Available online at www.sciencedirect.com
ScienceDirect
Journal of Molecular Graphics and Modelling 26 (2008) 1131–1144
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

# Exploring the P2 and P3 ligand binding features for Hepatitis C virus NS3 protease using some 3D QSAR techniques

Hsin-Yuan Wei, Chien-Sheng Lu, Thy-Hou Lin *

*Institute of Molecular Medicine & Department of Life Science, National Tsing Hua University, Hsinchu 30013, Taiwan*

Received 10 August 2007; received in revised form 28 September 2007; accepted 3 October 2007
Available online 9 October 2007

**Abstract**

Several three-dimensional quantitative structure–activity relationship (3D-QSAR) models have been constructed using the comparative molecular field analysis (CoMFA), comparative molecular similarity indices analysis (CoMSIA), and Catalyst pharmacophore feature building programs for a series of 26 truncated ketoacid inhibitors designed particularly for exploring the P2 and P3 binding pockets of HCV NS3 protease. The structures of these inhibitors were built from a structure template extracted from the crystal structure of HCV NS3 protease. The structures were aligned through docking each inhibitor into the NS3 active site using program GOLD. The best CoMSIA model was identified from the stepwise analysis results and the corresponding pharmacophore features derived were used for constructing a pharmacophore hypothesis by the Catalyst program. Pharmacophore features obtained by CoMFA and CoMSIA are found to be in accord with each other and are both mapped onto the molecular 5 K surface of NS3 active site. These pharmacophore features were also compared with those obtained by the Catalyst program and mapped onto the same NS3 molecular surface. The pharmacophore building process was also performed for 20 boronic acid based NS3 inhibitors characterized by a long hydrophobic side chain attached at position P2. This latter pharmacophore hypothesis built by the Catalyst program was also mapped onto the molecular surface of NS3 active site to define a second hydrophobic feature at position P2. The possibility of using the pharmacophore features mapped P2 and P3 binding pocket to design more potent depeptidized NS3 inhibitors was discussed.


*Keywords:* 3D QSAR; CoMFA; CoMSIA; Catalyst; Structure alignment; GOLD; Molecular docking; Pharmacophore model

## 1. Introduction

The hepatitis C virus (HCV) infects approximately 170,000,000 people worldwide [1]. The virus is transmitted primarily by blood and blood products. Most of the infected individuals have either received blood transfusions or have used intravenous drugs [1,2]. The HCV infection is more common in sexually promiscuous individuals but rare in monogamous couples [3]. Perinatal transmission from mother to fetus or infant is also relatively low but possible (less than 10%) [2,3]. Many individuals infected with HCV have no obvious risk factors [3]. Most of these persons have probably been inadvertently exposed to contaminated blood or blood products [3].

As a member of *Flaviviridae*, the complete genomes of various HCV isolates were cloned and sequenced by several groups [4–7]. The HCV genome is a positive, single-stranded RNA of approximately 10,000 nucleotides and encodes a single polyprotein of about 3010 amino acids [8,9]. The polyprotein is processed by host cell and two viral proteases NS2/3 and NS3 into ten different products, with the structural proteins (core (C), E1 and E2) located in the N-terminal third and the nonstructural (NS2-5) replicative proteins in the remainder [10–12]. The role of NS2/3 protease appears to be limited to the autoproteolytic cleavage of the NS2-NS3 junction in *cis* [13]. The amino-terminal 180-amino-acid sequence of NS3 protein encodes a serine protease which cleaves at the NS3/4A junction in *cis*, which is followed by cleavage at the NS4/4B, NS4B/5A, and NS5A/B sites in *trans* [13]. The NS3 serine protease requires an accessory viral protein, NS4A, for optimal cleavage activity [14]. The contribution of NS4A to NS3 protease activity can be mimicked by a synthetic peptide encompassing amino acid residues 21–34 of NS4 [14,15].

The NS3 protease has been considered as one of the most attractive targets for anti-HCV therapy because it is essential

* Corresponding author. Fax: +886 3 571 5934.
*E-mail address:* thlin@life.nthu.edu.tw (T.-H. Lin).

1093-3263/$ – see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2007.10.005

for viral replication and formation of infectious viral particles [5,7]. A 2.5 Å resolution structure of the NS3:NS4A complex reveals that the HCV NS3 serine protease domain adopts a chymotrypsin-like fold. The complex consists of two structural domains, each containing a twisted β sheet incorporating a ''Greek key'' motif while the C-terminal domain (residues 1120–1206) contains the conventional six-stranded β barrel, common to nearly all members of the chymotrypsin family, followed by a structurally conserved α helix [13]. The catalytic triad of His57, Asp81 and Ser139 is located in a crevice between the two domains [13]. The geometrical arrangement of the catalytic triad is similar to that of other serine proteases. Furthermore, a Zn ion is tetrahedrally coordinated with Cys97, Cys99, Cys145 and His149 residues of a site located opposite to the active site [13]. The Zn ion is believed to play a structural role because the removal of the ion causes unfolding and precipitation of the protein [13].

