Sequence analysis of hepatitis C virus nonstructural protein 3-4A serine protease and prediction of conserved B and T cell epitopes

AYESHA NAEEM[1] and YASIR WAHEED[1,2]

[1]Atta-ur-Rahman School of Applied Biosciences, National University of Sciences and Technology;
[2]Foundation University Medical College, Foundation University Islamabad, DHA-I, Islamabad 44000, Pakistan

Abstract. The hepatitis C virus (HCV) is a global health issue. The nonstructural protein 3 (NS3)-4Agene of HCV is responsible for serine protease activity. The aim of the present study was to develop a global consensus sequence of HCV serine protease, analyze conserved residues, and predict highly conserved B- and T-cell binding epitopes in the NS3-4A protein. A total of 160 NS3-4A sequences from the six genotypes of HCV were refracted in the current study. The amino acid sequences were aligned to obtain a global consensus sequence. The location of possible B- and T-cell epitopes were predicted in the HCV NS3-4A consensus sequence by employing bioinformatics tools. Despite the high mutation rate of HCV, the functionally important residues are highly conserved. These include residues that form the catalytic triad (His57, Asp81 and Ser139), the S1 and S6 pocket, zinc-binding site (Cys97, Cys99, Cys145 and His149) and the substrate binding groove. The epitopes B1, B8 and B9 are predicted to be ideal candidates for B-cell-based vaccine and are >95% conserved across six major HCV genotypes. The major histocompatibility complex (MHC) class I epitopes, M4, M5, M7 and M10, and MHC class II epitopes, T5, T7 and T10 are ideal epitopes for vaccine development with high antigenicity scores and high conservancy across major HCV genotypes. The predicted B- and T-cell epitopes are ideal targets for vaccine development, and are capable of producing a strong immune response for all major genotypes of HCV.

Introduction

The hepatitis C virus (HCV) is a plus stranded RNA virus belonged to the *Flaviviridae* family. HCV causes acute and chronic hepatitis, and more than half of HCV patients develop liver cirrhosis or hepatocellular carcinoma (1). Globally, 130-150 million people are living with HCV (2).

Hepatitis C has six major genotypes, which demonstrate many variations in geographic distribution, response to therapy and disease progression (3). The most prevalent genotypes of HCV are 1, 2 and 3, which occur around the globe. HCV genotype 4 is predominant in Africa and the Middle East, while in South Africa and Hong Kong genotypes 5 and 6, respectively are predominant (4). The most prevalent genotype in Pakistan is 3 (3).

There is currently no vaccine available for protection from HCV. Between 2001 and 2011, interferon and ribavirin were administered for HCV treatment. These types of medication resulted in limited responses with many adverse effects. Numerous interferon-free therapeutic strategies are at various stages of development and few of these produce a high response rate with minimum adverse effects (5,6). The prevalence of HCV is particularly high in the multitransfused patient population and in individuals who inject drugs (7,8).

The genome of HCV is comprised of 9,600 nucleotides and encoded as a single polyprotein. The genome is composed of four structural proteins and six non-structural proteins. HCV nonstructural protein 3 (NS3) protein is a 631-amino acid long dual-function protein, with a serine protease domain in the N terminal one-third, while the two-thirds region of the C terminus is RNA helicase domain (9,10). The NS4A region is a cofactor for the activation of the protease subunit of HCV (11). HCV protease has proven to be crucial for viral replication and is considered to be the best target for the development of anti-HCV therapeutic strategies (12). Thus, the aim of the present study was to develop a global consensus sequence of the HCV serine protease and predict conserved B- and T-cell binding epitopes.

Materials and methods

Sequence extraction and translation. A total of 160 complete genome sequences of HCV were extracted from the National Center for Biotechnology Information database (https://www.ncbi.nlm.nih.gov/nuccore/), including all six genotypes of HCV. The HCV genome sequences were trimmed using the NS3-4A HCV-H isolate as the reference sequence. The resulting NS3-4A nucleotide sequences were than translated using CLC main workbench software v.8 (Qiagen GmbH, Hilden, Germany) to their corresponding amino acid sequences.

Correspondence to: Dr Yasir Waheed, Foundation University Medical College, Foundation University Islamabad, Defence Avenue, DHA-I, Islamabad 44000, Pakistan
E-mail: yasir_waheed_199@hotmail.com

Table I. Position and sequences of the peptides that may be used for development of peptide vaccines.

