Research Article **ISSN: 2349 – 7114**

Asian Journal of Research in Pharmaceutical Sciences and Biotechnology

Journal home page: www.ajrpsb.com

DIFFERENCES BETWEEN HCV NS3/4A PROTEASE FROM GENOTYPES 1-6: SUBSTRATE SPECIFICITY, SEQUENTIAL AND STRUCTURAL DIFFERENCES OF ACTIVE SITE STRUCTURE. IS IT POSSIBLE TO DESIGN A UNIVERSAL HCV PROTEASE INHIBITOR?

Tomasz Florczak[1], Iga Karolina Jodłowska[1], Klaudia Izabella Szmejda[1], Michał Benedykt Kaczmarek[1], Aneta Monika Białkowska[1*]

[1*]Institute of Technical Biochemistry, Department of Biotechnology and Food Sciences, Technical University of Lodz, Lodz, Poland.

ABSTRACT

The increasing knowledge about the hepatitis C virus (HCV) gives rise to the development of novel and promising antiviral therapies. Despite apparent advances in this field, the common inhibitor of the NS3/4A protease from all HCV genotypes has not been designed yet. The main cause of the lack of such an inhibitor are sequential and structural differences among NS3/4A protease variants from existing HCV subtypes. This article presents differences and similarities in the structure and substrate specificity of selected enzyme isoforms, which may help to synthesise a specific inhibitor against all of them. The analysis performed by the authors showed that the identity of RNA sequences encoding the NS3 proteins from these HCV subtypes ranges from 69 to 91% (73.4% on average) while the identity of corresponding amino acid sequences is much higher and ranges from 80 to 95% (85.5% on average). Despite some differences between the peptide cofactors (NS4A) of the proteases from the selected HCV genotypes, their structure is similar and the sequence identity of genes encoding the NS4A peptides from the selected HCV genotypes ranges from 59% to 91% (77.7% on average). The alignment of amino acid sequences of the NS4A peptide variants revealed that its middle fragment, which is linked to the NS3 protein, is the least conserved region. The knowledge of differences in the structure of substrate binding pockets of the NS3/4A protease isoforms, affecting the strength of interactions with the substrates and inhibitors, will help to design the universal inhibitor of these enzymes.

KEYWORDS

HCV protease, Sequential alignment, Universal inhibitor and Structural differences.

Author for Correspondence:
Aneta Monika Bialkowska,
Institute of Technical Biochemistry,
Department of Biotechnology and Food Sciences,
Lodz University of Technology, Lodz, Poland.
Email: aneta.bialkowska@p.lodz.pl

INTRODUCTION

According to the WHO data[1], approximately 130 - 150 million people suffer from the chronic hepatitis C infection, and 500 000 of them die each year from hepatitis C-related liver diseases. Despite the growing effectiveness of Direct-Acting Antivirals, the high prevalence of HCV infections has been still a serious global problem. The diversity of HCV

subtypes causes that elimination of infections and their treatment are very difficult. The known HCV variants were classified[2] into 7 genotypes that were divided into 67 subtypes. Noteworthy, only 6 HCV genotypes were presented in the majority of papers that were published before 2014 while the genotype 7 was mentioned in only a few works. The RNA heterogeneity of these genotypes and subtypes reaches 30-35%[3] and 25%[4], respectively. The prevalence of individual genotypes depends on the level of safety in healthcare and geographical conditions. According to WHO, in 2013 the genotypes 1 and 2 were most widely distributed round the world, and caused 46.2% and 9.1% of all infections, respectively. Genotype 3 was the second most prevalent (30.1% infections) and occurred mainly in Southern Asia (3/4 of the cases, mainly in India). Because of the prevalence, the subtypes 1a, 1b, 2a and 3a are regarded as epidemic ones, while the other subtypes are endemic since the area of relevant infections does not increase. The genotypes 4, 5, and 6 are mainly found in Central Africa (cause 8.3% infections), South Africa (<1%) and Southeast Asia (5.4%)[5].

HCV STRUCTURE AND REPLICATION CYCLE

The structure and replication cycle of all HCV genotypes are the same despite differences in the RNA sequences that are the basis of the HCV classification. The virus is built of a protein capsid covered by a lipid envelope which contains glycoproteins E1 and E2 that are responsible for the infectivity of the virus and are characterized by the highest sequential heterogeneity among all HCV proteins. The lipid envelope surrounds the protein capsid encapsulating a single strand of RNA. The HCV genome is a positively polarized ssRNA, approximately 9500 bp long, which consists of one long Open Reading Frame located between the two regions that do not undergo expression such as 5'NTR and 3'NTR[6]. The genome translation results in a polypeptide chain consisting of approximately 3037 amino acid residues, which contains 10 viral proteins in the following order: NH2-C-E1-E2-p7-NS2-NS3-NS4A-NS4B-NS5A-NS5B-COOH[7].

These proteins are either structural or non-structural. The structural proteins are located at the N terminus of the polypeptide chain. The envelope protein (C) and the glycoproteins E1 and E2 are building blocks of the envelope and the capsid of the HCV virus. The non-structural proteins located at the C terminus of the polypeptide chain are involved in the RNA replication, virus infectivity, proteolytic cleavage of the polypeptide chain, and the host cells' modification[8]. One of the most important viral protein is NS5B, playing the role of RNA-dependent RNA polymerase, catalysing the process of formation of a daughter RNA strand. The proteins from NS3 to NS5B take part in the virus RNA replication process, yet, for most of them this is not the only function. Among the non-structural proteins (NS), only p7 and NS2 are not involved in RNA replication. These two proteins are responsible for the ion channel formation and the autolytic removal of NS2, respectively, and it is known that they can interact with each other[7,9]. The protein NS3 exhibits the activity of RNA/DTP helicase, yet in the complex with NS4A peptide (NS3/4A) it shows the activity of serine protease[6]. NS4B is an endoplasmic protein which reorganizes the membrane. The protein NS5A is a phosphoprotein which is mainly involved in the RNA replication but its function is not fully specified[10].

The cleavage of the polypeptide chain after translation causes the release of the listed above viral proteins. The proteins located between the N-terminus and the protein P7 are liberated by the host's proteases. The protein NS2 has the activity of cysteine protease that enables its auto-removal. The remaining proteins are released by the NS3/4A protease, showing the auto-Proteolytic activity[4].

Recent advances in analytic techniques enabled characterization of HCV proteins that became potential targets for antiviral therapies. For instance, the applied for over a decade interferon Alpha (INF) and the pegylated interferon Alpha 2a (PEGINF) that were used along with RBV (Ribavirin) have been increasingly replaced by oral directly acting antivirals (DAAs). The DAAs are applied in novel 3D therapies, involving inhibitors of NS3/4A protease, NS5B polymerase and NS5A protein. However, in developing countries the majority of HCV infections are cured with a combination of

PEGINF and RBV, causing various side effects, from slight ones, like nausea and headaches, to severe ones, like vomiting, anaemia and depression. On the other hand, it is one of the most universal therapies, acting against all types and subtypes of the HCV virus. Therefore, it is so important to design a universal inhibitor that would act against all genotypes in a short time and with few side effects.

The virus replication cycle is relatively well understood[11] although there are still certain details, like interactions of viral proteins with each other or details concerning HCV entry, which are of a hypothetical nature. From the point of view of the antiviral therapy, the most important of them are: interactions of the virus with hepatocyte receptors[12,] endocytose and pH-dependent fusion[9], synthesis of precursor Polyprotein, Proteolytic cleavage of a polypeptide chain, viral RNA replication, capsid assembly, virion formation and exocytosis.

NS3/4A PROTEASE STRUCTURE, FORMATION

The release of non-structural proteins (NS) from the HCV polypeptide, which is catalysed by the NS3 and NS3/4A enzymes, is one of most important steps, deciding of the development of HCV infection. NS3 is a multifunctional protein, containing 631 amino acid residues. Its N-terminal serine protease domain consists of 180 amino acid residues while the C-terminal fragment contains the helicase and NTPase domains[4]. The attachment of a peptide cofactor - a fragment of NS4A protein - to the NS3 protein causes that the activity of helicase is lost while the proteolytic activity is increased tenfold[13]. From the structural point of view, the NS3/4A protease belongs to the family of trypsin/chymotrypsin-like proteases[14]. This protein is responsible for the hydrolytic cleavage of the HCV polypeptide in four sites: NS3-NS4A, NS4A-NS4B, NS4B-NS5A, and NS5A-NS5B. The hydrolysis of the peptide bond between NS3 and NS4A proteins is the intramolecular process. The hydrolytic cleavage of the linkages in the 3 remaining sites was shown to be specific for the trans conformation of the peptide bond. It was also found that this protease is capable of carrying out proteolysis of some of the host's proteins, including MAVS (mitochondrial antiviral-signalling protein, known as IPS1) and TRIF (TRI domain-containing adapter inducing INFβ) that are responsible for induction of immune response[4].

The activity of NS3/4A protease depends on the peptide cofactor (Figure No.1a) and a zinc atom (Figure No.1b). NS4A is a peptide consisting of 54 amino acid residues. The first 20 amino acid residues are highly hydrophobic and form a trans-membrane alpha helix, which is responsible for anchoring the NS3/4A complex in the membrane of endoplasmic reticulum. The central domain of NS4A peptide (amino acid residues 21-34) directly interacts with NS3 protein that causes the rise in Proteolytic activity[15]. The C-terminal fragment of NS4A is responsible for interacting with the helicase domain and other viral proteins[4].