The NS3-dependent cleavage sites have been mapped by using amino-terminal sequence analysis of mature proteins expressed in cell culture [16]. The substrate specificity of the proteinase has been identified by comparing the cleavage sites on the HCV sequence. Sequences of these cleavage sites have been generalized as D/E-X-X-X-X-Cys/Thr↓Ser/Ala-X-X-L/W/Y, with cleavage occurring after cysteine or threonine [17]. According to the nomenclature of Schechter and Berger, [18] the newly generated carboxy terminus, after cleavage of the peptide bond, is designated P1, and it is preceded by the P2 residue, etc.; the newly generated amino terminus is designated P1', and it is followed by P2', etc. [19]. The preferred P1 residue is cysteine with an exception at the intramolecular NS3/4A junction where a threonine residue is preferred. Other conserved features are a negatively charged residue in the P6 position, an alanine or a serine in P1' and a hydrophobic residue in P4' position [19]. The corresponding binding subsites on the enzyme surface are denoted as S6 through S4' [20]. Many of the NS3 inhibitors designed are decapeptide or hexapeptide substrate analogues of the N-terminal cleavage (P6-P1) product [21]. Replacement of the P1 cysteine with other moieties, e.g. (S)-4,4-difluoro-2-aminobutyric acid and related 'serine trap' functional groups (cysteine mimetics) and substitution of the P3 unit with indoline-based structures (peptidomimetics) are the two most common approaches used for designing these inhibitors. Some rational designs such as the alternate P1 thiol chemistry in accordance with the lipophilicity of the S1 pocket formed by Leu135, Phe154 and Ala157, or the introduction of amphiphilic groups to indoline peptidomimetics in response to the presence of Arg161 and Lys136 proximal to the S3 site are conducted based on the NS3 structure determined [21]. The crystal structure of NS3/4A complex with either of two covalently bound α-ketoacid serine trap inhibitors determined by Di Marco et al. [22] reveals that the carbamate-protected tripeptide adducts bind in an extended conformation, occupying the S1–S4 sites and forming an antiparallel β-sheet with the E2 strand (Ala156-Val170). A similar overall binding configuration for a hexapeptide boronic acid inhibitor bound to NS3/4A was observed by NMR [23]. The boronic acid cysteine residue was found to interact with the active site serine residue such that the P1 region of the bound inhibitor mimics the transition state of substrate hydrolysis [23].

A novel class of NS3 protease inhibitors has been made based on the C-terminal tetrapeptide cleavage product (P1'–P4') [17]. However, the most potent inhibitors reported to date contain either a 4-substituted proline or a 3,4-disubstituted proline as P2 residue [24]. The potency of these inhibitors are further enhanced through a depeptidize process using 2-azabicyclo [2.2.1]-heptane carboxylic acid as a surrogate [24]. A bicyclic pyrimidinone-based P2–P3 dipeptide replacement has been incorporated into a peptide boronic acid inhibitor to increase the potency [25]. Oligopeptide derivatives containing α-ketoamide electrophilic trap are also potent inhibitors of NS3 as have been reported [26]. The potency of these P3-capped inhibitors is also increased through a depeptidization process [26]. Another series of highly potent inhibitors are designed based on the trisubstituted cyclopentane moiety at the P2 position [27]. The P2 benzene ring has been macrocyclized with the P3 capping group through an aryl-alkyl ether linkage to depeptize the P2–P3 moiety [28]. These are the conformationally preorganized inhibitors with better stability and strong potency [28–30]. Recently, a series of potent tripeptide truncated at the N-terminus from some hexapeptide ketoacid inhibitors are described to explore the P2–P3 binding features [31]. The $IC_{50}$ measured for the most potent inhibitor of this series is 0.38 μM. The N-terminal aminoacid of these α-ketotripeptide inhibitors are further replaced with α-hydroxy acid, leading to a series of capped dipeptide inhibitors [32]. The original P3 residue is replaced with a small hydroxylated one to alternate the original chirality at the P3 binding position from *S*-configuration to *R*-configuration [32]. Some even more potent β-sheet dipeptide mimetics aiming at the S2 binding pocket namely, 3-amino bicyclic pyrazinones, made from a hexapeptide boronic acid lead are also reported [25,33]. In general, the P2 binding feature is believed to be lipophilic while both hydrophobic and hydrophilic groups are accepted in P3. To characterize the P2–P3 binding features theoretically, we have employed several 3D QSAR techniques on the aligned structures of 26 truncated ketoacid inhibitors of NS3 designed by Nizi et al. [31] and Colarusso et al. [32] for constructing some 3D QSAR models for these compounds. The truncated ketoacid inhibitors are divided into the training and test set and the structures in each set are aligned and analyzed by the CoMFA [34] and CoMSIA [35] methods to derive the best 3D QSAR model for the truncated ketoacid inhibitors. Further, the pharmacophore features obtained from the best CoMSIA model are used to construct some pharmacophore hypotheses using the Catalyst 4.9 program [36]. The top hypothesis generated is mapped onto the structures of several highly active truncated ketoacid inhibitors selected from both the training and test sets. The predicted activities for both the training and test sets by the top hypothesis are found to be in good accord in statistics with those predicted by the best CoMSIA model. The pharmacophore hypotheses by Catalyst 4.9 program are also built for structures of 20 boronic acid based pyrimidinone and pyrazinone inhibitors made by Glunz et al. [25] and Zhang