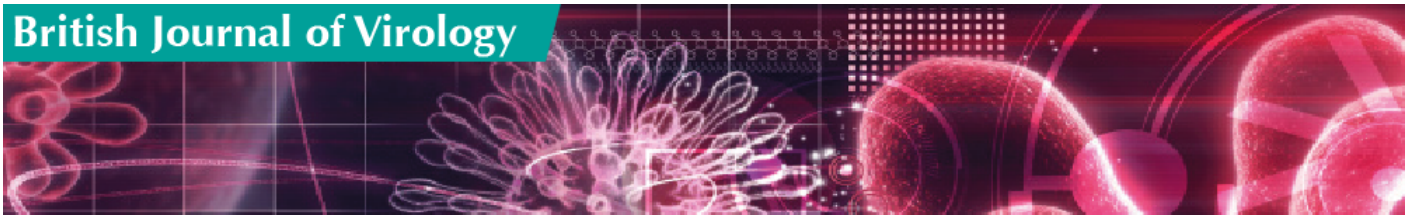


Large-scale analysis of Rotavirus A sequences reveals potential drug-target sites of non-structural protein 2

Journal	British Journal of Virology
Manuscript ID	SANDF20140401050408
Manuscript Type	Short Communication
Date Submitted by the Author	Tue, 01 Apr 2014, 05:27 AM
Complete List of Authors	■ Mr darapaneni vivek

Large-scale analysis of Rotavirus A sequences reveals potential drug-target sites of non-structural protein 2

VIVEK DARAPANENI

Department of computational biochemistry, Saket Institute for fundamental research, India

Running title: NSP2 of rotavirus as potential drug target

Correspondence:

Vivek Darapaneni, Department of computational biochemistry, Saket Institute for fundamental research, 55-18-1, LIG 57, HB colony, Visakhapatnam 530022, India

Tel.: +91 9866890432 E-Mail: darapanenivivek@gmail.com

Abstract

Non-structural protein 2 of rotavirus A plays a critical role in viral life cycle. The objective of this study was to determine the residue conservation in the Non-structural protein 2 and to ascertain the conserved regions of functional or structural significance, which may be exploited as potential target sites for drugs. The analysis was based on 1091 sequences. The degree of conservation was projected onto the experimental protein structure. Using the SITEHOUND binding site prediction algorithm, several conserved binding sites were identified. Additionally when the NSP2 structure was analysed in conjunction with conservation data, novel conserved sites were identified, namely Leu25-Ile27, Met30; Phe22, Ala26, Ala29 and Met30; Ile146-Thr148, Thr151; Lys59, Asn62-Asn64, Asn66 and Arg210, Lys301-Thr305. Totally, this work reveals novel conserved binding sites that are feasible prospects for protein-protein interactions and might act as foundation for designing universal anti-Rotaviral drugs.

Keywords: Non-structural protein, Rotavirus, Conservation, SITEHOUND, Drug, Novel, Protein-protein interactions

1. Introduction

Viruses will successfully replicate in the host, once it surmounts the antiviral response of the organism. RNA viruses whose genomes are small and coding space limited will often express proteins which are multifunctional, the example of which is rotavirus. Ever since the

discovery of rotaviruses in 1973 (Bishop et al., 1973), they have been associated with acute diarrhea often dehydrating in children (Parashar et al., 2003). Rotaviruses are accountable for 453,000 deaths per year globally, among children who are under the age of five years (Tate et al., 2012). The majority of rotaviral deaths occur in developing countries where accesses to the two effective and safe licensed vaccines against rotavirus are limited (Patton, 2012); however the infection is widespread in all the countries (Ruiz et al., 2009). Rotavirus A (RVA), a double stranded RNA (dsRNA), nonenveloped virus belongs to the family Reoviridiae (Estes and Kapikian, 2007). The segmented genome of RVA contain eleven RNA segments, which code for twelve proteins of which six are structural proteins and six are non-structural proteins (Pesavento et al., 2006).

