Peptide-based epitope design on non-structural proteins of SARS-CoV-2

R S Swathika^a, S Vimal^{2a}, E Bhagyashree^{3a}, Elakkiya Elumalai^b, Krishna Kant Gupta^{*4a}

a: School of chemical and Biotechnology, SASTRA Deemed University, Thanjavur, India

b: Center for Bioinformatics, Pondicherry University, Pondicherry, India

Email ID: krishnakant@scbt.sastra.ac.in

Contact No: +91-6268467956

Abstract

The SARS-CoV-2 virus has caused the severe pandemic, COVID19 and since then its been critical to produce a potent vaccine to prevent the quick transmission and also to avoid alarming deaths. Among all type of vaccines peptide based epitope design tend to outshine with respect to low cost production and more efficacy. Therefore, we started with obtaining the necessary protein sequences from NCBI database of SARS-CoV-2 virus and filtered with respect to antigenicity, virulency, pathogenicity and non- homologous nature with human proteome using different available online tools and servers. The promising proteins was checked for containing common B and T- cell epitopes. The structure for these proteins were modeled from I-TASSER server followed by its refinement and validation. The predicted common epitopes were mapped on modeled structures of proteins by using Pepitope server. The surface exposed epitopes were docked with the most common allele DRB1*0101 using the GalaxyPepDock server. The epitopes, **ELEGIQYGRS** from Leader protein (NSP1), **YGPFVDRQTA** from 3c-like proteinase (nsp5), **DLKWARFPKS** from NSP9 and **YQDVNCTEVP** from Surface glycoprotein (spike protein) are the epitopes which has more hydrogen bonds. Hence these four epitopes could be considered as a more promising epitopes and these epitopes can be used for future studies.

Keywords: epitope, Bcell, T cell, SARS-CoV, antigen, pathogen, docking

36

37

38 **Introduction**

39 In December 2019, COVID-19 was first reported in Wuhan, China. It is a severe
40 acute respiratory syndrome that causes respiratory disease which leads to death or
41 other serious complications such as heart disorders and multi organ failure [1]. The
42 virus was first transmitted from animals to humans. Majority of infected people
43 shows moderate symptoms but pathological effects of COVID-19 is heterogeneous.
44 Person to person transmission occurs through contact with infected secretion mainly
45 via respiratory droplets.

46 Currently, several old drugs such as chloroquine (antimalarial),
47 remdesivir(antiviral), azithromycin (antibiotic), tocilizumab (immunomodulatory
48 agent) and corticosteroids, are being used to counteract the symptoms of COVID-
49 19 [2,3,4]. The traditional vaccines have to many disadvantages such as allergic and
50 autoimmune reactions. They need storage at a cold temperature and have low
51 stability and. Therefore, peptide-based vaccines may be a promising strategy which
52 involves minimal microbial components to stimulate humoral and adaptive
53 immunity against a microorganism. These vaccines are able to target very specific
54 epitopes removing the risks associated with allergic and autoimmune responses [5].

55 The SARS CoV 2 contains four major structural proteins: Membrane protein, Spike
56 protein, Envelope protein and Nucleo-capsid protein. Early work on the new corona
57 virus have highlighted the importance of surface glycoprotein (spike protein), also
58 called as S proteins which act as the key mediator for the virus to enter the host cell
59 [6]. While comparing SARS COVID and SARS COVID2 the spike protein is found
60 to be binding to a receptor called angiotensin converting receptor ACE-2 which
61 helps in entering to the host cells. It has been reported that the spike protein of the
62 novel virus binds to ACE -2 more efficiently when compared to that of SARS
63 COVID. Therefore, it spreads more effortlessly and is also more contagious. Many

64 major pharmaceutical industries and also universities with research facilities are
65 working hard to find a potential vaccine to curb the global pandemic situation [7,8].

