Journal of Peptide Science
*J. Peptide Sci.* **7**: 2–14 (2001)

**REVIEW**

# A Personal Account of the Role of Peptide Research in Drug Discovery: the Case of Hepatitis C[1]

**ANTONELLO PESSI***

IRBM P. Angeletti, Via Pontina, Rome, Italy

*Received 3 November 2000*
*Accepted 4 November 2000*

Abstract: Although peptides themselves are not usually the end products of a drug discovery effort, peptide research often plays a key role in many aspects of this process. This will be illustrated by reviewing the experience of peptide research carried out at IRBM in the course of our study of hepatitis C virus (HCV). The target of our work is the NS3/4A protease, which is essential for maturation of the viral polyprotein. After a thorough examination of its substrate specificity we fine-tuned several substrate-derived peptides for enzymology studies, high-throughput screening and as fluorescent probes for secondary binding assays. In the course of these studies we made the key observation: that the protease is inhibited by its own cleavage products. Single analog and combinatorial optimization then derived potent peptide inhibitors. The crucial role of the NS4A cofactor was also addressed. NS4A is a small transmembrane protein, whose central domain is the minimal region sufficient for enzyme activation. Structural studies were performed with a peptide corresponding to the minimal activation domain, with a series of product inhibitors and with both. We found that NS3/4A is an induced fit enzyme, requiring both the cofactor *and* the substrate to acquire its bioactive conformation; this explained some puzzling results of 'serine-trap' type inhibitors. A more complete study on NS3 activation, however, requires the availability of the full-length NS4A protein. This was prepared by native chemical ligation, after sequence engineering to enhance its solubility; structural studies are in progress. Current work is focused on the P′ region of the substrate, which, at variance with the P region, is not used for ground state binding to the enzyme and might give rise to inhibitors showing novel interactions with the enzyme. Copyright © 2001 European Peptide Society and John Wiley & Sons, Ltd.

Keywords: drug discovery; hepatitis C

Antonello Pessi

* Correspondence to: Department of Biotechnology, Istituto di Ricerche di Biologia Molecolare P. Angeletti (IRBM), Via Pontina Km 30.600, 00040 Pomezia, Rome, Italy. E-mail: antonello_pessi@merck.com

[1] The Leonidas Zervas Award Lecture delivered at the 26th European Peptide Symposium, Montpellier, September 2000.

## INTRODUCTION

It is common experience that several barriers, notably low bioavailability and short half-lives, prevent the development and utilization of peptide drugs. This conclusion however – peptides are usually not the end result of a drug discovery effort – should not be conducive to another one, i.e. that peptides (and peptide research) have no, or at best a minor, role in drug discovery.

I would like to highlight the multifaceted role of peptide research in a drug discovery project by virtue of a personal account drawn from my own experience. The present paper will therefore limit itself to a single project, the search for inhibitors of the serine protease of human hepatitis C virus (HCV), and within this project, to the work carried out essentially in a single research institute. Moreover, the project is still in progress. Despite this strong bias, I believe that the conclusions drawn may be sufficiently illustrative for two reasons: first, most of our findings on HCV have been independently or successively confirmed in other laboratories, so that they have general value for this therapeutic target. Second, because of the importance of proteases as drug discovery targets. 2% of the gene products of the genomes that have been completely sequenced encode for proteases [1]. A rough estimate for a 100 kb human genome would yield approximately 1100 different (mostly new) candidate protease targets [2]. This figure of course ignores pathogens' genomes and their proteases.

## HEPATITIS C, HEPATITIS C VIRUS AND NS3

Hepatitis C is a slowly progressing disease caused by infection with HCV, which frequently results in end-stage liver disease and its complications, including hepatocellular carcinoma [3]. Today, hepatitis C is widely recognized as a major public health concern, with more than 300 million people infected worldwide, most of them unknowingly [4], while neither a generally effective treatment nor a preventive vaccine is available [5–7].

The genome of HCV is a 9.6-kb single-stranded positive sense RNA molecule, which contains a single open reading frame encoding for a polyprotein of approximately 3000 amino acids [8]. Replication of HCV depends on the proteolytic maturation of the polyprotein precursor and this maturation, in turn, is crucially dependent on the serine protease encoded in the viral protein NS3. Inhibition of this protease activity is thus currently considered a promising approach to an anti-HCV therapy [6,7]. A notable feature of the NS3 protease is that it is only weakly active on its own, requiring the formation of a non-covalent complex with another virally encoded protein, NS4A, for full activity [9–12]. Accordingly, the protease is frequently referred to as the NS3/4A protease [6].

NS3 is a multifunctional protein, containing both a serine protease and an RNA helicase [13–16]. The serine protease domain (henceforth NS3pro), which is contained within the first 180 amino acids of the NS3 protein, has been produced in enzymatically active form in both bacterial and eukaryotic expression systems [13–15,17–20]. A synthetic 14-residue peptide (Pep4A), corresponding to the central hydrophobic domain of the NS4A protein, has been shown to be both necessary and sufficient for activation of NS3pro [21–24]. Expression of the native complex between full-length NS3 and full-length NS4A was achieved later [25,26]. Single-chain forms of NS3 and NS3pro, where Pep4A is fused through a flexible linker to the *N*-terminus of the enzyme, have also been produced [27,28]. All the above-mentioned forms of the enzyme show similar activity towards polypeptide substrates. For essentially all the studies described below, we made use of the NS3pro/Pep4A complex [14,15].

## SUBSTRATE SPECIFICITY OF NS3pro

A comparative analysis of the sequences flanking the peptide bonds cleaved by NS3/4A within the HCV polyprotein yielded the following consensus ($P_6$–$P'_4$, according to the nomenclature of Schechter and Berger [29]): **Asp**/**Glu**-Xaa-Xaa-Xaa-Xaa-**Cys**/**Thr**↓**Ser**/**Ala**-Xaa-Xaa-Xaa. Cleavage occurred after a Cys residue in all *trans*-cleavage sites, whereas the intramolecular *cis* cleavage site (between NS3 and NS4A) was unique in this respect by having a Thr residue in the $P_1$ position. Other conserved features were a negatively charged residue in $P_6$, and Ser or Ala in $P'_1$. In addition, the amino acid present in the $P'_4$ position of all sites, although not strictly conserved, always possessed a bulky hydrophobic side-chain.

We initially determined the minimal substrate length by preparing a series of *N*- and *C*-terminally truncated peptides derived from the NS4A/NS4B cleavage site. The choice for this site was determined by the fact that the corresponding junction was cleaved with a relatively high efficiency in the

 

context of the polyprotein, while the synthesis was anticipated to be easier than the other two *trans*-cleavage sites, which had both one more cysteine residue. All the findings reported below, however, also apply to the NS5A/NS5B cleavage site [30]. Starting from a 20-residue peptide spanning positions $P_{10}$–$P'_{10}$, we found that a decapeptide spanning $P_6$–$P'_4$ was the optimum compromise between number of residues, cleavage efficiency and ease of detection by a standard HPLC method [20]. We then investigated the substrate specificity of NS3pro in more detail by introducing several modifications into this decapeptide [31].

The initial focus was on $P_1$-cysteine substitutes, using both natural and non-natural amino acids. In addition to interest in $P_1$ as the main determinant of specificity in serine proteases, in our case substitution of cysteine would have removed a serious obstacle for medicinal chemistry. For example, the thiol nucleophile complicated the preparation of classical mechanism-based serine proteinase inhibitors ('serine-traps') due to its incompatibility with a serine-targeting electrophile present within the same molecule. Unfortunately, no suitable substitute for Cys was identified in our [31] or others' laboratory [30,32]. In fact, most $P_1$ substitutions resulted in uncleaved substrates.

Among the natural amino acids, only threonine was accepted in $P_1$, with a $>30$-fold lower $k_{cat}/K_m$ value. This finding was strongly suggestive that factors, different from cleavage efficiency, must have acted in the selection of Thr in $P_1$ for the NS3/NS4A junction. A reasonable explanation for the choice of a suboptimal residue, however, came only later (see below) with the discovery of product inhibition [33–35] and the appearance of the crystal structure of the full-length NS3 protein [16].

Among the non-natural amino acids, the best cysteine substitutes turned out to be homocysteine (hCys, 5-fold decreased efficiency) with the side chain increased by one methylene unit. Replacement of the SH group by a hydroxyl group (Ser) abolished cleavage, as did incorporation of the SH group in a thiophene ring or its carboxymethylation. The most common isosteric replacement of cysteine, Abu, was compatible with cleavage of the resulting decamer, but at the expense of a 15-fold decrease in $k_{cat}/K_m$.

We could rationalize the strong preference for Cys in $P_1$ on the basis of the structure of the enzyme $S_1$ specificity pocket [13–15]. In fact, based on a homology model elaborated before the appearance of the crystal structure, this preference had even been predicted [41]. The specificity pocket of NS3pro was found to be small and lipophilic, lined by the hydrophobic side-chains of $Leu^{135}$, $Phe^{154}$ and $Ala^{157}$. The shape of the relatively small and lipophilic Cys side-chain was thus complementary to it. In addition to this, the positively polarized –SH hydrogen could favorably interact with the $\pi$-cloud of the aromatic ring of $Phe^{154}$ (Plate 1). This interaction gave also a plausible explanation for hCys being the best Cys substitute. Similar contacts had been observed in a number of protein and small molecule crystal structures [42,43].