Non-structural protein 2 (NSP2) is a basic protein of length 317 amino acid coded by the eighth genome segment and has molecular weight of 35kDa (Pesavento et al., 2006). This protein forms doughnut shaped octamer by interaction which is tail-tail between two tetramers (Jayaram et al., 2002). Each monomer of NSP2 consists of two domains which are distinct, namely N-terminal and C-terminal domains (Vasquez-Del Caprio et al., 2006). These domains are separated by 25Å deep electropositive cleft, which is 30Å wide (Jayaram et al., 2002; Vasquez-Del Caprio et al., 2006). The C-terminal domain (CTD) of NSP2 in spite of lacking signature HIT pattern (His-X-His-X-His-XX) shows structural similarity to HIT (Histadine Triad Protein) which is unexpected where as N-terminal domain (NTD) has novel fold (Jayaram et al., 2002; Vasquez-Del Caprio et al., 2006). NSP2 is a multifunctional protein, which has important functions such as single stranded RNA binding (Taraporewala et al., 1999), nucleic acid helix destabilizing (Taraporewala and Patton, 2001), Mg^{2+} dependent nucleoside triphosphatase (Kumar et al., 2007), nucleoside diphosphate kinase (Taraporewala

and Patton, 2001) and RNA triphosphatase (Vasquez-Del Caprio et al., 2006). Recently NSP2 was also shown to sequestrate tubulin which induces microtubule depolymerization during infection (Martin et al., 2010). NSP2 interacts with non-structural protein 5 (NSP5) to form viroplasms, where rotavirus genome replication and assembly take place (Fabbretti et al., 1999; Eichwald et al., 2004; Kumar et al., 2007).

At present there are no licensed drugs available for targeting NSP2, which plays a critical role in the infectious life cycle of RAV. To identify the degree of conservation among all RAV from different host, so as to identify potential drug binding sites were the aims of the present study. Together with binding site analysis, this study suggests potential binding sites which provide the basis for designing new anti-rotaviral drugs and also define novel in vivo interaction sites.

2. Methods

2.1 Sequence analysis and structure retrieval

The NSP2 sequences were collected from NCBI protein database. Sequences of full length were only selected. Multiple sequence alignment (MSA) was performed using MUSCLE version 3.8 (Edgar, 2004). The experimental structure of NSP2 was collected from Protein Data Bank (PDB) (www.pdb.org) (Berman et al., 2000).

2.2 Conserved region identification and binding site analysis

By giving MSA file and protein structure file as an input, conserved regions in NSP2 protein were identified by using ConSurf server (http://consurf.tau.ac.il/) (Celniker et al., 2013; Ashkenazy et al., 2010; Berezin et al., 2004). The conservation score given by this online server is divided into scale of nine grades which are given for the purpose of visualization. Most variable positions in the protein are placed in grade one (turquoise) and most conserved

positions are placed in grade nine (maroon). Ligand binding sites (LBS) were predicted using SITEHOUND–web server (http://scbx.mssm.edu/sitehound /sitehound-web/Input.html) (Hernandez et al., 2009) which uses energy based method to find regions with high potential for ligand interactions (Hernandez et al., 2009). The input to this server is structure of the protein which is characterized by different probes for identification of binding sites of different types (Hernandez et al., 2009).

3. Results and discussion

3.1 Multiple alignment, structure retrieval and conserved residues

For NSP2 sequence alignment, 1091 sequences from different hosts of rotavirus A were collected and analyzed. On the whole, MSA of collected sequences showed a higher degree of similarity. In particular, at residual positions 75, 82, 175, 200, 249, 253, 254 and 255 of NSP2 sequences showed lower degree of conservation. N-terminal region showed more conservation than C-terminal region of NSP2. The protein structure of NSP2 was obtained from Protein Data Bank (PDB) (www.pdb.org) (Berman et al., 2000) whose PDB ID is 1L9V (Jayaram et al., 2002). This experimental NSP2 structure is near complete structure with residues 1 to 313 experimentally solved while the region 314-317 has withstood all high-resolution experimental methods. Residues 1-140 form NTD and residues 156-313 form CTD. There are two sub-domains in NTD which are joined by a loop made up of residues 31-75 and the two domains are joined by a linker consisting of residues141-155. Altogether there are eleven α-helices and nine β-sheets (numbering has been done in a sequential order), of which six α-helices and four β-sheets are present in NTD, while the rest are in CTD.