66 Peptides have become one of the most important vaccine candidates given their
67 comparatively easy production and design, chemical stability of structure and non-
68 appearance of any possible infectious potential[9,10,11,12]. Hence we used
69 subtractive genomics and reverse vaccinology [13] approach to contribute towards
70 peptide based epitope design to curb this pandemic infection. All 38 proteins
71 (Genbank accession number: NC_045512) encoded by SARS CoV 2 genome was
72 screened for its Antigenicity, essentiality, virulency and non-homology with human
73 proteome. The final set of proteins were screened for B and T cell epitopes, mapped
74 on 3D protein structures and finally docked with DRB10101. Parallel bioinformatic
75 predictions identified a priori potential B and T cell epitopes for 2019-nCoV. We
76 hope that there is a high probability that these epitopes are targets for immune
77 recognition of 2019-nCoV.

78 **2. METHODOLOGY**

79 **2.1. Retrieval of proteome sequence data:**

80 The protein sequence of the SARS-CoV-19 NC_045512 was retrieved from NCBI
81 (Genbank accession number: NC_045512). It consists of the complete proteome
82 sequence of SARSCoV-19. The FASTA format for all proteins were retrieved.

83 **2.2. Comparative analysis with human genome:**

84 Each protein sequence was compared with human proteome by using **BLASTp** to
85 check the non-homogeneity to avoid auto immunity. BLASTP programs search
86 protein databases using a protein query.

87 **2.3. Antigenicity and virulence prediction:**

88 The non-homologous proteins must be antigenic for the stimulation of both B and
89 T cell immunity. Therefore, the antigenicity and virulency were predicted for those
90 proteins using the **Vaxijen2.0** [14] and **MP3** [15] servers respectively.

91 **2.4. Position of protein sequences in the transmembrane region:**

92 The **TMHMM** server [16] was used to analyse the trans-membrane position of
93 protein. It is a membrane protein topology prediction method based on a hidden
94 Markov model.

95 **2.5. Evaluation of essential proteins:**

96 The proteins were checked for its essentiality to the virus by using database of
97 essential gene (**DEG ver.15.2**). DEG [17] is a database that consists of data which
98 includes protein coding genes, on-coding RNA and other essential genomic
99 elements. **BLASTp** with default parameters were used to find essential proteins.

100 **2.6. Epitope prediction:**

101 B and T cells are the most responsible for inducing immune response in human body.
102 For the obtained protein sequences B cell epitopes were found using **ABCpred**
103 **server** [18]. To determine each sequences' binding affinity towards MHC-I and
104 MHC-II molecules, **MHC-I binding prediction tool**
105 (<http://tools.immuneepitope.org/mhci/>) and **MHC-II binding prediction tool**
106 (<http://tools.iedb.org/mhcii/>) were used respectively. A quantitative prediction
107 server MHCpred v.2.0 [19] was used to find half maximal (50%) inhibitory
108 concentration IC₅₀. The epitopes which had IC₅₀ values comparatively less than 100
109 nM were taken for further process.

110 **2.7. Prediction and analysis of 3D structure:**

111 The 3D structure of the target proteins was predicted using online software I-
112 TASSER [20]. It builds three dimensional atomic models from multiple threading
113 alignments and iterative structural assembly simulations. The obtained structures
114 were refined using GALAXYWEB server [21] and the appropriate models were
115 made choice by considering highest C-score value. The selected models were
116 validated using the SAVES server (<https://saves.mbi.ucla.edu/>) which includes
117 PROCHECK which estimates the dependability of the proteins by calculating the
118 proteins' stereo chemical quality which is done by constructing the Ramachandran
119 plot for the backbone structure of the protein.

2.8. Epitope exo-membrane topology evaluation:

The exo-membrane topology of the proteins were analysed using the Pepitope server (<http://pepitope.tau.ac.il/>) where position of each epitope in the predicted protein structure model was identified. Under epitope mapping algorithm, “**Combined**” option was selected. This is done to make sure that the targeted epitopes are exposed on the surface and ensure that the epitopes considered are not hidden within the protein globular structure.