The lack of a good alternative to cysteine, among the most commonly available amino acids, was later confirmed for the product inhibitors (see below) [34 and unpublished data]. Ultimately, a suitable replacement had to be designed [36] in the form of the $-CF_2H$ fragment of difluoroaminobutyric acid. This allowed the preparation of potent serine-trap inhibitors [36–40], whose covalent complex with NS3pro with [38] or without [39] the cofactor Pep4A have been recently disclosed.

It was clear from the beginning that electrostatics was playing a role in substrate cleavage. First, an acidic amino acid in $P_6$ was strictly conserved in all natural substrates. Second, the two sites with the highest turnover had an additional negative charge in $P_5$. Third, $K_m$ values for all substrates increased with an increase in ionic strength [20]. However, mutation of $P_6$-Asp to Asn or Ala in our decapeptide reduced turnover only 2-fold, and even charge inversion (Asp-to-Lys mutation) reduced turnover only 5-fold [31]. From the data of Zhang *et al.* [30], the effect was somewhat more pronounced (10-fold) for the NS5A/NS5B junction. Our results on polypeptide substrates — a preference rather than a stringent requirement for a negative charge in $P_6$ — fitted well with the findings of other authors, who had introduced point mutations in polyprotein substrates [44,45]. It appeared that, in the context of different cleavage sites, extensive mutagenesis of the $P_6$ position had little if any effect on cleavage efficiency. The absolute conservation of this residue, especially in light of the pronounced variability of the HCV genome, seemed thus to indicate that it served some more subtle function. More insight on this aspect of NS3pro activity would come with the availability of substrate-derived inhibitors (see below).

We attempted (but failed) to identify additional major determinants of substrate recognition by both 'classical' alanine scanning and by 'inverse alanine scanning' [31]. In the former experiment, $P_1$

 

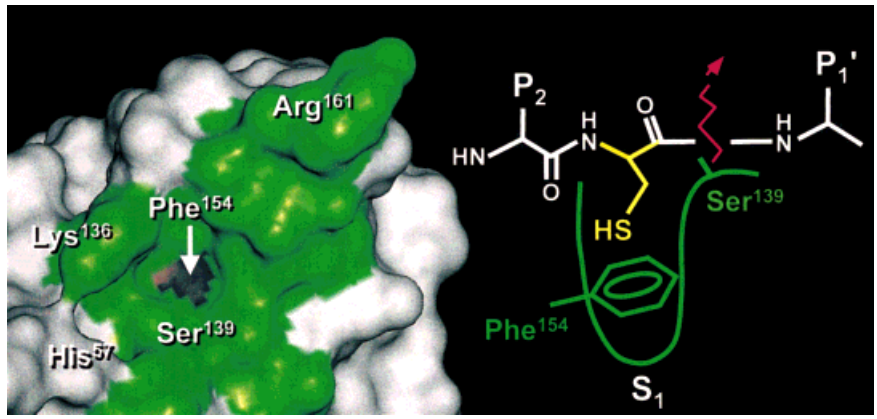


Plate 1 The specificity pocket of NS3pro. The $S_1$ subsite is small and lipophilic, lined by the hydrophobic side-chains of $Leu^{135}$ and $Ala^{157}$, and closed at the bottom by $Phe^{154}$. The site is complementary to a small and lipophilic side-chain. Highlighted is the aromatic ring of $Phe^{154}$, whose $\pi$-cloud could favorably interact with a positively polarized -SH hydrogen. This interaction is believed to explain the strong preference for cysteine in position $P_1$ of the substrate.

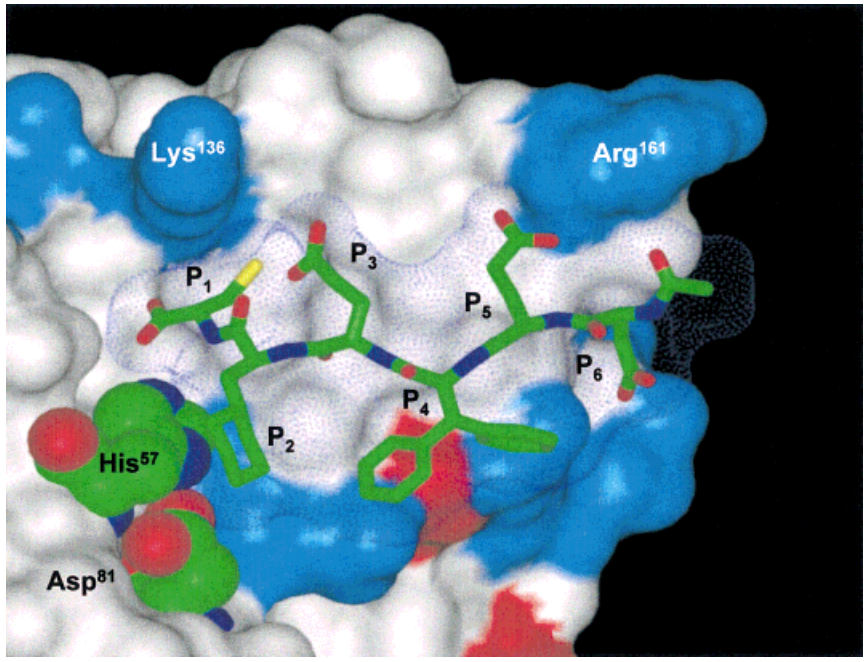


Plate 2 Influence of inhibitor/substrate binding on the catalytic diad of NS3/4A. The stick model of the product inhibitor Ac-Asp-Glu-Dif-Glu-Cha-Cys-OH (Dif = 3,3-diphenylalanine, Cha = $\beta$-cyclohexylalanine) is docked in the *S*-subsite of the protease on the basis of NMR data, as described in [46]. The solvent-accessible surface of the $His^{57}$-$Asp^{81}$ catalytic diad is shown as a CPK model, while the rest of the protease is shown as solvent-accessible surface (white), with the regions of higher electrostatic potential in red (negative) and blue (positive). The side chain of the $P_2$ residue (Cha) sits on top of the catalytic diad, shielding it from solvent. For more details see text.

was confirmed to be the only critical position for substrate processing, which could not be eliminated. $P_3$ had the greatest effect on $K_m$ and $P_6$ on $k_{cat}$, but both effects were modest. In the latter experiment, residues from the natural substrate were re-introduced, one at a time, in a 'minimalist' substrate having alanine in all but the most crucial positions ($P_1$, $P_3$, $P_6$), and the effect on cleavage efficiency was analyzed. We did not observe a sharp increase in cleavage efficiency at any position within the substrate, but rather a gradual and apparently synergistic effect.

In summary, our results showed a rather permissive substrate-binding site, where the only absolute requirement for cleavage was a small, hydrophobic $P_1$ residue. Several minor contributions arising from contacts with distal residues then cooperated in modulating substrate recognition and cleavage. The picture was consistent with what would be deduced from the structure of NS3pro [13–15]. The substrate-binding channel was found to be wide, shallow and almost completely solvent exposed. All the loops connecting the $\beta$ strands, which in other serine proteases of similar fold (e.g. chymotrypsin or elastase) form the $S_2$, $S_3$ and $S_4$ subsites, were shorter or absent in NS3pro. Accordingly, substrate modeling in the active site suggested that major binding contributions should come from the $P_6$–$P_2$ peptide backbone, with an apparent lack of $P_5$–$P_2$ side-chain-to-enzyme interaction [14,32]. Modeling also suggested that the two positively charged side-chains of $Arg^{161}$ and $Lys^{165}$ were engaged in specific interactions with the negatively charged residue in $P_6$ [16,46,47].

## SUBSTRATES FOR ENZYMOLOGY AND HIGH-THROUGHPUT SCREENING

The picture emerging from the initial studies on NS3pro did not augur well for the development of a small-molecule drug. On one side, little or no inhibition was observed with classical serine protease inhibitors, such as aprotinin, leupeptin, $N^{\alpha}$-*p*-tosyl-Lys-chloromethyl ketone [18,48,49], suggesting that specific rationally designed inhibitors had to be developed. On the other side, natural substrates did not offer a good starting point for rational drug design: NS3pro did not cleave small peptide substrates, and the binding interaction was spread along a large surface, with a major electrostatic component. These considerations heightened the need for a lead structure coming from high-throughput screening (HTS) and hence the requirement for a suitable substrate.

In the case of viral proteases, the identification of a high turnover substrate is usually difficult [50] since the kinetic parameters of synthetic peptides based on the natural cleavage sites are generally unfavorable: for example, values of $k_{cat}/K_m$ as low as 12–400 $M^{-1}$ $s^{-1}$ [51] and 17–37 $M^{-1}$ $s^{-1}$ [52] had been found for human cytomegalovirus and herpes simplex virus proteases, respectively. Indeed, when we started our work to produce a HTS substrate, our recombinant NS3pro produced in Baculovirus gave $k_{cat}/K_m = 100$ $M^{-1}$ $s^{-1}$ for the 20-residue peptide Fmoc-Tyr-Gln-Glu-Phe-Asp-Glu-Met-Glu-Glu-Cys-Ala-Ser-His-Leu-Pro-Tyr-Ile-Glu-Gln-Gly-$NH_2$ [19].