The 3D structure of NS3/4A serine protease of genotype 1a was resolved in 1996, using X-ray crystallography and NMR techniques. This enzyme is a Heterodimer and, like the other serine proteases from the trypsin/chymotrypsin family, has a structure of a double beta-barrel. The N-terminal subunit (the first 93 amino acid residues) consists of 8 beta strands, including seven strands of NS3 protein and one strand of NS4A cofactor. Strong interactions between NS3 and NS4A cause that the beta-barrel is tightly packed and the surface of the protease is increased. NS4A is non-covalently bound to the N-terminal subunit of NS3, fitting between the two beta strands of NS3 protein. The amino acid residues of the central part of NS4A, including Val23, Ile25, Ile/Val29 and Leu31, are involved in numerous hydrophobic interactions with many hydrophobic side chains of NS3. This complex is also stabilized by hydrogen bonds formed between the main chain carbonyl and amide groups of NS4A and NS3. The highly hydrophobic amino acid residues of peptide NS4A such as Ile25 and Ile/Val29, are essential for the interaction with peptide NS3 and for complete activation of serine protease[16].

The C-terminal sub domain (residues 96-180) consists of six beta strands and one conserved alpha helix (Figure No.1a). This domain binds the zinc atom, which is highly conserved in all closely related flavivirus proteases[16,17]. The zinc atom is tetrahedrally coordinated by Cys97, Cys99, Cys145, water molecule and His149 (Figure No.1b). The

zinc-binding pocket is located in a cleft between the sub domains, opposite the catalytic triad, that is very important for the catalytic function[18]. The catalytic triad (Figure No.1a) is located at the interface of the sub domains, and consists of: His57 and Asp81 (located in the N-terminal sub domain) and Ser139 (located in the C-terminal sub domain) [15].

BINDING OF A SUBSTRATE BY NS3/4A PROTEASE

The commonly applied model of protease's active centre, proposed by Schetcher and Berger[19], was used to describe interactions between the HCV protease and its substrates and inhibitors. The model substrate for NS3/4A protease is a decapeptide consisting of 6 amino acid residues at the N-terminus and four residues at the C-terminus, which can be presented as: NH2-P6-P5-P4-P3-P2-P1-P1'-P2'-P3'-P4'-COOH. The hydrolysis of this substrate gives two products, named the P-side product (NH2-P6-P5-P4-P3-P2-P1-OH) and the P'-side product (NH2-P1'-P2'-P3'-P4'-OH)[20]. The substrate specificity of NS3/4A is rather low and mainly determined by Cys in position P1 and a small side chain of the amino acid residue located in position P1'[4]. The study on the function of HCV protease showed that it recognized the following amino acid motif flanking the cleavage site: (Asp/Glu)-X-X-X-X-(Cys/Thr)‖(Ser/Ala)-X-X-X, where X means any amino acid residue[17]. The authors found that the most important amino acid residue, enabling recognition of the bond to be cleaved, is either Cys or Thr located in position P1 of the substrate. However, they have not proved that the specificity of the NS3/4A protease from all HCV genotypes is the same. Potential inhibitors of this enzyme are small, peptide-like compounds because the substrate-binding pocket is small and exposed to solvents. The interactions between NS3/4A protease from genotype 1a and its natural substrates were discussed in detail by Romano *et al*[21].

NS3/4A PROTEASE INHIBITORS - STRUCTURE AND INTERACTIONS

The analysis of the HCV replication cycle suggests that each of its steps can be the target of treatment. The successful application of HIV protease inhibitors caused that inhibition of the HCV protease became one of promising options. Despite the fact that also inhibitors of NS5B polymerase and NS5A proteins are used in anti-HCV therapies, the HCV protease inhibitors were a breakthrough in the search for a cure for this disease and opened a new era of oral DAAs. These inhibitors are Telaprevir and Boceprevir, both approved for use in 2011 by FDA. However, because of the low effectiveness, these inhibitors must be applied along with the pegylated interferon pegINF and RBV that enabled to increase the Sustained Virologic Response (SVR) against the HCV genotype 1 up to 75%. This therapy was also approved for "experienced" patients (who had already undergone an ineffective therapy) infected with the genotype 1. However, the enhanced side effects, a high genetic diversity of the virus, and numerous interactions with other drugs caused that this treatment requires further improvement.

Telaprevir and Boceprevir are alpha ketoamides, which reversibly form a covalent bond with Ser139. The resulting inactive enzyme-inhibitor complex dissociates very slowly. A common component of these two inhibitors is a heterocyclic derivative of proline in position P2, which in both cases has a similar spatial structure and is the most important structural element of the inhibitor's skeleton (Figure No.2a).

The second generation of NS3/4A protease inhibitors are compounds that non-covalently bind to this enzyme and are analogues of natural polypeptides or macrocyclic polypeptides. Their examples are: Simiprevir, Danoprevir, Asunaprevir, and Paritaprevir. These inhibitors contain a residue of hydroxyproline derivative in position P2. The substituted Hyp residue is larger than the heterocyclic proline derivative and better fits to the site S2 in the protease molecule (Figure No.2b).

The main problem encountered when the known NS3/4A protease inhibitors are administered to the patients is the increased tendency towards mutation within the protease gene, induced by the treatment. This results in its ineffectiveness, which is the reason why the inhibitors should be used in combination with other types of inhibitors/DAAs.

According to the accepted treatment regimens[11,22], the NS3/4A protease inhibitors are used only for HCV genotype 1. In the other cases, the following

combinations are applied: Sofosbuvir with Ribavirin (genotypes 1, 4, 5, and 6) or Sofosbuvir and Daclatasvir (Ribavirin can be additionally used) in therapy against genotypes 1, 2, and 3. These therapies are thus focused on the inhibition of NS5B protease or NS5A protein and polymerase.

METHODS

SEQUENCE ALIGNMENT

The sequence alignment was carried out by downloading files in FASTA format from Gen Bank, their conversion in accordance with the requirements of the software, generation of the alignment by the Clustal Omega Multiple sequence Alignment software for sequences of both the proteins and the nucleotides, and selection of the default settings. The generated alignment was visualized using the Jalview 2.9.0b2 software. The calculation of sequence identity was performed using the Blastp multiple alignment for the amino acid sequences or Blastn multiple alignment for the nucleotide sequences (National Centre for Biotechnology Information).

HOMOLOGY MODELLING

The protein homology models, based on the amino acid sequences of NS3 protease from the selected genotypes (in the form of FASTA files), were prepared using the protease from subtype 1a, i.e. 3LOX, as the matrix. The models were made using the server SWISS-MODEL - swiss model, expasy. org[23-26] in the automatic mode. The correctness of the models was checked using the software available on SWISS-MODEL, i.e. ModEval, ModFold, QMEAN server, ProQ2, and CAMEO.

RESULTS

Sequence differences between ns3/4a protease variants from the selected hcv genotypes, substrate binding amino acid residues

To design of a common inhibitor, interacting with the NS3/4A protease variants from different HCV genotypes, it is necessary to analyse nucleotide sequences identity in the genes encoding these proteins and analyse amino acid sequences identity as well. Location of the active site suggests that the sequence of the NS3 apoenzyme is most important in terms of the inhibitor binding. The results of analysis of identity of nucleotide sequences of RNA, encoding the NS3 protein variants from selected genotypes, such as (subtype name - Gen Bank number) 1a - EF154704.1, 1b - EF154705.1, 2a - EF154709.1, 3a - EF154713.1, 4a - EF154714.1, 5a - L29579.1, and 6a - DQ480523.1, are presented in Table No.1. The identity of these RNA sequences ranges from 69 to 91% (73.4% on average). Surprisingly, the identity of corresponding amino acid sequences of the NS3 proteins from these subtypes (1a - ABL96703.1, 1b - ABL96704.1, 2a - ABL96708.1, 3a - ABL96712.1, 4- ABL96713.1, 5 - AAA65789.1, and 6a - ABE98160.1) is much higher and ranges from 80 to 95% (85.5% on average).

Binding of the NS4A peptide to the NS3 protein is a prerequisite of the protease activity and therefore, differences in the sequences of this peptide variants were also analysed. The results of this analysis will help to design a synthetic peptide (an allosteric inhibitor), being the NS4A analogue, whose attachment to the NS3 protein will prevent folding to the active protease conformation. This in turn will be important for the anti-HCV therapy[27]. The results of analysis of the sequence identity of genes encoding the NS4A peptides from the selected HCV genotypes (sequence access numbers as in Figure No.3) are presented in Table No.2. The identity of gene sequences ranges from 59% to 91% (77.7% on average) while the alignment of amino acid sequences of the peptide variants (Figure No.3) enabled to find certain highly conserved regions, like the N- and C-terminal fragments of these peptides. However, the analysis of the NS3/4A protease structure (see chapter 2.2) showed that the middle fragment (amino acid residues 21-34) of the NS4A peptide binds to the NS3 protein. Surprisingly, this is the least conserved region of this peptide. Noteworthy, when the genotype 2 is not taken into consideration in the analysis, the percentages of gene and peptide sequences identity are significantly higher. The occurrence of conserved peptides in these genotypes (with an exception of genotype 2) may have therapeutic relevance (possibility of HCV treatment with the same allosteric inhibitor). The comparison of peptide sequences of genotype 2 and

the other ones, showed that the corresponding amino acid residues have a similar character, e.g. all the amino acid residues in position 29 (V, I, or L) have a branched aliphatic side group. The similar homology can be found in the other positions, with an exception of position No.30 in subtypes 2a and 3a, which contain a charged amino acid residue (either basic (2a) or acidic (3a)), in contrast to the other subtypes, which contain a neutral amino acid residue in this position.

