can be seen in *E. coli* total cell lysates from the IPTG induced culture (Figure 9 A. lane 2, arrowhead) which is absent from the non-induced culture (Figure 9 A. lane 1). The calculated molecular weight for the protein expressed

from the construct pT7-H6/HCVP is 21,232 daltons. The apparent molecular weight of the recombinant protein on SDS-PAGE is about 23 kilodaltons.

Recombinant HCV serine proteinase present in inclusion bodies

To obtain large quantities of protein, preparative scale expression (total volume of 2 liter culture) was carried out (Figure 9 B.). The cell pellet from each 500 ml of IPTG induced culture was resuspended in 35 ml buffer (0.1M NaH_2PO_4, 10 mM Tris-HCl pH8.0) and lysed by passing 3 times through a French press (>8,000 p.s.i.). The supernatant was separated from insoluble protein by centrifugation. Both the supernatant and the pellet were assayed on SDS-PAGE (Figure 9 B.), and the recombinant protein was found at high levels only in the pellet (Figure 9 B. lane 2, arrowhead).

It has been well documented that when exogenous proteins are overexpressed in *E. coli* they frequently accumulate intracellularly in an insoluble form known as inclusion bodies (61,98). These can only be dissolved with strong detergents or with dissociating agents such as 6 M guanidine hydrochloride or 8 M urea, and often this process also requires β-mercaptoethanol. When the HCV serine proteinase was expressed in *E. coli* , and the cells lysed by French press or by cycles of freeze-and-thaw, the recombinant protein was found almost exclusively in the pellet fraction after centrifugation. This simply indicates that the recombinant protein has accumulated in the form of inclusion bodies. These were found to be completely soluble in 6 M guanidine hydrochloride at all the times, while 8 M urea failed to dissolve the recombinant protein completely. The $(His)_6$ fusion protein was solubilized in 6 M guanidine hydrochloride, cell debris was

removed by centrifugation, and the protein was purified directly by nickel-chelate affinity chromatography.

Nickel-chelate affinity chromatography offers one step purification

Very often the purpose of overexpression of a recombinant protein is to obtain large quantities of pure protein. This approach is especially useful if the protein is available in only limited quantities from naturally occurring sources, and/or if it is difficult to isolate in pure form from these sources. A key step in obtaining useful recombinant proteins from *E. coli* or from other expression systems involves purification of the desired protein to eliminate other contaminating polypeptides. For example, it may be necessary to remove endogenous cellular proteinases from the recombinant protein in order to avoid degradation. As a simple rule of thumb, the fewer steps that are involved in purification, the less time and money is expended and the more likely the protein of interest is to be intact. A simple, inexpensive, and efficient approach is always preferred. In this project, we employed nickel-chelate affinity chromatography to purify the recombinant protein H6/HCVP. The approach involved loading and eluting under denaturing conditions, using a step gradient of different pH's (Figure 10). This method offered a simple and effective way to purify the protein with only a single step. After dialysis, the eluted fractions were pure enough for immunization and biochemical studies (Figure 11).

From a one liter culture of *E. coli* cells transformed with the construct pT7-H6/HCVP, at least 30 mg of purified recombinant HCV proteinase was obtained following nickel chelate affinity chromatography and dialysis. Protein express-ion was typically induced for 3.5 to 4 hours prior to harvest.

Extending the induction time at 37°C from 4 hours to overnight had no appreciable effect on the yield of recombinant protein. Overexpression of this initial HCV serine proteinase using the pT7 expression system produced ample amounts of protein for animal immunizations.

Recombinant proteinase has the expected N-terminus

The HCV serine proteinase recombinant protein H6/HCVP expressed from the construct pT7-H6/HCVP was purified and dialyzed, and the protein run on SDS-PAGE (Figure 11) and electro-transferred to a PVDF membrane. The membrane was stained with Coomassie brilliant blue and the protein band cut out for N-terminal amino acid sequencing. Data for the first 20 cycles gave a sequence of MHHHHHHLVPRGSGIRGLLG, which is exactly the same sequence as expected from the engineered DNA (Figure 8). The purified recombinant protein was dialyzed, refolded, lyophilized and used immunize laboratory animals. In this case 'refolding' was accomplished by simple dialysis against 1 mM DTT, which apparently made the purified HCV serine proteinase with the $(His)_6$ tag soluble in water. Unfortunately, when the protein was first obtained we did not know what peptide substrate to use to test the enzyme, since this was prior to the identification of the cleavage sites by Grakoui *et al.* (32). The protein could, however be used immediately to generate antisera.

Generation of antiserum against the H6/HCVP protein

Polyclonal antiserum against the HCV serine proteinase should provide a useful tool for protein purification. Such antiserum makes it possible to develop an immobilized antibody column for affinity purification, and to use standard assay techniques such as ELISA, Western blotting, immunostaining, and immunoprecipitation. Figure 12 shows the ELISA result for antiserum

from two rabbits (#D420 and #D669), two rats (#1 and #2), and a pool of four mice. For all these animals, purified, dialyzed, and lyophilized H6/HCVP protein was used as the immunogen. Immunization protocols are described in the Materials and Methods section. Rabbit #D420 was boosted 3 times, and rabbit #D669 2 times. Two weeks after the final boost antiserum was assayed by ELISA (Figure 12). Rabbit #D669 developed antibody titers of 1:100-1:1,000, while the rabbit #D420 showed virtually no response. Later on it became apparent that rabbit #D420 had become ill and was treated with steroids by the Health Science Laboratory Animal Services during the course of the immunization and this explains the minimal titers. These rabbit antisera were not harvested.

In total, rat #1 received 6 injections, and rat #2, 7 (initial immunizations plus boosting injections). The final antibody titers as assayed by ELISA were 1: 64,000 for rat #1, and 1: 8,000 for rat #2 (Figure 12). Both rat antisera were harvested and stored for future experiments. Antibody titers for mice assayed as a pool 2 weeks after the final immunization (total of 3 injections) were also disappointing. The ELISA assay showed only minimal titers in the mice (Figure 12).

Construction of pT7-H6/HCVPfix and production of a second recombinant HCV proteinase

The expression plasmid pT7-H6/HCVPfix was designed mid-way through this project, as soon as new information was released regarding the cleavage of NS2/3 site which defines the N-terminus of the HCV serine proteinase (31). This cleavage occurs between amino acids residue 1026/1027 (Leu-Ala) of the HCV polyprotein. The construct pT7-H6/HCVPfix was

engineered with the DNA sequence beginning at HCV polyprotein residue 1027 and ending at residue 1221(Figures 8). The 5' end of the gene was cloned into the *Eco* RI site of pT7-7His6 (Figure 7) with a short segment of DNA added which codes for an enterokinase cleavage sequence. The construct was engineered in such a manner that cleavage of the protein by enterokinase will release the defined N-terminus of the HCV serine proteinase (i.e. the residue 1027, Figure 8). The 5' oligonucleotide primer used in the PCR was made so as to correct the single dC deletion (Figure 5), to complete the HCV coding region back to residue Ala 1027, to add the enterokinase site, and to enable in-frame cloning into the *Eco* RI site (see Materials and Methods).

After the construct pT7-H6/HCVPfix was made and confirmed by DNA sequencing, it was transformed into *E.coli* BL21(DE3), and a pilot expression assay was conducted. Figure 13 shows a comparison of IPTG induced and non-induced cultures on SDS-PAGE, where a single induced band is seen (lane 1, arrowhead). The calculated molecular weight of the protein encoded by pT7-H6/HCVPfix is about 23 kilodaltons. The apparent molecular weight of the recombinant protein by SDS-PAGE is about 25 kilodaltons. The new recombinant protein was produced in a 1 liter culture, the cells processed by French press, and the supernatant and insoluble pellet assayed by SDS-PAGE (Figure 14). Again, the recombinant protein was found in the insoluble pellet (Figure 14, lane 2, arrowhead) indicating deposition of the protein in inclusion bodies. The recombinant protein was purified by nickel-chelate affinity chromatography (Figure 15). After dialysis against 1 mM DTT to refold the protein (the same as for H6/HCVP), the protein ran as a doublet on SDS-PAGE. A careful examination of the deduced amino acid sequence of H6/HCVPfix showed that a consensus sequence for the HCV serine proteinase cleavage

motif (32) had been inadvertently added at the N-terminus of the protein (Figure 8). It was hypothesized that once refolded, the serine proteinase might be active and intermolecular cleavage might occur at this N-terminal site. In order to rule out this possibility, the two protein bands (see Figure 16, lane 1) constituting the doublet were separated by SDS-PAGE, transferred to a PVDF membrane, and each bands sequenced separately by Edman degradation (40). The protein band which had the smaller apparent molecular weight on SDS-PAGE gel was also characterized by mass spectrometry. The Mass spectrum showed that this protein had a mass of 23066, which was in close agreement with the calculated mass of 23022 daltons for the full length recombinant $(His)_6$ fusion protein. This information suggested that the polypeptide was not cleaving itself. This result was confirmed by amino acid sequence analysis, which showed that both protein bands had the same N-terminus, namely MHHHHHHLV. Although these results indicated that the serine proteinase was not self cleaving, they did however, verify that overexpression of the new altered HCV serine proteinase in *E. coli* had been successful.

Purified recombinant protein produced from construct pT7-H6/HCVPfix (Figure 15) was refolded by dialysis against 1 mM DTT, 10 μM sodium acetate pH5.2. Two mg of this refolded and lyophilized protein was used as the starting material for enterokinase processing to produce proteinase with an authentic N-terminus. Figure 16 shows the result of enterokinase processing. Lane 1 shows the starting material, and lane 2 was the same material following overnight incubation with enterokinase. The enterokinase digested protein (lane 2) was repurified over a nickel-chelate affinity column, this time the follow through collected (lane 3). Presumably all protein in lane 3 has the $(His)_6$ segment removed, suggesting that the enterokinase digestion

was not very efficient. The longer band in lane 3 was hypothesized to be the $(His)_6$ tag cleaved full length HCV serine proteinase since its calculated molecular weight is 20.5 kilodaltons. In this case the $(His)_6$ tag segment on the original protein have apparently caused the protein run slower on the gel. The shorter fragments in lane 2 suggest that internal cleavage of the protein also occurred.

Production of *E. coli* thioredoxin fusion proteins

LaVallie and colleagues (1993) reported that expression of heterologous recombinant proteins in *E. coli* as fusions with *E. coli* thioredoxin circumvents the formation of inclusion bodies in the bacterial cytoplasm. To explore this possibility for the HCV proteinase, the *E. coli* thioredoxin gene was cloned by PCR and fused to the HCV serine proteinase (constructs pT7trx/HCVP and pT7trx/HCVPfix; Figures 7 and 8), using the same approach as LaVallie. We anticipated that this might make the HCV enzyme soluble, and induce proper folding. Although the HCV serine proteinase/thioredoxin fusion protein accumulated to high levels in IPTG induced cells (Figures 17 and 18), unfortunately the "dramatic increase in solubility of heterologous proteins synthesized in the *E. coli* cytoplasm" was not demonstrated when osmotic shock fractionation was conducted as described (55) (Figures 17 B. and 18). Cultures were also grown and induced at different temperatures in order to see if this would have an effect on protein solubility (Figure 17), but unfortunately in none of these experiments were we able to produce soluble protein. Osmotic shock buffers of different pH's were also tested to see if they would make the fusion protein more soluble. Again, no significant amount of soluble *E. coli* thioredoxin fusion protein was released from the *E. coli* cytoplasm by osmotic shock treatment (Figure 18). This suggests that use of the thioredoxin fusion

system to increase the solubility of heterologous proteins synthesized in *E. coli* is likely to be protein specific, since it did not circumvent the formation of inclusion bodies for the recombinant HCV serine proteinase. Given these results, production of the *E. coli* thioredoxin fusion with the HCV proteinase was not pursued further, particularly since there was no simple and effective means for purification of the recombinant material.

Generation of synthetic peptide substrate

The peptide substrate for testing the recombinant HCV serine proteinase expressed in *E.coli* was carefully modeled from the NS4A/NS4B junction of the HCV polyprotein (32). This junction has the sequence: Tyr-Gln-Glu-Phe-Asp-Glu-Met-Glu-Glu-Cys-Ser-Gln-His-Leu. This sequence is representative of the majority of the consensus cleavage recognition sequences at the P6, P1, and P1' positions (Table 2), using ten different HCV strains all aligned at the NS4A/NS4B, NS4B/NS5A, and NS5A/NS5B junctions (32). The amino acid sequence of the NS3/NS4A junction was not considered here since it is cleaved *in cis*. A 33-amino-acid segment from the NS4A protein was reported to be required for cleavage at the NS3/NS4A and NS4B/NS5A sites, and it accelerated cleavage at the NS5A/NS5B junction (26). However, cleavage at the NS4A/NS4B junction did not appear to require the presence of the NS4A C-terminal 33-amino-acid fragment, making a peptide segment from this junction the most ideal candidate for the substrate upon which to base our HCV serine proteinase activity assay.

Several modifications were made to the actual substrate peptide for technical reasons. Norleucine (Nle) was used in place of methionine, since methionine is known to have potential problems with oxidation to methionine sulfoxide, which shifts the elution point on reverse-phase HPLC during

purification (34). The peptide was synthesized by solid phase Fmoc methods (6) on Rink resin (77) to generate an amidated carboxyl terminus. This removes the charges from the carboxyl terminus making the whole structure more closely mimic a peptide bond. The same argument also applies to the acetylation of the N-terminus. Moreover, since we chose a colorimetric assay using 2,4,6-trinitrobenzenesulfonic acid (TNBS) to measure the proteolytically cleaved peptide (11,60), the N-terminus of the peptide substrate was also blocked in order to prevent this primary amino group from reacting with TNBS and contributing to a high background. Following peptide synthesis, the N-terminus was acetylated using acetic anhydride. This was done while the peptide was still on the resin, so that the side chain protecting groups are still present and the acetylation occurs only on the available N-terminal amino group. After cleavage and deprotection with 95% TFA and the appropriate free-radical scavengers (6), the TFA was removed and the crude peptide precipitated, washed, and dried in vacuum desiccator. The crude peptide was not easily dissolved in 5-20% glacial acetic acid, but could be dissolved in 60% acetic acid, to a concentration ≤4 mg/ml. After filtering through a 0.8 mm syringe filter, the crude peptides were purified by HPLC on a C-18 reversed-phase column (Figure 19). The major product peak was verified by mass spectroscopy. The mass spectra showed a major mass of 1811.3 which is in excellent agreement with the predicted mass of 1811.9 for the peptide.

Interestingly, the yield of the desired product accounted for only 50% (0.025 mmole) of the total synthesis. A second large peak (Figure 19) accounted for about 40% of the product, and had a mass of 2055. Since 2055-1811.9=243.1, this suggests that there was a trityl protecting group (molecular weight 242; present on His, Cys, and Gln Fmoc-amino acids) remaining on the peptide, and

that the material need further treatment with TFA. Given this reasonable assumption, it suggests that the original peptide synthesis yielded 90% (i.e. 0.045 mmole) of the desired protected peptide--which is quite reasonable--but that incomplete cleavage of protecting groups further reduced the final yield.

Enterokinase cleaved H6/HCVPfix is catalytically active

The goal of this project was to express a catalytically active recombinant HCV serine proteinase. In order to assay the enzyme we first had to construct and purify a peptide substrate. The enzyme used in the initial activity assay was first processed by enterokinase cleavage and passage over a Ni^{2+}-NTA column to remove the $(His)_6$ fusion peptide (see Material and Methods). This was followed by dialysis against 1 mM DTT in order to remove the urea and refold the protein. The substrate peptide (Ac-Tyr-Gln-Glu-Phe-Asp-Glu-Nle-Glu-Glu-Cys-Ser-Gln-His-Leu-NH_2) was expected to be cleaved into two parts: a N-terminal decapeptide (Ac-Tyr-Gln-Glu-Phe-Asp-Glu-Nle-Glu-Glu-Cys), and a C-terminal tetrapeptide (Ser-Gln-His-Leu-NH_2). The colorimetric assay used to measure proteinase activity depends in theory on the reaction of TNBS with the primary amino group of the tetrapeptide Ser-Gln-His-Leu-NH_2 generated by the enzymatic cleavage. The proteolysis assays were conducted at room temperature (23°C) for variable times using a biological buffer (0.1 M potassium phosphate, 6 mM sodium citrate, 2 mM EDTA, pH 7.5) as in the hepatitis A 3C proteinase study (60).