To solve the problem we first prepared the simple ester substrate Ac-Asp-Glu-Met-Glu-Glu-Abu-$OCH_3$, on the ground that the acyl–enzyme intermediate would form more readily as a result of the thermodynamically favored transesterification reaction [53]. This ester indeed gave a $k_{cat}/K_m$ of 1010 $M^{-1}$ $s^{-1}$. The corresponding ester with the cysteine in $P_1$ proved, however, very difficult to obtain in a reasonable yield. We then considered a depsipeptide substrate incorporating an internal ester bond at the cleavage site: alanine in $P'_1$ was substituted by lactic acid (HO–CH($CH_3$)COOH), yielding Ac-Asp-Glu-Met-Glu-Glu-Abu-$\psi$[COO]Ala-Ser-His-Leu-Pro-Tyr-Lys(Ac)-$NH_2$. The kinetic parameters for this substrate were dramatically improved ($k_{cat}/K_m =$ 13600 $M^{-1}$ $s^{-1}$) with respect to the corresponding amide substrate, even with the more favored Cys residue in $P_1$ [53]. Spontaneous hydrolysis of the ester bond was also found to be almost negligible, when compared with enzyme-catalyzed cleavage.

As is often the case for a work in progress, successive improvements in expression and purification of NS3pro, and especially in the conditions used for the enzyme assays had since substantially increased the kinetic parameters for the tridecapeptide substrate with a standard scissile bond, from $k_{cat}/K_m = 100$ $M^{-1}$ $s^{-1}$ [53] to $k_{cat}/K_m = 1460$ $M^{-1}$ $s^{-1}$ [31]. This was a sufficiently high value to allow the use of this substrate for a radiometric HTS assay [54].

Although the above depsipeptide was not actually used for our own HTS, the idea had general value and we used it again to prepare a substrate suitable for continuous monitoring of protease activity [55]. The peptide used was an internally quenched fluorogenic substrate based on resonance energy transfer (RET) with a $P_1$–$P'_1$ Abu-$\psi$[COO]Ala

 

depsipeptide bond. It contained a fluorescent donor, 5-[(2′-aminoethyl)amino]naphthalene sulfonic acid (Edans) near one end of the peptide and an acceptor group, 4-[[4′-(dimethylamino)phenyl]azo]benzoic acid (Dabcyl) near the other end. The fluorescence of this type of substrate is initially quenched by intramolecular RET between the donor and the acceptor, but as the enzyme cleaves the substrate, the products are released from RET quenching and the fluorescence of the donor becomes apparent. The result is an increase of fluorescence proportional to the amount of substrate cleaved. We introduced Edans and Dabcyl in positions $P_4$ and $P'_3$, respectively, which had been shown to exert little influence on substrate binding [31]. The resulting kinetic parameters were excellent ($k_{cat}/K_m = 345000$ $M^{-1}$ $s^{-1}$) and we were allowed to run the continuous assay both in a cuvette and in a microplate, with enzyme concentrations as low as 250 pM. The fluorogenic substrate found several applications in the course of our project (see for example References [31,37]) and is now also commercially available. Given its usefulness, we elaborated an improved scheme for the solid-phase synthesis of this type of peptides (with or without a depsipeptide bond) [56].

At the conclusion of this chapter, it must be said that unfortunately no convincing lead from HTS has so far emerged [57], or at least none has so far been disclosed. Rational design, therefore, still represents the main avenue to elaborate inhibitors of NS3/4A.

## PRODUCT INHIBITION

A major discovery, with a large impact on our subsequent work, came with the realization that NS3pro was subject to inhibition by its own cleavage products [33]. At about the same time, independent discovery of this phenomenon occurred in another laboratory [35]. Starting from the observation that during a time-course cleavage experiment of the NS4A/4B substrate (Ac-Asp-Glu-Met-Glu-Glu-Cys-Ala-Ser-His-Leu-Pro-Tyr-Lys(Ac)-$NH_2$) a considerable deviation from linearity was apparent already after 15% of substrate conversion, we fitted our data with an integrated form of the Michaelis–Menten equation, which took into account the possibility that one of the cleavage products was an inhibitor of the enzyme [58]. The data were indeed compatible with an inhibitor with a $K_i$ of 0.7 μM. We synthesized the peptides corresponding to the *N*- and *C*-terminal cleavage products of the NS4A/4B junction and found that the $P_6$–$P_1$ *N*-terminal product (Ac-Asp-Glu-Met-Glu-Glu-Cys-OH) inhibited the enzyme with $K_i = 0.6$ μM.

We quickly found that NS3pro underwent inhibition by the *N*-terminal cleavage products of substrate peptides corresponding to the NS4A/NS4B, NS4B/NS5A and NS5A/NS5B *trans* junctions, whereas no inhibition was observed with a cleavage product of the intramolecular NS3/NS4A *cis* cleavage site. By contrast, none of the *C*-terminal products had any inhibitory activity. The order of potency for the substrate-derived inhibitors was NS4A (Ac-Asp-Glu-Met-Glu-Glu-Cys-OH, $K_i = 0.6$ μM) > NS5A (Ac-Glu-Asp-Val-Val-Cys-Cys-OH, $K_i = 1.4$ μM) ≫ NS4B (Ac-Asp-Cys-Ser-Thr-Pro-Cys-OH, $K_i = 180$ μM). These values of $K_i$ were all lower than the $K_m$ of the corresponding substrates, which were 10 μM, 3.8 μM and > 1 mM, respectively, a clear signature of product inhibition. We chose the NS4A peptide for further studies.

Product inhibition of serine proteases, to the extent observed with NS3pro, is unusual. In general, serine proteases may be inhibited by high millimolar concentrations of their *C*-terminal cleavage products for thermodynamic reasons, the free amino group acting as a nucleophile in an acyltransfer reaction [59]. This mode of inhibition clearly differed from the mechanism of *N*-terminal product inhibition observed by us. A series of experiments [33,34] was carried out in parallel on the NS4A-derived product inhibitor and its *C*-terminally amidated analog, including pH titration and site-directed mutagenesis. They all pointed to the interaction of the $P_1$ α-carboxylate in the active site of NS3pro as the key factor for the high affinity of the product inhibitors. Modeling suggested that the carboxylate be positioned in the active site to form hydrogen bonds with the backbone amides of $Ser^{138}$ and $Gly^{137}$ (the 'oxyanion hole'), with the $N^{\varepsilon 2}$ of the catalytic histidine and with the side chain of the conserved $Lys^{136}$. This binding mode was analogous to published structures of serine protease-product complexes [60–62].

While product inhibition of the NS3pro offered an excellent opportunity for the development of more potent inhibitors, the question whether it had physiological relevance was open to debate. It was clear, however, that it offered a good explanation for the presence in the intramolecular cleavage site between NS3 and NS4A of a threonine residue, in sharp contrast with the $P_1$ preference of the protease (see Section 'Substrate Specificity of

 

NS3pro'). This threonine was conserved in all HCV isolates, suggesting a selective pressure to preserve it. Our studies had just told us that the resulting cleavage product did not inhibit the enzyme, in contrast with those derived from the *trans* cleavage sites. It could be argued, therefore, that a product inhibitor with the preferred Cys in $P_1$, gaining further potency by an intramolecular interaction, should be necessarily counter-selected, since it might prevent all the subsequent cleavages. A telling picture of such a potentially self-inactivating interaction was given by the crystal structure of the NS3 full-length protein [16]. The structure also confirmed the presence of hydrogen bonds between the *C*-terminal carboxylate, derived from cleavage at the NS3/4A *cis* junction and the $Ser^{138}$, $Gly^{137}$ and $His^{57}$ residues.

## OPTIMIZATION OF PRODUCT INHIBITORS

The SAR of the product inhibitors was studied in depth in both our [33,34] and Linàs-Brunet's [35,63] laboratories, reaching similar conclusions. Optimal binding was found to require a dual anchor: a '$P_1$ anchor' at the *C*-terminus of the molecule and an 'acid anchor' at the *N*-terminus. In addition to the already mentioned key role of the *C*-terminal $\alpha$-carboxylate, the $P_1$ residue gave a major contribution to binding also through its side-chain. Not surprisingly, the $P_1$ preferences were the same as those observed for the substrate, with cysteine being the side-chain providing optimal occupancy of the $S_1$ pocket of the enzyme, followed in decreasing order of potency by Abu, Val and Ser. As previously mentioned, a suitable replacement for Cys [36] was designed in the form of difluoroaminobutyric acid [36–40].