Position	Peptides
1-12	APITAYAQQTRG
110-121	ADVIPARRRG
153-165	FRAAVCTRGVAK
202-215	LHAPTGSGKSTKVP
225-235	VLVLNPSVAAT
266-273	TYSTYGKF
275-285	ADGGCSGGAYD
301-314	LGIGTVLDQAETAG
318-330	VLATATPPGST
360-368	KGGRHLIFC
369-378	HSKKKCDELA
411-420	TDALMTGYTG
418-429	TGDFDSVIDCN
436-446	VDFSLDPTFTI
452-469	PQDAVSRSQRRGRTGRGR
518-533	TPGLPVCQDHLEFWE
594-603	GPTPLLYRLG

Development of global consensus sequence. The amino acid sequences were aligned to obtain specific consensus sequences for all six genotypes of HCV. The respective consensus sequences for genotypes 1-6 of HCV NS3-4A were developed utilizing the CLC workbench software. The consensus sequences were subsequently aligned together to obtain a global consensus sequence (13). The global consensus sequence of HCV NS3-4A serine protease was analyzed for its variable residues and highly conserved domains that determine the activity of the serine protease. Short highly conserved peptides were selected from the consensus sequence of HCV NS3-4A.

B-cell and T-cell epitopes prediction. The location of possible B- and T-cell epitopes was mapped in the consensus sequence of the NS3-4A gene. Possible B-cell epitopes in NS3-4A were predicted for antibody binding using the Immune Epitope Database (IEDB) (http://www.iedb.org/). Similarly, target epitopes for T-lymphocytes in NS3-4A were identified for binding to major histocompatibility complex (MHC) class I and II using ProPred-I and ProPred software (http://crdd.osdd.net/raghava//propred/), respectively. The predicted B- and T-cell epitopes in HCV NS3-4A were subjected to conservation analysis in the IEDB epitope conservation analysis tool. The epitopes with 80-100% conservancy were selected and these epitopes were compared with human proteome to confirm that these peptides would not trigger autoimmunity.

Results

Development of global consensus sequence and selection of conserved peptides. A HCV NS3-4A consensus sequence for each genotype was separately drawn. The genotypic consensus sequences were used for the development of the global consensus sequence, which aided with analyzing the conserved amino acids among all the genotypes of HCV.

Small peptide fragments consisting of 9-18 amino acid residues were deduced from the highly conserved regions of the NS3-4A consensus sequence (Table I). These highly conserved residues offer potent target sites for peptide vaccine development or designing site-specific HCV inhibitors.

Prediction of B-cell epitopes. The location of possible B- and T-cell MHC class I and II epitopes was identified in the consensus sequence of the NS3-4A gene and the predicted epitopes were subjected to conservation analysis in the IEDB epitope conservation analysis tool. The selected epitopes were compared with human proteome to confirm that these peptides do not trigger autoimmunity.

Different B-cell epitopes were predicted by IEDB in the NS3-4A gene (Table II). Each epitope is given a distinctive

Table II. B-cell epitopes and their conservation in hepatitis C virus nonstructural protein 3-4A sequences from genotypes 1-6.

Name	B-cell epitopes	Position	Peptide length (amino acids)	Epitope conservancy (%)
B1	TAYAQ	4-8	5	100.0
B2	TGRDKNEV	22-29	8	82.75
B3	GAGSKTLA	58-65	8	78.6
B4	LVGWPAPPGAKSLTPCTC	82-99	18	83.0
B5	DVIPARRRGDTRGSLL	112-127	16	84.8
B6	GSSGGPLLCP	137-146	10	84.3
B7	HAPTGSGKSTKVPAAYV	203-219	17	95.0
B8	GGCSGG	277-282	6	100.0
B9	DQAETA	308-313	6	100.0
B10	TATPPGSVTVPHPNI	322-336	15	89.0
B11	LTPAET	504-509	6	86.0
B12	RAKAPPPSWDE	570-580	11	87.8
B13	TLHGP	591-595	5	91.5

Table III. T-cell class I MHC-specific epitopes and their conservation in hepatitis C virus nonstructural protein 3-4A sequences from genotypes 1-6.

Name	Class I MHC-specific T-cell epitopes	Position	Peptide length (amino acids)	Epitope conservancy (%)
M1	YHGAGSKTL	56-64	9	85.6
M2	SLTPCTCGS	93-101	9	80.0
M3	GHAVGIFRA	148-156	9	87.0
M4	AVCTRGVAK	157-165	9	98.5
M5	SGKSTKVPA	208-216	9	98.0
M6	KVLVLNPSV	224-232	9	81.0
M7	TYSTYGKFL	266-274	9	98.0
M8	QAETAGARL	309-317	9	95.0
M9	HPNIEEVAL	333-341	9	92.0
M10	FCHSKKKCD	367-375	9	100.0
M11	CLIRLKPTL	584-592	9	87.4
M12	TKYIMTCMS	616-624	9	92.0
M13	LVGGVLAAL	636-644	9	90.5
M14	LAALAAYCL	641-649	9	97.0

MHC, major histocompatibility complex.