1) Evidence from the colorimetric assay using TNBS

Figure 20 shows the results of the colorimetric assay using TNBS. The proteinase (0.19 µM) was incubated at room temperature with 2.5 mM peptide substrate and biological buffer, in a total volume of 200 µl. Aliquots were taken

from the reaction at 2, 3, 4, 5, 6, and 23hrs., and quenched in borate buffer. The following day, freshly prepared TNBS was added and the samples incubated at room temperature for 10 minutes. Color stabilization solution was added and the absorbance at 405 nm measured. Values obtained from the colorimetric assay (which measures the putative tetra peptide cleavage product) were converted into μM of amino groups released using a glycine standard curve. The data showed that with a low concentration of enzyme (0.19 μM) the conversion to product was linear with time (Figure 20). Amino groups were generated at a rate of 1.2 μM/min, well above the background generated with substrate or enzyme alone. Although there was some scatter in the base-line of the control reactions, enzyme alone or substrate alone had essentially zero slopes, suggesting that little or no product was released from either of these reactions. Addition of PMSF (7 μl of a 10 mM stock solution) to the 20 μl of enzyme solution 10 minutes before adding the substrate peptide significantly reduced the rate of generation of amino groups (Figure 20). In order to rule out the possibility of contamination by endogenous *E. coli* proteolytic activities, a recombinant malaria sexual stage surface antigen protein PfsY-c1 was also tested in the assay with our peptide substrate. The PfsY protein was expressed in the same *E. coli* host and purified in exactly the same way as was the H6/HCVPfix protein. The results using PfsY, suggested only random, scattered background values (Figure 20). Addition of 5 units of enterokinase to the assay also did not induce the release of any significant quantity of primary amino groups (Figure 20).

The initial velocity at which amino groups were generated increased as more enzyme was used (Figure 21). Furthermore, the increase in the velocity at which amino groups were generated is also linear in the range of enzyme

concentrations tested (Figure 21). Initial rate measurements yielded a k_{cat} of 6.6 ± 0.2 s^{-1}, assuming that the enzyme preparation contained 100% active molecules. This compares to a k_{cat} of 1.8 ± 0.1 s^{-1} for the Hepatitis A 3C proteinase (60), which is among the lowest values of k_{cat} reported for any proteinase. The k_{cat} for the HCV serine proteinase may actually be considerably higher, since it is highly unlikely that all of the recombinant molecules would be perfectly refolded and enzymatically active. The assay was also conducted in HEPES-buffered saline (0.8% NaCl, 0.037% KCl, 75 mM Na_2HPO_4, 2.1 mM HEPES, pH8.2). The results showed a small but not significant difference in velocity of the reaction between HEPES-buffered saline and phosphate buffer (Table 4).

2) HPLC purification and mass spectroscopy of the major cleavage product derived from the peptide substrate

The proteolysis reaction for this study was done in a total volume of 300 µl, and included 1.66 mM substrate peptide and 1.9 µM enterokinase cleaved and refolded H6/HCVPfix protein in potassium phosphate buffer pH 7.5. The reaction was incubated overnight at 37°C to maximize yield of product, and the entire 300 µl was subjected to HPLC separation on a C-18 reverse phase column (Figure 22). The eluded peaks were collected and dried in a Speedvac. The expected C-terminal tetrapeptide fragment (Ser-Gln-His-Leu-NH_2) is very polar and probably flows through the reverse phase column. On the other hand, the expected N-terminal decapeptide cleavage product (Ac-Tyr-Gln-Glu-Phe-Asp-Glu-Nle-Glu-Glu-Cys) is more polar than the substrate, and we anticipated that it would elute earlier than the substrate and could be collected. The HPLC profile of the cleavage reaction was compared with that of the identical reaction at t=0 under exactly the same conditions (Figure 22). The fraction containing

the differential peak (indicated by the arrow in Figure 22) was analyzed by plasma desorption mass spectroscopy (PDMS) (22,59) (Figure 23). The differential peak gave a mass at 1370.5, (likely the sodium adduct of the expected peptide), and a mass at 1386.5 (likely the potassium adduct of the same peptide). These type of adducts are often seen in mass spectra. In particular the reaction buffer contained potassium which could account for its presence in the product. Because the HPLC separation was performed at about pH3 (0.5% TFA) it is likely that only the terminal carboxyl group was unprotonated (i.e. the pKa of the acidic side chains would be higher than 3.0), thus accounting for the fact that only a single sodium or single potassium ion could become associated with the peptide. Since 1370.5-23 (the mass of sodium)=1347.5, and 1386.5-39.1(the mass of potassium)=1347.4, both values are in excellent agreement with the calculated average mass of 1347.4 for the N-terminal decapeptide (Ac-Tyr-Gln-Glu-Phe-Asp-Glu-Nle-Glu-Glu-Cys). This mass spectra data offered direct evidence that the substrate tetradecapeptide was actually cleaved specifically at the Cys-Ser bond, just as would be predicted for enzymatic cleavage by the HCV serine proteinase (32). It also supports the initial results obtained using the colorimetric/TNBS assay that the enterokinase cleaved, refolded HCV serine proteinase expressed in *E. coli* was enzymatically active.

CHAPTER FOUR

FINAL DISCUSSION AND CONCLUSIONS

The goal of this research project was to produce enzymatically active HCV serine proteinase. During viral replication, HCV initially produces a polyprotein which is subsequently cleaved to generate the various structural and non-structural (NS) proteins. Normally a single viral protein (NS3) expresses both serine proteinase and helicase activities, and we attempted to isolate the proteinase activity by expressing only a N-terminal subsegment of NS3. By comparing the HCV polyprotein sequence with amino acid sequences of other similar viral and non-viral serine proteinases, an initial hypothesis was made about the N-terminal and C-terminal boundaries of the HCV serine proteinase domain. The domain was initially assumed to extend from polyprotein residue 1007 (the G following RR) to 1221 (ending PQ). With this information in hand, the next step was to obtain or generate a DNA fragment encoding the desired protein.

Since cloned HCV genes were not readily available, but infected patient serum was, it was decided that it would be relatively simple to generate the necessary DNA fragment de novo. By trying several different reaction conditions, a cDNA fragment encompassing roughly the desired region of the HCV genome was obtained by RT-PCR of viral RNA obtained from patient serum. Since the HCV virus nucleotide sequence is highly polymorphic between strains, the initial boundaries of the cDNA fragment generated were dictated not only by assumptions about the serine proteinase domain, but also by the necessity of designing PCR primers which would recognize conserved sequences which are present in a variety of HCV strains (Figures 3 and 5).

The cDNA fragments generated independently from three different HCV patient isolates were cloned and characterized by DNA sequencing. Surprisingly, all three sequences showed the same deletion--a dC normally present at about nucleotide position 3440 in the HCV genome (polyprotein codon number 1033) was absent--and one sequence contained a second deletion as well (dG at about HCV nucleotide position 3360). The fact that all three independent cDNA isolates contained an identical deletion suggested that either the viral polymerase has a tendency to make a systematic error during viral replication (HCV is know to generate large quantities of defective viral particles), or that a systematic error occurred during either the reverse transcription or the PCR reactions. Another possible explanation for this result is that of a systematic sequencing error, since only one strand of the DNA was sequenced, but a careful review of the sequencing gels showed the data to be very clear, and free of any compression artifacts. Given that one clone had a second deletion mutation, and that such deletion mutations are relatively rare in RT-PCR, the theory that HCV RNA polymerase itself is responsible for the deletion mutations is most likely.

The occurrence of the dC deletion mutation at codon position 1033 of the HCV polyprotein lead to a re-evaluation of the assumptions about the N-terminal boundary of the HCV serine proteinase. Production of a recombinant protein starting at polyprotein residue 1007 would require site-specific mutagenesis of the cDNA at codon position 1033 to restore the reading frame, and this would require some time and effort. The reasons for choosing residue 1007 were relatively weak, and comparison with rat pancreatic trypsin (Figure 6) suggested that starting at polyprotein residue 1037 might be equally successful, and not require the site-specific mutagenesis. This chain of logic

lead to the production of the first recombinant protein, H6/HCVP, by making use of PCR to re-engineer the original cloned cDNA fragment shown in Figure 5.

To express H6/HCVP we used *E. coli*, since this was the simplest, cheapest, and most straightforward system available. Since it was assumed that expression of an active HCV proteinase might be toxic to *E. coli*, a T7 vector and T7 RNA polymerase driven expression system was used. This system allowed the DNA cloning to be completed in the absence of any possible protein expression, and it is reputed to give very high yields of recombinant protein in the appropriate host. The protein was expressed as a fusion protein with a short N-terminal $(His)_6$ tag and thrombin cleavage site (Figure 8), since this would facilitate purification of the protein under either denaturing or non-denaturing conditions, and the affinity tag could then be cleaved off using thrombin.

Although H6/HCVP was expressed at high levels using the T7 expression system in *E. coli*, it was unfortunately deposited in the *E. coli* cytoplasm as insoluble inclusion bodies. These could only be solubilized in guanidine-HCl, and this meant that the recombinant protein would be completely denatured and would not retain any enzymatic activity. Fortunately the $(His)_6$ tag allowed affinity purification of the protein under denaturing conditions using a nickel chelating-NTA column, and this method proved to be simple, straightforward, and reproducible.

With highly purified but denatured H6/HCVP protein in hand (in 8 M urea), the next task was to attempt to re-fold it and perhaps restore enzymatic activity. Since no assay for enzymatic activity was available, the simple

assumption was made that once the protein was properly refolded, it would remain soluble in water or physiological saline at approximately neutral pH. Since the H6/HCVP protein contained 7 cysteines, proper formation of intra-chain disulfide bonds appeared critical for the refolding process. Several refolding protocols were attempted, including rapid 100-fold dilution into and extended mixing in 'refolding buffer' which contained reduced and oxidized glutathione (see Materials and Methods). Empirically it was discovered that dialysis into water containing 1 mM DTT at a certain rate resulted in completely soluble protein, whereas dialysis into other buffers or at faster rates did not. This enabled the production of purified soluble H6/HCVP protein in mg quantities, with yields of up to 30 mg/l of bacterial culture. The purified protein was sequenced from the N-terminus, which demonstrated that the expected H6/HCVP polypeptide had been generated.

The soluble H6/HCVP protein was also dialyzed into dilute phosphate buffered saline, dried down, and used to immunize mice, rats, and rabbits. It was reasonable to begin immunization with the first available purified HCV proteinase, since the generation of polyclonal immune antisera can take some time. The rationale for generating immune sera was as follows: even if the H6/HCVP protein proved not to be enzymatically active, and even if slightly different boundaries for the HCV serine proteinase were defined in the future, the immune sera would provide a valuable reagent for Western blotting and monitoring the expression and purification of any future related recombinant proteins, with or without affinity tags, and whether expressed in prokaryotic or eucaryotic systems.

The immunization experiments were in general surprisingly unsuccessful, since a group of mice and one of the rabbits showed virtually no

immune response after repeated immunizations. A second rabbit gave a very minimal response as well. Fortunately both rats responded considerably better, and one rat in particular yielded 8 ml of immune serum with a titer of 1:64,000. This result demonstrates the value of trying a number of different animals, and a number of different species, when trying to generate polyclonal immune serum. Clearly rats would be the animal of choice for any future immunization experiments with this or related recombinant HCV proteins.

Once soluble H6/HCVP protein was obtained in quantity, and immunization experiments were underway, the design of an assay for the HCV serine proteinase was seriously considered. During the first year of this project, the exact cleavage sites on the HCV polyprotein were not known. This meant specifically that the N-terminal boundary of the NS3 region (i.e. serine proteinase) was not defined, and in general the nature of the cleavage recognition site for the viral serine proteinase was unknown. However, when the nature of the cleavage sites suddenly became available from a conference report, it became apparent that the H6/HCVP protein was missing the 10 N-terminal amino acids present on the native HCV protein. Therefore it was decided to construct and purify a second recombinant protein, H6/HCVPfix, which would incorporate the 10 additional N-terminal amino acids, rather than expending effort at that time to develop a peptide based HCV serine proteinase assay to test the original H6/HCVP protein. It was also anticipated that with the experience gained with the first recombinant protein, the construction, purification, and refolding of a second protein would go much more quickly.

Other information which became available at the same time demonstrated that the NS2-NS3 junction was first cleaved *in cis* by a viral metalloproteinase present in the NS2 region, and suggested that this release of

the N-terminus of the NS3 region was a key even in activation of the viral serine proteinase. This idea and consideration of a series of other serine proteinases which are synthesized as inactive precursors and which are activated by cleavage at the N-terminus lead us to focus on a method to generate a recombinant protein with an N-terminus precisely the same as the wild-type viral proteinase. Thus the H6/HCVPfix construct was designed in such a way that the precursor could be affinity purified over an NTA-Ni^{2+} chelating column, and then cleaved with enterokinase--a relatively specific proteinase which cleaves immediately C-terminal to its recognition sequence--to generate the precise N-terminus of the viral protein (Figure 8).

The expression and purification of the H6/HCVPfix protein was done in an identical fashion as was the previous H6/HCVP protein. Again the protein was expressed in *E. coli* as insoluble inclusion bodies, and again denaturing agents were required during the purification process. Since it would have been highly desirable to have the recombinant protein expressed in a soluble form, attempts were also made at that time to express both the HCVP and HCVPfix proteins as fusions with *E. coli* thioredoxin. All constructs gave comparable yields, but unfortunately none of these proteins were soluble in *E. coli*, even when cultures were grown at lower temperatures, and these attempts were abandoned.

The final purified, denatured H6/HCVPfix protein could also be 'refolded' (i.e. result in a soluble protein) by dialysis against 1 mM DTT using the same protocol which had been successful for the H6/HCVP protein. Yields of the two proteins were also comparable. One rather surprising result was that after dialysis the H6/HCVPfix protein began to run as a doublet on SDS-PAGE. This raised the possibility that the refolded protein might be cleaving

itself, either *in cis* or most likely *in trans*. Examination of the expected amino acid sequence near the N-terminus of the protein and in the vicinity of the enterokinase cleavage site demonstrated that a HCV serine protease cleavage recognition sequence had indeed serendipitously been added in this region (Figure 8.), and that self-cleavage might be possible. To confirm this possibility, the two bands present in the doublet were separated on a long SDS-PAGE gel, transferred to a membrane, and subject to separate N-terminal sequencing. Contrary to expectation, both bands had exactly the same N-terminal sequence, suggesting that some factor other than length (e.g. folding and/or disulfide bond formation) was accounting for the difference in mobility between the two proteins. Another possibility is that the longer form represents a read-through of the first stop-codon, but this seems less likely since the original denatured protein ran as a single band.

To generate recombinant protein which had the same N-terminus as the wild-type HCV serine proteinase, the purified and refolded H6/HCVPfix protein was subjected to cleavage by enterokinase. This cleavage reaction would be expected to cleave the $(His)_6$ tag from the N-terminus of the protein, so that any cleaved protein would now run through a NTA-Ni^{2+} chelating column. The cleavage reaction appeared to be extremely inefficient, since the vast majority of the H6/HCVPfix protein still bound to the nickel column after overnight incubation with enterokinase (Figure 16, lane 2). The small quantity of protein which did flow through the column (Figure 16, lane 3) contains two major protein bands which run at about 15 and 21 KDa. Both bands are shorter than expected if the H6/HCVPfix protein had been cleaved at the enterokinase site, which would remove only 21 amino acids or about 2.3 KDa from the original protein. Perhaps the minute quantity of full-length enterokinase

cleaved protein which is present cannot be detected with the relatively insensitive Coomassie brilliant blue staining method. On the other hand, the calculated molecular weight of enterokinase cleaved H6/HCVPfix protein is about 20.5 KDa, and it may be that the larger band seen in lane 3 is the full-length recombinant enzyme. If this were the case, it would suggest that the N-terminal 21 amino acids present on the H6/HCVPfix protein caused a disproportionate decrease in mobility of the original protein. This question could be resolved by sequencing the N-terminus of the 21 KDa band seen in lane 3, but because of the small quantities available, in the present project all of this material was devoted to enzymatic assays.

The enterokinase cleavage buffer contained 4 M urea, and although recommended in some protocols, the presence of this denaturant might in part be responsible for the low activity of the enterokinase. Future experiments are planned which will test enterokinase cleavage in the absence of urea, and also determine the optimal enterokinase:substrate ratio, the optimal time of incubation, and the optimal commercial source for the enzyme.

Material from the enterokinase cleavage reaction which flowed through the nickel chelating column under denaturing conditions was refolded by dialysis against 1 mM DTT, 10 mM sodium acetate (pH measured: 6.3) and then concentrated. This was tested in a cleavage assay using a purified synthetic peptide which was modeled on the NS4A-NS4B junction. Rationale for the use of this particular peptide, and details of the proteinase assay are contained in the previous chapter. The TNBS assay suffered from high and sometimes fluctuating backgrounds, but in general demonstrated that the enterokinase cleaved H6/HCVPfix was the only protein tested which gave a continual increase in primary amino groups over a 23 hour period. This activity was

totally inhibited by PMSF (a classical inhibitor of serine proteinases). A control protein which was expressed and purified by the same method gave no activity, nor did enterokinase alone. As well, the initial velocity of the reaction (i.e. the rate of generation of primary amino groups) increased linearly with increasing concentrations of the enterokinase cleaved H6/HCVPfix protein.

When the assay reaction containing the substrate peptide and enterokinase cleaved H6/HCVPfix protein was analyzed on HPLC at t=0 and t=16 hours, a differential peak was observed in the t=16 hour material. When the molecules constituting this differential peak were analyzed by plasma desorption mass spectroscopy (PDMS), a mass corresponding precisely to the sodium and potassium adducts of the expected decapeptide cleavage product was obtained. This provided proof that the substrate peptide was being cleaved at the exact HCV serine proteinase recognition site. Although it is possible that another proteinase might also be catalyzing this cleavage, this would appear highly unlikely in the present set of experiments. The most likely possibility is that the enterokinase cleaved H6/HCVPfix protein preparation contains at least a small quantity of catalytically active recombinant HCV serine proteinase.