The second anchor of the product inhibitors resided in the $P_6$–$P_5$ acidic pair, whose simultaneous deletion yielded a $>100$-fold decrease in activity. This interaction was found to be quite permissive on the exact nature of the negatively charged moiety, with simple $C_4$ or $C_5$ carbon chain diacids being good substitutes for the acidic amino acids. This was qualitatively similar to the substrate preferences in the $P_6$ and $P_5$ [30,31] but the magnitude of the effect was much more pronounced and allowed better appreciation of the importance of the *N*-terminal anchor for substrate/inhibitor binding. This was best underlined by the finding that the $P_1$-deleted pentapeptides and the *C*-terminally amidated hexapeptides, both lacking a crucial component of the $P_1$ anchor, were still competitive inhibitors of NS3pro, albeit with much reduced potency. As observed for the $K_m$ of the substrate [20], the $K_i$ of product inhibitors for NS3/4A also increased with an increase in the ionic strength of the medium, the effect being greater the greater the charge density of the inhibitor [34]. Analysis of the electrostatic potential of NS3pro showed a charge distribution rather unusual for a serine protease. Beyond the already mentioned $Arg^{161}$ and $Lys^{165}$, other basic amino acids ($Arg^{117}$, $Arg^{119}$, $Arg^{123}$ and $Arg^{130}$) contributed to the positive electrostatic potential in the $S_6/S_5$ subsite area [46]. A strong positive potential was also apparent in the vicinity of the active site, with $Lys^{136}$ being the main player [33]. The electrostatic potential of the inhibitor was opposite (and complementary) to the protease potential in both regions [82]. The high degree of electrostatic complementarity in the $S_6/S_5$ sub-sites seemed therefore to represent a compensatory mechanism for the lack of well-defined binding pockets in the $S_4$–$S_2$ subsites. Moreover, during the maturational cleavage of the polyprotein, it could serve as electrostatic guidance for the correct docking of the scissile junctions onto the NS3 active site [16]. Interestingly, limited proteolysis experiments, which compared the ternary complex NS3pro/Pep4A with inhibitors spanning either $P_6$–$P_1$ or $P_5$–$P_1$, showed that $Lys^{165}$, which had been suggested to make a specific interaction with the $P_6$ residue of the inhibitor [16,46,47], had comparable accessibility to proteolysis in the two complexes [64], a finding that was more in line with the residue playing a role in substrate docking.

Having defined the general features of binding to NS3, our goal was now to optimize the sequence of the product inhibitors [34]. To this aim, we decided to target the positions that were less critical to binding and therefore less likely to have already been optimized by evolution. We chose as a starting point the NS4A-derived product inhibitor, Ac-DE-MEEC-OH, because of the higher potency ($IC_{50} = 1$ μM) and the much better solubility, and prepared combinatorial peptide libraries targeting positions $P_5$–$P_2$. To have a compilation as complete as possible of the SAR for each subsite, we used a large set of amino acids, including coded and non-coded amino acids, in both L- and D-configuration.

$P_2$ showed a preference for either negatively charged or hydrophobic residues, while polar and positively charged residues were not accepted; the best $P_2$ residue was Cha ($\beta$-cyclohexylalanine), followed by Leu. $P_4$ showed a strong preference for

 

hydrophobic amino acids, the best residue being Dif (3,3-diphenylalanine); The Cha/Dif combination for $P_2/P_4$ gave a potent inhibitor of the NS3 proteinase, with a $K_i$ of 40 nM. The results for position $P_3$ were very clear-cut: only two residues yielded a potency comparable with Glu in the $P_3$ position: Val and Ile. The preference for hydrophobic amino acids in $P_3$ could be explained by the lipophilic nature of the $S_3$ region, formed by the side-chains of $Val^{132}$, $Leu^{135}$, $Ala^{157}$ and $Cys^{159}$ [7], while the effect of Dif in $P_4$ seemed to derive from the entropy benefit of preorganizing the peptide backbone, prior to complex formation [46].

However, the preference for hydrophobic amino acids, such as Leu and Cha, in $P_2$ was not immediately obvious. Once again, a plausible explanation required more studies, in particular the structural analysis of enzyme–inhibitor complexes [38,39, 46,64]. These studies showed that the $S_2$ subsite is mainly formed by the $His^{57}$-$Asp^{81}$ catalytic diad. One of the members of the diad in particular, $Asp^{81}$, appeared to be completely solvent-exposed in the free enzyme, thus compromising its ability to form the hydrogen bond with $His^{57}$, which is essential for the functioning of the enzyme catalytic machinery. As shown in Plate 2, the inhibitor (and by analogy the substrate) was well positioned to provide the necessary shield from solvent, via its $P_2$ residue [46].

Further optimization found a D-amino acid as preferred in position $P_5$, and brought the potency of the best inhibitor (Ac-Asp-(D)-Gla-Leu-Ile-Cha-Cys-OH, Gla = $\gamma$-carboxyglutamic acid) to an $IC_{50}$ of 1.5 nM [34]. Modeling here suggested that the exchange in chirality changed the backbone conformation, thus approximately exchanging the positions of the $P_5$ and $P_6$ side chains relative to NS3pro.

Apart from its value to guide the design of small molecule drugs, the SAR information obtained on product inhibitors allowed for the design of an active site fluorescent probe [40]. Active site occupancy could now be monitored by following the fluorescence energy transfer between a Dansyl fluorophore, in position $P_4$ of an hexapeptide product inhibitor, and tryptophan residues of NS3pro, which are in close vicinity of the active site ($Trp^{53}$ and $Trp^{85}$, distance from modeling $< 20$ Å).

## NS3/4A AS AN INDUCED-FIT PROTEINASE

I have already quoted results coming from the structural analysis of the covalent complex between NS3pro and a serine-trap inhibitor [38,39]. However, the first results with serine traps were totally disappointing. When peptide aldehydes were prepared based on the sequence of natural substrates, the resulting inhibitors were only marginally more potent than the corresponding peptide acids [Narjes F *et al*. Unpublished data]. Similar data have since appeared in the literature [32,63,65,66]. For example, the hexapeptide aldehydes Ac-Glu-Asp-Val-Val-Abu-Val-H [32], Ac-Asp-Asp-Ile-Val-Pro-Nva-H [63] and Ac-Glu-Asp-Val-Val-Abu-Nva-H [66] showed $IC_{50}$s of 50, 10 and 12 μM, respectively. For comparison, simple di- or tripeptide aldehydes like Z-Val-Pro-H [67] or Boc-(D)Phe-Pro-Arg-H [68] had low nanomolar potency against their target enzymes. Moreover, the substrate-derived serine traps were found to be much less selective towards NS3 than the parent peptide acids [63]. Therefore, not only active site-directed small molecule protease inhibitors (chymostatin, Pefabloc, etc.), and other classical serine protease inhibitors (aprotinin, leupeptin, etc.) failed to significantly inhibit NS3pro [18,48,49], but also specifically designed substrate-derived inhibitors did so.

The discovery and optimization of product inhibitors was key to an explanation of this puzzling results. We had already learned by CD studies that NS3pro underwent major conformational changes upon binding of the cofactor Pep4A [68]. This was interpreted as the result of flexibility in the *N*-terminal domain of the enzyme which, in the absence of the cofactor, was completely disordered in the 1–23 region [70]. Accordingly, kinetic analysis had suggested that cofactor binding induced formation of the S′ subsites in the protease *N*-terminal domain [32]. Structural data then showed that in the complex with NS3pro, Pep4A formed a $\beta$-strand intercalating between two $\beta$-strands contributed by the *N*-terminal domain of the enzyme, resulting in a more ordered structure [14,15]. This conformational change led to a rearrangement of the catalytic triad, which was proposed as being the activation mechanism of NS3 by the cofactor.

With the availability of a series of competitive product inhibitors, spanning a potency range of 1–1000 nM [34] we could now study binding to the S region of the enzyme [64]. This is located in the *C*-terminal domain of NS3pro, while the active site is in a crevice between the *N*- and *C*-terminal domains. Near-UV CD showed that the binding of each inhibitor induced a change in the tertiary structure of NS3pro, a change that varied among different inhibitors. By contrast, the near-UV CD

 

spectrum of the ternary complex NS3pro/Pep4A/inhibitor was identical for all the peptides, despite the large difference in their potency [64].

We compounded the spectral results with limited proteolysis experiments, a simple yet powerful way to investigate the dynamics of protein–protein interaction [64,71–73]. In excellent agreement with the CD data, we observed that (i) upon Pep4A binding, NS3pro became less susceptible to proteolysis. With the exception of a few specific sites, mostly located at the extreme *N*-terminal region, hydrolysis at previously sensitive sites [73] became slower, or was completely prevented, including residues located in the *C*-terminal domain and the inter-domain loop. Therefore, binding of the cofactor to the *N*-terminal domain extended its effect to the active site and the *C*-terminal domain. (ii) An even more dramatic stabilization toward proteolytic attack was apparent in the ternary complex NS3pro/Pep4A/inhibitor, and the level of stabilization paralleled the potency of the inhibitor. Here again occupancy of the S region of the enzyme did not only protect residues directly implicated in substrate binding, such as $Phe^{154}$ and $Cys^{159}$, but also (completely or partially) residues in the *N*-terminal domain, including the 1–26 region [64]. In line with these reciprocal effects, it has been recently shown that the affinity of Pep4A for NS3pro is increased in the presence of the substrate, and *vice versa* [73].