Residues which are conserved were identified using ConSurf server (Celniker et al., 2013; Ashkenazy et al., 2010; Berezin et al., 2004) and these residues are shown in Table 1. The

variable residues of grades 1-3 and conserved residue of grades 7-9 are grouped together in Table 1. The conservation scores were projected onto the backbone and spacefill model of the protein in figure 1. The conserved residues in the secondary structures of NSP2 protein are shown in table 2. The conserved residues identified on NSP2 might be of structural importance or functionally significant (Schueler-Furman and Baker, 2003). In the NTD of NSP2, residues Lys37, Lys38, Lys58-Arg60 and Arg68 are implicated in tubulin binding (Martin et al., 2010). Residues Lys38, Lys59 and Arg60 are conserved residues and they may play an important role in tubulin binding. The regions Asn64-Arg68, Ile179-Asn183, Ser232-Thr251 and Gln291-Gly302 were found to be involved in the interaction between NSP2 and NSP5 (Jiang et al., 2006). Asp65-Ser67, Ile179, Asp181, Asn183, Ser232-Val234, Asn236-Val241, Ala243, Asn248, Lys250, Gln291, Lys292, Lys294, Phe300 and Gly302 were found to be conserved residues and these residues might have functional importance in the NSP2-NSP5 interaction. Residues Phe43, Phe184, Lys230, Tyr242, Phe245 and Phe300 are implicated in preferential binding to single stranded RNA (Jiang et al., 2006) of which Phe43, Phe184, Lys230 and Phe300 are conserved residues and they might be of functional importance in RNA binding.

Residues Lys188, His221, Lys223, His225 and Arg227 are implicated in catalytic activity of NSP2 (Kumar et al., 2007). All the residues implicated in catalytic activity of NSP2 are conserved residues implying that this activity is critical function of NSP2. Residues Asn64, Asp65, Asn183, Trp211, Asn214, Ser232, Phe290, Met309 and Glu311 were found to participate in two-fold contact (Jayaram et al., 2002). Asp65, Asn183, Trp211, Asn214, Ser232, Phe290 and Glu311 were found to be conserved residues. It can be postulated that Asp65, Asn183, Trp211, Phe290 and Glu311 are actually involved in two-fold contact and

residues Asn214 and Ser232 most probably are structure stabilizing residues. Asn23-Leu25, Thr32, Ala33, Tyr44, Ser46-Tyr49, Gln122-Ile124, Ile150, Gln160, and Glu204 are the residues which are implicated in four-fold contact (Jayaram et al., 2002). Residues Asn23, Leu25, Thr32, Ala33, Tyr44, Ser46, Ile47, Tyr49, Gln122, Asp123, Gln160 and Glu204 are conserved residues. It can be postulated that residues Asn23, Thr32, Ala33, Tyr44, Gln122, Asp123, Gln160 and Glu204 are actually involved in four-fold contact while residues Leu25, Ser46, Ile47 and Tyr49 most probably be structure stabilizing residue.

3.2 Predicted binding site

Among the ten preeminent LBS reported by SITEHOUND (Hernandez et al., 2009), five LBS were selected on the basis of conserved residues present in the predicted LBS and are shown in table 3 and figure 2A. Note that numbering of the sites are not sequentially numbered, but are numbered by rank reported by SITEHOUND (Hernandez et al., 2009). In NSP2, site 3 located across both NTD and CTD contains twelve conserved residues, namely ASN113, ARG117, ASN120, GLN122, ASP123, Thr149, Glu153- Ile157 and Trp167. Site 5 is located across both domains of NSP2 and all the residues in this predicted site are conserved. Site 7 is located in CTD and contains eight conserved residues, namely Tyr231, Val234, GLN267, ASN268 and Leu289-Lys292. Site 8 is located across the two sub-domains of NTD which contains ten conserved residues, namely Ala5, Pro10, Asp41-Phe43, Pro54, Gln56, Phe57, Leu102 and Ser103. Site 9 is located across the first sub-domain and loop that joins the two sub domains of NTD containing conserved residues, namely Lys28, Leu31, Thr32, Asn62, Thr63, ASN71 and GLU73. By targeting site 3 with a suitable ligand, it is possible to interfere in NSP2-NSP5 interaction and there by disrupting the association of the two proteins, while four-fold contact of NSP2 can also be disrupted by targeting this site. By

targeting site 5, it is possible to inhibit NSP2 catalytic activity, NSP2-NSP5 and NSP2 preferential binding to single stranded RNA. By targeting site7, site 8 and site 9, it is possible to inhibit NSP2-NSP5 interaction, NSP2 preferential binding to single stranded RNA and four-fold contact of NSP2.