Although this project was successful in expressing a minimal quantity of enzymatically active recombinant HCV serine proteinase, obviously more work needs to be done to confirm these results. For example the differential peaks obtained after cleavage could also be analyzed by protein sequencing to confirm their identity. Antisera from patients with HCV, as well as our own rat polyclonal antisera could be tested for specific blocking of the enzyme activity. Perhaps even more importantly, the various experiments should be repeated once more substrate peptide becomes available. Since the final yields of

enterokinase cleaved proteins were quite poor, this cleavage reaction should be further optimized. Western blots could also be performed on the enterokinase cleaved material which flows through the nickel chelating column--using our rat antisera and sera from HCV patients--to confirm size and identity of the flow through products.

Since this work was started, HCV serine proteinase activity has been reported by other groups when the polyprotein residues 1027-1207 (57), 1049-1215 (101), or 1007-1237 (26) were expressed using eucaryotic systems. This information suggests that our original assignment of the C-terminal boundary of the HCV serine proteinase domain at polyprotein residue 1221 is quite reasonable. Furthermore, it also suggests that having the precise N-terminal boundary at residue 1027 may not be absolutely necessary for at least some protein activity.

In terms of mechanisms of activation after N-terminal cleavage, Brayer, Delbaere and James suggested that activation of chymotrypsin requires stabilization of the active site by the formation of a permanent salt-bridge with Asp 194, (adjacent to the active site serine 195; ref. 102). In this case, N-terminal cleavage of the zymogen apparently releases the N-terminal α-amino group of Ile 16 which forms the necessary salt bridge. However, in the case of the HCV serine proteinase there is a Ser residue adjacent to the active site serine, and therefore no similar salt bridge could be formed. This is at least consistent with the observation that active HCV serine proteinases have been expressed with a variety of N-termini. Since the H6/HCVPfix protein had only 21 additional residues (i.e. His_6, etc.) appended to the natural N-terminus of the viral proteinase, it is also possible that this protein is active. One preliminary experiment with the refolded H6/HCVP protein (data not shown)

demonstrated some proteinase activity, but this was perhaps 100-fold less than that obtained with the enterokinase cleaved H6/HCVPfix protein.

In the future, once additional substrate peptide is available, the original refolded H6/HCVP and refolded but uncleaved H6/HCVPfix protein can also be carefully tested for enzymatic activity, and compared to optimally cleaved H6/HCVPfix protein. Once any quantity of active enzyme has been produced and tested using the existing peptide substrate and TNBS assay, a new and more sensitive fluorescence assay--based on the cleavage of a substrate peptide with a fluorescence group on one end and a fluorescence quenching group on the other--should be rapidly developed. With the development of this assay and with quantities of refolded, active protein in hand additional enzymatic and kinetic studies can be carried out. These studies would include complete characterization in terms of *k*cat and *k*m. Attempts to demonstrate what proportion of the molecules were properly folded and enzymatically active could be made using suicide substrates as active-site titrants such as radio-labeled peptide chloroketon derivatives. Design of optimal peptide substrates using phage display or random peptide strategies, and construction and testing of peptide based or other inhibitors could then be undertaken.

The major limitation of the present *E. coli* expression system is that the recombinant HCV proteinase must be denatured and then eventually refolded to generate enzymatically active material. This process is very inefficient, and means that properly folded molecules will be contaminated with disordered proteins. Although such material could be used to screen proteinase inhibitors, it would not be suitable for structural studies such as X-ray crystallography. The use of other expression systems such as *Pichia pastoris*, baculovirus, or mammalian cells may yield soluble, active recombinant protein which would

be highly advantageous. This work would be facilitated by the rat polyclonal antisera which was generated in the present project. However, preliminary reports from Glaxo Inc. suggest that a similar recombinant NS3-derived protein is also insoluble when expressed in baculovirus. Therefore, use of the *E. coli* expression system described in the present thesis may still be a efficient way to proceed to produce active enzyme, particularly if cleavage and refolding protocols can be further optimized. Another straightforward possibility would be to try and express the protein with a prokaryotic signal sequence, either in *E. coli* or in another bacteria such as *Bacillus subtilis*. Should sufficient soluble protein be obtained, and providing that properly folded molecules could be purified away from the unfolded material, crystallization and X-ray diffraction analysis is also a real possibility. In the future, perhaps highly specific non-peptide or peptide-based inhibitors can be designed using the information obtained in the structural studies. On a more practical note, existing libraries of proteinase inhibitor molecules (peptide based or otherwise) can be screened immediately, even with relatively crude preparations of enzyme such as the one produced in the present thesis.

Table 1. HCV-encoded proteins and proposed functions.

Protein	Position in Polyprotein	Molecular Weight	Proposed Functions	Notes
C	aa 1- 191	21 kD	Nucleo-Capsid	RNA Binding
E1	aa 192 - 383	31 kD	Envelope	Glycoprotein
E2	aa 384 - 809	70kD	Envelope	Glycoprotein
NS2	aa 810 - 1026	23 kD	NS2-NS3 Metallo-proteinase	Zn^{2+} Denpendent
NS3	aa 1027 - 1657	70 kD	NS2-NS3 Metallo-proteinase + Ser Proteinase + Helicase	Polyprotein Processing
NS4A	aa 1658 - 1711	8 kD	?	Membrane Binding
NS4B	aa 1712 - 1972	27kD	?	Membrane Binding
NS5A	aa 1973 - 2420	58 kD	Polymerase (NS5)	Replicase Domain A
NS5B	aa 2421 - 3011	68 kD	Polymerase (NS5)	Replicase Domain B

* Apparent molecular weight by SDS-PAGE.

It is reported recently that a p7 protein (aa 747-809) was generated by incomplete cleavage, leading to the production of two E2-specific species: E2 and E2-p7 (Lin, et al., 1994).

Table 2. Proteinase recognition sites in HCV polyprotein known to be cleaved by HCV viral proteinases (from Gracoui *et al.*, 1993).

Cleavage Site	Position in Polyprotein	Cleavage Senquence# $P_6P_5P_4P_3P_2P_1\downarrow P'_1P'_2$	Proteinase Which Cleaves
NS2 - NS3	aa 1021 - 1029	KGWRLL↓AP*	Metalloproteinase
NS3 - NS4A	aa 1651 - 1658	DLEVVT↓ST*	Serine Proteinase
NS4A - NS4B	aa 1706 - 1714	DEMEEC↓SQ DEMEEC↓AS	Serine Proteinase
NS4B - NS5A	aa 1967 - 1974	DCSTPC↓SG	Serine Proteinase
NS5A - NS5B	aa 2415 - 2422	EDVVCC↓SM DSVVCC↓SM	Serine Proteinase

NOTE:

HCV serine proteinase has consensus (bold) cleavage sequence as:

$P_6P_5P_4P_3P_2P_1\downarrow P'_1P'_2$

D C S

E T A

* Cleavage in *cis* .

Table 3. Other sites in the HCV polyprotein which have the consensus serine Proteinase recognition sequence but which are not cleaved.

CONSENSUS SEQUENCE for HCV Serine Proteinase		$P_6P_5P_4P_3P_2P_1\downarrow P'_1P'_2$ D C S E T A
SIT E	POSITION	SEQUENCE $P_6P_5P_4P_3P_2P_1P'_1P'_2$
NS5A #1	aa 2167 - 2174	**DVAVLTSM**
NS5A #2	aa 2177 - 2184	**DPSHITAE**
NS5A #3	aa 2337 - 2344	**ESTLSTAL**
NS5B #1	aa 2779 - 2786	**DLELITSC**
NS5B #2	aa 2781 - 2788	**ELITSCSS**

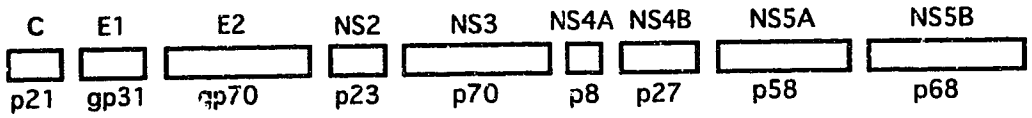
NOTE: Amino acids residue in consensus position are in bold.

Table 4. Rate at which glycine-equivalent primary amino groups are released from the substrate peptide in the presence of refolded HCV serine proteinase (lane 3, Figure 16) under two different buffer conditions.

Buffer Conditions	Velocity (μM / min.) *
Potassium Phosphate (pH 7.5)	1.12 ± 0.05
HEPES-Buffered Saline (pH 8.2)	0.89 ± 0.05

* Calculated after converting the absorbance of TNP-peptide at 405 nm into glycine equivalents using the Linear Regression Pro- computer software written by Gerard Hammond, 1993. The reaction contained 0.19 μM enzyme, 2.5 mM peptide substrate, in 0.1 M potassium phosphate, 6 mM citrate, and 2 mM EDTA (pH 7.5). It was incubated at 23°C and sampled at 2, 3, 4, 5, 6, and 23 hrs. Incubation of buffer plus substrate alone, or enzyme alone, gives nearly 0 velocity under identical condition.

A. HCV Polyprotein



B. NS3 Region & Constructs

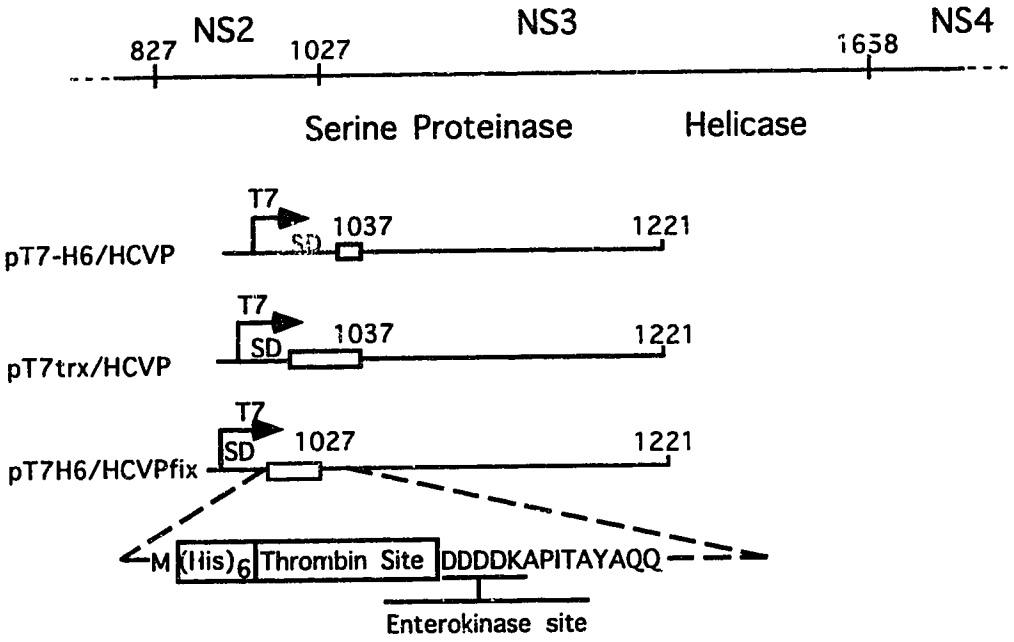


Figure 1. Schematic representation of the HCV polyprotein cleavage products, and relationship to the recombinant proteins expressed in this study. A. The virus polyprotein cleavage products are shown as open boxes; the name of each protein and its molecular size are indicated. B. The NS3 region and the expression constructs are shown . The name of each plasmid is shown on the left. The T7 promoter is shown as an arrow; the ribosome binding/ Shine-Delgarno sequence indicated by SD. The fusion peptides, either 6 histidines and a thrombin cleavage site or E. coli thiordoxin A and a enterokinase cleavage site, are shown as open box. The amino acid residues are numbered as they are in HCV polyprotein. The amino acid sequence from 1027 to 1221 is APITAYAQ .....SSPPAVPQ, and that from 1037 to 1221 is RGLLGCII.....SSPPAVPQ.

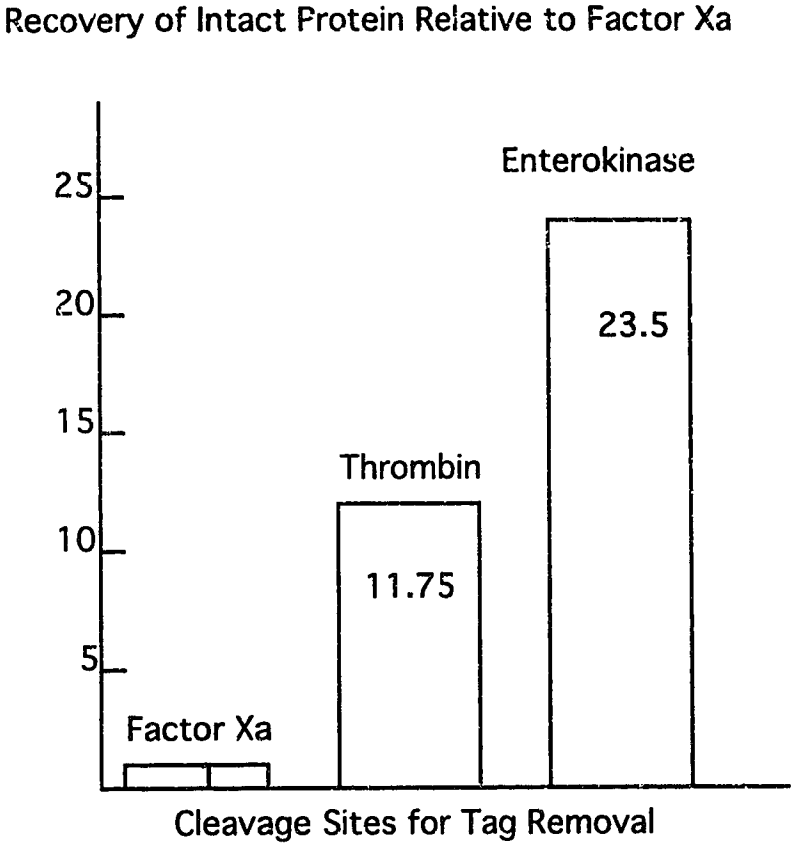


Figure 2. The probability of recovery of an intact protein after ceavage of affinity tag by enterokinase is 23.5 times greater than with Factor Xa, and 2 times greater than with thrombin (data from the IBI FLAG™ Biosystem, Kodak, New Haven, CT).

```
HPCPLYPRE 3305 AGATACCGCCGCGTGCGGTGACATCATCAACGGCTTGCCtGTtTCcGCCCGcAGG·····
  HPCCGAA 3305 AGATACCGCCGCGTGCGGTGACATCATCAACGGCTTGCCCGTCTCtGCCCGtAGG·····
   HPCJTA 3305 GGACACCGCGGCGTGTGGGGACATCATCTTGGGACTGCCCGTCTCCGCCCGAAGG·····
   HPCJTB 3305 GGACACCGCGGCGTGTGGGGACATCATCTTGGGACTGCCCGTCTCCGCCCGAAGG·····
   HPCJCG 3293 aGACACCGCGGCGTGTGGGGACATCATCTcGGGtCTaCCaGTCTCCGCCCGAAGG·····
   HPCJ8G 3317 tGAgACaGTGGCGTGTGGaGACATCcTgcatGGcCTcCCgGTCTCCGCgaGgcta·····
 consensus      aGAcACcGcgGCGTGtGGgGACATCaTctagGGccTgCCcGTcTCcGCccGaagg·····
```

5'-ACCGCGGCGTGTGGGGACATCATC-3'

Oligonucleotides Primer (5'): HCV-NS3-B

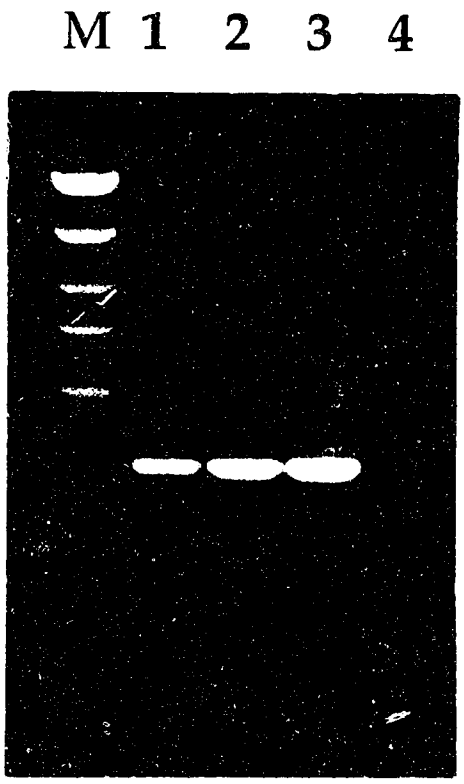
```
·····CtCACCTcCATGCT·CCACaGGCAGCGGcAAaAGCACCAAGGTCCCGGCTGCaTAt 4074 HPCPLYPRE
·····CCCACCTgCATGCTCCCACcGGCAGCGGtAAGAGCACCAAGGTCCCGGCTGCGTAC 4074 HPCCGAA
·····CCCATCTACACGCTCCCACTGGCAGCGGCAAGAGCACTAAAGTGCCGGCTGCGTAC 4074 HPCJTA
·····CCCATCTACACGCTCCCACTGGCAGCGGCAAaAGCACTAAAGTGCCGGCTGCGTAC 4074 HPCJTB
·····CaCATTTACACGCTCCCACTGGCAGCGGCAAGAGCACCAAAGTGCCGGCTGCaTAT 4062 HPCJCG
·····gttAcTTgCACGCaCCaACaGGCAGCGGaAAGAGCACCAAgGTcCCtGCcGCgTAT 4086 HPCJ8G
·····cccA-cTaCAcGCtCCcACtGGCAGCGGcAAgAGCACCcAA-GT-CCgGCtGCgTA-     consensus
```

3'-CCGTCGCCGTTCTCGTG$^{G}_{A}$TT-5'

Oligonucleotides Primer (3'): HCV-NS3-2A

Figure 3. Viral cDNA sequences flanking the putative HCV serine proteinase domain. The identity of the HCV strains and the position in their genome (first and the last bases inclusive) are listed next to each sequence. The relative consensus regions from which the PCR cloning primers were designed is underlined, and the actual primers used are listed below this. cDNA position 3305-3307 corresponds to residue 988 in the polyprotein, and cDNA position 4074 corresponds to residue 1244 in the polyprotein.