SUBSTRATE SPECIFICITY OF NS3/4A PROTEASES FROM DIFFERENT GENOTYPES

To design a competitive inhibitor of NS3/4A protease from all HCV genotypes it is necessary to determine the substrate specificity of each enzyme variant and find these amino acid residues that are the same at specific substrate positions. The substrate specificities of NS3 and NS3/4A proteases from the selected HCV subtypes are shown in figure no. 4a and 4b-4d, respectively (the natural cleavage sites in the polypeptides from the selected virus subtypes were shown using their fragments consisting of 20 amino acid residues, including the sequence recognized by the enzyme). The presented data suggest that the amino acid residues that are recognized by NS3 in peptide between NS3 and NS4A are highly conserved both at the carboxyl and the amino side of the scissile bond, with an exception of positions P2, P8 and P6' (the two latter positions are not recognized by the enzyme). Interestingly, certain non-conserved amino acid residues are homological in these positions. For instance, residues of S or Q that contain uncharged side groups are located in position P8, residues of I, V, or M (hydrophobic side groups) are located in position P2, while residues of V, A, or L (aliphatic amino acids) are located in position P6'. The opposite situation is observed in case of the bonds cleaved by the NS3/4A protease, which recognizes various conserved regions. In case of the cleavage site NS4A-NS4B, the conserved amino acid residues are P7-P1' (in fact, only P6-P1' are recognized). In the cleavage site NS4B-NS5A, the sites that are the same in known genotypes are P1, P2, P6, and P4', as well as P9, P10, P9', and P5', which are not recognized by the protease.

A small number of conserved amino acid residues in the cleavage site may decide of the strength of interactions between this peptide and the NS3/4A protease. For instance, apart from the S and A residues that were previously described in the literature, also D, T and N residues were found in position P1'. This was confirmed by the study of Romano *et al*[21]. Who found that the strength of interactions between the peptide, which is located at the interface of proteins NS4B and NS5A, and the protease is ten times lower than the strength of interactions with the other polypeptide fragments from the genotype 1. The opposite location of conserved amino acid residues than in case of the cleavage site NS4A-NS4B is observed in the cleavage site NS5A-NS5B. In the latter case, the conserved amino acid residues are located on the opposite side of the cleavage site, i.e. between position P3 and P10'. All these findings suggest that the universal inhibitor of NS3/4A protease should contain the conserved regions of substrates that are recognized by this enzyme. Furthermore, this fragment of the inhibitor, which is located on the acyl side of the peptide cleavage site (P1-P6) should be similar to the site NS4A-4B while the fragment interacting with S1' should be alike the P1' site (taking into account its surrounding) in the polypeptide region NS5A-5B.

The comparison of amino acid residues that are conserved in the known HCV polypeptide cleavage sites (Figure No.5) revealed that either the same or similar amino acid residues are located in certain positions. In all the cases, serine residue is located in position P1'. Therefore, this fragment of the universal inhibitor, which will interact with the site S1' of protease, should be a residue of either this amino acid or a compound of a similar nature. All peptides recognized by the NS3/4A protease contain cysteine residue in position P1, but in molecules of the currently used inhibitors (Figure No.5, examples are Boceprevir and Asuneprevir) the site S1 is occupied by residues of more hydrophobic compounds, which only fit into the pocket S1 but do not generate any interactions. Noteworthy, the site P6 in the known polypeptides is always occupied by

a residue of an acidic amino acid. This suggests that efficient inhibitors of the protease should interact also with the site S6. However, the currently applied inhibitors interact with the NS3/4A protease only up to the site S4. The position P4 of known polypeptides (with an exception the cleavage site between NS5A and NS5B) and the position P5 (with an exception of the cleavage site NS3-NS4A) are occupied by polar amino acid residues such as: S, C, M, and E. Interestingly, the positions P3' and P4' are occupied by aromatic amino acid residues, containing either nitrogen (Trp and His) or oxygen (Tyr) atom. Their analogues may be incorporated into the molecule of a universal efficient inhibitor.

DIFFERENCES IN THE STRUCTURE OF NS3/4A ACTIVE SITE FROM GENOTYPES 1-6

To provide evidence that counterparts of the mentioned above, conserved amino acid residues contained in the HCV polypeptides should be contained in the universal HCV protease inhibitor, it is necessary to compare the structures of the enzyme's active site from the known genotypes (both the amino acid sequences and the 3D structure). This may help to model the position of the substrate in the catalytic pocket of proteases from different HCV genotypes and estimate the strength of substrate-enzyme interactions. This is particularly important when certain positions in different natural peptides digested by this enzyme are occupied by not the same but homologous amino acid residues. Because also the amino acid sequence of the enzyme may decide of interactions with the substrate, the differences between the HCV genotypes have to be taken into consideration (Figure No.6). The objective of this analysis is to find differences in the structure of NS3/4A protease variants that in turn will help to design an inhibitor maintaining the naturally occurring interactions with the enzyme, generating an additional site or/and more exactly fitting into the substrate binding sites in the enzymes from all known HCV genotypes.

As it was mentioned in chapter 4.1, despite some differences between the peptide cofactors (NS4A) of proteases from the known HCV genotypes, their structure is similar. Particularly important are the Ile and Ile/Val residues that always occupy positions 25 and 29 of the peptide cofactor, respectively, which are crucial for the NS4A - NS3 interactions. This suggests that the homologous structures generated for the NS3 proteins from genotypes 2-6, using the NS3/4A complex from subtype 1a as a matrix, will be folded into the active conformation of serine protease like the matrix protein. Because it is likely that the subtle differences in the structure of NS4A peptides may affect the complex structures, the complete analysis of substrates/inhibitors docking in the generated models was omitted in this work. However, the differences in the structure of the substrate binding pockets and their impact on the putative interactions with the substrate amino acid residues were discussed (Figure No.7).

Amino acid residues Thr38 and Ala39 are contained in the NS3/A4 protease from all selected subtypes, with an exception of 4a, in which they are replaced by Ser38 and Val39. Although the corresponding amino acid residues have the same character, the side groups of Thr and Val are larger than those of Ser and Ala. The model presented in Figure No.7 shows that the change of Thr/Ser does not affect the shape of the catalytic pocket since the methyl group differentiating these amino acids is directed to the interior of the protein molecule and the hydroxyl group is exactly in the same site. However, the change of Ala/Val renders the 4a subtype protein surface more convex in the site (Figure No.7b), which is located at the end of the substrate recognition pocket (site S4'). Therefore, the latter changes may not affect interactions with the substrates/inhibitors. The opposite situation is observed in case of the pocket S2' because the enzyme variants from subtypes 1b and 4a contain Ser42 while the counterparts from the other analysed subtypes contain Thr42. The analysis of Figure No.7c shows that the surface of enzyme molecule containing Thr42 is more convex with an exposed oxygen atom. This structural change may affect formation of hydrogen bonds in this site. The significant changes in the 3D structure are caused by the replacement of Arg123 with Thr123, observed only in the subtype 3a, presented in Figure No.7d. This mutation causes that the convex surface of the alkaline character and bearing the positive charge

changes into the convex surface with an exposed neutral oxygen atom. According to Romano *et al*[21], depending on the substrate, Arg123 interacts either indirectly through Asp168 with the site P1 of the substrate or directly with the acidic substrate substituent in site P4. It may also interact along with Lys165 with the site P6. This example shows how complex is the nature of enzyme-substrate interactions is and how difficult is to predict the consequences of enzyme mutations.

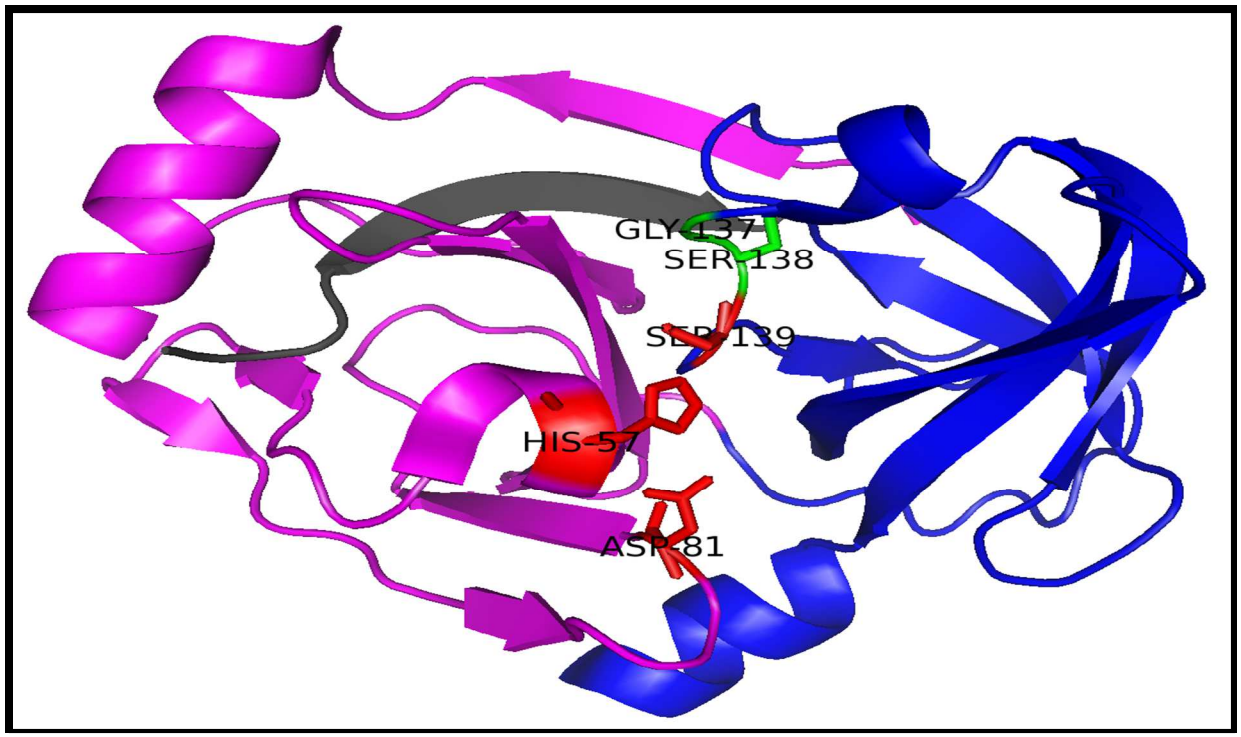
This analysis is more difficult in case of the subtype 3a, in which Asp168 is replaced with Arg168. The protease variants from the subtype 1a and 3a differ in amino acid residues in positions 123 and 168. This in turn strongly affects interactions with an acidic substituent in position P6 and may influence interactions with the substituent in position P1 because Asp168 stabilizes a range of amino acid residues in the binding site S1. The change observed in the subtype 5a, in which Asp168 was replaced by Glu, seems to have no impact on the 3D structure and enzyme-substrate interactions.