Separately when the NSP2 structure was analysed in conjunction with conservation data, novel conserved putative binding sites were identified and are shown in table 4 and figure 2B. Site 1 and site 2 are located in first sub-domain of NTD, site 3 and site 4 are located in the loop region that joins the sub-domains in NTD, while site 5 is located in CTD of NSP2. Site1, site 2 and site 3 are completely conserved binding sites, while in site 4 except Asn64 which is variable residues, the remaining residues are conserved. Site 5 contains conserved residues Arg210, Gly302, Ser304 and intermediately conserved residues Lys301, Leu303 and Thr305. These binding sites might have still unknown functional importance as these sites have residues which are conserved and there is an urgent need to experimentally elucidate the functions of these sites in the activity of NSP2.

4. Conclusions

In conclusion, this study found that NSP2 showed a range of variable and conserved residues among all RVA subtypes. Binding sites identified in this study are potential target for developing universal anti-rotaviral drugs. The drugs thus identified are unlikely to become ineffective as a consequence of a mutation of the virus into a drug-resistant form. Further studies emerging from this work should characterize the function of the novel binding sites which are conserved as well as to develop lead compounds that target both putative and novel binding sites.

Acknowledgements

This work was funded by the author personally. The author is grateful to Kandula Alekhya for useful comments and helpful discussions.

Conflict of interest

The author declare that there is no conflict of interests regarding the publication of this article

References

H. Ashkenazy, E. Erez, E. Martz, T. Pupko, and N. Ben-Tal, 2010: ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res. 38: W529-W533.

R.F. Bishop, G.P. Davidson, I.H. Holmes, and B.J. Ruck, 1973: Virus particles in epithelial cells of duodenal mucosa from children with acute non-bacterial gastroenteritis. Lancet 2: 1281–1283.

H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, and P.E. Bourne PE, 2000: The Protein Data Bank. Nucleic Acids Res. 28: 235-242.

C. Berezin, F. Glaser, J. Rosenberg, I. Paz, T. Pupko, P. Fariselli, R. Casadio and N. Ben-Tal, 2004: ConSeq: The identification of functionally and structurally important residues in protein sequences. *Bioinformatics.* Vol. 20 1322-1324.

G. Celniker, G. Nimrod, H. Ashkenazy, F. Glaser, E. Martz, I. Mayrose, T. Pupko and N. Ben-Tal, 2013: ConSurf: Using Evolutionary Data to Raise Testable Hypotheses about Protein Function. *Isr. J. Chem.* doi: 10.1002/ijch.201200096.

R.C. Edgar, 2004: MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004, 32: 1792-1797.

C. Eichwald, J.F. Rodriguez, and O.R. Burrone, 2004: Characterization of rotavirus NSP2/NSP5 interactions and the dynamics of viroplasm formation. J Gen Virol. 85: 625-634.

M.K. Estes., and A.Z. Kapikian, 2007: Rotaviruses. In: D.M. Knipe, and P.E. Howley (eds) Fields Virology, 5th edition. Chapter 53, Wolters Kluwer-Lippincott Williams and Wilkins, Philadelphia.

E. Fabbretti, I. Afrikanova, F. Vascotto, and O.R. Burrone, 1999: Two non-structural rotavirus proteins, NSP2 and NSP5, form viroplasm-like structures in vivo. J Gen Virol. 1999, 80: 333-339.

M. Hernandez, D. Ghersi, and R. Sanchez, 2009: SITEHOUND-web: a server for ligand binding Site identification in protein structures. Nucleic Acids Res.37: W413-W416.

H. Jayaram, Z. Taraporewala, J.T Patton, and B.V. Prasad, 2002: Rotavirus protein involved in genome replication and packaging exhibits a HIT-like fold. Nature. 417: 311-315.

X. Jiang, H. Jayaram, M. Kumar, S.J. Ludtke, M.K. Estes, and B.V. Prasad, 2006: Cryoelectron microscopy structures of rotavirus NSP2-NSP5 and NSP2-RNA complexes: implications for genome replication. J Virol. 80: 10829-10835.