Our conclusion was the cofactor and the inhibitor (and by extension the substrate) played a mutually stabilizing role, and that *both* were required for the protease to achieve a stable, fully functional state. Unlike other proteases of its family, NS3pro apparently exists in solution as an ensemble of many rapidly interconverting structures, and this flexibility needs to be reduced to achieve effective catalysis. The cofactor and the substrate/inhibitor synergistically guide the movement and rigidification of a relatively large portion of the protein around them. This reciprocal influence might be exerted through the stabilization of the *N*-terminal and *C*-terminal domains after binding of the cofactor and the inhibitor, respectively, and subsequent tightening of the interdomain interaction (Plate 3).

The above qualified NS3/4A as an induced-fit protease, using the terminology of LaPlante *et al.* [74] for the serine protease of human cytomegalovirus. According to these authors, the meaning of induced fit is that the strength of binding of the substrate is reduced by a factor that reflects the energy cost of converting the enzyme to its activated state. This offered an explanation for the poor performance of the serine traps prepared so far. Mechanism-based inhibitors require an optimal functioning of the catalytic machinery, and now we knew that this in turn depended on the cofactor and the substrate/inhibitor. Even with the cofactor already bound, the substrate/inhibitor had to: (i) complete stabilization of the active site; and (ii) shield the $His^{57}$-$Asp^{81}$ diad from solvent exposure through its $P_2$ residue (see section 'Optimization of Product Inhibitors'). If the peptidyl portion of the serine trap was not good in this role, the reactive electrophilic moiety attached to it could not react effectively to form a reversible covalent complex mimicking the Transition State. The weak aldehydes described above all had a relatively unoptimized $P_6$–$P_2$ peptidyl moiety. Accordingly, when serine traps based on the sequence of optimized hexapeptide product inhibitors were prepared, they inhibited NS3/4A with the expected subnanomolar potency [36–38,40]. The presence of a bulky $P_2$ residue with its diad-shielding effect, coupled to the designed $P_1$ Cys mimetic, and to optimization of the electrophilic moiety, allowed to finally obtain a tripeptide $\alpha$-chetoacid with a $K_i$ of 27 nM [36], whose complex with the protease was characterized [38,39].

Moreover, induced-fit offered an explanation for the surprisingly high $K_d$ (6.8 $\mu$M) that we had determined for the NS3pro/Pep4A complex [69]. This value was indicative of a loose complex, and poorly compatible with the view offered by the crystal structure, where more than 2000 $Å^2$ of surface were buried in the complex. By applying the same argument of the substrate, the observed $K_d$ would incorporate the energetic cost of reducing NS3pro flexibility. Accordingly, it was later shown [73] that this $K_d$ could vary by more than two orders of magnitude, depending on physicochemical conditions, such as concentration of glycerol, salt and detergent. An analogous antichaotropic effect has been associated with an induced-fit mechanism of action for cytomegalovirus protease [74].

To shine further light on this aspect of NS3 activation, it would have been desirable to repeat some of these studies using the full-length NS4Acofactor ($NS4A_{fl}$). However, despite its small size (54 amino acids (aa)) $NS4A_{fl}$ proved very difficult to prepare. The protein was predicted to be a type I transmembrane (TM) protein [75], with the TM domain located in the 20 *N*-terminal residues (Plate 4). Any attempt at preparing $NS4A_{fl}$ by recDNA failed, while stepwise chemical synthesis, even with optimized protocols, produced a highly heterogeneous crude

 

product [75]. This was not surprising since NS4A is the prototype of a 'difficult sequence' according to Kent's definition [76]. For both techniques the key problem was the extreme insolubility and aggregation potential of the protein, even in media like DMSO or 6M guanidine hydrochloride, which prevented any chromatographic purification. We therefore decided to engineer $NS4A_{fl}$ into a more soluble protein by addition of a lysine tail, a strategy that we had successfully pursued in the past [77]. In this case to obtain a well-behaved protein it was necessary to derivatize the protein at both sides (Plate 4). Engineered $NS4A_{fl}$ was now a 67-aa protein, which was monomeric at a concentration higher than 100 μM in 10 mM Tris buffer, pH 7 [75]. The yield from stepwise assembly of engineered $NS4A_{fl}$, however, was still very low, mostly due to failure to efficiently incorporate the 19 *N*-terminal residues. We then realized that native chemical ligation [78] was ideally suited for our purpose, since the required cysteine residue was conveniently located close to one end of the *N*-terminal TM domain. We could thus produce $NS4A_{fl}$ in good yield by ligating a 19-aa fragment corresponding to TM helix with and a 48-aa fragment corresponding to the rest of the molecule [75] (Plate 3). We have already reported the initial characterization of the NS3pro/$NS4A_{fl}$ complex [75] and further studies are in progress.

## SUMMARY AND PERSPECTIVES

Clearly, the maturational cleavage of HCV polyprotein does not fit the picture of an 'active' enzyme (NS3) cleaving a 'passive' substrate. Unlike cellular enzymes like trypsin or elastase, the substrate(s), the cofactor protein (NS4A) and the minimal catalytic entity (NS3pro) collaborate in an elaborate scheme to achieve optimal activity. In our current view, NS3pro is extremely inefficient as an enzyme, being only capable to cleave intramolecularly at the NS3/NS4A junction thanks to optimal substrate positioning within the folded NS3 protease–helicase. This optimal positioning also serves to compensate for the presence of the suboptimal Thr in the $P_1$ position of the substrate. However, it is precisely this lack of intrinsically good features as a substrate that allows the NS3 *C*-terminus, once cleaved, to leave the active site free for the *trans* cleavages. The freed NS4A cofactor can now do two things: (i) anchor the newly formed NS3/4A complex to the ER membrane; and (ii) complete the folding of the *N*-terminal domain of NS3pro, forming the much more stable (and protease resistant) NS3/4A protease. Unlike NS3pro, NS3/4A has a correctly oriented $His^{57}$-$Asp^{81}$ catalytic diad. This diad, however, is essentially all solvent-exposed, which limits the ability of these residues to form the H-bond, critical to activate the catalytic machinery. Shielding is thus provided by the substrate, more specifically by its $P_2$ residue. The surface exposed area is reduced from 86 $Å^2$ in NS3/4A to 45 $Å^2$ in the ternary NS3/4A–inhibitor complex [Koch U. Unpublished]. The effect of NS4A and the substrate(s) on NS3pro are synergic, with the net effect of reducing the conformational flexibility of the protease to a level compatible with full enzymatic activity. Docking of the yet-to-be-cleaved junctions to the NS3/4A active site is mainly electrostatically driven.

Within this background, it is difficult to define the perspectives for the successful design of NS3-targeted drugs. The increasing understanding of the way in which HCV orchestrates the maturation of its polyprotein reinforces on one side the pessimism on the ability to block it with a small MW molecule. The most linear direction for drug design, i.e. to study how the substrate binds and then improve on it, seems almost precluded. On the other side, alternative targets within the enzyme, not yet been fully exploited, have been suggested, such as the NS3/NS4A interaction and the zinc-binding site [79,80].

Our most recent work is directed towards the prime side of NS3/4A, since it represents, in our opinion, an untapped opportunity for drug design [81]. Inspection of the structure of NS3pro with or without Pep4A shows the presence of binding pockets in the S′ region, which might be exploited for binding of active site-directed inhibitors. The potential interaction with the prime side of the enzyme however, is not used by the substrate. We know in fact that, in contrast to the P region, the P′ region of the substrate is important for catalysis but makes little contribution to ground state binding [30–32]. Accordingly, peptides based on the P′ regions of the natural substrates (spanning residues $P'_1$ up to $P'_{10}$) do not inhibit NS3/4A to any significant extent [33,47]. We have reasoned therefore that a ligand, taking advantage of S′-binding, could display a range of interactions with the enzyme different from the ones evolved by the virus, and represent a novel class of HCV protease inhibitors. S′-binding was optimized in the context of non-cleavable decapeptides spanning $P_6$–$P'_4$. It was sequentially increased by introducing the previously optimized P-region [33], changing the $P'_4$ residue, and combinatorially

 

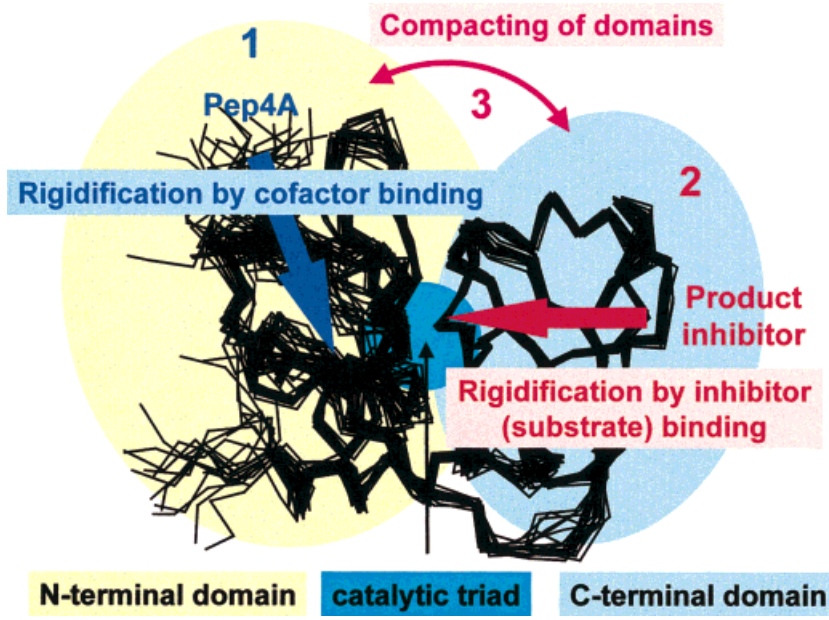


Plate 3 Hypothetical mechanism for substrate- and cofactor-driven induced fit activation of the catalytic machinery of NS3 protease. See text for details. The NS3 structure shown as $C^{\alpha}$ trace is taken from [70].