Figure 4. Non-denaturing agarose gel electrophoresis of RT-PCR products stained with ethidium bromide and photographed under uv illumination. cDNA fragments encoding the putative hepatitis C virus serine proteinase gene were amplified originally from viral RNA isolated from the serum of three different HCV patient: PG, ZA, and BF (lanes 1, 2, 3), and these resulted in clones #3 (747 bp), #5 (748 bp) and #6 (748 bp) respectively. Lane 4 is the negative control without the target cDNA. Lane M contains DNA size markers of 4132, 2835, 1942, 1543, 1276, 1043, 812, 636, 525, 413, and 400 base pairs (top to bottom). The markers were derived from pCDM8 (Brian Seed) which has been digested in separate reactions with *Sty* I, *Pvu* II, *Xho* I, and *Hae* III and then pooled following inactivation of restriction enzymes by heating (70°C, 15 min.) in 25 mM EDTA.


M 1 2 3 4

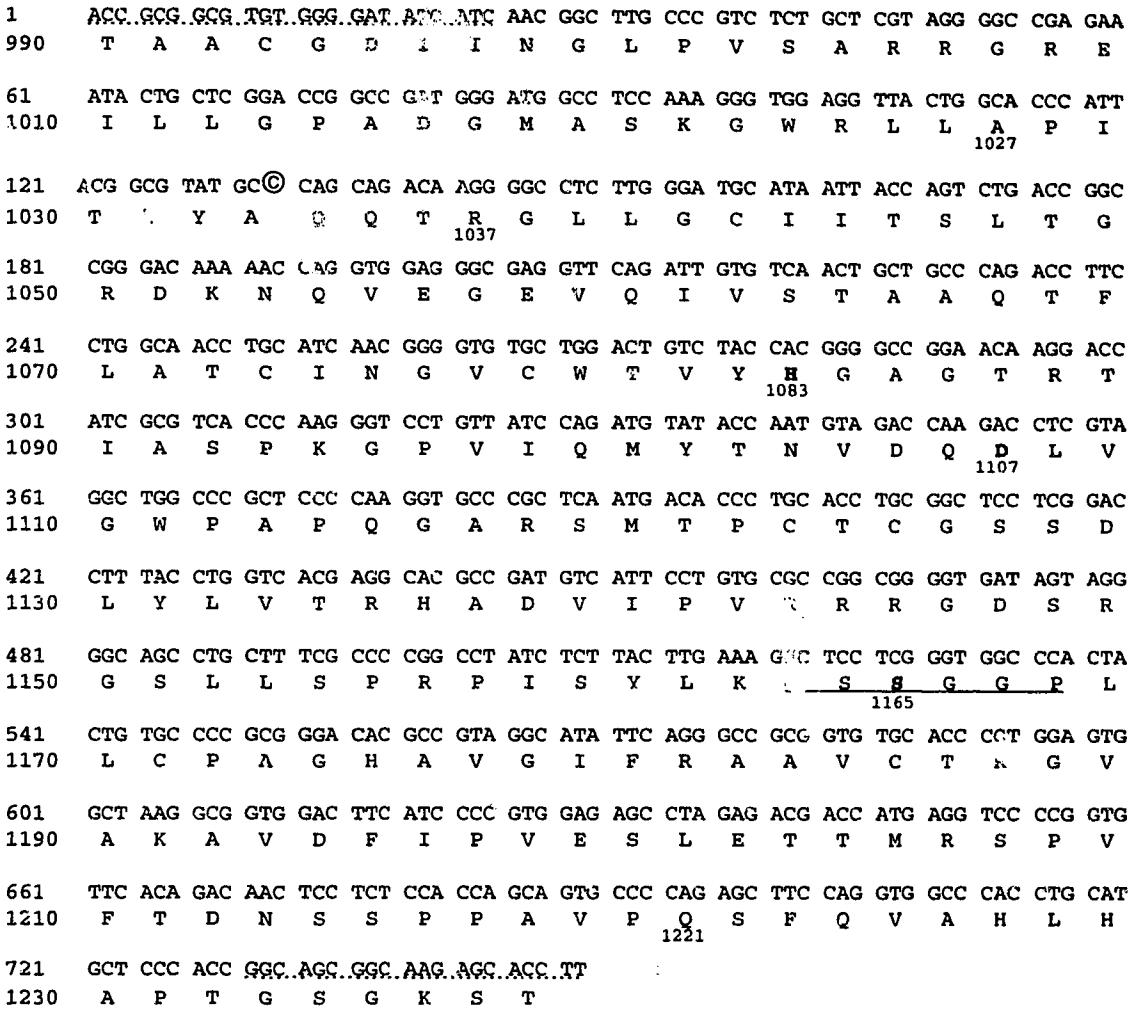


Figure 5. Complete nucleotide sequence and deduced amino acid sequence of the clone #6 cDNA insert (Figure 4). A single dC deletion was found at nucleotide 132, which is added as a circled C in order to maintain the open reading frame. Therefore the entire DNA sequence presented here totals 749 instead of 748. DNA sequences contributed by the PCR primers are underlined with dotted lines. The putative HCV serine proteinase catalytic triad includes histidine 1083, aspartate 1107, and serine 1165, all shown in bold. The GXSGXP motif around the active site serine residue is underlined. The amino acid numbers correspond to the residue number in the HCV polyprotein.

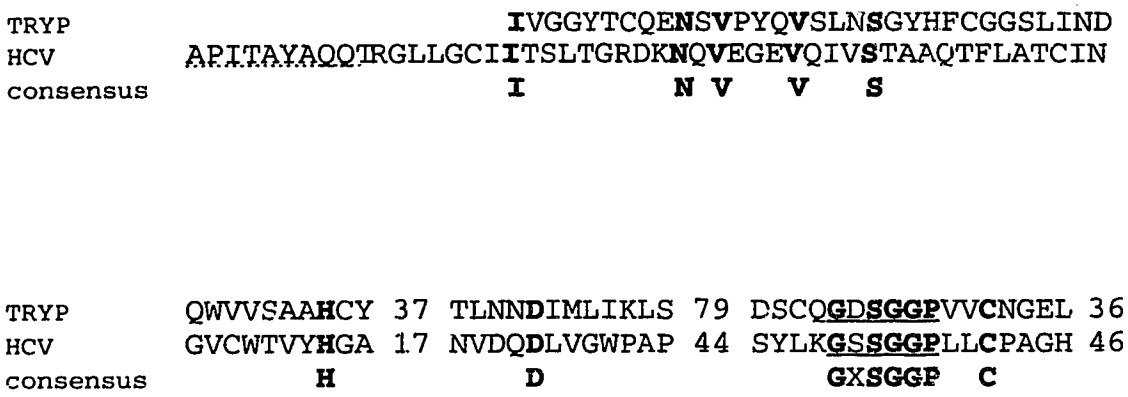


Figure 6. Comparison of the protein sequence of rat pancreatic trypsin II with that of the HCV serine proteinase, using the N-terminus of the HCV serine proteinase molecule published during the course of this work (Grakoui *et al.*, 1993). This N-terminus was used in the final construct pT7-H6/HCVPfix, whereas the N-terminal 10 amino acid residues (underlined with dots, ending in QQT) are absent in constructs pT7-H6/HCVP and pT7trx/HCVP. The N-terminal A corresponds to residue 1027 in the HCV polyprotein. The residues of the catalytic triad found in both proteinases are in bold, and the GXSGXP motifs surrounding the active site serines are underlined. Numbers between the blocks of sequence indicate the number of amino acids present in that region of the polypeptide; numbers at the end indicate the number of residues to the real (trypsin) or assumed (HCV) C-terminus.

A.

```
cat ATG GCT AGA ATT CGC GCC CGG GGA TCC TCT AGA GTC GAC CTG CAG CCC AAG CTT
    M   A   R   I   R   A   R   G   S   S   R   V   D   L   Q   P   K   L

ATC GAT GAT AAG
I   D   D   K
```

B.

```
cat ATG CAC CAC CAC CAC CAC CAC CTG GTT CCG CGT GGT TCC GGA ATT CGC GCC CGG
    M   H   H   H   H   H   H   L   V   P   R   G   S   G   I   R   A   R

GGA TCC TCT AGA GTC GAC CTG CAG CCC AAG CTT ATC GAT GAT AAG
G   S   S   R   V   D   L   Q   P   K   L   I   D   D   K
```

C.

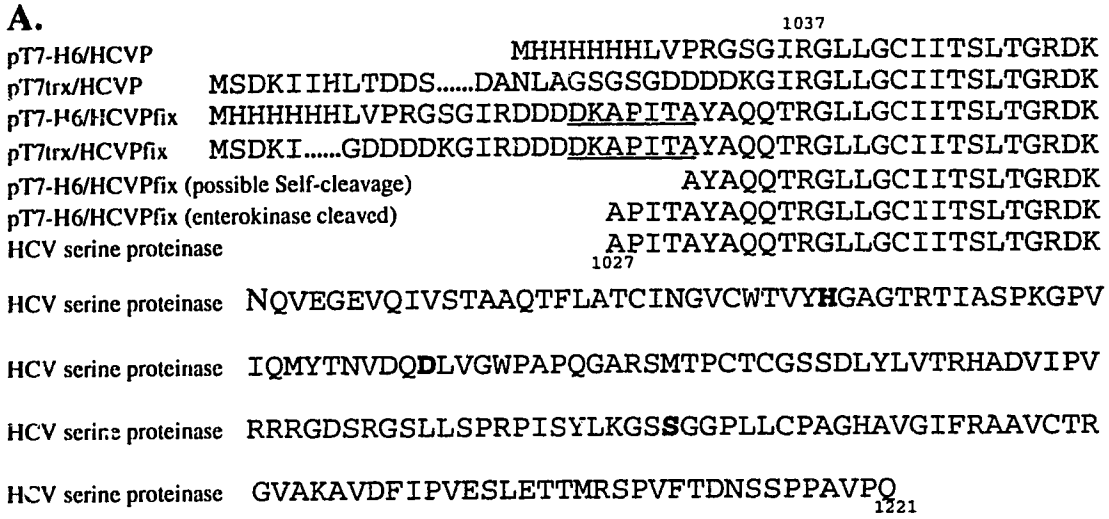
```
cat ATG AGC GAT AAA ATT ATT CAC CTG ACT GAC GAC AGT TTT GAC ACG GAT GTA CTC AAA
    M   S   D   K   I   I   H   L   T   D   D   S   F   D   T   D   V   L   K

GCG GAC GGG GCG ATC CTC GTC GAT TTC TGG GCA GAG TGG TGC GGT CCG TGC AAA ATG ATC
A   D   G   A   I   L   V   D   F   W   A   E   W   C   G   P   C   K   M   I

GCC CCG ATT CTG GAT GAA ATC GCT GAC GAA TAT CAG GGC AAA CTG ACC GTT GCA AAA CTG
A   P   I   L   D   E   I   A   D   E   Y   Q   G   K   L   T   V   A   K   L

AAC ATC GAT CAA AAC CCT GGC ACT          AAA TAT GGC ATC CGT GGT ATC CCG ACT CTG
N   I   D   Q   N   P   G                    Y   G   I   R   G   I   P   T   L

CTG CTG TTC AAA AAC GGT GA                   AAA GTG GGT GCA CTG TCT AAA GGT CAG
L   L   F   K   N   G                        K   V   G   A   L   S   K   G   Q

TTG AAA GAG TTC CTC                          TCT GGT TCT GGT GAT GAC GAT GAC AAG
L   K   E   F   L                            S   G   S   G   D   D   D   D   K

GGA ATT CGC GCC CGG                        G CAG CCC AAG CTT ATC GAT GAT AAG
G   I   R   A   R                            Q   P   K   L   I   D   D   K
```

Figure 7. DNA sequence and the deduced protein sequence in the polylinker of expression vectors which were used in this project. pT7-7(**A**), pT7-7His6 (**B**), and pT7-7trx (**C**) are included. Unique restriction sites are underlined and occur in the following order: *Nde* I, *Eco* RI, *Sma* I, *Bam* HI, *Sal* I, *Pst* I, *Hind* III, and *Cla* I.

Figure 8. A. Amino acid sequences of the various recombinant proteins expressed in the course of this project. The name of each plasmid construct from which the protein was made is indicated on the left. The putative catalytic triad of H1083, D1107, and S1165 are indicated in boldface type. The sequence of a possible HCV serine proteinase cleavage motif (Grakoui *et al.* 1993) which was serendipitously added to the pT7-H6/HCVPfix and pT7trx/HCVPfix constructs is underlined. The proteins which would result from possible self cleavage at this site, or from cleavage at the enterokinase site are also indicated. Except for the different N-terminal sequences, all proteins would be the same and hence are indicated by a single sequence labeled "HCV serine protease'.

B. DNA sequences and the deduced amino acid sequence at 5' end and 3' end of expression constructs including pT7H6/HCVP, pT7trx/HCVP, pT7H6/HCVPfix, and pT7trx/HCVPfix. Restriction sites *Eco* RI and *Sal* I used to clone the cDNAs into the expression vectors are underlined.

A.



B.

pT7-H6/HCVP 5'·····CCG CGT GGT TCC GGA ATT CGG GGC CTC TTG GGA ·····TCC TCT CCA
P R G S G I R G L L G S S P

CCA GCA GTG CCC CAG tga gtc gac ctc -3'
P A V P Q Stop

pT7trx/HCVP 5'·····TCT GGT TCT GGT GAT GAC GAT GAC AAG GGA ATT CGG GGC CTC TTG
S G S G D D D D K G I R G L L

GGA·····TCC TCT CCA CCA GCA GTG CCC CAG tga gtc gac ctc -3'
G S S P P A V P Q Stop

pT7-H6/HCVPfix 5'·····CCG CGT GGT TCC GGA ATT CGC GAC GAC GAT GAC AAG GCA CCC ATT
P R G S G I R D D D D K A P I

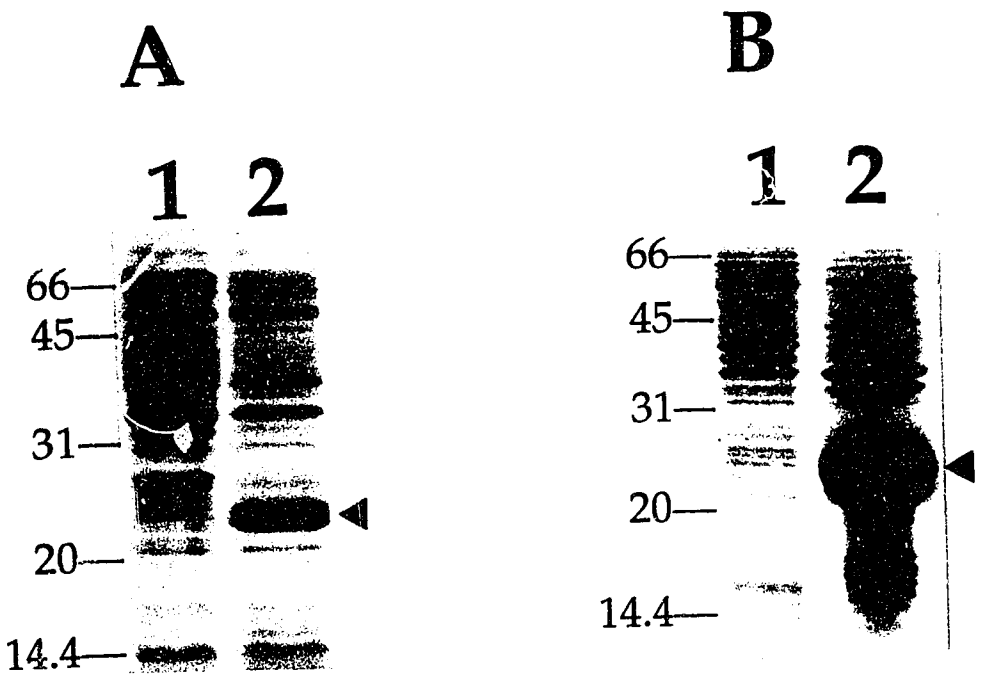
ACG G·····TCC TCT CCA CCA GCA GTG CCC CAG tga gtc gac ctc -3'
T S S P P A V P Q Stop

pT7trx/HCVPfix 5'···GGT TCT GGT GAT GAC GAT GAC AAG GGA ATT CGC GAC GAC GAT GAC
G S G D D D D K G I R D D D D

AAG GCA CCC ATT ACG G·····TCC TCT CCA CCA GCA GTG CCC CAG tga
K A P I T S S P P A V P Q Stop

gtc gac-3'

Figure 9. Coomassie brilliant blue stained SDS-PAGE of recombinant HCV serine proteinase expressed from construct pT7-H6/HCVP in *E. coli* BL21(DE3). A. Total cell lysate from an IPTG induced culture (lane 2) is compared with that from a non-induced culture (lane 1). B. Soluble supernatant (lane 1) is compared with in-soluble pellet (lane 2) after the cell lysis using a French press. The recombinant HCV serine proteinase is indicated by a arrowhead. Molecular weight standards (shown left) are in kilodaltons.