Table IV. T-cell class II MHC-specific epitopes and their conservation in hepatitis C virus nonstructural protein 3-4A sequences from genotypes 1-6.

Name	Class II MHC-specific T-cell epitopes	Position	Peptide length (amino acids)	Epitope conservancy (%)
T1	YHGAGSKTL	55-64	10	87.0
T2	IQMYTNVDQ	72-80	9	82.67
T3	VGHLHAPTG	199-207	9	92.0
T4	IRTGVRTIT	252-260	9	85.78
T5	YSTYGKFLA	267-275	8	98.25
T6	IICDECHST	287-295	9	86.5
T7	ILGIGTVLD	300-308	9	100.0
T8	LVVLATATP	317-325	9	93.6
T9	VVLCECYDA	489-497	9	93.67
T10	FTGLTHIDA	536-544	9	96.8

MHC, major histocompatibility complex.

name, from B1 to B13. The positions of the residues, the length of the peptide and the percentage conservancy of the epitope are refracted in Table II.

Among the B-cell epitopes, B1, B7, B8 and B9 are considered to be conserved among all the six genotypes of HCV. These epitopes developed from the global consensus sequence are capable of producing strong neutralizing antibodies against all six genotypes of HCV.

Prediction of MHC class I and II epitopes. In total, 38 different MHC class I epitopes were predicted by ProPred-I. The epitopes with 80-100% conservancy are presented in Table III. Among these epitopes, M4, M5, M7 and M10 were highly conserved in the HCV NS3-4A consensus sequence and all genotypes.

Various MHC class II epitopes were predicted from the HCV NS3-4A gene using ProPred software. Certain important epitopes are presented in Table IV. The epitopes T5, T7 and T10 were identified to be 95% conserved among genotypes 1-6 of HCV.

Discussion

The NS3-4A gene of HCV has a highly conserved catalytic triad comprised of His57, Asp81 and Ser139 residues. The catalytic triad is essentially required for the proteolysis of the HCV polyprotein (9). Replacing any of the catalytic triad amino acids, histidine, aspartate or serine with any other amino acid eliminated the proteolytic cleavage by NS3 (14).

The consensus sequence alignment demonstrates that the residues His57, Asp81, and Ser139 remained conserved across all HCV genotypes.

Previous X-ray observations and computational modeling analysis have revealed that a zinc-binding site is present opposite to the catalytic triad of HCV protease (11). In the present study, results from the consensus sequence analysis indicated that the zinc-binding site amino acids are well conserved among all HCV genotypes.

NS4A is a 54-amino acid long protein forming a non-covalent heterodimer with protease domain of NS3 (11). Mutation analysis reveals that the N-terminus 22 amino acids of NS3 are involved in the interaction with the central region 21-34 residues of the NS4A protein. Mutations affecting the non-covalent bonding between these two proteins cause a significant decline or inhibition of protease activity, confirming that the configuration of the bonded complex is vital for protease function (15). Numerous amino acids located in the middle section of the NS4A protein develop elaborate hydrophobic interactions with various hydrophobic side-chains in two β strands of NS3 serine protease, forming a sandwich-like configuration between the β-barrels of NS3 and the NS4A cofactor. These hydrophobic amino acid residues of NS4A primarily include Val23, Ile25, Ile29 and Leu31. The consensus sequence analysis reveals that the residues, Val23 and Ile25 are conserved among all HCV genotypes. The residue Ile29 has been replaced by Leu and Val in genotype 2 and 4, respectively, which are similar branched-chain amino acids. However, the residue, Leu31 has been mutated to Thr in genotype 6, which is a significant mutation.

Highly conserved B- and T-cell binding epitopes were predicted from the consensus sequence of HCV. Among B-cell epitopes, certain epitopes demonstrated 8%-100% conservation among all six genotypes of HCV. Various MHC class I and II predicted epitopes exhibited maximum allele-binding affinity confirming them as potential T-cell epitopes. The epitopes B1, B8 and B9 are considered to be the best targets for B cell-based vaccine development and are >95% conserved across six major HCV genotypes. Similarly, M4, M5, M7 and M10 are the best MHC class I epitopes to be adopted as synthetic vaccines against multi-isotypes of HCV. Additionally, epitopes T5, T7 and T10 are ideal MHC class II specific epitopes with high antigenicity scores and high conservancy across major genotypes. In comparison to the epitopes derived from highly variable genome regions, the use of conserved epitopes from the consensus sequence may provide broader protection against HCV. Therefore, these predicted epitopes may be invoked as effective vaccine candidates against major genotypes of HCV.