Table No.1: Percentage identity of NS3 nucleotide/amino acid sequences from selected HCV genotypes; Blastn multiple alignment (identity [%])

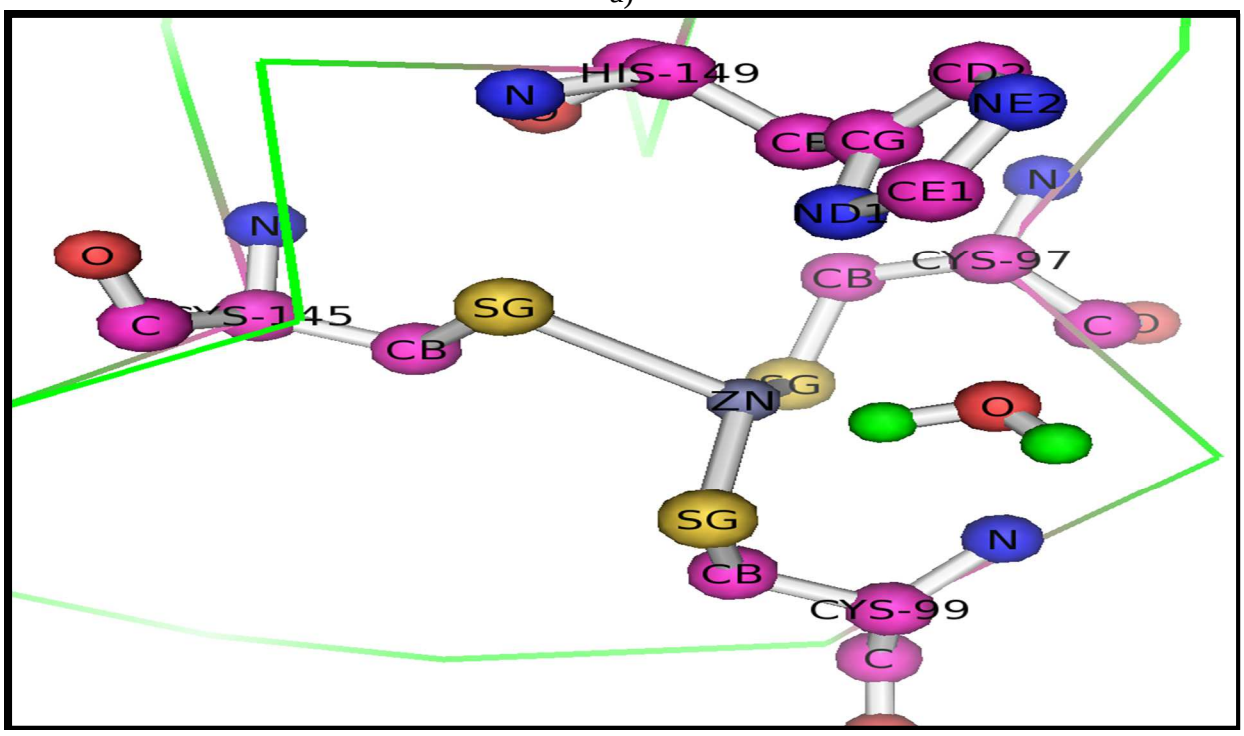
1b	2a	3a	4a	5a	6a	NS3 form genotype
79/92	79/91	69/81	74/86	74/86	71/83	1a
	91/95	71/83	75/88	73/86	71/85	1b
		70/81	74/86	73/86	71/84	2a
			69/80	73/84	70/81	3a
				74/88	71/84	4a
					70/85	5a

Table No.2: Percentage identity of NS4A peptide sequences from selected HCV genotypes

1b	2a	3a	4a	5a	6a	NS4 form genotype
89	91	83	80	76	70	1a
	87	87	85	76	67	1b
		67	67	67	59	2a
			81	78	72	3a
				85	85	4a
					80	5a

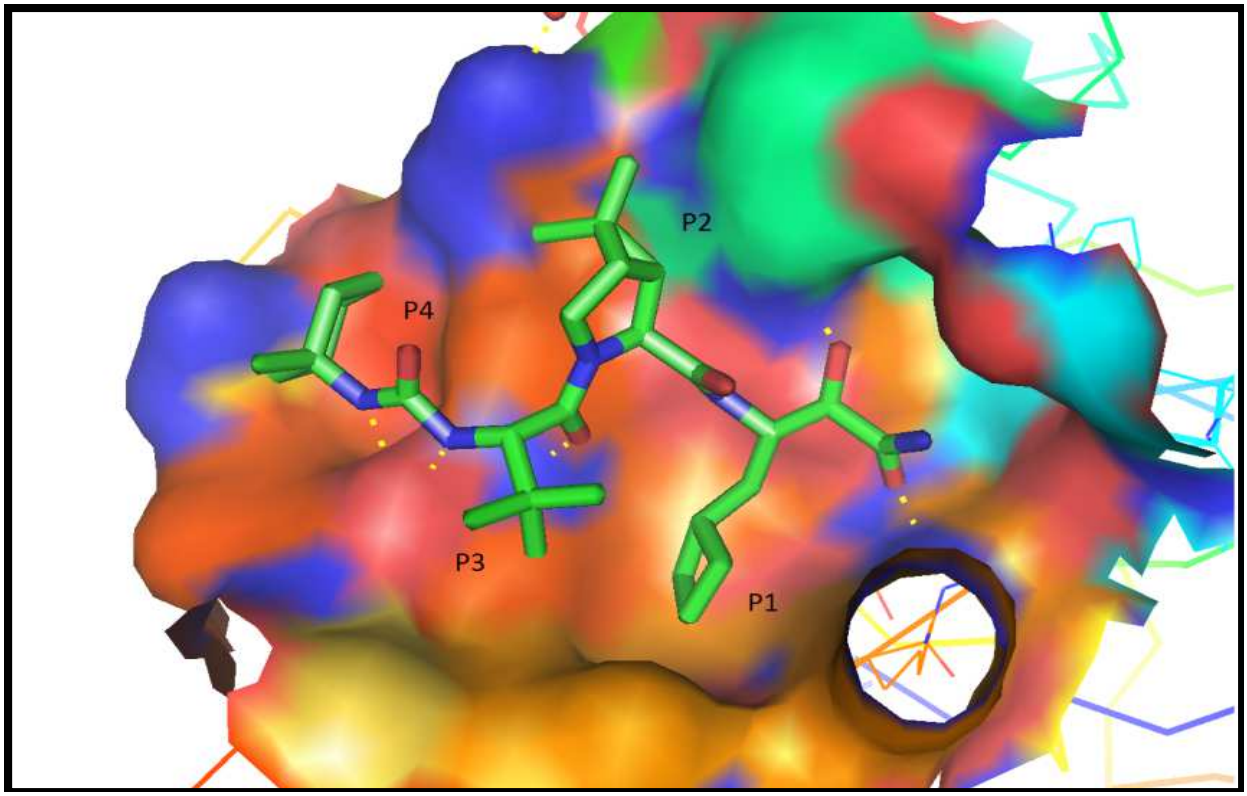


a)

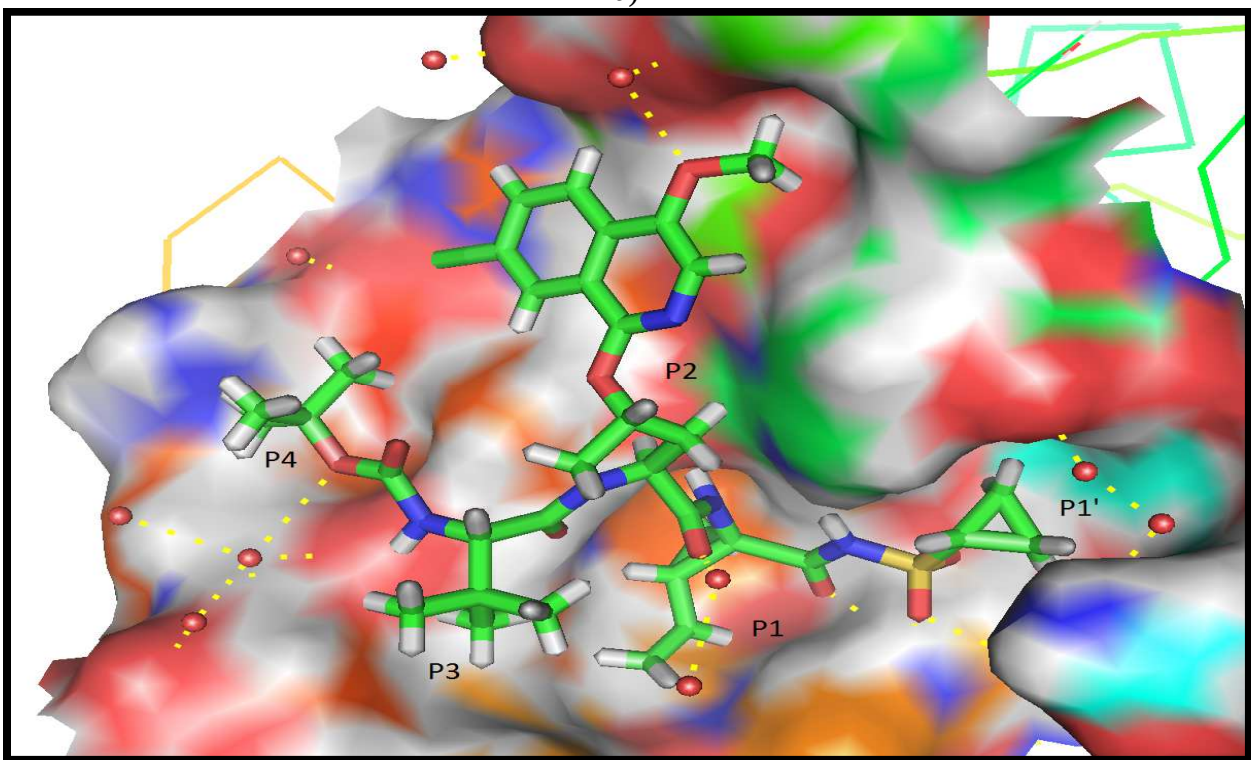


b)