A
1
2
66
45
31
20
14.4
B
1
2
66
45
31
20
14.4

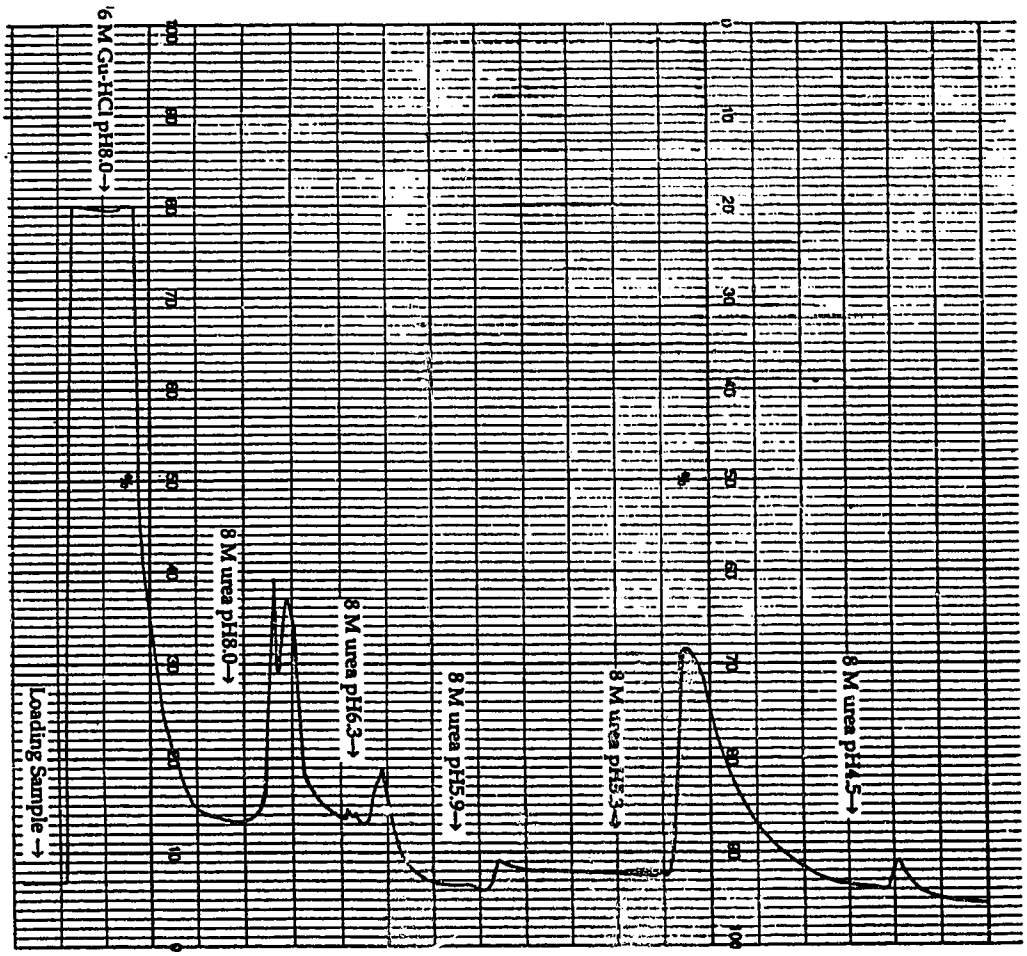
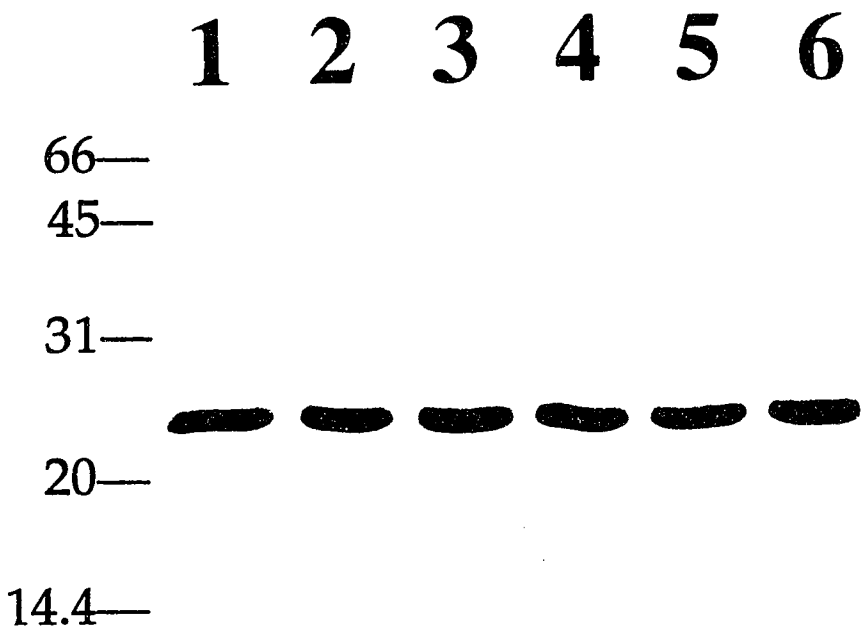


Figure 10. Affinity purification of recombinant protein H6/HCVP by nickel chelate column chromatography. This recombinant protein with a $(His)_6$ tag was purified over a Ni^{2+}-NTA column under denaturing conditions. The starting point of the loading, washing, and eluting events are indicated. The inclusion bodies containing the recombinant protein were dissolved in 6 M guanidine-HCl, 0.1 M NaH_2PO_4, 0.01 M Tris, 1 mM β-mercaptoethanol (pH8.0) (Gu-HCl buffer) and loaded onto the Ni^{2+}-NTA column which had been prewashed and equilibrated with Gu-HCl buffer. The column was washed first with Gu-HCl buffer, then with 8 M urea, 0.1 M NaH_2PO_4 , 0.01 M Tris, 1 mM β-mercapto-ethanol (pH8.0) followed by the same buffer but with the pH adjusted to 6.3 using HCl. The recombinant protein was eluted with 8 M urea, 0.1 M NaH_2PO_4 , 0.01 M Tris, 1 mM β-mercaptoethanol at the following pHs: 5.9, 5.3 and 4.5.

Figure 11. Protein purified by Ni^{2+}-NTA affinity chromatography. Protein expressed from the construct pT7-H6/HCVP was eluted from the nickel-chelating resin at low pH. Fractions 20 to 25 eluted at pH5.9 correspond to lanes 1 to 6 above. The protein was dialyzed, lyophilized, and used to immunize animals. Molecular weight standards (left) are shown in kilodaltons.


1 2 3 4 5 6
66—
45—
31—
20—
14.4—

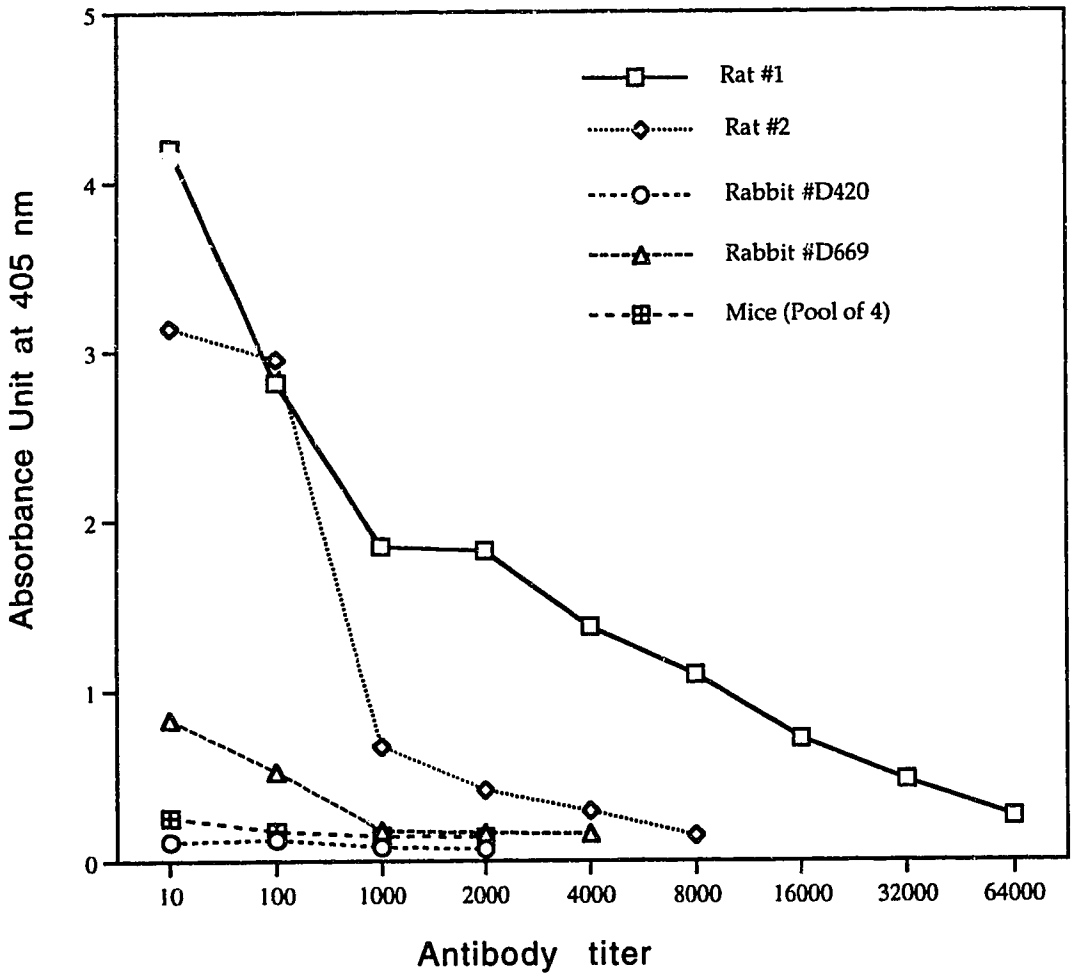


Figure 12. Antibody titers for anti-H6/HCVP antisera raised in a variety of laboratory animals. Animals were immunized with the purified H6/HCVP protein (Figure 10) dialyzed against water, and using complete and incomplete Freund's adjuvants as discribed in the Materials and Methods. Antibodies were measured by ELISA assay, using plates coated with H6/HCVP protein. The secondary antibodies used in this assay are goat anti-rat, rabbit, or mouse conjugated to peroxidase. This reacts with ABTS/H_2O_2 to give a green product that absorbs at 405 nm. Plates were measured on a ELISA plate reader and are expressed in OD_{405} units.

Figure 13. Coomassie brilliant blue stained SDS-PAGE of expression trial from construct pT7-H6/HCVPfix in *E. coli* BL21(DE3). Cells were pelleted and lysed by the freeze-and-thaw method, and the total lysate from an IPTG induced culture (lane 1) was compared with that from a non-induced culture (lane 2). The recombinant protein is indicated by the arrow head. Position of molecular weight standards are shown at left, with weights given in kilodaltons.

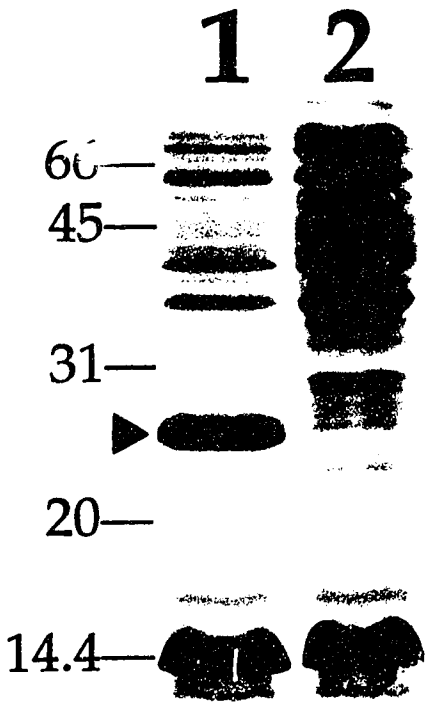

1
2
60
45
31
20
14.4

Figure 14. Coomassie brilliant blue stained SDS-PAGE of recombinant HCV serine proteinase released from the cell pellet after lysis using a French press. The recombinant protein expressed by the construct pT7-H6/HCVPfix in *E. coli* BL21(DE3) was found in the pellet (lane 2, arrowhead), and not in the supernatant (lane 1). Molecular weight standards (left) are shown in kilodaltons.

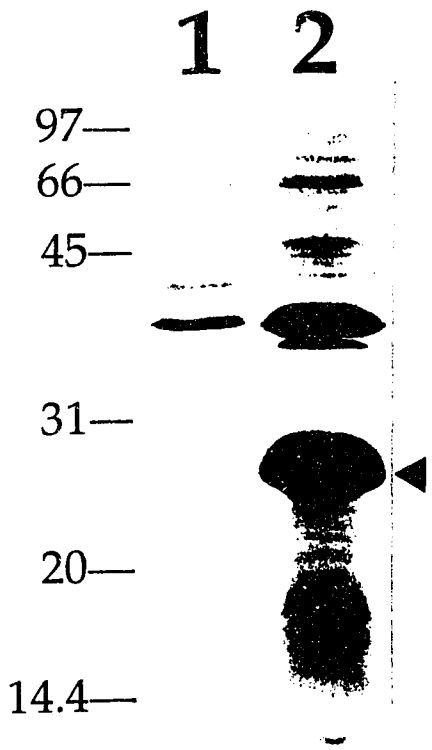

1
2
97—
66—
45—
31—
20—
14.4—

Figure 15. The recombinant $(His)_6$ fusion protein produced from the construct pT7-H6/HCVPfix was purified by nickel-chelating affinity chromatography as in Figure 9. Fractions which eluted at pH5.9 (lanes 1 to 4), pH5.3 (lanes 5 to 7), and pH4.5 (lanes 8 to 9) are shown. Molecular weight standards (left) are in kilodaltons.

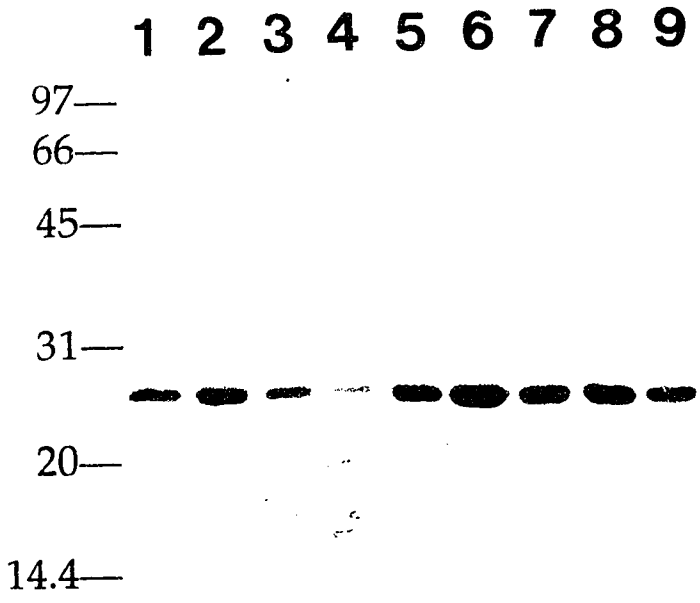

1 2 3 4 5 6 7 8 9
97—
66—
45—
31—
20—
14.4—

Figure 16. Coomassie brilliant blue stained SDS-PAGE of recombinant HCV serine proteinase processed by enterokinase. Purified recombinant protein produced from construct pT7-H6/HCVPfix (Figure 15) was refolded by dialysis against 1 mM DTT. 2 mg of refolded, lyophilized, recombinant protein was digested by incubation at 37°C overnight with 100 units of enterokinase (Sigma) in 10 ml of 50 mM sodium acetate pH5.0, 15 mM $CaCl_2$, 0.1% triton X-100, and 4 M urea. Lane 1 shows an aliquot of the starting material, and lane 2 shows the same material following incubation with enterokinase. The reaction mixture was brought to a final concentration of 8 M urea and 2 mM DTT, and the $(His)_6^+$ protein /peptide was removed by nickel-chelate affinity column (about 1 ml resin bed). The purified HCV serine proteinase (lane 3) was dialyzed against 1 mM DTT for 4 hours, 0.2 mM DTT overnight, 0.05 X PBS for 10 hours, and finally 0.02 X PBS overnight. The dialyzed proteinase was concentrated 2 fold by centricon-10 spin column (Amicon) and then used in the proteolysis assay.

