Receptor modification as a therapeutic approach against viral diseases

Rabia Farid, Mohammad Haroon Khan & Hamid Rashid*

Department of Bioinformatics, Muhammad Ali Jinnah University Islamabad, Pakistan; Hamid Rashid - Email: drhamid@jinnah.edu.pk; *Corresponding author

Abstract:
Poliovirus causes flaccid paralysis through the destruction of motor neurons in the CNS. Susceptibility to its infection is mainly due to the interaction in between the surface capsid proteins and its receptors on the host cell surface, important for binding, penetration and other necessary events during early infection. Receptor modification is a new approach to treat viral diseases by the modification of target proteins structure. Binding domains are modified in an effective way to make it difficult for the virus to recognize it. In this study, tolerant and intolerant induced mutations in the poliovirus receptor, VP1 and VP2 were identified and substituted in the seed sequence to get the modified versions. Substitutions causing changes in initial folding were short listed and further analyzed for high level folding, physiochemical properties and interactions. Highest RMSD values were observed in between the seed and the mutant K90F (3.265 Å) and Q130W (3.270Å) respectively. The proposed substitutions were found to have low functional impact and thus can be further tested and validated by the experimental researchers. Interactions analyses proved most of the substitutions having decreased affinity for both the VP1 and VP2 and thus are of significant importance against poliovirus. This study will play an important role for bridging computational biology to other fields of applied biology and also will provide an insight to develop resistance against viral diseases. It is also expected that same approach can also be applicable against other viruses like HCV, HIV and other in near future.

Keywords: Poliovirus, receptor, modification, substitution, structure prediction, docking

Background:
Poliomyelitis is a disease of major public health importance [1] and is one of the important emergent viral diseases of the twentieth century [2]. Pakistan is one of the four countries worldwide where polio is still endemic despite general success in eradicating polio from most of the globe [3]. Poliovirus is a member of the genus Enterovirus in the family Picornaviridae, having a single-stranded 7500 nucleotide RNA genome of positive-sense strand [4]. There are three serotypes of poliovirus, PV1, PV2, and PV3, each with a slightly different capsid defining its receptor specificity and antigenicity [5]. The Poliovirus cause flaccid paralysis through selective destruction of motor neurons in the central nervous system [6]. Susceptibility to poliovirus infection is determined mainly by the interaction between the surface capsid proteins (VP1, VP2, and VP3) of poliovirus and the poliovirus receptor protein (CD155) on the host cell surface [7].

CD155 is a transmembrane protein with 3 Ig-like extracellular domains, D1-D3, where D1 is recognized by the polio virus [8]. The Ig-like folds have predominant function related to cellular adhesion and activation [9, 10]. They activate natural killer cells [11, 12] and also play a role in cell motility and tumor cell invasion [13]. Membrane bound PVRs are considered to play important roles in the early steps of infection, such as binding of the virus to the cell surface, penetration of the cell and uncoating of the virus [14]. The virus–receptor interactions for PVs have a greater dependence on hydrophobic interactions [8].

BIOINFORMATION

Binding to a specific receptor is the most crucial step in viral infectious cycle. Blocking these receptors can prevent infection. Entry inhibitors are gaining interest as promising antiviral therapeutics which acts by preventing binding and entry of the virion. Viral infections can be inhibited by blocking the host cell receptors used by viruses to gain entry into the cells. Various strategies have been proposed for blocking the receptors. The purpose of this study was to predict induced mutations which can alter the binding domain of CD155 receptor in a significant way to make it unattachable and unrecognizable for the polio virus and thus blocking the viral entry into the cell which will prevent infection.