Figure No.1: 3D structure of NS3/4A protease (PDB- 4k8b). a) The catalytic triad is coloured with red (His57, Asp81, Ser139), amino acid residues stabilizing oxyanion-hole are coloured with green, N-terminal part - magenta, C-terminal - blue, NS4A fragment interacting with NS3 shaded with grey; b) interaction of zinc ion with amino acid residues Cys97, Cys99, Cys145, His149 and a water molecule



a)



b)

Figure No.2: Structure of NS3/4A protease active site in a complex with the inhibitor a) with Boceprevir (PDB-3lox), b) with Asunaprevir (PDB-4wf8)

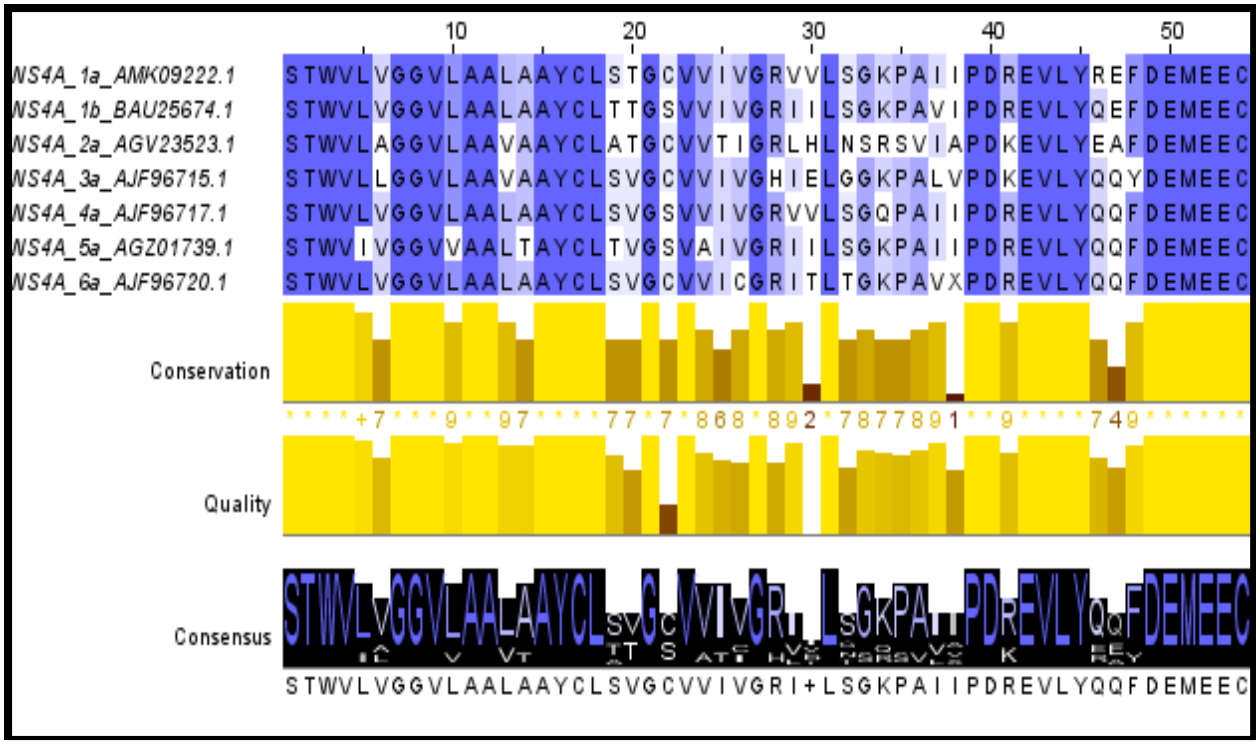
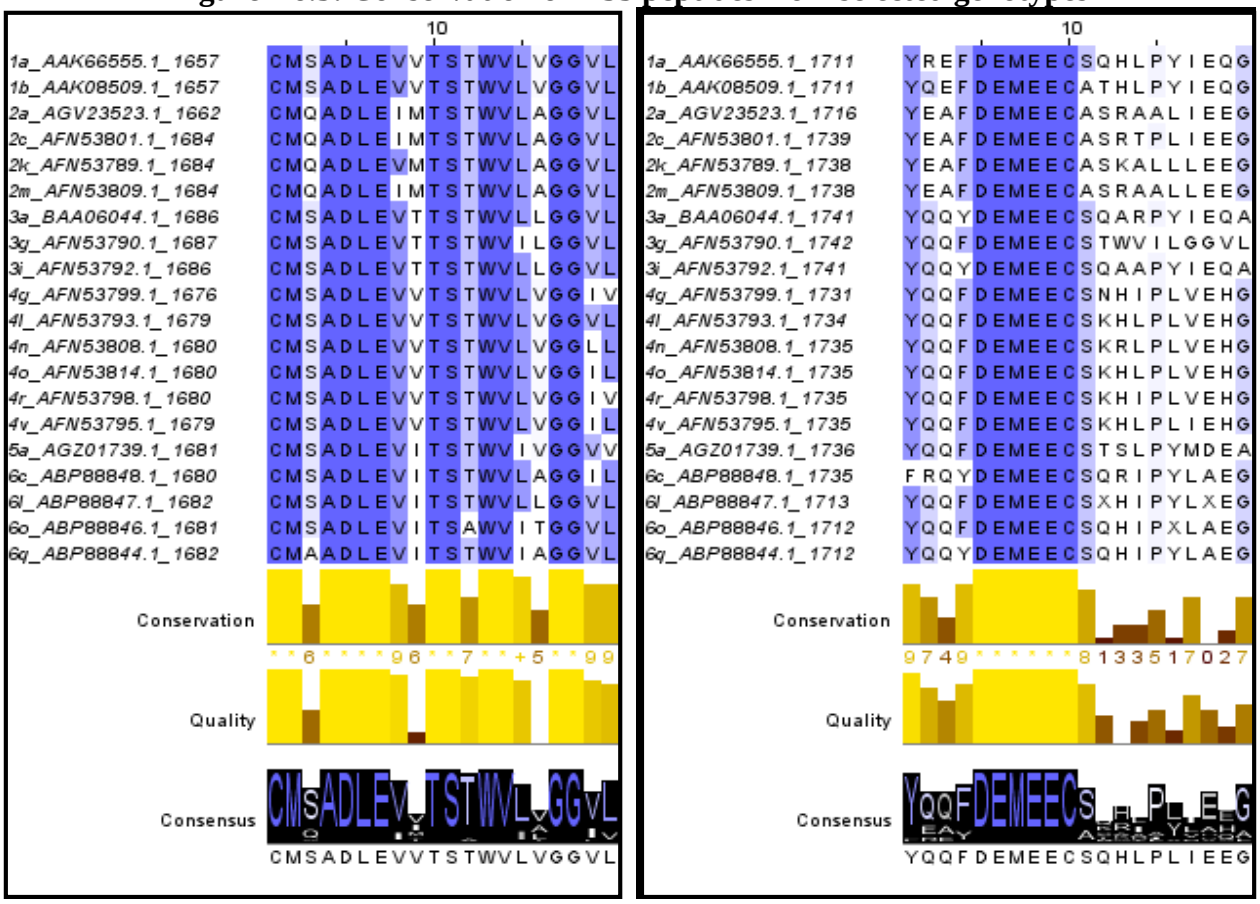
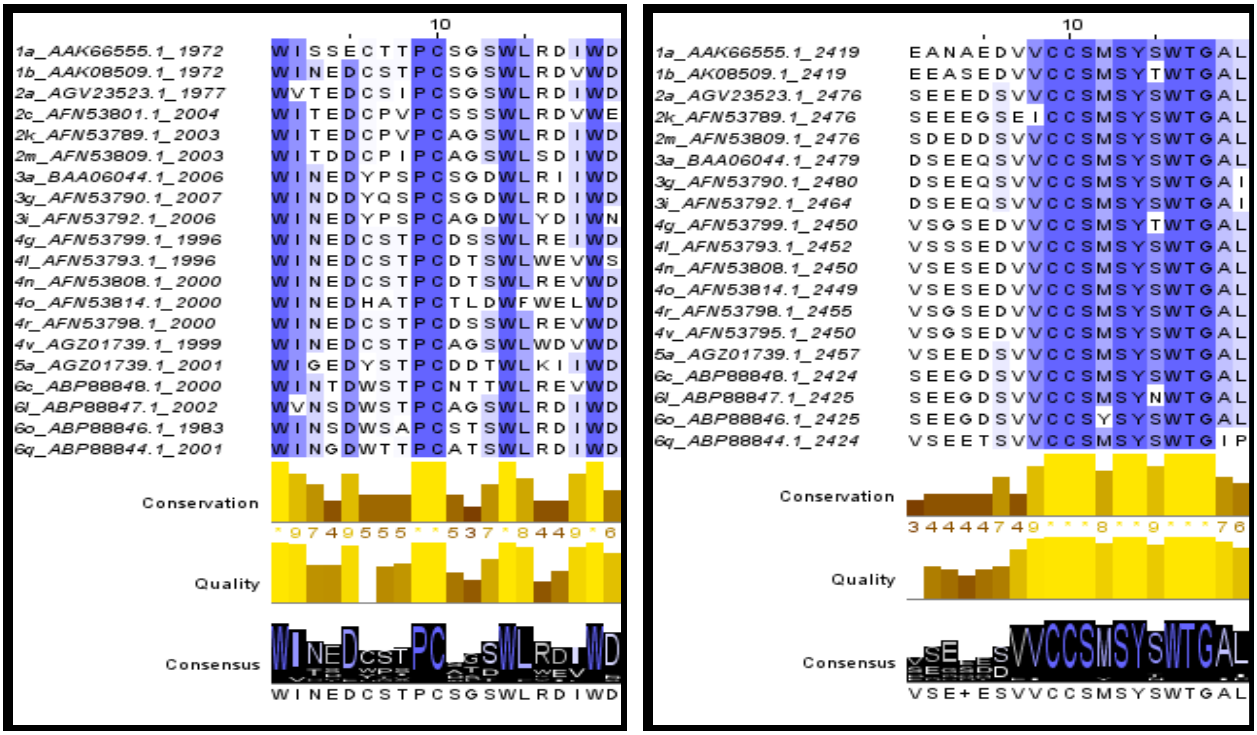