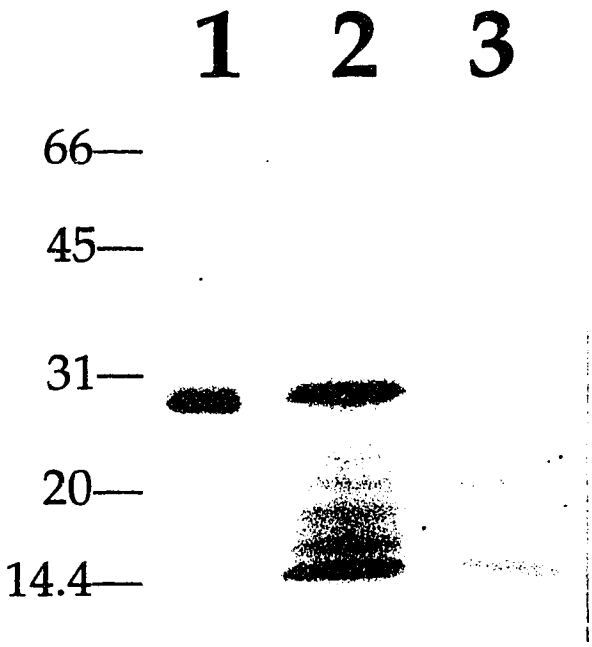

1 2 3
66—
45—
31—
20—
14.4—

Figure 17. Coomassie brilliant blue stained SDS-PAGE of total cellular proteins from non-induced (lane 1) and IPTG induced (lanes 2 and 3) cultures grown at various temperatures. Protein was expressed from the construct pT7trx/HCVPfix in BL21(DE3) cells, and cells were grown at the following temperatures: (A) at 23°C, (B) 30°C, and (C) 37°C. Almost no recombinant protein is expressed at 23°C (A), whereas high levels of expression are seen at 37°C (C, lane 2, arrowhead). Growth and expression at 30°C (B) gives intermediate level of expression, with most of the recombinant protein in the insoluble pellet (lane 3, arrowhead), and almost none in the supernatant (lane 2). Osmotic shock treatment (LaVallie *et al.*) was used to release the protein from the cells in (B). Molecular weight standards (left of each gel) are in kilodaltons.

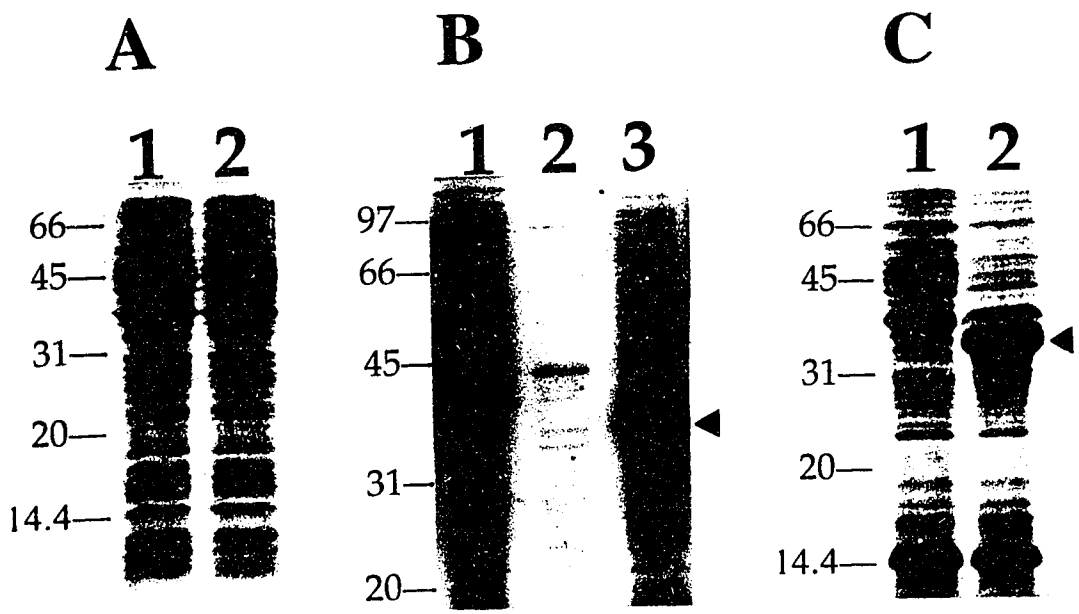
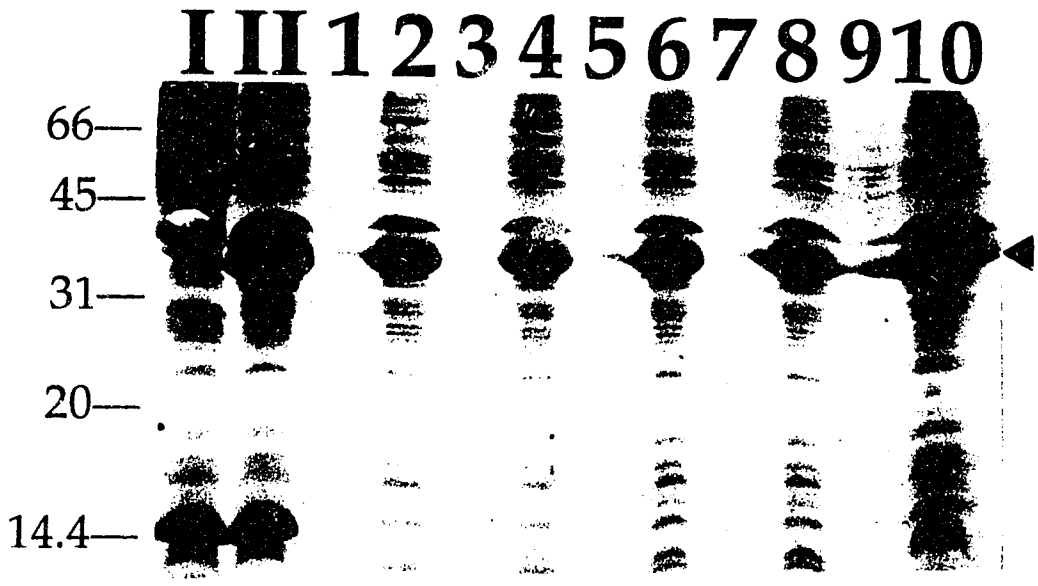

A
1 2
66
45
31
20
14.4
B
1 2 3
97
66
45
31
20
C
1 2
66
45
31
20
14.4

Figure 18. Coomassie brilliant blue stained SDS-PAGE of proteins from pT7trx/HCVPfix expressed in *E. coli* BL21(DE3), with cells lysed by osmotic shock in a variety of different solutions. The final supernatant (lanes 1, 3, 5, 7, and 9), and pellet fractions (lanes 2, 4, 6, 8, and 10) run along with the original total cell lysate with (lane II) and without (lane I) induction by IPTG. The osmotic-shock solutions tested include: 20 mM Tris-HCl, pH7.8 (lanes 1 , 2), 20 mM Tris-HCl, pH 8.0 (lanes 3, 4), TE pH8.0 (lanes 5, 6), 20 mM Tris-HCl, pH 8.55 (lanes 7, 8), and pure H_2O (lanes 9, 10). *Note*: None of the solutions contained additional salts or detergents. In all cases virtually all of the recombinant thioredoxin/HCVP fusion protein (Position indicated by arrow head) remained in the insoluble pellet. Molecular weight standards (left) are shown in kilodaltons.


I II 1 2 3 4 5 6 7 8 9 10
66—
45—
31—
20—
14.4—

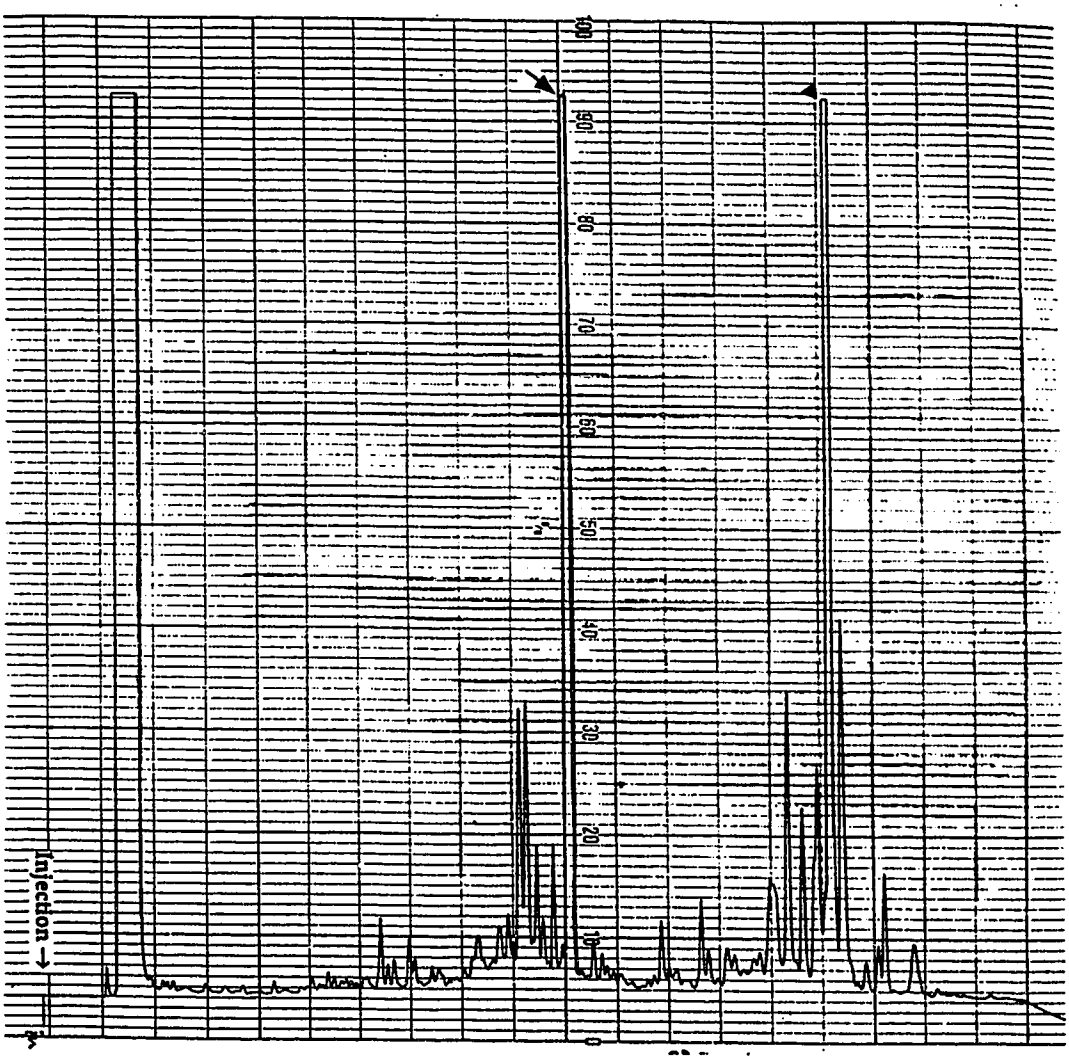


Figure 19. Purification of the synthetic peptide Ac-Tyr-Gln-Glu-Phe-Asp-Glu-Nle-Glu-Glu-Cys-Ser-Gln-His-Leu-amide by HPLC. The products of peptide synthesis, deprotection, and cleavage were purified over a C-18 reverse phase column (250 mm X 1/2") at a flow rate of 2 ml/min. using a 15-70% acetonitrile gradient (0.5%/min.) containing 0.05% TFA. The major product peak is indicated by the arrow. Mass spectroscopy analysis assigned the peak a molecular weight of 1811.3, which is in agreement with the calculated molecular weight of 1811.9. The second large peak (arrowhead), representing ~40% of the product , has a mass of 2055.

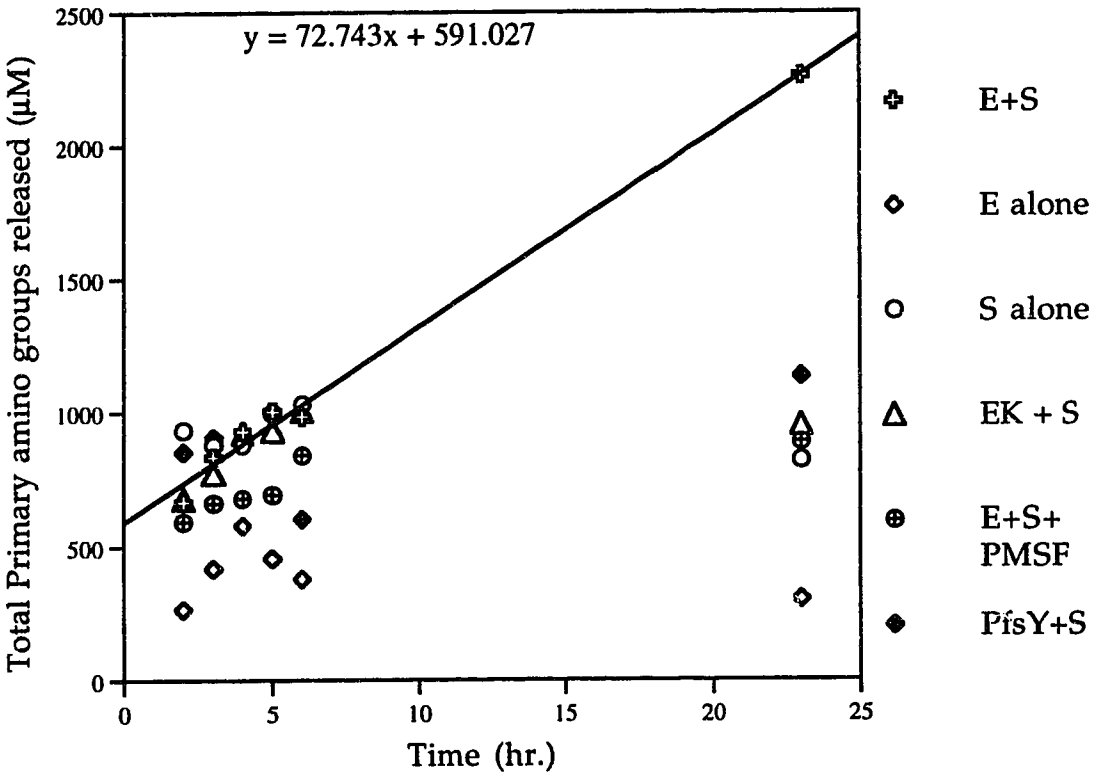


Figure 20. Generation of primary amino groups by the enterokinase cleaved, refolded H6/HCVPfix protein (protein shown in lane 3, Figure 16). 0.19 μM of refolded protein was incubated with 2.5 mM substrate peptide in 0.1 M potassium phosphate, 6 mM citrate, 2 mM EDTA (pH7.5) with (⊕) or without (✚) the inhibitor PMSF (350μg/ml). Controls include substrate alone (O) and enzyme alone (◇) incubated in the same buffer system over the same time period. Also included are enterokinase alone (Δ), and a recombinant malaria control protein PfsY (♦) (see text), both incubated with the peptide substrate. Aliquots of the reaction were taken at various times, and the accumulation of primary amino groups was assayed by conversion of TNBS to a yellow compound TNP-peptide, which absorbs at 405 nm (Satake et al. 1959). Absorbance at 405 was converted to glycine equivalents of total primary amino groups released, using a standard curve generated under the same assay conditions.