Methodology:
Virus receptors are important for viral binding, penetration and other necessary events during early infection. Exploring their functional regions and pattern of interactions in each step is essential for better understanding of the molecular mechanisms of host-virus interaction. Sequences of PVR, VP1 and VP2 proteins were retrieved from the UniProt database (www.uniprot.org) and were searched for binding domains through pfam (pfam.sanger.ac.uk). Among the identified domains, the one important for viral attachment were identified and confirmed from the literature. The next step was to predict significant substitutions in the binding domains which can cause significant structural and interacting variations without enforcing functional impacts. Sorting Intolerant From Tolerant (SIFT) was used (http://blocks.fhcrc.org/sift/SIFT.html) for predicting the impact of the proposed substitution and protein conservation analysis. Protein evolution and function are inter related, variations at conserved regions are expected to be less tolerated than those at the diverse positions. Its algorithm is using a modified version of the PSIBLAST [15] and Dirichlet mixture regularization [16] for the proteins global multiple sequence alignment against the query sequence. This program generates alignments with a large number of homologous sequences and assigns scores to each residue, ranging from 0 to 1. Scores ≤ 0.05 are predicted to be intolerant or deleterious, whereas scores >0.05 are considered tolerant. Higher the tolerance index of a particular substitution, lesser is its likely impact.

Better understanding of the functional impacts of residue substitution, potential causes and molecular basis depends on the relevant information related to sequences, structures and solvent accessibility. Secondary structures for the seed and the mutants of the PVR and VP proteins were predicted by using the Hierarchal Neural Network (HNN) (http://npsa-pbil.ibcp.fr/cgi\bin/npsa_automat.pl?page=/NPSA/npsa_hnn.html) and PRALINE (http://www.ibi.vu.nl/programs/pralinewww) was used to observe the variations in secondary structure elements. A detail insight of the protein three-dimensional structure is essential for understanding its functionality and thus was predicted for all the seeds and mutants by using LOOPP server (http://clsb.ices.utexas.edu/loopp/web). It performs multiple sequence alignment, sequence profiling, threading and then combines the results to generate a 3D model. All the predicted models were refined with Chiron: Rapid Protein Energy Minimization Server through discrete molecular dynamics for each residue. The structural parameters were further validated for Z-Score through ProSA-web (https://prosa.services.came.sbg.ac.at/prosa.php); VERIFY 3D [17] and

WHAT IF [18]. The refined and validated 3D structures were compared for similarity by PDBeFOLD (http://www.ebi.ac.uk/msd-srv/ssm), an interactive service for protein 3D structures comparison.

Amino acid substitution can cause physiochemical differences, which in turn can affect the protein interactions, physiochemical properties were predicted through ProtParam (http://au.expasy.org/tools/protparam.html). Functional impacts of the induced mutations were investigated through mutation assessor (http://mutationassessor.org). The functional impact was assessed based on evolutionary conservation of the affected residues in protein homologs. Protein-protein interactions were analyzed through HEX (http://hexserver.loria.fr) for the PVR with both the VP1 and VP2 to further understand their complex structure and also to analyze the impact of proposed substitutions on the virus-receptor interaction. HEX is an interactive molecular graphics program for calculating and displaying feasible docking modes of pairs of protein and DNA molecules.

Discission:
Retrieval of Protein Sequence and domain identification
Polio Virus Receptor is the product of PVR gene and belongs to the family of immunoglobins. It is composed of three immunoglobin domains (D1 immunoglobin V-Set domain from residues 28-142, D2 CD80 like C2-set immunoglobin domain from residues 143-242 and D3 immunoglobin domain from residues 243-330), a transmembrane domain from 331-355 and a cytoplasmic tail from residues 356-417. D1 immunoglobin V-Set domain is actually involved in viral attachment.

Mutation identification and substitution
Receptor modification is a new approach to block viruses by inducing substitutions in the residues at the binding domain to make it difficult for the virus to recognize it. The intolerant mutations were substituted in the seed sequence, out of 2166 mutations only 7 were predicted intolerable which are of primary importance for the proceeding study. At the residue position 90, W, C and F were predicted to be intolerable substitutions in place of the original K. Similarly at the position 124, W in place of L and at 130, W, Y and F in place of Q were predicted to be intolerable.
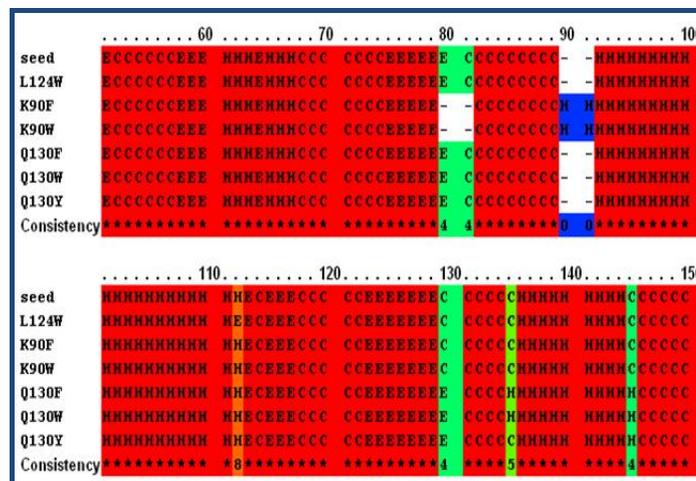
Figure 1: Secondary structure elements alignment of seed and mutated PVR proteins.