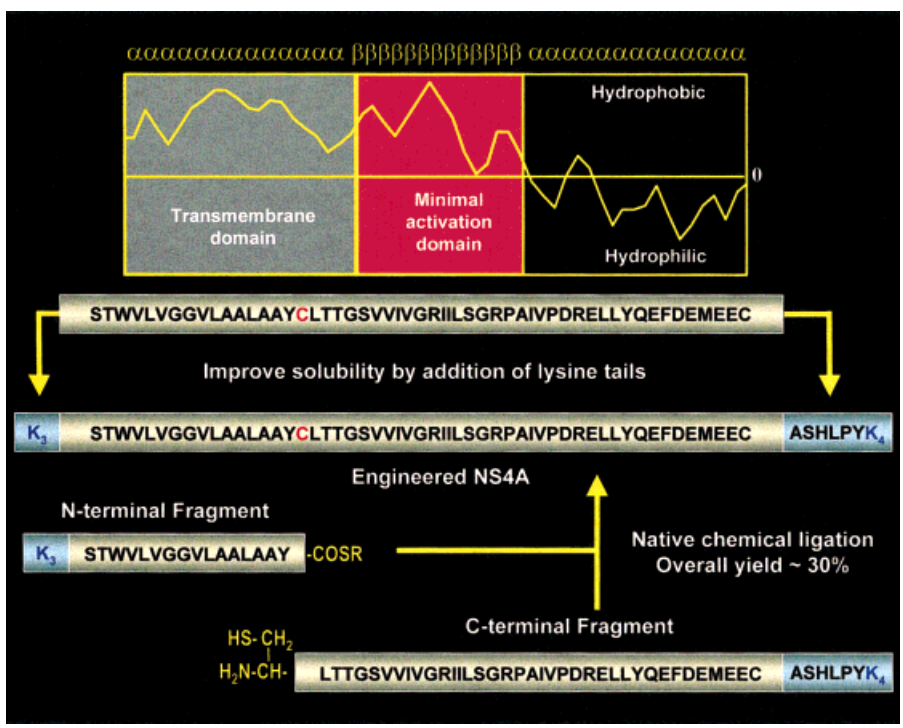


Plate 4 Engineering and chemical synthesis of full-length NS4A. Top: hydrophobicity plot and secondary structure prediction. The minimal activation domain (aa 21–34) and the putative transmembrane region (aa 1–20) are indicated. The protein was first engineered for improved solubility and then synthesized by native chemical ligation. For more explanation see text and for full details see [75].

optimizing positions $P'_2$–$P'_3$. The overall process led to an increase in binding of more than three orders of magnitude, with the best decapeptide showing $IC_{50} < 200$ pM [81].

In line with our assumptions, we found that the binding mode of the decapeptide inhibitors shared features with the natural substrates, but we found also a new and unexpected preference for an acidic residue in position $P'_3$. Since natural amino acids (Asp and Glu) are most effective here, we are tempted to speculate that the virus should necessarily avoid them, to allow the P′ substrate region to leave the enzyme after cleavage. Accordingly, we found that substrate peptides, derived from the sequence of the NS5A/5B junction but incorporating the optimized $P'_2$–$P'_4$ sequence -Cha-Asp-Leu- showed $k_{cat}$ values 10–20-fold lower than the corresponding wild-type substrates [81].

## CONCLUSION

From the above, I hope it is apparent how peptide research has played a major role in our efforts to design a suitable inhibitor of the NS3 serine protease from HCV, acting as an interface with the many disciplines integrated in this drug discovery effort. In parallel with screening, peptide research is still being pursued as one of the most promising ways to identify a weak point of this difficult drug target. I anticipate that this course of events will frequently occur for many other (especially viral) protease targets.

## Acknowledgements

By reviewing the vast amount of work performed at IRBM, I acted as spokesman of a large group of people, which would be impossible to thank properly here. I am indebted to them all and of course, they are all dutifully present in the author's list of the quoted literature. I would like, however, to express here my special gratitude to the person who has been the key player of the entire peptide-related work, namely Elisabetta Bianchi. I am also deeply grateful to Paolo Ingallinella, Raffaele De Francesco, Christian Steinkühler and Sergio Altamura, who were more intimately involved in the work described here, and to Riccardo Cortese, for the constant encouragement, advice and expert discussions throughout our work. Last, but definitely not the least, I wish to thank, on this special occasion, my wife Rossana for the perspective and balance brought into my life, which were critical to any professional achievement.

## REFERENCES

1. Barrett AJ, Rawlings ND, Woessner JF. *Handbook of Proteolytic Enzymes*. Academic Press: London, 1998.
2. Southan C. Assessing the protease and protease inhibitor content of the human genome. *J. Peptide Sci.* 2000; **6**: 453–458.
3. Houghton M. Hepatitis C viruses. In *Field's Virology* (3rd edn), Fields BN, Knipe DM, Howley PM (eds). Lippincott-Raven: Philadelphia, PA, 1996; 1035–1058.
4. Van der Poel CL. Hepatitis C virus: epidemiology, transmission and prevention. In *Current Studies in Hematology and Blood Transfusion*, Reesink HW (ed.). Karger: Basel, 1994; 137–163.
5. Hoofnagle JH. Hepatitis C: the clinical spectrum of disease. *Hepatology* 1997; **26**(Suppl. 1): 15S–20.
6. Kwong AD. Hepatitis C virus NS3/4A protease. *Curr. Opin. Infect. Dis.* 1997; **10**: 485–490.
7. Steinkühler C, Koch U, De Francesco R, Pessi A. The proteinases encoded by hepatitis C virus as therapeutic targets. In *Handbook of Experimental Pharmacology*, Von der Helm K, Korant BD, Cheronis JC (eds). Springer: Berlin, 2000; 75–94.
8. Major ME, Feinstone SM. The molecular virology of hepatitis C. *Hepatology* 1997; **25**: 1527–1538.
9. Failla C, Tomei L, De Francesco R. Both NS3 and NS4A are required for proteolytic processing of hepatitis C virus nonstructural proteins. *J. Virol.* 1994; **68**: 3753–3760.
10. Bartenschlager R, Ahlborn-Laake L, Mous J, Jacobsen H. Kinetic and structural analyses of hepatitis C virus polyprotein processing. *J. Virol.* 1994; **68**: 5045–5055.
11. Lin C, Thomson JA, Rice CM. A central region in the hepatitis C virus NS4A protein allows formation of an active NS3–NS4A serine proteinase complex *in vivo* and *in vitro*. *J. Virol.* 1995; **69**: 4373–4380.
12. Tanji Y, Hijikata M, Satoh S, Kaneko T, Shimotohno K. Hepatitis C virus-encoded nonstructural protein NS4A has versatile functions in viral protein processing. *J. Virol.* 1995; **69**: 1575–1581.
13. Love RA, Parge HE, Wickersham JA, Hostomsky Z, Habuka N, Moomaw EW, Adachi T, Homstomska Z. The crystal structure of hepatitis C virus NS3 proteinase reveals a trypsin-like fold and a structural zinc binding site. *Cell* 1996; **87**: 331–342.
14. Kim JL, Morgenstern KA, Lin C, Fox T, Dwyer MD, Landro JA, Chambers SP, Markland W, Lepre CA, O'Malley ET, Harbeson SL, Rice CM, Murcko MA, Caron PR, Thomson JA. Crystal structure of the hepatitis virus NS3 proteinase domain complexed with a synthetic NS4A cofactor peptide. *Cell* 1996; **87**: 343–355.