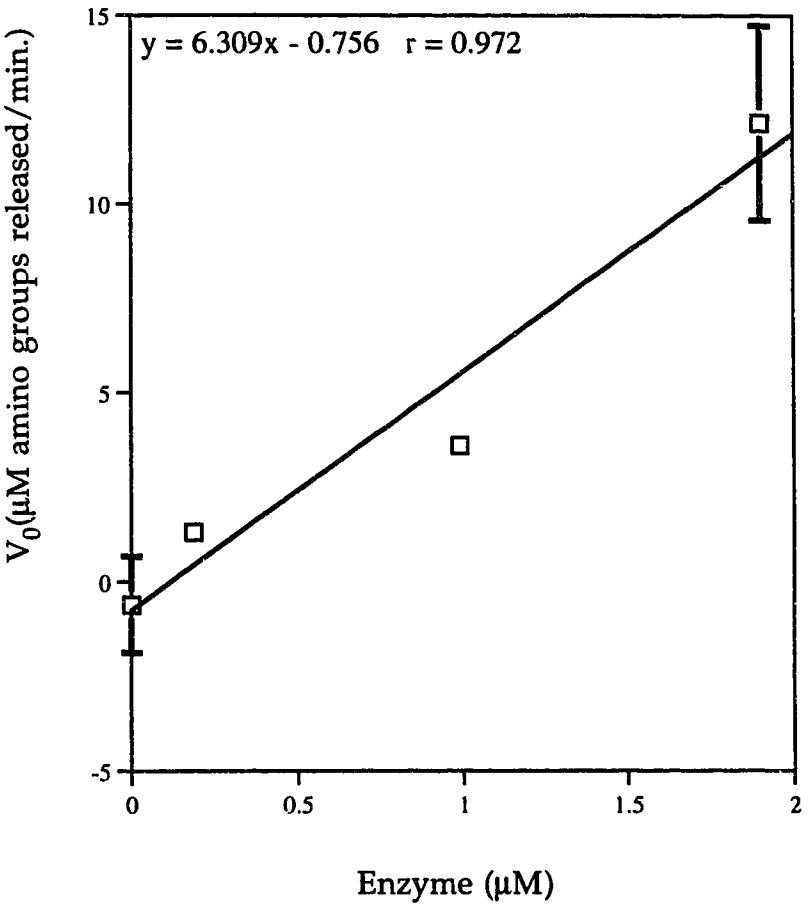
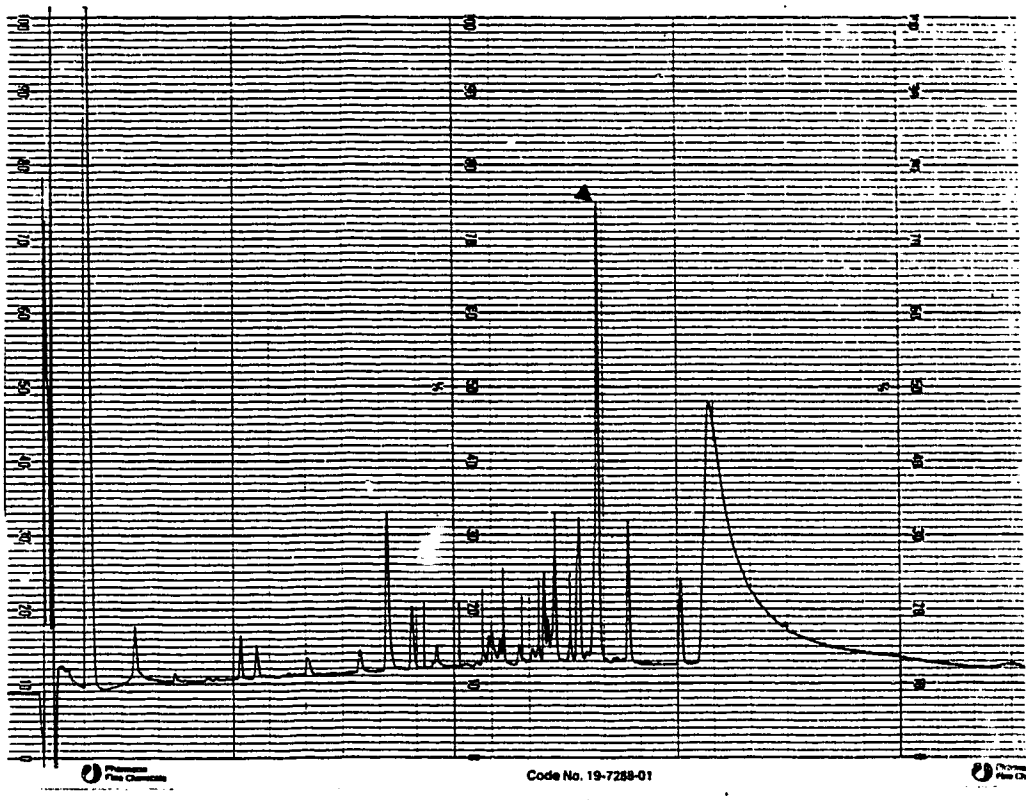
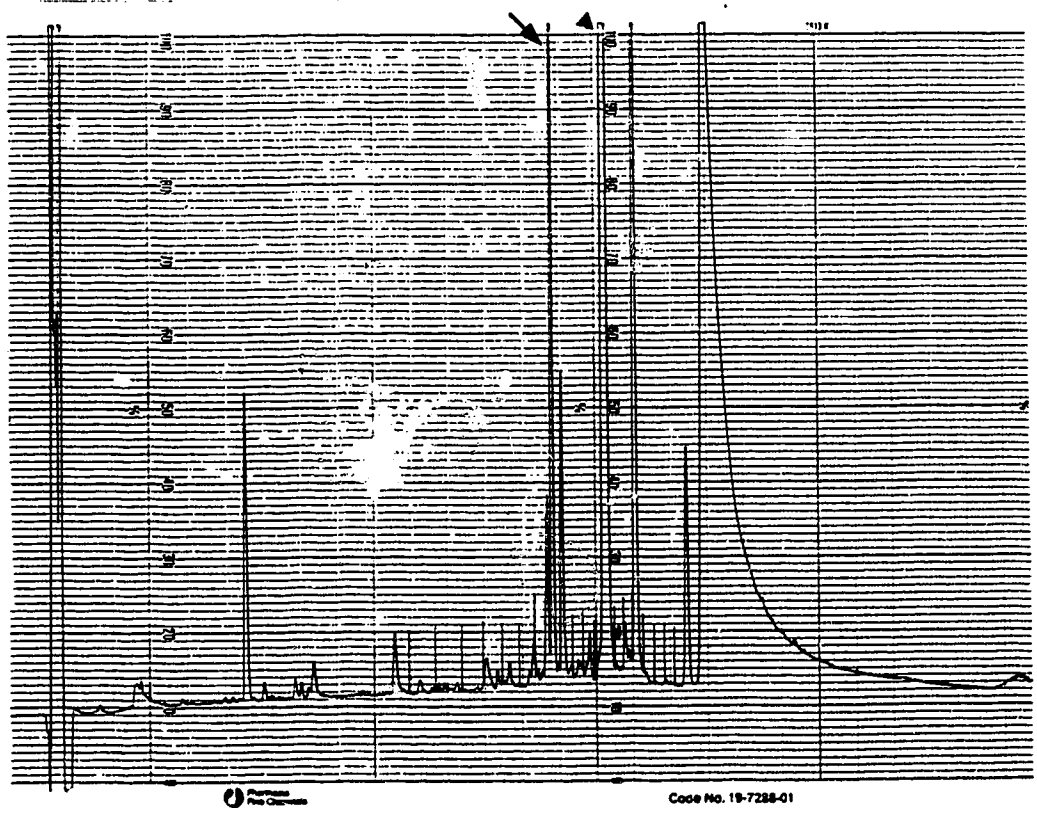


Figure 21. Initial velocity at which primary amino groups are generated when incubating the substrate peptide with various concentrations of the enterokinase cleaved, refolded protein H6/HCVPfix. The reaction conditions are as in Figure 20. Aliquots of the reaction were taken at time 0, 10, 20, 30, 40, 60, and 80 min. and assayed as in Figure 20. Velocity was assumed to be linear over this short time period.

Figure 22. HPLC profile of the proteolyzed substrate peptide (bottom) compared with that of HCV serine proteinase assay mixture at t=0 (top). Enterokinase cleaved, refolded protein H6/HCVPfix (lane 3, Figure 16) was incubated at 37°C at a concentration of 0.067 μM. The reaction included 1.6 mM peptide substrate in 0.1 M potassium phosphate, 6 mM citrate, 2 mM EDTA (pH7.5). Separation of the reaction mixtures at t=0 and t=16 hrs. were done over a C-18 column (same as in Figure 18) t a flow rate of 2 ml/min. and using a 0-70% acetonitrile, 0.05% TFA gradient (0.5%/min.). The major differential peak generated by proteolysis is indicated by the arrow. The substrate peak present in both traces is indicated by the arrowhead.


Code No. 19-7288-01


Code No. 19-7288-01

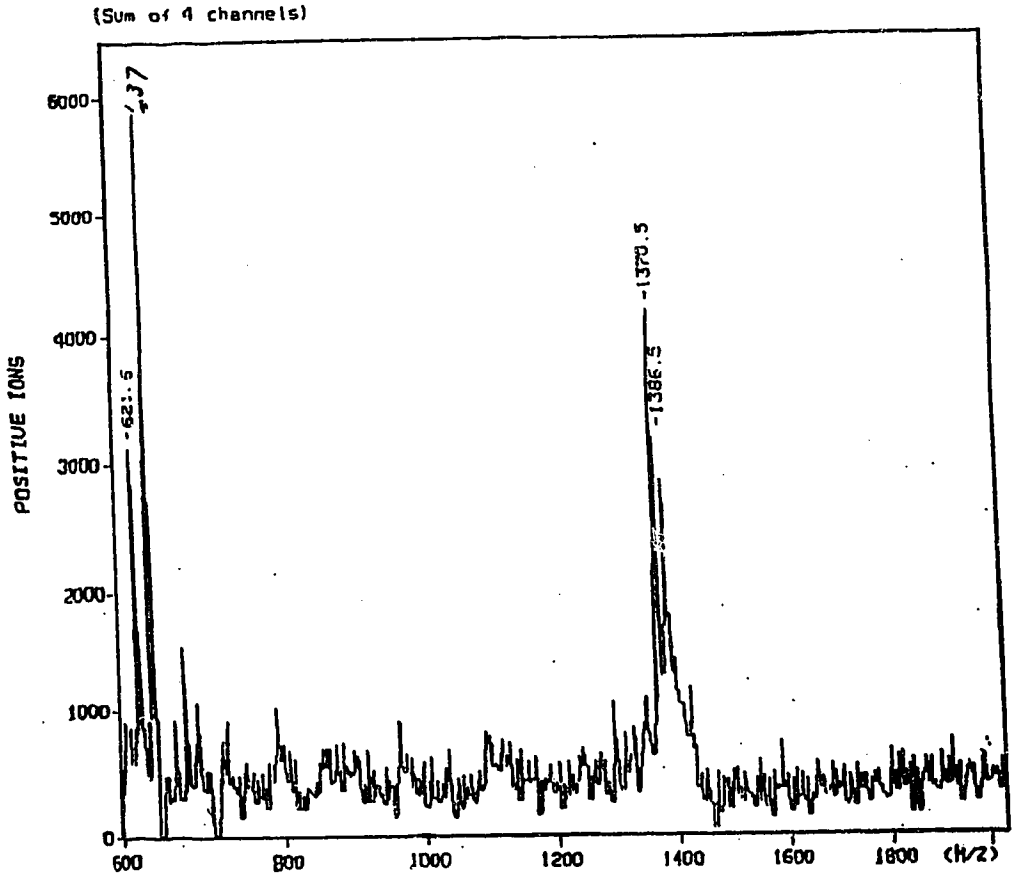


Figure 23. Plasma desorption mass spectra (PDMS) of the differential peak shown in Figure 22. The mass at 1370.5 is likely the sodium adduct of the decapeptides and the mass at 1386.5 suggests a potassium adduct of the same peptide (see text for the details).

BIBLIOGRAPHY

1. **Alter, H. J.** 1991. Decartes before the horse: I clone, therefore I am: The hepatitis C virus in current perspective. Ann. Intern. Med. **115**:644-649.

2. **Alter, H. J., R. H. Purcell, P. V. Holland, and et al.** 1975. Clinical and serological analysis of transfusion-associated hapatitis. Lancet **2**:834-841.

3. **Alter, M. J.** 1989. Non-A, non-B hepatitis: Sorting through a diagnosis of exclusion. Ann. Intern. Med. **110**:583-585.

4. **Alter, M. J., C. S. Hadler, D. P. Francis, and J. E. Maynard.** 1985. The epidemiology of non-A, non-B hepatitis in the United States, p. 71-79. In R. Y. Dodd and L. F. Barker (eds.), Infection, Immunity, and Blood Transfusion. Alan R. Liss, New York.

5. **Alter, M. J., C. S. Hadler, H. S. Margolis, and et al.** 1988. The changing epidemiology of non-A, non-B hepatitis in the United States: Relationship to transfusions. Transfusion-Associated Infections and Immune Response III **1**:14.(Abstract)

6. **Atherton, E. and R. C. Sheppard.** 1989. Solid Phase Peptide Synthesis. Oxford University Press, Osford.

7. **Bartenschlager, R., L. Ahlborn-Laake, J. Mous, and H. Jacobsen.** 1993. Nonstructural protein 3 of the hepatitis C virus encoded a serine-type proteinase required for cleavage at the NS3/4 and NS4/5 junctions. J. Virol. **67**:3835-3844.

8. **Bartenschlager, R., L. Ahlborn-Laake, J. Mous, and H. Jacobsen.** 1994. Kinetic and structural analyses of hepatitis C virus polyprotein processing. J. Virol. **68**:5045-5055.

9. **Beach, M. J. and D. W. Bradley.** 1991. Analysis of the putative nonstructural gene region of hepatitis C virus, p. 376-381. In F. B. Hollinger, S. Lemon, and H. Margolis (eds.), Viral hepatitis and liver disease. Williams and Wilkins, Baltimore.

10. **Billich, A., F. Hammerschmid, and G. Winkler.** 1990. Purification , assay and kinetic features of HIV-1 proteinase. Biol. Chem. Hoppe. Seyler. **371**:265-272.

11. **Billich, A. and G. Winkler.** 1990. Colorimetric assay of HIV-1 proteinase suitable for high-capacity screening. Peptide Research **3**:274.

12. **Billich, S., M. -T. Knoop, J. Hansen, P. Strip, J. Sedlacek, R. Mertz, and K. Moelling.** 1988. Synthetic peptides as substrates and inhibitors of human immune deficiency virus-1 protease. J. Biol. Chem. **263**:17905-17908.

13. **Bradley, D. M.** 1985. The agenst of non-A, non-B viral hepatitis. J. Virol. Methods **10**:307-319.

14. **Bradley, D. W., K. Krawczynski, M. J. Beach, and M. A. Purdy.** 1991. Non-A, Non-B hepatitis: Toward the discovery of hepatitis C and E viruses. Seminars in Liver Disease **11**:128-146.

15. **Chambers, T. J., D. W. McCourt, and C. M. Rice.** 1989. Yellow fever virus proteins NS2A, NS2B, and NS4B: Identification and partial N-terminal amino acid sequence analysis. Virology **169**:100-109.

16. **Chen, P. -J., M. -H. Lin, K. -F. Tai, P. -C. Liu, and D. -S. Chen.** 1992. The Taiwanese hepatitis C virus gemone: Sequence determination and mapping the 5' termini of viral genomic and antigenomic RNA. Virology **188**:102-113.

17. **Chien, D., Q. Choo, A. Tabrizi, C. Kuo, and et al.** 1987. Diagnosis of hepatitis C virus (HCV) infection using an immunodominant chimeric polyprotein to capture circulating antibodies: reevaluation of the role of HCV in liver disease. Anal. Biochem. **162**:156-159.

18. **Chomczynski, P. and N. Sacchi.** 1987. Single-step method of RNA isolation by acid guanidinium Thiocyanate-Phenol-chloroform extraction. Anal. Biochem. **162**:156-159.

19. **Choo, Q. -L., G. Kuo, A. J. Weiner, L. R. Overby, D. W. Bradley, and M. Houghton.** 1989. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. Science **244**:359.

20. **Choo, Q. -L., K. H. Richman, J. H. Han, K. Berger, C. Lee, C. Dong, C. Gallegos, D. Coit, A. Medina-Selby, P. J. Barr, A. J. Weiner, D. W. Bradley, G. Kuo, and M. Houghton.** 1991. Genetic organization and diversity of the hepatitis C virus. Proc. Natl. Acad. Sci. USA **88**:2451-2455.

21. **Colombo, M., G. Kuo, Q. -L. Choo, M. F. Donato, E. D. Ninno, M. A. Tommasini, N. Dioguardi, and M. Houghton.** 1989. Prevalence of antibodies to hepatitis C virus in Italian patients with hepatocellular cancinoma. Lancet **ii**:1006-1008.

22. **Cotter, R. J.** 1988. Plasma desorption mass spectormetry comes of age. Anal. Chem **60**:781A-793A.

23. **Craske, J., N. Dilling, and D. Stern.** 1983. An outbreak of hepatitis associated intravenous injection of Factor VIII concentrate. Lancet **85**:439-462.

24. **Dienstag, J. L.** 1983. Non-A, non-B hepatitis. Gastroenterolgy **85**:439-462.

25. **Eckart, M., M. Selby, F. Masiarz, and et al.** 1993. The hepatitis C virus encodes a serine protease involved in processing of the putative nonstructural proteins from the viral polyprotein precursor. Biochem. Biophys. Res. Commun. **192**:399-406.

26. **Failla, C., L. Tomei, and R. de Francesco.** 1994. Both NS3 and NS4A are required for proteolytic processing of hepatitis C virus nonstructural proteins. J. Virol. **68**:3753-3760.

27. **Genesca, J., J. I. Esteban, and H. J. Alter.** 1991. Blood-borne non-A, non-B hepatitis: Hepatitis C. Seminars in Liver Disease **11**:147.

28. **Geoghegan, K. F., R. W. Spencer, D. E. Danley, L. G. Contillo, and G. C. Andrews.** 1990. Fluorescence-based continuous assay for the aspartyl protease of human immunodeficiency virus-1. FEBS Lett. **262**:119-122.