Structure prediction, comparison and characterization

Secondary structures can give a useful insight into the folding pattern and function. The residues in the seed and mutants were analyzed for secondary structure to check the protein stability. It was observed that there were 84 helices, 114 - sheets and 219 coils in the seed PVR. This number either decreased or increased whenever a mutation was incorporated which means these substitutions may affect the key residues and thus the folding pattern **(Figure 1).** Substitutions may also affect the protein high order structure which determines protein functions. It is therefore better to use the three-dimensional protein knowledge for better and in-depth understandings. All the predicted models were analyzed for structurally variable regions (SVR) by superimposing in 3D. The highest RMSD values were observed in between the seed and the mutant K90F (3.265 Å) and Q130W (3.270Å), while the least was recorded in between the seed and Q130Y (1.374Å). A higher RMSD value means more deviation between the structures and thus may have more functional impacts. The highest Q-score was recorded for the K90F and Q130W i.e. 0.63 while for K90W and Q130F, the scores recorded were 0.112. The lowest Q-score was for L124W i.e. 0.110. The sequence identity was observed as 0.924 for L124W, Q130F and Q130Y, 0.101 for K90F and Q130W and 0.911 for K90W. According to the predicted physiochemical properties, pI was observed changed for mutants K90F and K90W from the normal 6.07 to 5.94 while change in GRAVY was recorded as the most affected parameter, deviated the most from the seed (-0.048) to Q130Y (0.033) to K90F (-0.032).

Functional analysis

The biochemical differences, nature and location of amino acid substitution can affect the protein in various ways and is therefore important to determine whether it can alter the protein function. It was observed from the results that functional impact score lied in between 0.83 and 0.95 for all the proposed substitutions i.e. 0.83 for mutations at position 90, 0.91 and 0.95 for the mutations at position 130. The functional impact score lies below 1 it means, mutations have low functional impact and thus can be further tested and validated by the experimental researchers for further research and future therapeutics against the polio virus.

Protein-protein Interaction analysis

Docking aims to determine the three dimensional structures of the protein-ligand complexes [19]. Protein-protein interactions are important for all biological processes [20] but are always difficult to predict due to the different complexities like thermodynamics and topography etc. Protein-protein docking results of the PVR seed along with its mutated versions with VP1 and VP2 proteins are shown in **Table 1 & 2 (see supplementary material)**. Most of the PVR mutants in complex with VP1 were observed with decreased number of bonding interactions. Mutant K90W-VP1 was the only complex having more than the normal hydrogen bond (7) interaction with a salt bridge while mutant Q130Y-VP1 having no interaction at all and thus is the most suitable substitution against polio virus. In case of Seed-VP2 complex, 8 hydrogen bond interactions were observed. When mutants were compared to the Seed-VP2 complex, it was noticed that with the exception of Q130Y-VP2 (having 8 interactions), all the mutants were having decreased number of interactions. From cluster analysis of the docked complexes, it is clear that, substitution Q130Y is the most

significant against VP1 while Q130F against VP2. It was further observed that substitutions L124W and Q130W are significant against both the VP1 and VP2 **(Table 1 & 2; Figure 2).**
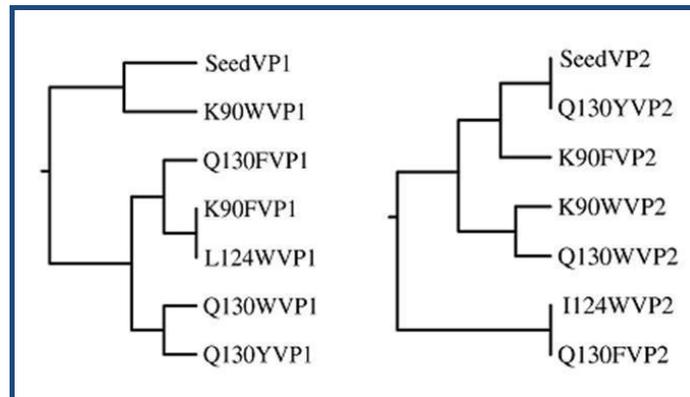