 

15. Yan Y, Li Y, Munshi S, Sardana V, Cole J, Sardana M, Steinkühler C, Tomei L, De Francesco R, Kuo L, Chen Z. Complex of NS3 proteinase and NS4A peptide of BK strain hepatitis C virus: a 2.2 Å resolution structure in a hexagonal crystal form. *Protein Sci*. 1998; **7**: 837–847.
16. Yao N, Reichert P, Taremi SS, Prosise WW, Weber PC. Molecular views of viral polyprotein processing revealed by the crystal structure of the hepatitis C virus bifunctional protease–helicase. *Structure* 1999; **7**: 1353–1363.
17. Hirowatari Y, Hijikata M, Tanji Y, Nyunoya H, Mizushima H, Kimura K, Tanaka T, Kato N, Shimotohno K. Two proteinase activities in HCV polypeptide expressed in insect cells using baculovirus vector. *Arch. Virol*. 1993; **133**: 349–356.
18. D'Souza ED, Grace AK, Sangar DV, Rowlands DJ, Clarke BE. *In vitro* cleavage of hepatitis C virus polyprotein substrates by purified recombinant NS3 protease. *J. Gen. Virol*. 1995; **76**: 1729–1736.
19. Steinkühler C, Tomei L, De Francesco R. *In vitro* activity of hepatitis C virus protease NS3 purified from recombinant baculovirus-infected Sf9 cells. *J. Biol. Chem*. 1996; **271**: 6367–6373.
20. Steinkühler C, Urbani A, Tomei L, Biasiol G, Sardana M, Bianchi E, Pessi A, De Francesco R. Activity of purified hepatitis C virus proteinase NS3 on peptide substrates. *J. Virol*. 1996; **70**: 6694–6700.
21. Lin C, Thomson JA, Rice CM. A central region in the hepatitis C virus NS4A protein allows formation of an active NS3-NS4A serine proteinase complex *in vivo* and *in vitro*. *J. Virol*. 1995; **69**: 4373–4380.
22. Koch JO, Lohmann V, Herian U, Bartenschlager R. In vitro studies on the activation of the hepatitis C virus NS3 proteinase by the NS4A cofactor. *Virology* 1996; **221**: 54–66.
23. Butkiewicz NJ, Wendel M, Zhang R, Jubin R, Pichardo J, Smith EB, Hart AM, Ingram R, Durkin J, Mui PW, Murray MG, Ramanathan L, Dasmahapatra B. Enhancement of hepatitis C virus NS3 proteinase activity by association with NS4A-specific synthetic peptides: identification of sequence and critical residues of NS4A for the cofactor activity. *Virology* 1996; **225**: 328–338.
24. Shimizu Y, Yamaji K, Masuho Y, Yokota T, Inoue H, Sudo K, Satoh S, Shimotohno K. Identification of the sequence of NS4A required for enhanced cleavage of the NS5A/5B site by hepatitis C virus NS3 proteinase. *J. Virol*. 1996; **70**: 127–132.
25. Sali DL, Ingram R, Wendel M, Gupta D, McNemar C, Tsarbopoulos A, Chen JW, Hong Z, Risano C, Zhang R, Yao N, Kwong AD, Ramanathan L, Le HV, Weber PC. Serine protease of hepatitis C virus expressed in insect cells as the NS3/4A complex. *Biochemistry* 1998; **37**: 3392–3401.
26. Gallinari P, Brennan D, Nardi C, Brunetti M, Tomei L, Steinkühler C, De Francesco R. Multiple enzymatic activities associated with recombinant NS3 protein of hepatitis C virus. *J. Virol*. 1998; **72**: 6758–6769.
27. Dimasi N, Pasquo A, Martin F, Di Marco S, Steinkühler C, Cortese R, Sollazzo M. Engineering, characterization and phage-display of hepatitis C virus NS3 protease and NS4A cofactor peptide as a single chain protein. *Prot. Eng*. 1998; **11**: 1257–1265.
28. Pasquo A, Nardi MC, Dimasi N, Tomei L, Steinkühler C, Delmastro P, Tramontano A, De Francesco R. Rational design and functional expression of a constitutively active single-chain NS4A-NS3 proteinase. *Folding Des*. 1998; **3**: 433–441.
29. We follow the nomenclature of Schechter and Berger [Schechter I, Berger A. On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun*. 1967; **27**: 157–162] in designating the cleavage sites as $P_6$–$P_5$–$P_4$–$P_3$–$P_2$–$P_1$↓$P'_1$–$P'_2$–$P'_3$–$P'_4$, with the scissile bond between $P_1$ and $P'_1$ and the *C*-terminus of the substrate on the prime site.
30. Zhang R, Durkin J, Windsor WT, McNemar C, Ramanathan L, Le HV. Probing the substrate specificity of hepatitis C virus NS3 serine proteinase by using synthetic peptides. *J. Virol*. 1997; **71**: 6208–6213.
31. Urbani A, Bianchi E, Narjes F, Tramontano A, De Francesco R, Steinkühler C, Pessi A. Substrate specificity of the hepatitis C virus serine proteinase NS3. *J. Biol. Chem*. 1997; **272**: 9204–9209.
32. Landro JA, Raybuck S A, Luong YPC, O'Malley ET, Harbeson SL, Morgenstern KA, Rao G, Livingston DJ. Mechanistic role of an NS4A peptide cofactor with the truncated NS3 proteinase of hepatitis C virus: elucidation of the NS4A stimulatory effect via kinetic analysis and inhibitor mapping. *Biochemistry* 1997; **36**: 9340–9348.
33. Steinkühler C, Biasiol G, Brunetti M, Urbani U, Koch U, Cortese R, Pessi A, De Francesco R. Product inhibition of the hepatitis C virus NS3 proteinase. *Biochemistry* 1998; **37**: 8899–8905.
34. Ingallinella P, Altamura S, Bianchi E, Taliani M, Ingenito R, Cortese R, De Francesco R, Steinkühler C, Pessi A. Potent peptide inhibitors of human hepatitis C virus NS3 proteinase are obtained by optimizing the cleavage products. *Biochemistry* 1998; **37**: 8906–8914.
35. Llinàs-Brunet M, Bailey M, Fazal G, Goulet S, Halmos T, Laplante S, Maurice R, Poirier M, Poupart MA, Thibeault D, Wernic D, Lamarre D. Peptide-based inhibitors of the hepatitis C virus serine protease. *Bioorg. Med. Chem. Lett*. 1998; **8**: 1713–1718.
36. Narjes F, Koehler C, Steinkühler C, Brunetti M, Altamura S, De Francesco R, Pessi A, Koch U, Matassa VG. A designed mimetic of the $P_1$ cysteine for substrate and inhibitors of the hepatitis C virus NS3 serine protease. Potent, reversible, covalent and non-covalent competitive inhibitors. Submitted, 2000.
37. Narjes F, Brunetti M, Colarusso S, Gerlach B, Biasiol G, Fattori D, De Francesco R, Matassa VG, Steinkühler C. Chetoacids are potent slow binding inhibitors of

 