29. **Gerber, M. A.** 1993. Relation of hepatitis C virus to hepatocellular carcinoma. J. Hepatol. **17 (Suppl. 3)**:108-111.

30. **Gorbalenya, A. E., A. P. Donchenko, and E. V. Koonin.** 1989. N-terminal domains of putative helicases of flavi- and pestiviruses may be serine protease. Nucleic Acids Res. **17**:3889-3897.

31. **Grakoui, A., D. W. McCourt, C. Wychowshi, S. M. Feinstone, and C. M. Rice.** 1993. A second hepatitis C virus-encoded proteinase. Proc. Natl. Acad. Sci. USA **90**:10583-10587.

32. **Grakoui, A., D. W. McCourt, C. Wychowski, S. M. Feinstone, and C. M. Rice.** 1993. Characterization of the hepatitis C virus-encoded serine proteinase: Determination of proteinase-dependent polyprotein cleavage sites. J. Virol. **67**:2832-2843.

33. **Grakoui, A., C. Wychowski, C. Lin, S. M. Feinstone, and C. M. Rice.** 1993. Expression and identification of hepatitis C virus polyprotein cleavage products. J. Virol. **67**:1385-1395.

34. **Grant, G. A.** 1992. Synthetic Peptides A User's Guide. p.185-195. Freeman, New York.

35. **Grant, M. R., H. V. Colot, L. Guarente, and M. Robash.** 1982. Open reading frame cloning: Identification, cloning, and expression of open reading frame DNA. Proc. Natl. Acad. Sci. USA **79**:6598-6602.

36. **Han, J. H., V. Shyamala, K. H. Richman, M. J. Brauer, B. Irvine, M. S. Urdea, P. Tekamp-Olson, G. Kuo, Q. -L. Choo, and M. Houghton.** 1991. Characterization of the terminal regions of hepatitis C viral RNA: Identification of conserved sequences in the 5' untranslated region and poly(A) tails at the 3' end. Proc. Natl. Acad. Sci. USA **88**:1711-1715.

37. **Harlow, E. and D. Lane.** 1988. Antibodies --A Laboratory Manual. p.53-138. Cold Spring Harbor Laboratory, Cold Spring, N.Y..

38. **Hellings, L. A., J. Van der Veen-du Prie, and P. Boender.** 1988. Viral Hepatitis and Liver Diseases. p.543-549. Liss, New York.

39. **Hemdan, E. S., Y. -j. Zhao, E. Sulkowski, and J. Porath.** 1989. Surface topography of histidine residues: A facile probe by immobilized metal ion affinity chromatography. Proc. Natl. Acad. Sci. USA **86**:1811-1815.

40. **Hewick, R. M., M. W. Hunkapiller, L. E. Hood, and W. J. Dreyer.** 1981. A gas-liquid solid phase peptide and protein sequenator. J. Biol. Chem. **256**:7990-7997.

41. **Hijikata, M., N. Kato, Y. Ootusyama, M. Nakagawa, and K. Shimotohno.** 1991. Gene mapping of the putative structural region of the hepatitis C virus genome by in vitro processing analysis. Proc. Natl. Acad. Sci. USA **88**:5547-5551.

42. **Hijikata, M., H. Mizushima, T. Akagi, S. Mori, N. Kakiuchi, N. Kato, T. Tanaka, K. Kimura, and K. Shimotohno.** 1993. Two distinct proteinase activities required for the processing of a putative nonstructural precursor protein of hepatitis C virus. J. Virol. **67**:4665-4675.

43. **Hirel, P. H., F. Parker, J. Boizeau, G. Jung, D. Outerovitch, A. Dugue, C. Peltiers, C. Giuliacci, R. Boulay, Y. Levievre, B. Cambou, J. F. Mayaux, and IT. Cartwright.** 1990. HIV-1 aspartic proteinase: high-level production and automated fluorometric screening assay for inhibitors. Antiviral Chem. Chemother. **1**:9-15.

44. **Hochuli, E.** 1990. Purification of recombinant proteins with metal chelate adsorbent. Genetic Engineering **12**:87-98.

45. **Hollinger, F. B.** 1990. Non-A, non-B hepatitis viruses, p. 2239-2273. In B. N. Fields (ed.), Virology. Raven Press, Ltd., New York.

46. **Inchauspe, G., S. Zebedee, D. -H. Lee, M. Sugitani, M. Nasoff, and A. M. Prince.** 1990. Genomic structure of the human prototype strain H of hepatitis C virus: comparison with the American and Japanese isolates. Proc. Natl. Acad. Sci. USA **87**:9524-9528.

47. **Kato, N., M. Hijikata, Y. Ootsuyama, M. Nakagawa, S. Ohkoshi, T. Sugimura, and K. Shimotohno.** 1990. Molecular cloning of the human hepatitis C virus genome from Japenese patients with non-A, non-B hepatitis. Proc. Natl. Acad. Sci. USA **87**:9524-9528.

48. **Kellerman, O. K. and T. Ferenci.** 1982. Maltose binding protein from *E. coli.* Methods Enzymol. **90**:459-463.

49. **Kiyosawa, K., T. Sodeyama, and E. Tanaka.** 1990. Interrelationship of blood transfusion, non-A, non-B hepatitis and hepatocellular carcinoma. Analysis by detection of antibody to hepatitis C virus. Hepatology **12**:671-675.

50. **Knott, J. A., D. C. Orr, D. S. Montgomery, C. A. Sullivan, and A. Weston.** 1989. The expression and purification of human rhinovirus protease 3C. Eur. J. Biochem. **182**:547-555.

51. **Koerner, T. J., J. E. Hill, A. M. Myers, and A. Tzagoloff.** 1990. High-expression vectors with multiple cloning sites for construction of trpE-fusion genes: pATH vectors. Methods Enzymol. **194**:477-490.

52. **Kotler, M., W. Danho, R. A. Katz, J. Lris, and M. Skalka.** 1989. Avian retroviral protease are distinguished by activities on peptide substrates. J. Biol. Chem. **264**:3428-3435.

53. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227**:680-685.

54. **Langer, T., C. Lu, H. Echols, J. Flanagan, M. K. Hayer, and F. U. Hartl.** 1992. Successive action of DnaK, DnaJ and GroEL along the pathway of chaperone-mediated protein folding. Nature **356**:683-689.

55. **LaVallie, E. R., E. A. DiBlasio, S. Kovacic, K. L. Grant, P. F. Schendel, and J. M. McCoy.** 1993. A thioredoxin gene fusion expression system that circumvents inclusion body formaiton in the *E. coli* cytoplasm. Bio/Technology **11**:187-193.

56. **Libby, R. T., D. Cosman, M. K. Cooney, J. E. Merriam, C. J. March, and T. P. Hopp.** 1988. Human rhinovirus 3C protease: Cloning and expression of an active form in *Escherichia coli.* Biochemistry **27**:6262-6268.

57. **Lin, C., B. M. Pragai, A. Grakoui, J. Xu, and C. M. Rice.** 1994. Hepatitis C virus NS3 serine proteinase: *trans* -cleavage requirements and processing kinetics. J. Virol. **68**:8147-8157.

58. **Loeb, D. D., R. Swanstrom, L. Everitt, M. Manchester, S. E. Stamper, and C. A. III Hutchison.** 1989. Complete mutagenesis of the HIV-1 protease. Nature **340**:397.

59. **MacFarlane, R. D., R. P. Skowronski, and D. F. Torgenson.** 1974. New approach to the mass spectroscopy of non-volatile compounds. Biochem. Biophys. Res. Commun. **60**:616-621.

60. **Malcolm, B. A., S. M. Chin, D. A. Jewell, J. R. Stratton-Thomas, K. B. Thudium, R. Ralston, and S. Rosenberg.** 1992. Expression and characterization of recombinant hepatitis A virus 3C proteinase. Biochem. **31**:3358-3363.

61. **Marston, F. A.** 1986. The purification of eukaryotic polypeptides synthesized in *Escherichia coli.* Biochem. J. **240**:1-12.

62. **Matayoshi, E. D., G. T. Wang, G. A. Krafft, and J. Erickson.** 1990. Novel fluorogenic substrates for assaying retroviral protease by resonance energy transfer. Science **247**:954-958.

63. **McKeever, B. M., M. A. Navia, P. M. Fitzgerald, J. P. Springer, C. -T. Leu, J. C. Heimbach, W. K. Herber, I. S. Sigal, and P. L. Darke.** 1989. Crystallization of the aspartylprotease from the human immunodeficiency virus, HIV-1. J. Biol. Chem. **264**:1919-1921.

64. **Meek, T. D., B. D. Dayton, B. W. Metcalf, G. B. Dreyer, J. E. Strickler, J. G. Gorniak, M. Rosenberg, M. L. Moore, V. W. Magaard, and C. Debouk.** 1989. Human immunodeficiency virus 1 protease expressed in *Escherichia coli* behaves as a dimeric aspartic protease. Proc. Natl. Acad. Sci. USA **86**:1841-1845.

65. **Miller, R. H. and R. H. Purcell.** 1990. Hepatitis C virus shares amino acid sequence similarity with pestiviruses and flaviviruses as well as members of two plant virus supergroups. Proc. Natl. Acad. Sci. USA **87**:2057-2061.

66. **Nashed, N. T., J. M. Luois, J. M. Sayer, E. M. Wondrak, P. T. Mora, S. Oroszlan, and D. M. Jerina.** 1989. Continuous spectrophotometric assay for retroviral proteases of HIV-1 and AMV. Biochem. Biophys. Res. Commun. **163**:1079-1085.

67. **Nicklin, M. J., K. S. Harris, P. V. Pallai, and E. Wimmer.** 1988. Poliovirus proteinase 3C: Large-scale expression, purification, and specific cleavage activity on natural and synthetic substrates in vitro. J. Virol. **62**:4586-4593.

68. **Nishioka, K.** 1990. Posttransfusion hepatitis C. Proceedings of the Second International Symposium on HCV **November**:36.

69. **Okamoto, H., S. Okada, Y. Sugiyama, and et al.**. 1991. Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human carrier: comparison with reported isolates for conserved and divergent regions. J. Gen. Virol. **72**:2697-2704.

70. **Okuyama, T. and K. Satake.** 1960. On the preparation and properties of 2,4,6,-trinitrophenyl-amino acids and -peptides. J. Biochem. **47**:454-466.

71. **Pallai, P. V., F. Burkhardt, M. Skoog, K. Schreiner, P. Bax, K. A. Cohen, G. hansen, DE. Palladino, K. S. Harris, M. J. Nicklin, and E. Wimmer.** 1989. Cleavage of synthetic peptides by purified poliovirus 3C proteinase. J. Biol. Chem. **264**:9738-9741.

72. **Phillips, M. A., R. J. Fletterick, D. E. Palladino, K. S. Harris, M. J. Nicklin, and E. Wimmer.** 1992. Protease. Curr. Opin. Struct. Biol. **2**:713-720.

73. **Porath, J., J. Carlsson, I. Olsson, and G. Belfrage.** 1975. Metal chelate affinity chromatography, a new approach to protein fractionation. Nature **258**:598-599.

74. **Preugschat, F., C. -W. Yao, and J. H. Strauss.** 1990. In vitro processing of Dengue virus type 2 nonstructural proteins NS2A, NS2B, and NS3. J. Virol. **64**:4364-4374.

75. **Prince, A. M., B. Brotman, G. F. Grady, and et al.**. 1974. Long-incubation post-transfusion hepatitis without serological evidence of exposure to hepatitis-B virus. Lancet **2**:241-246.

76. **Richards, A. D., L. H. Phylip, W. G. Farmerie, P. E. Scarborough, A. Alvarez, B. M. Dunn, P. -H. Hirel, J. Konbalinka, P. Strop, L. Pavlickova, V. Kostka, and J. Kay.** 1990. Sensitive, soluble chromogenic substrates for HIV-1 proteinase. J. Biol. Chem. **265**:7733-7736.

77. **Rink, H.** 1987. Solid-phase synthesis of protected peptide fragments using a trialkoxy-diphenyl-methylester resin. Tetrahedron Lett. **28**:3789-3791.

78. **Roberts, N. A., J. A. Martin, D. Kinchington, A. V. Broadhurst, J. C. Ctaig, I. B. Duncan, S. A. Galpin, B. K. Handa, J. Kay, A. Krohn, R. W. Lambert, J. H. Merrett, J. S. Mills, K. E. Parkes, S. Redshae, A. J. Ritchie, D. L. Taylor, G. J. THomas, and P. J. Machin.** 1990. Rational design of peptide-based HIV proteinase inhibitors. Science **248**:358-361.

79. **Ruther, U. and B. Muller-Hill.** 1983. Easy identification of cDNA clones. EMBO J. 2:1791-1794.

80. **Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich.** 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **239**:487-491.

81. **Saito, I., T. Miyamura, A. Ohbayashi, H. Harada, T. Katayama, S. Kikuchi, S. Watanabe, S. Koi, M. Onji, Y. Ohta, Q. -L. Choo, M. Houghton, and G. Kuo.** 1990. Hepatitis C virus infection is associated with the development of hepatocellular carcinoma. Proc. Natl. Acad. Sci. USA **87**:6547-6549.

82. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular Cloning: A Laboratory Manual . Cold Spring Harbour Press, Cold Spring Harbour, New York.

83. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U. S. A. **74**:5463-5467.

84. **Santolini, E., G. Migliacchio, and N. la Monica.** 1994. Biosynthesis and biochemical properties of the hepatitis C virus core protein. J. Virol. **68**:3631-3641.

85. **Satake, K., T. Okuyama, M. Ohashi, and T. Shinoda.** 1960. The spectrophotometric determination of amine, amino acid and peptide with 2,4,6-trinitrobenzene 1-sulfonic acid. J. Biochem. **47**:654-660.

86. **Shimizu, Y. K., A. Iwamoto, M. Hijikata, R. H. Purcell, and H. Yoshikura.** 1992. Evidence for *in vitro* replication of hepatitis C virus genome in a human T-cell line. Proc. Natl. Acad. Sci. USA **89**:5477-5481.

87. **Short, J. M., J. M. Fernandez, J. A. Sorge, and W. D. Huse.** 1988. l ZAP: a bacteriophage l expression vector with in vivo excision properties. Nucleic Acids Res. **16, No.15**:7583-7600.

88. **Smith, D. B. and K. S. Johnson.** 1988. Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase. Gene **67**:31-40.

89. **Studier, F. W., A. H. Rosenberg, J. J. Dunn, and J. W. Dubendorff.** 1990. Use of T7 RNA polymerase to direct expression of cloned genes. Methods Enzymol. **185**:60-89.

90. **Tabor, S. and C. C. Richardson.** 1985. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. Proc. Natl. Acad. Sci. USA **82**:1074-1078.

91. **Tabor, S., A>R. Shatzman, T. Hoey, P. Riggs, D. B. Smith, and et al.** 1994. Current Protocols in Molecular Biology. John WIley and Sons, Inc.,

92. **Takamizawa, A., C. Mori, I. Fuke, and et al.** 1991. Structure and organization of the hepatitis C virus genome isolated from human carriers. J. Virol. **65**:1105-1113.

93. **Tanaka, T., T. Hirohata, N. Kato, and et al.** 1991. Hepatitis C and hepatitis B in the etiology of hepatocellular carcinoma in the Japanese population. Cancer Res. **51**:2842-2847.

94. **Tanaka, T., N. Kato, M. Nakagawa, and et al.** 1992. Molecular cloning of hepatitis C virus genome from a single Japanese carrier: sequence variation within the same individual and among infected individuals. Virus Res. **23**:39-53.

95. **Tomaszek, T. A., V. W. Magaard, H. G. Bryan, M. L. Moore, and T. D. Meek.** 1990. Chromophic peptide substrates for the spectrophotometric assay of HIV-1 protease. Biochem. Biophys. Res. Commun. **168**:274-280.

96. **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.

97. **van Doorn, L. -J.** 1994. Review: Molecular biology of the hepatitis C virus. J. Med. Virol. **43**:345-356.

98. **Williams, D. C., R. M. Van Frank, W. L. Muth, and J. P. Burnett.** 1982. Cytoplasmic inclusion bodies in *Escherichia coli* producing biosynthetic human insulin proteins. Science **215**:687-689.

99. **Wondrak, E. M., T. D. Copeland, J. M. Louis, and S. Oroszlan.** 1990. A solid phase assay for the protease of human immunodeficiency virus. Anal. Biochem. **188**:82-85.

100. **Komoda, Y., M. Hijikata, Y. Tanji, Y. Hirowatari, H. Mizushima, K. Kimura, and K. Shimotohno.** 1994. Processing of hepatitis C viral polyprotein in *Escherichia coli* . Gene **145**:221-226.

101. **Tanji, Y., M. Hijikata, Y. Hirowatari, and K. Shimotohno.** 1994. Identification of the domain required for trans-cleavage activity of hepatitis C viral serine proteinase. Gene **145**:215-219.

102. **Brayer, G.D., L. T. J. Delbaere, and M. N. G. James.** 1978. Molecular structure of crystalline Streptomyces griseus protease A at 2.8 Å resolution II.

Molecular conformation, comparison with a-chymotrypsin and active-site geometry. J. Mol. Biol. 124:261-283.