M. Kumar, H. Jayaram, R. Vasquez-Del Caprio, X. Jiang, Z.F. Taraporewala, R.H. Jacobson, J.T. Patton, and B.V. Prasad, 2007: Crystallographic and biochemical analysis of rotavirus

NSP2 with nucleotides reveals a nucleoside diphosphate kinase-like activity. J Virol. 81: 12272-12284.
D. Martin, M. Duarte, J. Lepault, and D. Poncet, 2010: Sequestration of Free Tubulin Molecules by the Viral Protein NSP2 Induces Microtubule Depolymerization during Rotavirus Infection. J Virol. 84(5): 2522-2532.
U.D. Parashar, E.G. Hummelman, J.S. Bresee, M.A. Miller, and R.I. Glass, 2003: Global illness and deaths caused by rotavirus disease in children. Emerg Infect Dis 9:565–572.
J.T. Patton, 2012: Rotavirus diversity and evolution in the post-vaccine world. Discov Med 13: 85–97.
J.B. Pesavento, S.E. Crawford, M.K. Estes, and B.V. Prasad, 2006: Rotavirus proteins: structure and assembly. *Curr. Top. Microbiol. Immunol*. 309, 189–219.
M.C. Ruiz, T. Leon, Y. Diaz, and F. Michelangeli, 2009: Molecular biology of rotavirus entry and replication. Scientific World Journal. 9: 1476-1497.
O. Schueler-Furman, and D. Baker, 2003: Conserved residue clustering and protein structure prediction. Proteins. 52: 225-235.
Z.F. Taraporewala, D. Chen, and J.T. Patton, 1999: Multimers formed by the rotavirus nonstructural protein NSP2 bind to RNA and have nucleoside triphosphatase activity. J Virol. 73: 9934-9943.
Z.F. Taraporewala, and J.T. Patton, 2001: Identification and characterization of the helix-destabilizing activity of rotavirus nonstructural protein NSP2. J Virol. 75: 4519-4527.
J.E. Tate, A.H. Burton, C. Boschi-Pinto, A.D. Steele, J. Duque, et al., 2012: 2008 estimate of worldwide rotavirus-associated mortality in children younger than 5 years before the introduction of universal rotavirus vaccination programmes: a systematic review and meta-analysis. Lancet Infect Dis 12: 136–141.
R. Vasquez-Del Caprio, F.D. Gonzalez-Nilo, G. Riadi, Z.F. Taraporewala, and J.T. Patton, 2006: Histidine triad-like motif of the rotavirus NSP2 octamer mediates both RTPase and NTPase activities. J Mol Biol. 362: 539-554.

Figure 1: The conservation scores were projected onto the spacefill (A) and backbone (B) model of the non-structural protein.

Figure 2: Predicted binding sites obtained from SITEHOUND (A) and novel binding sites (B). Site 3 (A) and site1 (B) is shown green; Site 5 (A) and site 2 (B) is shown red; Site 7 (A) and site 3 (B) is shown cyan; Site 8 (A) and site 4 (B) is shown blue; Site 9 (A) and site 5 (B) is shown violet. Binding sites are shown in stick format.

Table 1: Variable residues and conserved residues of NSP2 identified using ConSurf server (Celniker et al., 2013; Ashkenazy et al., 2010; Berezin et al., 2004).