2.9. Epitope binding mode and interaction analysis:

The selected epitopes which fulfilled all the constrains were then subjected to docking with the 3-D structure of most dominant allele DRB1*0101 (1AQD). To perform docking **GALAXY Pepdock server** [22] was used. From the obtained peptide-protein complex, the best complex was considered based on hydrogen bonding pattern, similarity score and accuracy score. The binding mode of the complex and binding interaction was visualised using the **LIGPLOT tool** [23].

3. RESULTS AND DISCUSSION

3.1 Sequence retrieval and evaluation of protein sequences:

The Primary genomic sequence data of SARS-CoV-19 was fetched from the NCBI database. From the obtained Genomic data 38 protein sequences were collected in FASTA format.

3.2 Virulency:

Virulency is considered as one of the main factor for studying the pathogenesis and the MP3 tool uses a combined SVM-HMM approach to yield increased efficiency and reliability to predict pathogenic proteins. Out of 38 proteins, 11 proteins were found to be virulent (Table 1 & Supplementary Table1).

3.3 Antigenicity:

For the candidate protein sequences, antigenicity was evaluated using the Vaxijen server. Antigenicity is the ability of a chemical structure to bind explicitly with a group of products that have adaptive immunity. The VaxiJen 2.0 server uses an

incipient alignment independent method for antigen prognostication predication. Protein sequences that showed antigenicity value (>0.4) were identified as antigenic and were subjected to further studies. Similarly, 11 proteins were predicted to be antigenic. The nsp9 protein was most antigenic (Table 1 & Supplementary Table1).

3.4 Human genome comparison:

It is necessary to predict the similarity with human genome because as any similarity between human and virus genome may lead to side effects including auto immunity issues. Hence all 38 protein sequences were subjected to comparison with human proteome by using BLASTp tool with default parameters. “**No significant similarity found**” hits were considered as non-homologous proteins. All 11 proteins were found to be non-homologous to human proteome (Table 1 & Supplementary Table1).

3.5 Membrane topology analysis:

The presence of these protein sequences either on the Transmembrane helix or outside the membrane were identified using the TMHMM server 2.0. This is necessary because proteins present in the core are less exposed and not viable. Hence the proteins which are present outside or in the Trans-membrane region were considered. In Table 1 & Supplementary Table1 , all 11 proteins were shown to be present outside of trans membrane region.

3.6 Essentiality of protein sequences:

The protein sequences were screened for essentiality using the DEG version 15.2 database which is a database that consists of the details of essential genes. Here for the submitted protein sequences BLAST is performed against the DEG database and hits are obtained which concludes that all 11 structural proteins are essential for the virus growth and survival. The protein sequences which fulfilled all the above mentioned filtration process were tabulated separately.

178

Table1: Final set of nCoV proteins for epitope prediction

Locus	Protein name	Human protein Similarity	Virulence	Antigenicity	TM Helix/Outside region	Essentiality of protein
YP_009725297	Leader protein(nsp1)	No	P	0.4064	Outside	Yes
YP_009725301	3c-like proteinase(nsp5)	No	P	0.4159	Outside	Yes
YP_009725304	nsp8	No	P	0.4008	Outside	Yes
YP_009725305	nsp9	No	P	0.6476	Outside	Yes
YP_009725306	nsp10	No	P	0.4039	Outside	Yes
YP_009742608	Leader protein	No	P	0.4064	Outside	Yes
YP_009742612	3c-like proteinase	No	P	0.4078	Outside	Yes
YP_009742615	nsp8	No	P	0.4107	Outside	Yes
YP_009742616	nsp9	No	P	0.6226	Outside	Yes
YP_009742617	nsp10	No	P	0.4149	Outside	Yes
YP_009724390	Surface glycoprotein(spike protein)	No	P	0.4661	Outside/TM helix	Yes

179

180

181

182

3.7. B-cell Epitope prediction:

After obtaining final set of 11 proteins, B-cell epitopes were predicted for each of these sequences. B-cell epitope is a part of the protein that will bind with the antibody. Hence it is necessary to find the B-cell epitopes for these 11 proteins. The B-cell epitopes predicted for each 11 proteins are given in Supplementary Table 2.