the hepatitis C virus NS3 protease. *Biochemistry* 2000; **39**: 1849–1861.
38. Di Marco S, Rizzi M, Volpari C, Walsh M, Narjes F, Colarusso S, De Francesco R, Matassa VG, Sollazzo M. Inhibition of the hepatitis C virus NS3/4A protease. The crystal structure of two protease-inhibitor complexes. *J. Biol. Chem.* 2000; **275**: 7152–7157.
39. Barbato G, Cicero DO, Cordier F, Narjes F, Gerlach G, Sambucini S, Grzesiek S, Matassa VG, De Francesco R, Bazzo R. Inhibitor binding induces active site stabilization of the HCV NS3 protein serine protease domain. *EMBO J.* 2000; **19**: 1195–1206.
40. Fattori D, Urbani A, Brunetti M, Ingenito R, Pessi A, Prendergast K, Narjes F, Matassa VG, De Francesco R, Steinkühler C. Probing the active site of the hepatitis C virus serine protease by fluorescence resonance energy transfer. *J. Biol. Chem.* 2000; **275**: 15106–15113.
41. Pizzi E, Tramontano A, Tomei L, La Monica N, Failla C, Sardana M, Wood T, De Francesco R. Molecular model of the specificity pocket of the hepatitis C virus proteinase: implications for substrate recognition. *Proc. Natl. Acad. Sci. USA* 1994; **91**: 888–892.
42. Burley SK, Petsko GA. Amino-aromatic interactions in proteins. *FEBS Lett.* 1986; **203**: 139–143.
43. Levitt M, Perutz MF. Aromatic rings act as hydrogen bond acceptors. *J. Mol. Biol.* 1988; **201**: 751–754.
44. Kolykhalov AA, Agapov EV, Rice CM. Specificity of the hepatitis C virus NS3 serine protease: effects of substitutions at the 3/4A, 4A/4B, 4B/5A, and 5A/5B cleavage sites on polyprotein processing. *J. Virol.* 1994; **68**: 7525–7533.
45. Bartenschlager R, Ahlborn-Laake L, Yasargil K, Mous J, Jacobsen H. Substrate determinants for cleavage in cis and in trans by the hepatitis C virus NS3 proteinase. *J. Virol.* 1995; **69**: 198–205.
46. Cicero DO, Barbato G, Koch U, Ingallinella P, Bianchi E, Nardi MC, Steinkühler C, Cortese R, Matassa VG, De Francesco R, Pessi A, Bazzo R. Structural characterization of the interactions of optimized product inhibitors with the *N*-terminal proteinase domain of the hepatitis C virus (HCV) NS3 protein by NMR and modeling studies. *J. Mol. Biol.* 1999; **289**: 385–396.
47. LaPlante SR, Cameron DR, Aubry N, Lefebvre S, Kukolj G, Maurice R, Thibeault D, Lamarre D, Llinàs-Brunet M. Solution structure of substrate-based ligands when bound to hepatitis C virus NS3 protease domain. *J. Biol. Chem.* 1999; **274**: 18618–18624.
48. Hahm B, Han DS, Back SH, Song O-K, Cho M-J, Kim C-J, Shimotohno K, Jang SK. NS3-4A of hepatitis C virus is a chymotrypsin-like protease. *J. Virol.* 1995; **69**: 2534–2539.
49. Bouffard P, Bartenschlager R, Ahlborn-Laake L, Mous J, Roberts N, Jacobsen H. An *in vitro* assay for hepatitis C virus NS3 serine proteinase. *Virology* 1995; **209**: 52–59.
50. Kassel DB, Green MD, Wehbie RS, Swanstrom R, Berman J. HIV-1 protease specificity derived from a complex mixture of synthetic substrates. *Anal. Biochem.* 1995; **228**: 259–266.
51. Sardana VV, Wolfgang JA, Veloski CA, Long WJ, LeGrow K, Wolanski B, Emini EA, LaFemina RL. Peptide substrate cleavage specificity of the human cytomegalovirus protease. *J. Biol. Chem.* 1994; **269**: 14337–14340.
52. Darke PL, Chen E, Hall DL, Sardana MK, Veloski CA, LaFemina RL, Shafer JA, Kuo LC. Purification of active herpes simplex virus-1 protease expressed in *Escherichia coli*. *J. Biol. Chem.* 1994; **269**: 18708–18711.
53. Bianchi E, Steinkühler C, Taliani M, Urbani A, De Francesco R, Pessi A. Synthetic depsipeptide substrates for the assay of human hepatitis C virus protease. *Anal. Biochem.* 1996; **237**: 239–244.
54. Cerretani M, Di Renzo L, Serafini S, Vitelli A, Gennari N, Bianchi E, Pessi A, Urbani A, Colloca S, De Francesco R, Steinkühler C, Altamura S. A high-throughput radiometric assay for hepatitis C virus NS3 protease. *Anal. Biochem.* 1999; **266**: 192–197.
55. Taliani M, Bianchi E, Narjes F, Fossatelli M, Urbani A, Steinkühler C, De Francesco R, Pessi A. A continuous assay of hepatitis C virus protease based on resonance energy transfer depsipeptide substrates. *Anal. Biochem.* 1996; **240**: 60–67.
56. Taliani M, Bianchi E, Pessi A. An improved solid-phase synthesis of resonance energy transfer fluorescent peptides and depsipeptides employing the EDANS/DABCYL donor–acceptor pair. *Lett. Peptide Sci.* 1997; **4**: 101–106.
57. Walker MA. Hepatitis C virus: an overview of current approaches and progress. *Drug Discov. Today* 1999; **4**: 518–529.
58. Segel IH. *Enzyme Kinetics*. Wiley: New York, 1993.
59. Schellenberger V, Turck CW, Rutter WJ. Role of the S′ subsites in serine protease catalysis. Active-site mapping of rat chymotrypsin, rat trypsin, $\alpha$-lytic protease, and cercarial protease from *Schistosoma mansoni*. *Biochemistry* 1994; **33**: 4251–4257.
60. James MN, Sielecki AR, Brayer GD, Delbaere LT, Bauer CA. Structures of product and inhibitor complexes of Streptomyces griseus protease A at 1.8 Å resolution. A model for serine protease catalysis. *J. Mol. Biol.* 1980; **144**: 43–88.
61. Tong L, Wengler G, Rossmann MG. Refined structure of Sindbis virus core protein and comparison with other chymotrypsin-like serine proteinase structures. *J. Mol. Biol.* 1993; **230**: 228–247.
62. Nienaber VL, Mersinger LJ, Kettner CA. Structure-based understanding of ligand affinity using human thrombin as a model system. *Biochemistry* 1996; **35**: 9690–9699.
63. Llinàs-Brunet M, Bailey M, Déziel R, Fazal G, Gorys V, Goulet S, Halmos T, Maurice R, Poirier M, Poupart M-A, Rancourt J, Thibeault D, Wernic D, Lamarre D. Studies on the C-terminal of hexapeptide inhibitors of the hepatitis C virus serine protease. *Bioorg. Med. Chem. Lett.* 1998; **8**: 2719–2724.

 

64. Bianchi E, Orrù S, Dal Piaz F, Ingenito R, Casbarra A, Biasiol G, Koch U, Pucci P, Pessi A. Conformational changes in human hepatitis C virus NS3 protease upon binding of product-based inhibitors. *Biochemistry* 1999; **38**: 13844–13852.
65. Han W, Hu Z, Jiang X, Decicco CP. α-chetoamides, α-chetoesters and α-dichetones as HCV NS3 protease inhibitors. *Bioorg. Med. Chem. Lett.* 2000; **10**: 711–713.
66. Ede NJ, Eagle SN, Wickham G, Bray A, Warne B, Shoemaker K, Rosenberg S. Solid phase synthesis of peptide aldehyde protease inhibitors. Probing the proteolytic sites of hepatitis C virus polyprotein. *J. Peptide Sci.* 2000; **6**: 11–18.
67. Yokosawa H, Nishikata M, Ishii S. *N*-Benzyloxycarbonyl-valyl-prolinal, a potent inhibitor of post-proline cleaving enzyme. *J. Biochem.* 1984; **95**: 1819–1821.
68. Shuman RT, Rothenberger RB, Campbell CS, Smith GF, Gifford-Moore DS, Gesellchen PD. Highly selective tripeptide thrombin inhibitors. *J. Med. Chem.* 1993; **36**: 314–319.
69. Bianchi E, Urbani A, Biasiol G, Brunetti M, Pessi A, De Francesco R, Steinkühler C. Complex formation between the hepatitis C virus serine protease and a synthetic NS4A cofactor peptide. *Biochemistry* 1997; **36**: 7890–7897.
70. Barbato G, Cicero DO, Nardi MC, Steinkühler C, Cortese R, De Francesco R, Bazzo R. The solution structure of the *N*-terminal proteinase domain of the hepatitis C virus (HCV) NS3 protein provides new insights into its activation and catalytic mechanism. *J. Mol. Biol.* 1999; **289**: 370–384.
71. Scaloni A, Miraglia N, Orrù S, Amodeo P, Motta A, Marino G, Pucci P. Topology of the calmodulin-melittin complex. *J. Mol. Biol.* 1998; **277**: 945–958.
72. Orrù S, Dal Piaz F, Casbarra A, Biasiol G, De Francesco R, Steinkühler C, Pucci P. Conformational changes in the NS3 protease from hepatitis C virus strain Bk monitored by limited proteolysis and mass spectrometry. *Protein Sci.* 1999; **8**: 1445–1454.
73. Urbani A, Biasiol G, Brunetti M, Volpari C, Di Marco S, Sollazzo M, Orrù S, Piaz FD, Casbarra A, Pucci P, Nardi C, Gallinari P, De Francesco R, Steinkühler C. Multiple determinants influence complex formation of the hepatitis C virus NS3 protease domain with its NS4A cofactor peptide. *Biochemistry* 1999; **38**: 5206–5215.
74. LaPlante SR, Bonneau PR, Aubry N, Cameron DR, Déziel R, Grand-Maître C, Plouffe C, Tong L, Kawai SH. Characterization of the human cytomegalovirus protease as an induced-fit serine protease and the implications to the design of mechanism-based inhibitors. *J. Am. Chem. Soc.* 1999; **121**: 2974–2986.
75. Bianchi E, Ingenito R, Simon RJ, Pessi A. Engineering and chemical synthesis of a transmembrane protein: the HCV protease cofactor NS4A. *J. Am. Chem. Soc.* 1999; **121**: 7698–7699.
76. Kent SBH. Chemical synthesis of peptides and proteins. *Annu. Rev. Biochem.* 1988; **57**: 957–989.
77. Bianchi E, Venturini S, Pessi A, Tramontano A, Sollazzo M. High level expression and rational mutagenesis of a designed protein, the minibody. *J. Mol. Biol.* 1994; **236**: 649–659.
78. Dawson PE, Muir TW, Clark-Lewis I, Kent SBH. Synthesis of proteins by native chemical ligation. *Science* 1994; **266**: 776–779.
79. De Francesco R, Urbani A, Nardi MC, Tomei L, Steinkühler C, Tramontano A. A zinc binding site in viral serine proteinases. *Biochemistry* 1996; **35**: 13282–13287.
80. Urbani A, Bazzo R, Nardi MC, Cicero D, De Francesco R, Steinkühler C, Barbato G. The metal binding site of the hepatitis C virus NS3 proteinase. A spectroscopic investigation. *J. Biol. Chem.* 1998; **273**: 18760–18769.
81. Ingallinella P, Bianchi E, Ingenito R, Koch U, Steinkühler C, Altamura S, Pessi A. Optimization of the P′-region of peptide inhibitors of hepatitis C virus NS3/4A protease. *Biochemistry* 2000; **39**: 12898–12906.
82. Koch U, Biasiol G, Brunetti M, Fattori D, Pallaoro M, Steinkühler C. The role of charged residues in the catalytic mechanism of the hepatitis C virus NS3 protease: electrostatic pre-collision guidance and transition state stabilization. *Biochemistry* 2000, in press.

 