Residues*	**NSP2**
Conserved	MET1-CYS6, CYS8-ASP15, TYR17, PHE19-ASN23, LEU25-ILE27, ALA29-LYS34, LYS38, ASP39, ASP41-ILE47, TYR49-ALA52, PRO54, GLN56, PHE57, LYS59-THR63, ASP65-SER67, GLY69, ASN71, GLU73, PHE77, LYS79-ALA81, CYS85-LEU88, SER90, THR94, GLN95, LEU102, SER103, SER107, HIS110, GLU112, ASN113, VAL115, ARG117, ASN120, GLN122, ASP123, LEU125-SER128, LEU131, LEU132, ILE138-ILE140, GLU145-THR149, THR151-VAL158, GLN160, ASN161, PHE164-TRP167, LEU169, TYR171, HIS174, LEU176, PRO178, ILE179, ASP181, ASN183, PHE184, GLU186, LYS188, THR190, ASN192, GLU193, LYS195, PRO196, SER198, GLU204, GLU208-ASN214, ALA217, ILE219-ARG227, LYS230-VAL234, ASN236-VAL241, ALA243, ASN248, LYS250, PHE257, LEU260-ARG263, ILE265, GLN267, ASN268, TYR270, ALA271, PHE272, SER274-MET276, GLN278, ASN280, CYS285, LYS286, LEU288-LYS292, LYS294, PHE300, GLY302, SER304, ARG307, LYS308, ASP310-SER313, GLY316
Variable	LYS18, ASN24, VAL35-LYS37, ILE48, PRO55, LYS58, ASN64, PHE72, THR74, ILE75, THR78, MET82, ILE84, LEU91-VAL93, ALA96-ASN100, ARG104, ILE116, LYS118, PRO121, LYS129, ASP130, LEU133-THR136, GLY141-SER143, ALA162, LYS168, THR170, GLU173, GLN175, ILE185, LEU191, VAL200-LYS203, LEU205, VAL206, LYS215, VAL218, VAL229, PHE245, SER247, VAL249, THR251-ASP256, ASN258, THR273, THR281, LEU282, VAL284, ARG287, MET293, GLU296, LYS297, PRO299, VAL315, VAL317

* Variable residues (grades 1-3) and conserved residue (grades 7-9)

Table 2: Secondary structures conservation of NSP2 identified using ConSurf server (Celniker et al., 2013; Ashkenazy et al., 2010; Berezin et al., 2004).

NTD	**Conserved residues**	**CTD**	**Conserved residues**
β 1 (residues 2-4)	2-4	β 5 (residues 156-161)	156-158, 160, 161
β 2 (residues 7-13)	8-13	β 6 (residues 163-169)	164-167, 169
β 3 (residues 16-22)	17, 19-22	β 7 (residues 186-192)	186, 188, 190, 192
α 1 (residues 24-30)	25-27, 29, 30	α 7 (residues 199-212)	204, 208-212
α 2 (residues 76-89)	77, 79-81, 85-88	β 8 (residues 217-219)	217, 219
α 3 (residues 95-103)	95, 102, 103	β 9 (residues 226-229)	226, 227
β 4 (residues 104-106)	-	α 8 (residues 234-247)	234, 236-241, 243
α 4 (residues 108-118)	110, 112, 113, 115, 117	α 9 (residues 267-275)	267, 268, 270, 271, 272, 274
α 5 (residues 125-126)	125-126	α 10 (residues 282-290)	285, 286, 288-290
α 6 (residues 130-140)	131, 132, 138-140	α 12 (residues 303-311)	304, 307, 308, 310, 311

Table 3: Ligand Binding Sites predicted by SITEHOUND server (Hernandez et al., 2009).

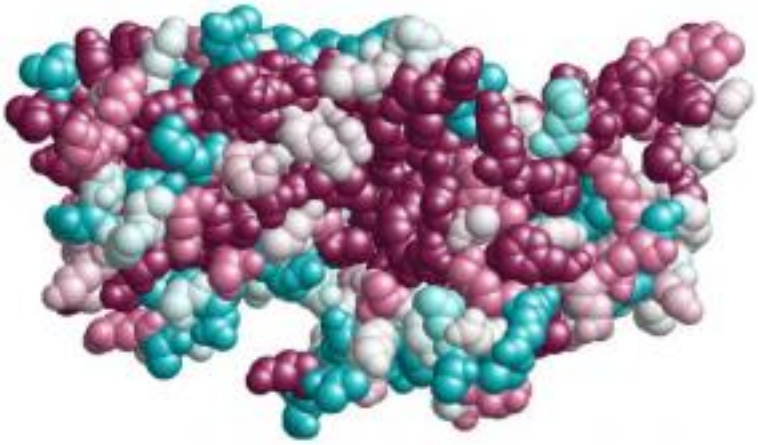
Site *	Residues in ligand binding site of NSP2	Site volume ($Å^3$)
3	Asn113, Ile116, Arg117, Asn120, Gln122- Ile124, Thr149, Glu153-Ile157, Trp167, Thr170	63
5	Met1, Ser107, His110, Glu153, Leu169, Tyr171, Ile179, Phe184, Glu186, Lys188, His225, Arg227, His237, Arg240	44
7	Tyr231, Val234, Ala235, Trp266-Asn268, Leu289- Met293	38
8	Ala5, Pro10, Asp41-Phe43, Pro54, Gln56, Phe57, Val98-Asn100, Leu102, Ser103	37
9	Lys28, Leu31, Thr32, Asn62, Thr63, Met70-Thr74	33

*The numbering of the sites is according to the rank obtained from SITEHOUND.