3.8. T-cell Epitope prediction:

Next we predicted MHC Class I and MHC Class II epitopes. All the epitopes are given in Supplementary Table 2. We looked for those epitopes that was common to both B-cell and Tcell epitopes. The list of common epitopes are given in Supplementary Table 2.

3.9. Inhibitory concentration:

The common epitopes were checked for the inhibitory concentration (IC_{50}) values that predicts epitope binding to major histocompatibility complexes (MHCs). For this, MHCpred tool was used. In this server common epitope was checked for its binding affinity with allele **DRB1*0101** as it is considered as immunodominant allele. All those common epitopes were selected whose IC_{50} value was <200 nM. Similarly, the epitopes with IC_{50} values less than 200 nM considered for epitope mapping on protein structure. We considered only those epitopes whose predicted IC_{50} values was less than 100 nM (Table2).

Table 2: IC_{50} values of common epitopes

YP_009742616	B-cell	MHC-I	MHC-II	IC_{50} values
	KGPKVKYLYFIKGLNN	KGPKVKYLYF	KGPKVKYLYF	
	TGTIYTELEPPCRFVT	TGTIYTELE	TGTIYTELE	
	AGTTQTACTDDNALAY	AGTTQTACTD	AGTTQTACTD	37.76

	FPKSDGTGTIYTELEP	FPKSDGTGTI	FPKSDGTGTI	
	DLKWARFPKSDGTGTI	DLKWARFPKS	DLKWARFPKS	15.6
	ACTDDNALAYYNTTKG	ACTDDNALAY	ACTDDNALAY	1.55
	LAYYNTTKGGRFVLAL	LAYYNTTKGG	LAYYNTTKGG	
	KGLNNLNRMVGLSLA	KGLNNLNRM	KGLNNLNRM	46.88
	GGRFVLALLSDLQDLK	GGRFVLALLS	GGRFVLALLS	2.55

3.10. Prediction of 3-D structure and its validation:

All proteins except spike protein (PDB ID: 6M0J and 6LVN) were modeled and validated (Supplementary Table 3) and used for epitope mapping. The final epitopes were found on six proteins namely Leader protein(nsp1)(Figure 1), 3c-like proteinase(nsp5)(Figure 2), Nsp8 (Figure 3), Nsp9 (Figure 4), Nsp9 (isoform)(Figure 5), Nsp10 (Figure 6) and Surface glycoprotein (spike protein).

After refinement the selected models were validated using SAVES server which helps in evaluating the dependability of the protein. SAVES version 5.0 server from NIH MBI laboratory is mostly useful for validation analysis. The quality was evaluated using VERIFY3D and PROCHECK (Supplementary Table 3). From this server, Ramachandran plot was also visualized for each protein (Figure 1-6). If a protein sequence is present in more disallowed region, then that protein is not considered.

3.11 Peptide mapping

Out of the 11 protein sequences, after validating the structure from the Ramachandran plot only 6 proteins (**nsp1, nsp5, nsp8, nsp9, nsp9(isoform) and nsp10**) were having final epitopes. The epitopes were mapped and again antigenicity of those epitopes were predicted. The eight surface exposed and high antigenicity epitopes (Fig 7 a-h) were selected for docking study with “DRB1*0101” (PDB ID “1AQD”).