In conclusion, regardless of numerous variations in the NS3-4A gene sequences from different genotypes of HCV, the functionally important residues of the serine protease and helicase are highly conserved. These regions of the NS3-4A sequence may be useful in developing antiviral agents or peptide vaccines against HCV. Prediction of epitope immunogenicity and characterization on the basis of peptide sequences is important in developing a potent peptide vaccine for HCV. Though as the antigens predicted in the present study were based on computer software analysis, the antigenic potential of the peptides should be further characterized in animal models.

References

1. Hoofnagle JH: Hepatitis C: The clinical spectrum of disease. Hepatology 26 (Suppl 1): 15S-20S, 1997.
2. World Health Organization: Global health sector strategies on viral hepatitis 2016-2021. Accessed August 29, 2017.
3. Safi SZ, Badshah Y, Waheed Y, Fatima K, Tahir S, Shinwari A and Qadri I: Distribution of hepatitis C virus genotypes, hepatic steatosis and their correlation with clinical and virological factors in Pakistan. Asian Biomed 4: 253-262, 2010.
4. Smith DB and Simmonds P: Review: Molecular epidemiology of hepatitis C virus. J Gastroenterol Hepatol 12: 522-527, 1997.
5. Waheed Y: Effect of interferon plus ribavirin therapy on hepatitis C virus genotype 3 patients from Pakistan: Treatment response, side effects and future prospective. Asian Pac J Trop Med 8: 85-89, 2015.
6. Waheed Y: Ledipasvir and sofosbuvir: Interferon free therapy for hepatitis C virus genotype 1 infection. World J Virol 4: 33-35, 2015.
7. Waheed Y, Najmi MH, Aziz H, Waheed H, Imran M and Safi SZ: Prevalence of hepatitis C in people who inject drugs in the cities of Rawalpindi and Islamabad, Pakistan. Biomed Rep 7: 263-266, 2017.
8. Saeed U, Waheed Y, Ashraf M, Waheed U, Anjum S and Afzal MS: Estimation of hepatitis B virus, hepatitis C virus, and different clinical parameters in the thalaseemic population of capital twin cities of Pakistan. Virology (Auckl) 6: 11-16, 2015.
9. Bartenschlager R, Ahlborn-Laake L, Mous J and Jacobsen H: Nonstructural protein 3 of the hepatitis C virus encodes a serine-type proteinase required for cleavage at the NS3/4 and NS4/5 junctions. J Virol 67: 3835-3844, 1993.
10. Failla C, Tomei L and De Francesco R: An amino-terminal domain of the hepatitis C virus NS3 protease is essential for interaction with NS4A. J Virol 69: 1769-1777, 1995.
11. Kim JL, Morgenstern KA, Lin C, Fox T, Dwyer MD, Landro JA, Chambers SP, Markland W, Lepre CA, O'Malley ET, et al: Crystal structure of the hepatitis C virus NS3 protease domain complexed with a synthetic NS4A cofactor peptide. Cell 87: 343-355, 1996.
12. Grakoui A, McCourt DW, Wychowski C, Feinstone SM and Rice CM: Characterization of the hepatitis C virus-encoded serine proteinase: Determination of proteinase-dependent polyprotein cleavage sites. J Virol 67: 2832-2843, 1993.
13. Waheed Y, Saeed U, Anjum S, Afzal MS and Ashraf M: Development of global consensus sequence and analysis of highly conserved domains of the HCV NS5B protein. Hepat Mon 12: e6142, 2012.
14. Eckart MR, Selby M, Masiarz F, Lee C, Berger K, Crawford K, Kuo C, Kuo G, Houghton M and Choo QL: The hepatitis C virus encodes a serine protease involved in processing of the putative nonstructural proteins from the viral polyprotein precursor. Biochem Biophys Res Commun 192: 399-406, 1993.
15. Shimizu Y, Yamaji K, Masuho Y, Yokota T, Inoue H, Sudo K, Satoh S and Shimotohno K: Identification of the sequence on NS4A required for enhanced cleavage of the NS5A/5B site by hepatitis C virus NS3 protease. J Virol 70: 127-132, 1996.