Figure No.3: Conservation of NS3 peptides from selected genotypes



a) b)



c) d)

Figure No.4: Identity of amino acid sequences in the selected HCV genotypes that are recognized by either the NS3 protein (a), or NS3/4A protease (b, c, d). a) autolytic cleavage between NS3 and NS4A proteins, b) cleavage between NS4A and NS4B proteins, c) between NS4B and NS5A proteins, d) between NS5A and NS5B proteins. The cleavage site is between 10th and 11th amino acid residues in the presented sequences. The sequence name consists of (in this order): HCV subtype name, Gen Bank access number, and cleavage site in polypeptide

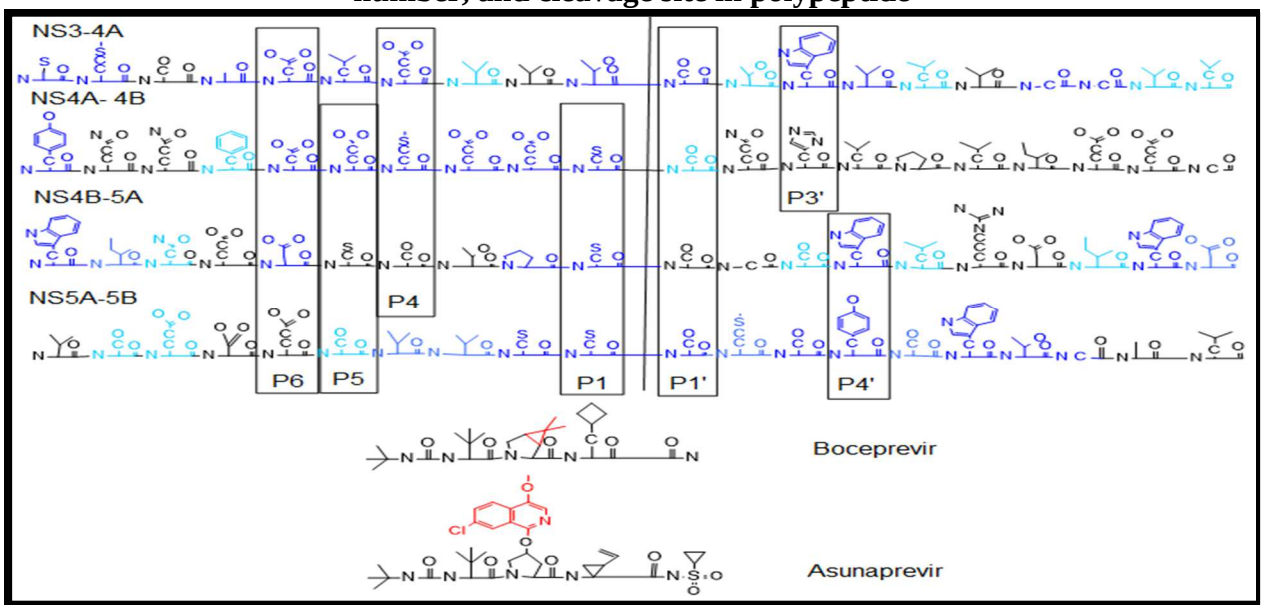


Figure No.5: HCV polypeptide cleavage consensus for the selected HCV genotypes. The highly conserved amino acid residues were coloured with blue, the fragments of inhibitor interacting with the deeper part of S2 pocket with red. The frames show these amino acid residues which can be the basis for designing a universal inhibitor of NS3/4A protease from all HCV genotypes

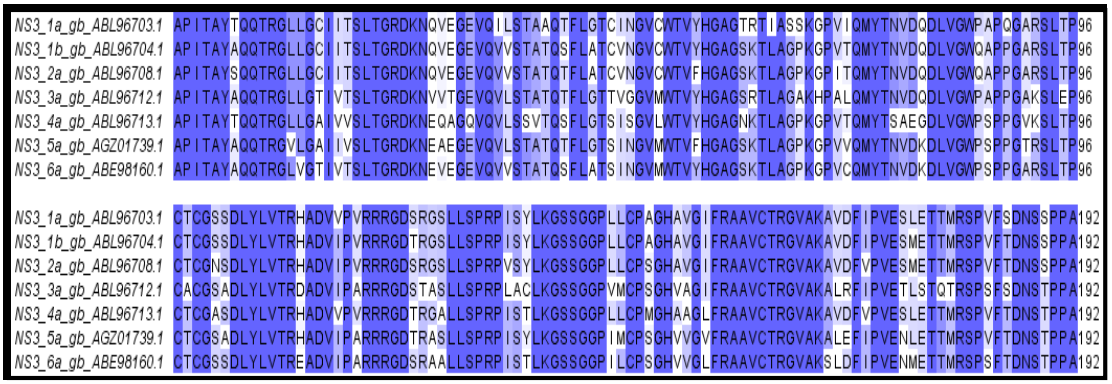
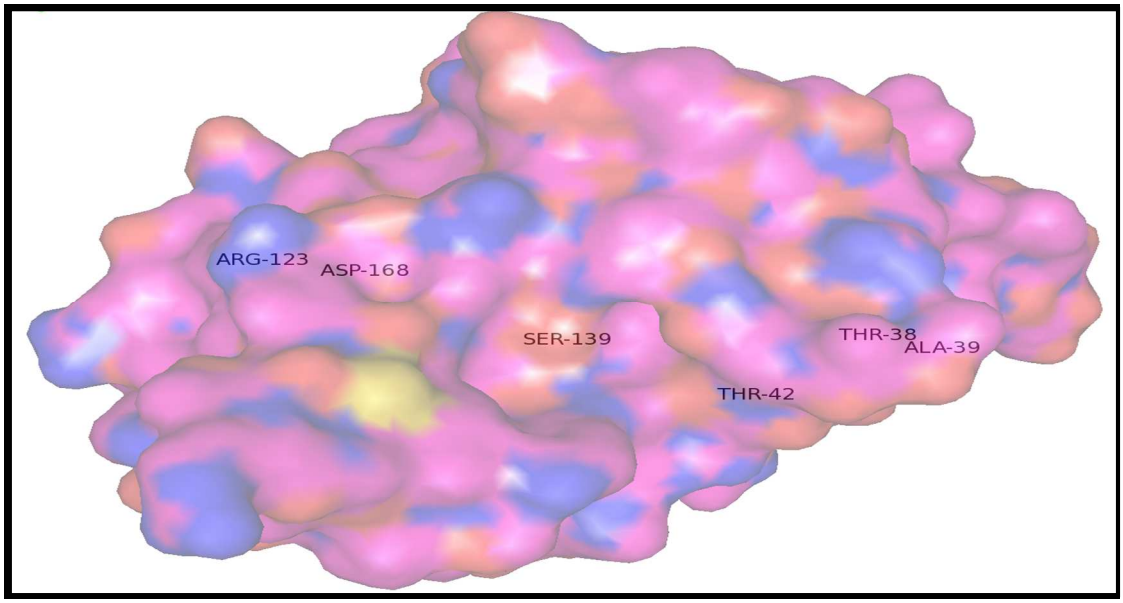
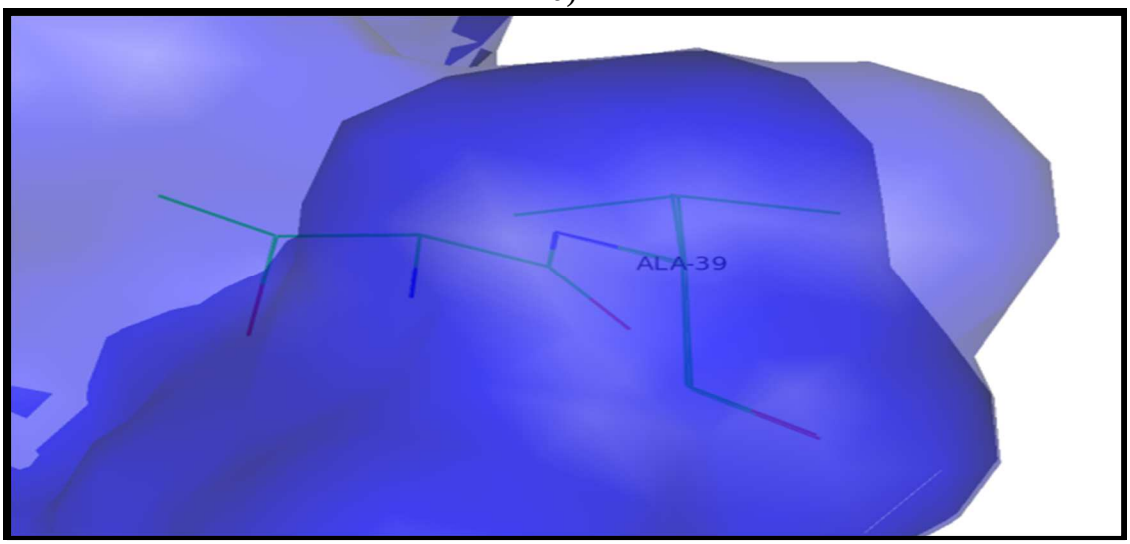