PROTEIN ID: YP_009725297

EPITOPE: ELEGIQYGRS (Red color)

Fig 7a Position of epitope in YP_009725297

PROTEIN ID: YP_009725301
 EPITOPE: YGPFVDRQTA (Red color)

Fig 7b Position of epitope in YP_009725301

PROTEIN ID: YP_009725304

243 EPITOPE: **ARDGCVPLNI** (Red color)

Fig 7c Position of epitope in YP_009725304

248 PROTEIN ID: **YP_009725305**

249 EPITOPE: **DLKWARFPKS** (Red color)

Fig 7d Position of epitope in YP_009725305

254 PROTEIN ID: **YP_009742616**

255 EPITOPE: **DLKWARFPKS** (Red color)

256

Fig 7e Position of epitope in YP_009742616

257

258

259

260

261

262

263 PROTEIN ID: YP_009742617

264 EPIPOPE: ASCCLYCRCH (Red color)

Fig 7f Position of epitope in YP_009742617

265

266

267

268 PROTEIN ID: YP_009724390

269 EPIPOPE: YQDVNCTEVP (Red color)(6LVN)

270

Fig 7g Position of epitope in YP_009724390

EPITOPE: **ESLIDLQELG** (blue color) (6MOJ)

Fig 7h Position of epitope in YP_009724390

3.12 Analysis of epitope binding and interaction:

The best peptide-protein complexes were ranked based on interaction parameters which include hydrogen-bonding pattern, similarity score and accuracy score. For interpreting the binding mode of the complex Ligplot server was used and results are shown in Supplementary Figure file 1(Fig 8 a-g).

3.13 Analysis of hydrogen bonding pattern:

From the obtained hydrogen-bonding interaction the one with the maximum number of hydrogen bonds were taken into consideration as this results in stability and more energetically favourable interactions. The epitopes with more number of hydrogen bonds (**ELEGIQYGRS** and **YGPFDVDRQTA**) can be considered as more promising epitopes (Table 3).

Table 3: H bond between epitope atom and protein atom

Epitopes	Peptide atom	Protein atom	Hydrogen Bond Length (Å)
ELEGIQYGRS	N	O	2.94
	N	O	3.25
	O	OG	2.98
	O	N	2.87
	N	O	3
	O	NE2	3.08
	N	OE1	3.32
	N	OD1	2.95
YGPFDVDRQTA	N	O	2.76
	N	O	2.98
	N	OH	3.05
	N	OH	2.91
	NH1	OD2	2.78
	NE2	O	2.96
	NE2	OD1	3.05
ARDGCVPLNI	N	OD1	2.79
	OD2	NH2	3.02
	O	ND2	2.81
	N	O	3.3
DLKWARFPKS	N	OD2	3.13
	O	N	3.14
	O	N	2.89
	N	O	2.61
	OC	N	3.03

	NZ	OD1	2.85
ASCCLYCRCH	N	OE2	2.94
	OH	OE2	2.97
	NH2	OE1	2.85
	NH2	OD1	3.03
YQDVNCTEVP	N	OD2	2.82
	NE	OD1	2.88
	OG1	O	2.96
	NH1	OE2	3.22
	NH2	OE2	2.81
	O	OH	2.92
ESLIDLQELG	N	OD1	2.96

297

298

299

300 **ELEGIQYGRS** from Leader protein (NSP1), **YGPVDRQTA** from 3c-like
301 proteinase (nsp5), **DLKWARFPKS** from NSP9 and **YQDVNCTEVP** from Surface
302 glycoprotein (spike protein) are the epitopes which has more hydrogen bonds. Hence
303 these four epitopes could be considered as a more promising epitopes and these epitopes
304 could be used for further clinical studies.

CONCLUSION

Initially from the 38 protein sequences considered from the whole genome of SARS-CoV19, after fulfilling the necessary constraints which include antigenicity, virulency, nonhomologous to human proteome, presence of the epitope on the membrane and essentiality of protein only 11 protein sequences remained. For these 11 proteins linear B-cell and T-cell epitope prediction were done and the common epitopes were found. The protein structure for these sequences were modelled using I-TASSER and these models were subjected to refinement and validation processes after which only 6 protein were remaining. For the spike protein available protein model from PDB were taken into consideration. Now for the 7 protein models their epitope's antigenicity were evaluated and the one with highest value for each protein was further considered for the docking process. These epitopes were docked with DRB1*0101 allele using GalaxyPepDock server and the docked structures obtained were analysed in LigPlot for hydrogen bonding interaction, similarity and accuracy score. As a result the hydrogen bonding pattern between the epitopes and the allele were analysed and the epitopes with more number of hydrogen bonds were considered to be more stable and energetically favourable. These epitopes could be used for further studies.