Figure 2: Cluster analysis of Protein-protein interactions analysis of seed and mutated PVRs with VP1 and VP2

Conclusion:

Our current analysis focuses on receptor modifications against polio virus, and our findings could significantly explain the types of substitution which can structurally modify the polio virus receptors, making them unrecognizable for them. Six mutations (K90F, K90W, L124W, Q130F, Q130W and Q130Y) were predicted and substituted in the CD155 protein key residues which are promising against polio virus. This approach might also be applied against certain other viruses which are otherwise difficult to block. This study bridges computational biology to molecular, structural biology and experimental biology, which may deepen our understanding towards novel antiviral therapeutics. Structure and sequence based computations were systematically evaluated and were applied for the induced substitutions in PVR1 and PVR2 proteins and have provided a comprehensive structural explanations for impacts of proposed substitutions. It is obvious from the results that this novel bioinformatics approach can be significantly used against HCV, HIV etc which are the major and difficult to handle agents threatening humanity in general. Substitution Q130Y and Q130F were found the most effective against VP1 and VP2 respectively while substitutions L124W and Q130W are effective against both the VP1 and VP2. Such techniques should be developed through which these mutations can be induced *in vitro* to investigate their effectiveness at the first stage and later on *in vivo*.

References:

[1] Dutta A *et al. Vaccine.* 2008 **26**: 5767 [PMID: 18755232].
[2] Trevelyan B *et al. Ann Assoc Am Geogr*. 2005 **95**: 269 [PMID: 16741562].
[3] Afzal O & Rai MA, *Vaccine*. 2009 **27**: 5431 [PMID: 19643214].

BIOINFORMATION

[4] Hogle JM, *Annu Rev Microbiol.* 2002 **56**: 677 [PMID: 12142481].

[5] Blondel B, *J Neurovirol.* 1998 **4**: 1 [PMID: 9531008].

[6] Suzuki Y, *Gene.* 2006 **373**: 16 [PMID: 16500041].

[7] Mendelsohn CL *et al*. *Cell.* 1989 **56**: 855 [PMID: 2538245].

[8] Zhang P *et al*. *Proc Natl Acad Sci U S A.* 2008 **105**: 18284 [PMID: 19011098].

[9] Mueller S & Wimmer E, *J Biol Chem*. 2003 **278**: 31251 [PMID: 12759359].

[10] Ogita H & Takai Y, *IUBMB Life* 2006 **58**: 334[PMID: 16754328].

[11] Bottino C *et al*. *J Exp Med*. 2003 **198**: 557 [PMID: 12913096].

[12] Fuchs A *et al*. *J Immunol*. 2004 **172**: 3994 [PMID: 15034010].

[13] Sloan KE *et al*. *BMC Cancer*. 2004 **4**: 73 [PMID: 15471548].

[14] Ohka S *et al*. *J Virol*. 2004 **13**: 7186 [PMID: 15194795].

[15] Altschul SF *et al*. *Nucleic Acids Res.* 1997 **25**: 3389 [PMID: 9254694].

[16] Sjolander K *et al*. *Comput Appl Biosci*. 1996 **12**: 327 [PMID: 8902360].

[17] Eisenberg D *et al*. *Methods in Enzymol*. 1997 **277**: 396 [PMID: 9379925].

[18] Vriend G, *J Mol Graph*. 1990 **8**: 52 [PMID: 2268628].

[19] Chen R *et al*. *Proteins*. 2003 **52**: 88 [PMID: 12784372].