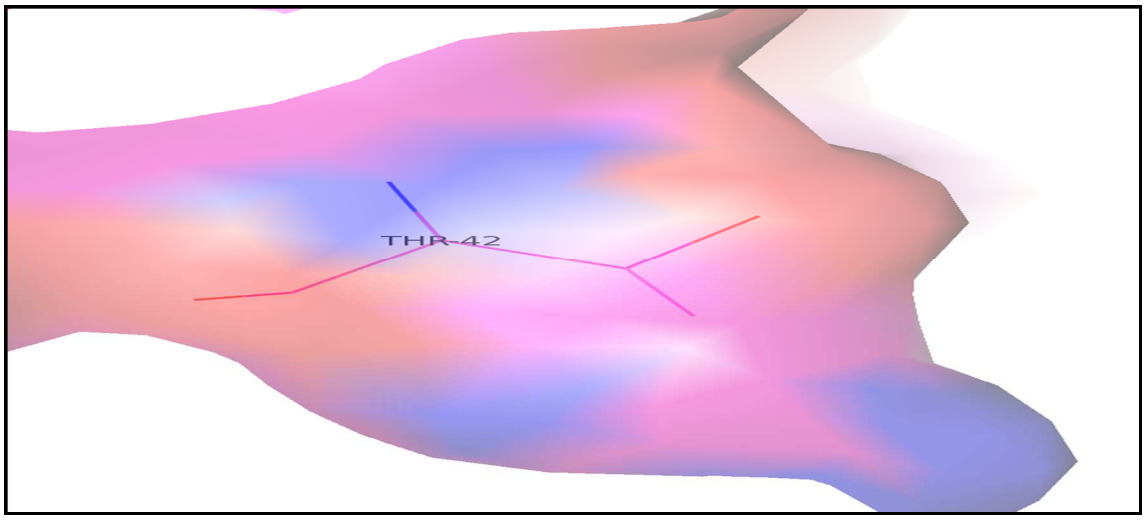
Figure No.6: Comparison of a fragment of NS3 protein sequence from genotypes 1-6. The conserved amino acid residues are highlighted with navy blue. The frames include the amino acid residues that interact with the substrate (according to 9, 21). Serine 139 contained in the active site is marked with an asterisk

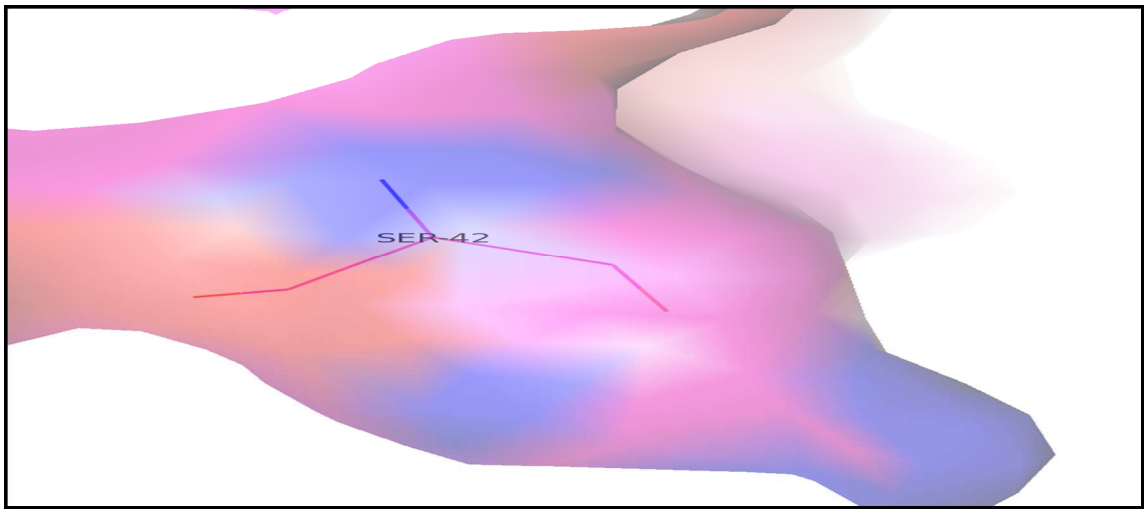


a)

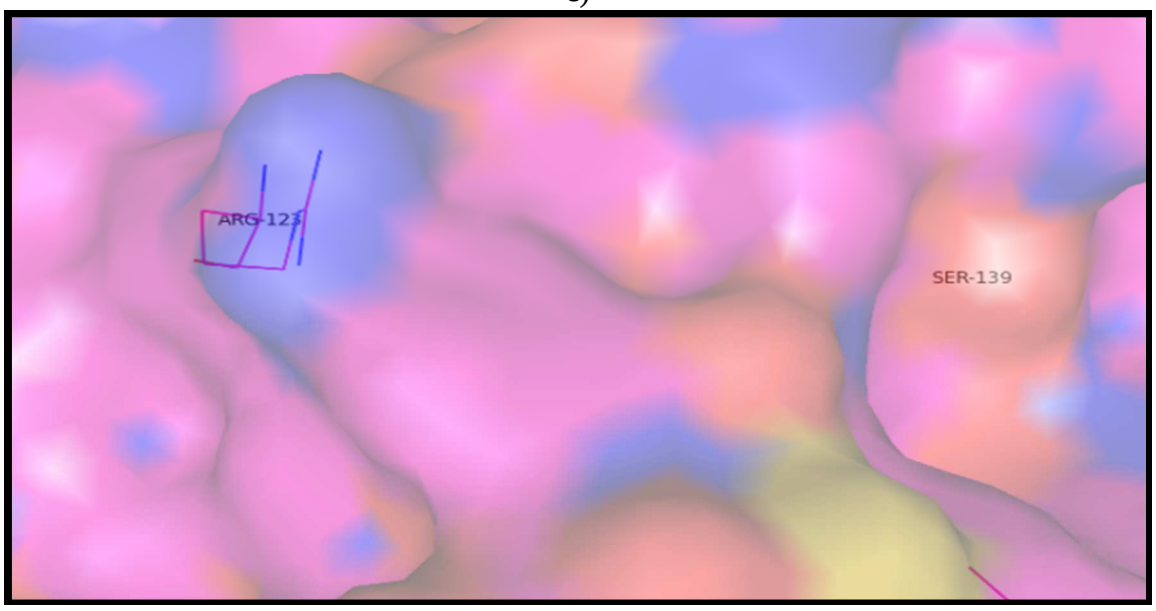


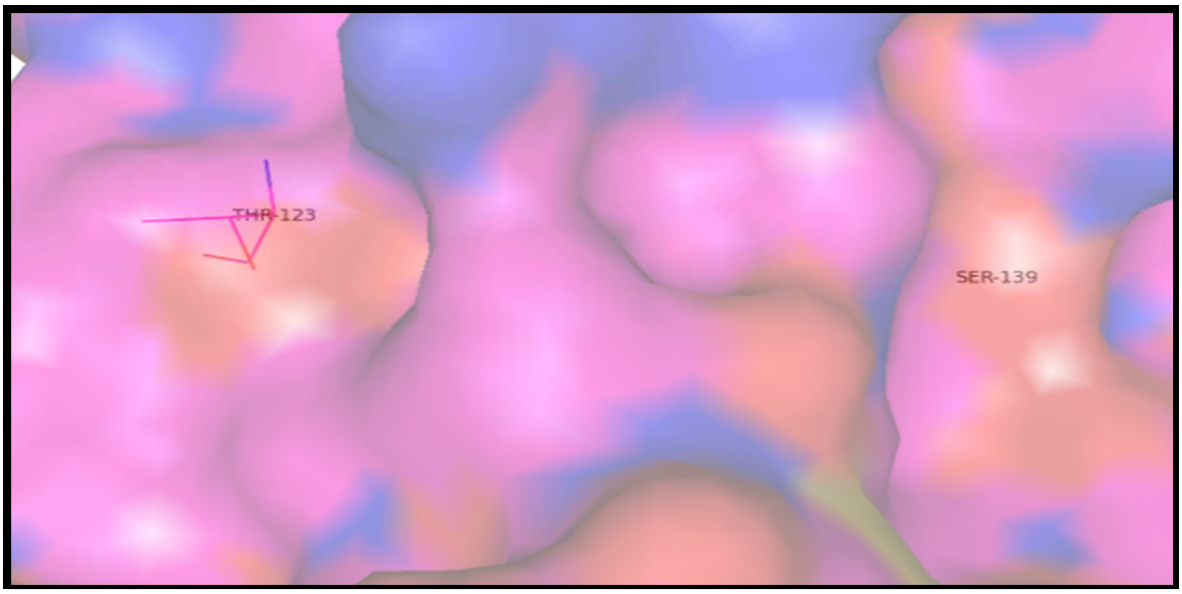
b)





c)





d)

Figure No.7: Differences in the active site structure of HCV protease from selected subtypes. a) localization of substrate binding amino acids residues in relation to active Ser139 (undersigned for subtype 1a), selected based on the analysis of Figure No.6, b) structural differences resulting from a change of Ala39Val, c) induced by a change of Thr42Ser, d) induced by a change of Arg123Thr