Acknowledgement

We sincerely acknowledge SASTRA Deemed To be University for providing computational facility.

Conflict of Interest

None

References

- 1- Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol.* 2020 Apr;5(4):536-544.
- 2- Khuroo MS. Chloroquine and hydroxychloroquine in coronavirus disease 2019 (COVID-19). Facts, fiction and the hype: a critical appraisal. *Int J Antimicrob Agents.* 2020 Sep;56(3):106101.
- 3- Popescu CP, Fischer PR. Hydroxychloroquine-azithromycin for COVID-19 - Warranted or dangerous? *Travel Med Infect Dis.* 2020 May-Jun;35:101764.
- 4- Wang M, Cao R, Zhang L, Yang X, Liu J, Xu M, Shi Z, Hu Z, Zhong W, Xiao G. Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. *Cell Res.* 2020 Mar;30(3):269-271.
- 5- Li W, Joshi MD, Singhanian S, Ramsey KH, Murthy AK. Peptide Vaccine: Progress and Challenges. *Vaccines (Basel).* 2014 Jul 2;2(3):515-36.
- 6- Huang Y, Yang C, Xu XF, Xu W, Liu SW. Structural and functional properties of SARS-CoV-2 spike protein: potential antiviral drug development for COVID-19. *Acta Pharmacol Sin.* 2020 Sep;41(9):1141-1149.
- 7- Barh D et al. "Exoproteome and secretome derived broad spectrum novel drug and vaccine candidates in *Vibrio cholerae* targeted by Piper betel derived compounds." *PloS one* vol. 8,1 (2013): e52773.
- 8- Singh A, Thakur M, Sharma LK, Chandra K. Designing a multi-epitope peptide based vaccine against SARS-CoV-2. *Sci Rep.* 2020 Oct 1;10(1):16219.
- 9-Singh, A., Thakur, M., Sharma, L.K. et al. Designing a multi-epitope peptide based vaccine against SARS-CoV-2. *Sci Rep* 10, 16219 (2020). *Scientific Reports* volume 10, Article number: 16219 (2020)

- 10-Murat TOPUZOĞULLARI, Tayfun ACAR, Pelin PELİT ARAYICI et al. An insight into the epitope-based peptide vaccine design strategy and studies against COVID19. Turk J Biol. 2020; 44(3): 215–227.
- 11-Niloofar khairkhan, Mohammad Reza Aghasadeghi, Ali namvar, Azam Bolhassini. Design of novel multiepitope constructs-based peptide vaccine against the structural S, N and M proteins of human COVID-19 using immunoinformatics analysis. (2020).
- 12-Khan MA, Hossain MU, Rakib-Uz-Zaman SM, Morshed MN. Epitope-based peptide vaccine design and target site depiction against Ebola viruses: an immunoinformatics study. Scand J Immunol. 2015 Jul; 82(1):25-34.
- 13-Moxon R, Reche PA, Rappuoli R. Editorial: Reverse Vaccinology. Front Immunol. 2019 Dec 3;10:2776.
- 14-Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics. 2007 Jan 5;8:4.
- 15-Gupta A, Kapil R, Dhakan DB, Sharma VK. MP3: a software tool for the prediction of pathogenic proteins in genomic and metagenomic data. PLoS One. 2014 Apr 15;9(4):e93907.
- 16-Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001 Jan 19;305(3):567-80.
- 17- Zhang R, Ou HY, Zhang CT. DEG: a database of essential genes. Nucleic Acids Res. 2004 Jan 1;32(Database issue):D271-2.
- 18-El-Manzalawy Y, Dobbs D, Honavar V. Predicting linear B-cell epitopes using string kernels. J Mol Recognit. 2008 Jul-Aug;21(4):243-55.