[20] Stelzl U *et al*. *Cell*. 2005 **122**: 957 [PMID: 16169070]

BIOINFORMATION

Supplementary material:

Table 1: Protein-protein interactions analysis of seed and mutated PVR's with VP1

Complex	NIR	IA (Å²)	NHB	NSB	NNC	RRI		Distance
Seed-VP1	A= 28	1951	5	0	261	GLU(101)	SER(44)	3.24
						VAL (40)	THR(46)	2.44
						PRO(100)	ARG (109)	2.19
						PRO(100)	ARG (109)	2.12
	B = 33	1928				SER (146)	GLN (176)	2.20
K90F-VP1	A = 31	1569	2	0	371	ASN(168)	GLY (41)	2.79
	B = 32	1532				ARG (62)	LYS (115)	3.04
K90W-VP1	A = 37	1754	7	1	366	ASP (67)	ARG (111)	1.24
						ASN (204)	ASN (118)	3.07
						VAL (13)	ARG (136)	1.86
						GLN (111)	THR (140)	2.90
						THR (22)	THR (140)	2.47
						THR (65)	GLY (141)	2.73
	B = 38	1788				LYS (116)	THR (206)	3.09
I124W-VP1	A = 37	1947	2	0	406	LYS (147)	ALA (114)	3.00
	B = 35	2029				PRO (95)	PHE (199)	2.44
Q130W-VP1	A = 28	1602	1	0	303	VAL (40)	HIS (196)	2.51
	B = 29	1640						
Q130F-VP1	A = 35	1685	3	0	304	LYS (36)	ARG (80)	3.30
						LYS (147)	PRO (116)	2.79
	B = 32	1721				LYS (147)	GLN (117)	2.93
Q130Y-VP1	A=29	1518	0	0	260	0	0	0
	B = 27	1434						

NIR = No. of interface residues,   IA = Interface area
NHB = No. of hydrogen bonds   NSB = No. of salt bridges
NNC = No. of non-bonded contacts   RRI = Residue-Residue interaction

Table 2: Protein-protein interactions analysis of seed and mutated PVR's with VP2

Complex	NIR	IA (Å²)	NHB	NSB	NNC	RRI		Distance
Seed-VP2	A=26	1294	8	0	314	ARG 87	GLU 71	2.44
						ARG 87	GLU 71	1.19
						ASP 67	ARG 114	2.84
						MET 89	GLY 119	2.85
						GLN 94	ASP 136	3.04
						THR 252	ARG 172	2.69
	B = 25	1396				THR 24	ALA 175	3.15
						ASN 113	LEU 196	2.87
K90F-VP2	A=37	1754	7	1	366	ASP 67	ARG 111	1.24
						ASN 204	ASN 118	3.07
						VAL 13	ARG 136	1.86
						GLN 111	THR 140	2.90
						THR 22	THR 140	2.47
	B = 38	1788				THR 65	GLY 141	2.73
						LYS 116	THR 206	3.09
K90W-VP2	A=33	1649	5	1	405	GLY 188	GLN 7	2.62
						LEU 234	GLN 7	2.70
						ASN 238	ASP 107	2.87
	B = 29	1599				THR 158	ARG 111	2.71
						ASN 137	LEU 205	2.80
I124W-VP2	A=31	1569	2	0	371	ASN 168	GLY 41	2.79
	B = 32	1532				ARG 62	LYS 115	3.04
Q130W-VP2	A=32	1690	5	0	287	THR 24	ARG 80	2.23
						THR 24	ARG 80	2.43
						ARG 87	GLU 195	2.61
						ASP 88	HIS 196	2.27
	B = 27	1694				ASP 88	HIS 196	3.31
Q130F-VP2	A=31	1745	2	0	327	SER 45	VAL 2	2.93
	B = 31	1665				GLU 40	TRP 109	2.25
Q130Y-VP2	A=44	1937	8	0	497	TYR 57	SER 169	2.96
						ARG 85	GLN 272	2.58
						THR 149	ASN 113	2.53
						SER 161	HIS 109	2.61
						SER 161	THR 24	2.95
						GLN 162	THR 252	2.32
						PRO 164	ARG 62	3.30
	B = 41	1912				VAL 171	GLU 152	2.94

NIR = No. of interface residues,   IA = Interface area
NHB = No. of hydrogen bonds   NSB = No. of salt bridges
NNC = No. of non-bonded contacts   RRI = Residue-Residue interaction