SUMMARY

The high costs of HCV therapies cause that the majority of patients receive the treatment of the 1st and 2nd generation, based on NS3/4A protease inhibitors. In case of patients infected by HCV genotypes different than 1, only the newest generation therapies, involving the inhibitors of NS5A RNA polymerase and NS5B protein, and in some cases also ribavirin, are efficient. Therefore, a universal, unnatural peptide effectively inhibiting NS3/4A proteases from all HCV genotypes has been searched for. The knowledge about natural substrates recognized by the NS3/4A enzyme is crucial for designing this inhibitor. It is known that the natural substrates are pentapeptides (P6-P4') having certain common properties, i.e. an acidic amino acid residue in position P6 (with an exception of proline in TRIF protein[28], a polar amino acid residue in position P5, cysteine residue in position P1, and serine residue in position P1'. However, the currently applied inhibitors of both 1st and 2nd generation (Figure No.5) do not interact with the pockets S6 and S5. The fragments of certain inhibitors, such as Asunaprevir and Boceprevir, that interact with the pockets S4 and S3, contain trim ethyl groups, which have the similar character as Leu residue contained in natural peptides. However, their occurrence does not ensure the sufficient specificity of the inhibitor. Although the S2 site of the NS3/4A protease is highly conserved, it was found[29] that a slight change, such as R155K or D168A mutation, is enough to make the inhibitor ineffective. The analysis presented in this work shows that the amino acid residue in position 168 is crucial for the interactions with both the inhibitor and other amino acid residues located in the catalytic pocket. Noteworthy, the common component of all currently applied inhibitors of the NS3/4A protease from all HCV genotypes is a chiral proline derivative in position P2.

The comparison of sequences of both the natural substrates and active sites of the NS3/4A protease suggests that the inhibitor only fitting into the active site but do not interacting like the natural substrates, may not effectively block the active site of the enzymes from all HCV genotypes. Mapping of natural interactions, with maintaining only the structural features characteristic of the enzyme from all genotypes (e.g. proline derivative in site P2), can firstly ensure an appropriate level of inhibition and secondly reduce a chance for virus mutations induced by the treatment.

Another option of HCV therapy is allosteric inhibition caused by the NS4A peptide replacement with a synthetic compound that will block the active conformation of NS3 apoenzyme. However, the competition of the artificial inhibitor with the natural cofactor is nearly impossible because the NS3/4A complex is formed during the translation of the HCV genome.

ACKNOWLEDGEMENT

The authors wish to express their sincere gratitude to Institute of Technical Biochemistry, Department of Biotechnology and Food Sciences, Technical University of Lodz, Lodz, Poland for providing necessary facilities to carry out this research work.

CONFLICT OF INTEREST

We declare that we have no conflict of interest.

BIBLIOGRAPHY

1. WHO | Hepatitis C, http://www.who.int/mediacentre/factsheets/fs164/en/ (accessed Sep 22, 2016).
2. Smith D B, Bukh J, Kuiken C, Muerhoff A S, Rice C M, Stapleton J T. Simmonds P. Expanded Classification of Hepatitis C Virus into 7 Genotypes and 67 Subtypes: Updated Criteria and Genotype Assignment Web Resource, *Hepatology*, 59(1), 2014, 318-327.
3. Irshad M, Mankotia D S, Irshad K. An Insight into the Diagnosis and Pathogenesis of Hepatitis C Virus Infection, *World J. Gastroenterol. WJG*, 19(44), 2013, 7896-7909.
4. Bartenschlager R, Lohmann V, Penin F. The Molecular and Structural Basis of Advanced Antiviral Therapy for Hepatitis C Virus Infection, *Nat. Rev. Microbiol*, 11(7), 2013, 482-496.
5. Messina J P, Humphreys I, Flaxman A, Brown A, Cooke G S, Pybus O G, Barnes E. Global Distribution and Prevalence of Hepatitis C Virus Genotypes, *Hepatology*, 61(1), 2015, 77-87.
6. Paredes A M, Blight K J. A Genetic Interaction between Hepatitis C Virus NS4B and NS3 Is Important for RNA Replication, *J. Virol*, 82(21), 2008, 10671-10683.
7. Jones C T, Murray C L, Eastman D K, Tassello J, Rice C M. Hepatitis C Virus p7 and NS2 Proteins Are Essential for Production of Infectious Virus, *J. Virol*, 81(16), 2007, 8374-8383.
8. Lindenbach B D, Rice C M. The Ins and Outs of Hepatitis C Virus Entry and Assembly, *Nat. Rev. Microbiol*, 11(10), 2013, 688-700.
9. Atoom A M, Taylor N G A, Russell R S. The Elusive Function of the Hepatitis C Virus p7 Protein, *Virology*, 462-463(2), 2014, 377-387.
10. Buhler S, Bartenschlager R. New Targets for Antiviral Therapy of Chronic Hepatitis C, *Liver Int*, 32(1), 2012, 9-16.
11. De Lemos A S, Chung R T. Hepatitis C Treatment: An Incipient Therapeutic Revolution, *Trends Mol. Med*, 20(6), 2014, 315-321.
12. Zeisel M B, Fofana I, Fafi-Kremer S, Baumert T F. Hepatitis C Virus Entry into Hepatocytes: Molecular Mechanisms and Targets for Antiviral Therapies, *J. Hepatol*, 54(3), 2011, 566-576.
13. Urbani A, Bianchi E, Narjes F, Tramontano A, Francesco R D, Steinkuhler C, Pessi A. Substrate Specificity of the Hepatitis C Virus Serine Protease NS3, *J. Biol. Chem*, 272(14), 1997, 9204-9209.
14. Love R A, Parge H E, Wickersham J A, Hostomsky Z, Habuka N, Moomaw E W, Adachi T, Hostomska Z. The Crystal Structure of Hepatitis C Virus NS3 Proteinase Reveals a Trypsin-like Fold and a Structural Zinc Binding Site, *Cell*, 87(2), 1996, 331-342.
15. De Francesco R, Carfi A. Advances in the Development of New Therapeutic Agents Targeting the NS3-4A Serine Protease or the NS5B RNA-Dependent RNA Polymerase of the Hepatitis C Virus, *Adv. Drug Deliv. Rev*, 59(12), 2007, 1242-1262.
16. Kim J L, Morgenstern K A, Lin C, Fox T, Dwyer M D, Landro J A, Chambers S P, Markland W, Lepre C A, O'Malley E T,

Harbeson S L, Rice C M, Murcko M A, Caron P R, Thomson J A. Crystal Structure of the Hepatitis C Virus NS3 Protease Domain Complexed with a Synthetic NS4A Cofactor Peptide, *Cell*, 87(2), 1996, 343-355.

17. Grakoui A, McCourt D W, Wychowski C, Feinstone S M, Rice C M. Characterization of the Hepatitis C Virus-Encoded Serine Proteinase: Determination of Proteinase-Dependent Polyprotein Cleavage Sites, *J. Virol*, 67(5), 1993, 2832-2843.
18. De Francesco R, Urbani A, Nardi M C, Tomei L, Steinkühler C, Tramontano A. A Zinc Binding Site in Viral Serine Proteinases, *Biochemistry (Mosc.)*, 35(3), 1996, 13282-13287.
19. Schechter I, Berger A. On the Size of the Active Site in Proteases, I. Papain, *Biochem. Biophys. Res. Commun*, 27(2), 1967, 157-162.
20. Zhang R, Durkin J, Windsor W T, McNemar C, Ramanathan L, Le H V. Probing the Substrate Specificity of Hepatitis C Virus NS3 Serine Protease by Using Synthetic Peptides, *J. Virol*, 71(8), 1997, 6208-6213.
21. Romano K P, Laine J M, Deveau L M, Cao H, Massi F, Schiffer C A. Molecular Mechanisms of Viral and Host Cell Substrate Recognition by Hepatitis C Virus NS3/4A Protease▽, *J. Virol*, 85(13), 2011, 6106-6116.
22. Webster D P, Klenerman P, Dusheiko G M. Hepatitis C, *The Lancet*, 385(9973), 2015, 1124-1135.
23. Guex N, Peitsch M C, Schwede T. Automated Comparative Protein Structure Modeling with SWISS-MODEL and Swiss-Pdb Viewer: A Historical Perspective, *Electrophoresis*, 30(1), 2009, S162-173.
24. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Cassarino T G, Bertoni M, Bordoli L, Schwede T. SWISS-MODEL: Modelling Protein Tertiary and Quaternary Structure Using Evolutionary Information, *Nucleic Acids Res*, 42(Web Server issue), 2014, W252-W258.
25. Kiefer F, Arnold K, Kunzli M, Bordoli L, Schwede T. The SWISS-MODEL Repository and Associated Resources, *Nucleic Acids Res*, 37(Database issue), 2009, D387-D392.
26. Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL Workspace: A Web-Based Environment for Protein Structure Homology Modelling, *Bioinforma. Oxf. Engl*, 22(2), 2006, 195-201.
27. Lin C. HCV NS3-4A Serine Protease, In *Hepatitis C Viruses: Genomes and Molecular Biology; Tan, S.-L, Ed, Horizon Bioscience: Norfolk (UK)*, 2006.
28. Raney K D, Sharm S D, Moustafa I M, Cameron C E. Hepatitis C Virus Non-Structural Protein 3 (HCV NS3): A Multifunctional Antiviral Target, *J. Biol. Chem*, 285(30), 2010, 22725-22731.
29. Soumana D I, Ali A, Schiffer C A. Structural Analysis of Asunaprevir Resistance in HCV NS3/4A Protease, *ACS Chem. Biol*, 9(11), 2014, 2485-2490.

Please cite this article in press as: Florczak Tomasz. *et al*. Differences between HCV ns3/4a protease from genotypes 1-6: substrate specificity, sequential and structural differences of active site structure, Is it possible to design a universal HCV protease inhibitor?, *Asian Journal of Research in Pharmaceutical Sciences and Biotechnology*, 4(4), 2016, 71-87.