- 19- Guan P, Doytchinova IA, Zygouri C, Flower DR. MHCpred: bringing a quantitative dimension to the online prediction of MHC binding. *Appl Bioinformatics*. 2003;2(1):63-6.
- 20- Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc*. 2010 Apr;5(4):725-38.
- 21- Ko J, Park H, Heo L, Seok C. GalaxyWEB server for protein structure prediction and refinement. *Nucleic Acids Res*. 2012 Jul;40(Web Server issue):W294-7.
- 22- Lee H, Heo L, Lee MS, Seok C. GalaxyPepDock: a protein-peptide docking tool based on interaction similarity and energy optimization. *Nucleic Acids Res*. 2015 Jul 1;43(W1):W431-5.
- 23- Wallace AC, Laskowski RA, Thornton JM. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng*. 1995 Feb;8(2):127-34.

Ramachandran Plot

2661497

Plot statistics

Residues in most favoured regions [A,B,L]	87	90.6%
Residues in additional allowed regions [a,b,l,p]	8	8.3%
Residues in generously allowed regions [~a,~b,~l,~p]	1	1.0%
Residues in disallowed regions	0	0.0%
-----		-----
Number of non-glycine and non-proline residues	96	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	9	
Number of proline residues	6	
----		----
Total number of residues	113	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

Ramachandran Plot

1786878

Residues in most favoured regions [A,B,L]	106	91.4%
Residues in additional allowed regions [a,b,l,p]	9	7.8%
Residues in generously allowed regions [~a,~b,~l,~p]	1	0.9%
Residues in disallowed regions	0	0.0%
-----		-----
Number of non-glycine and non-proline residues	116	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	13	
Number of proline residues	8	
-----		-----
Total number of residues	139	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

Ramachandran Plot

7030135

Residues in most favoured regions [A,B,L]	87	90.6%
Residues in additional allowed regions [a,b,l,p]	8	8.3%
Residues in generously allowed regions [~a,~b,~l,~p]	0	0.0%
Residues in disallowed regions	1	1.0%
<hr/>		
Number of non-glycine and non-proline residues	96	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	9	
Number of proline residues	6	
<hr/>		
Total number of residues	113	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

Ramachandran Plot

3954600

Plot statistics

Residues in most favoured regions [A,B,L]	89	92.7%
Residues in additional allowed regions [a,b,l,p]	5	5.2%
Residues in generously allowed regions [~a,~b,~l,~p]	1	1.0%
Residues in disallowed regions	1	1.0%
<hr/>		
Number of non-glycine and non-proline residues	96	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	9	
Number of proline residues	6	
<hr/>		
Total number of residues	113	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

Ramachandran Plot

saves

Plot statistics		
Residues in most favoured regions [A,B,L]	89	92.7%
Residues in additional allowed regions [a,b,l,p]	6	6.2%
Residues in generously allowed regions [~a,~b,~l,~p]	0	0.0%
Residues in disallowed regions	1	1.0%
<hr/>		
Number of non-glycine and non-proline residues	96	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	9	
Number of proline residues	6	
<hr/>		
Total number of residues	113	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

Ramachandran Plot

2903312

Plot statistics

Residues in most favoured regions [A,B,L]	105	90.5%
Residues in additional allowed regions [a,b,l,p]	10	8.6%
Residues in generously allowed regions [~a,~b,~l,~p]	1	0.9%
Residues in disallowed regions	0	0.0%
<hr/>		
Number of non-glycine and non-proline residues	116	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	13	
Number of proline residues	8	
<hr/>		
Total number of residues	139	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

Fig 7a Position of epitope in YP_009725297

Fig 7b Position of epitope in YP_009725301

Fig 7c Position of epitope in YP_009725304

Fig 7d Position of epitope in YP_009725305

Fig 7e Position of epitope in YP_009742616

Fig 7f Position of epitope in YP_009742617

Fig 7g Position of epitope in YP_009724390

Fig 7h Position of epitope in YP_009724390