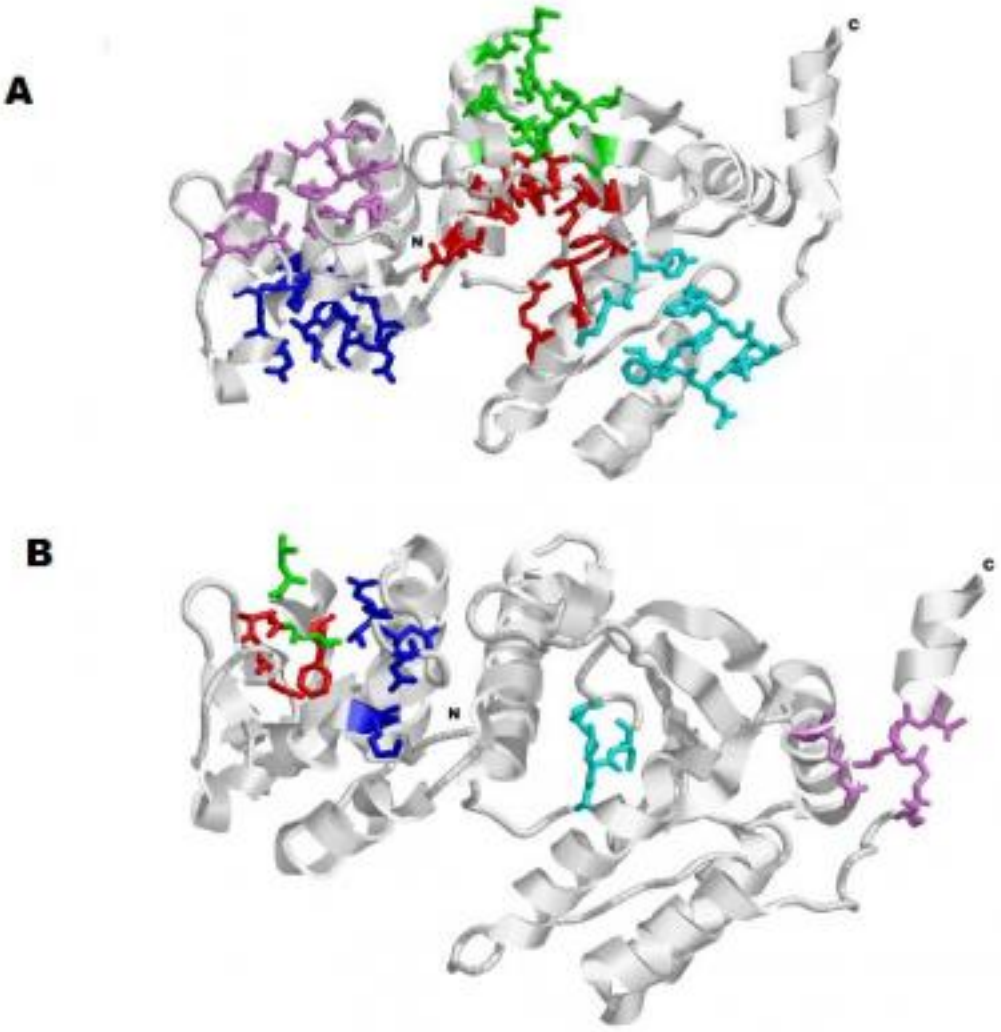
Table 4: Novel Binding Sites identified in NSP2.

Site	Residues in putative binding site
1	Leu25-Ile27, Met30
2	Phe22, Ala26, Ala29 and Met30
3	Ile146-Thr148, Thr151
4	Lys59, Asn62-Asn64, Asn66
5	Arg210, Lys301-Thr305

A



B


A
C
N
B
C
N