Original Research Article **Bioinformatics**

International Journal of Pharma and Bio Sciences

ISSN 0975-6299

STRUCTURE PREDICTION OF THE PROTEIN CCR5 DELTA 32 AND STRUCTURE COMPARISON WITH CCR5

GOPALUNI SAI AKASH

Department of Biotechnology, Sreenidhi Institute of Science And Technology, Yamnampet, Ghatkesar, Telangana, India-501301

ABSTRACT

Human immunodeficiency virus (HIV) is an obligate intracellular parasite of the $CD4^+$ T-cells. This virus targets and kills helper T cells that play a vital role in immune response. Studies have shown that the interaction between the viral glycoprotein and host proteins like CD4 and Chemokine receptor type 5 (CCR5) enable the virus to get into the helper T cells. There exists a mutant form of the protein CCR5 called the CCR5 delta 32 or CCR5-D32 which does not interact with the viral glycoprotein in the same way as CCR5 does and hence, individuals carrying this mutation are resistant to HIV infection. Here I have predicted the structure of CCR5-D32 using Modeller9.16. Structure alignment and active site analysis revealed that, though both CCR5 and CCR5-D32 had a similar structure, their active site makeup was different and thus their functions can be different. It was also observed that the mutant lacked a few amino acids like Tyr 187, Gln 188, Lys 191, Gln 194 and Thr 195 in its active site which were lost due to the frame-shift mutation.

KEYWORDS: CCR5-D32, Homology Modelling, Energy Minimization, Molecular Dynamic Simulations, Structural alignment.

GOPALUNI SAI AKASH
Department of Biotechnology, Sreenidhi Institute of Science And Technology, Yamnampet, Ghatkesar, Telangana, India-501301

INTRODUCTION

Human immunodeficiency virus belongs to the Retroviridae family. The virus has a '+' sense ssRNA (RT) as its genetic material which is approximately 9Kb long.[1] The virus has two species HIV-1 and HIV-2 of which HIV-1 is more virulent and is spread throughout the planet. HIV-1 is enveloped by host glycoprotein and is also known to show the lysogenic mode of replication. This virus specifically infects $CD4^+$ T-cells (T helper cells), which play a major role in immune response.[2] As the HIV virus kills these cells, the diseased gets immune compromised and often dies of opportunistic infections.[3] To enter $CD4^+$ T-cells, he virus uses its glycoprotein GP120 to interact with the CD4 receptor and later with CCR5 to enter the host cell by fusing its envelop with the host cell membrane.The wild type CCR5 protein is 352 amino acids long and is a co-receptor for the entry of HIV into the $CD4^+$ T-cells. There exists a mutant form of the gene called the *CCR5-D32* which has a deletion of 32 nucleotides from the original gene. Hence the mutant form of CCR5 (CCR5-D32) protein has a sequence with deletion of 137 amino acids from CCR5 and has 215 amino acids.[4] As a result of this deletion, the mutant CCR5-D32 can be expected to have a significant change in its protein structure and function compared to CCR5. This change in the structure and function might be responsible for the resistance of mutant $CD4^+$ T-cells to HIV infection by not allowing the virus inside the host.[5] Individuals having two alleles of this mutation are resistant to HIV and do not develop AIDS. This mutation is less likely to be found in the African and Asian population but 20% of the Caucasian population is heterozygous and 1% of the Caucasian population is homozygous to CCR5-D32 mutation.[6] Here I used bioinformatics and computational biology to predict the in-silico structure of the protein CCR5-D32 with accuracy. The protein sequence of CCR5-D32 was retrieved from UniProt and PDB structure of CCR5 from RCSB.org. The crystal structure of CCR5 (obtained from PDB) was used as a template for homology modelling of CCR5-D32 on Modeller 9.16. The generated structure of CCR5-D32 was verified by loop modelling through SPDBV, energy minimization and molecular dynamic simulations using GROMACS 5.1.2. Then the structures of CCR5 and CCR5-D32 can be compared by PyMol and their active sites were analyses using CASTp.

MATERIALS AND METHODS

Sequence and Structure Retrieval

The sequence of the protein CCR5-D32 was retrieved from UniProt, which is a manually curated database for high quality protein sequences. The crystal structure of the protein CCR5 was retrieved from RCSB protein databank.

Tools and Materials

Homology modelling of the protein CCR5-D32 was performed using Modeller 9.16.[7] Validation of the CCR5-D32 was done using loop modelling function of SPDBV.[8] PDB-sum was used to generate the list of amino acids that lie in the disallowed region of the Ramachandran plot.[9] Energy minimization and molecular dynamic studies were carried out on the CCR5-D32 receptor using GROMACS 5.1.2.[10] Structural alignment of the two proteins was done using the 'align structure' function in PyMol. Active site analysis of the proteins was done using CASTp to find changes in the active site residues.[11]

Homology Modeling

The 3D structure of the protein CCR5-D32 was predicted using Modeller 9.16. This tool takes the alignment file and structure of the template as input and predicts the structure of the query protein. The predicted protein structure can be then validated using loop modelling, energy minimization, and molecular dynamic simulations.

Protein Structure validation

The predicted structure might have some of its residues in the outlier region of the Ramachandran plot, which can be found out by submitting the structure to PDB-sum. PROCHECK is a part of PDB-sum that generates a Ramachandran plot, which shows the residues that are in the outlier region. These residues in the outlier region can cause steric hindrances and hence have to be looped into any of the three allowed regions. This can be done using loop modeling function of SPDBV. The loop modelled structure can be submitted to GROMACS for energy minimization, and molecular dynamic simulations, which validates and stabilizes the energy of the predicted protein structure.

Structural alignment and active site comparison

The proteins were submitted to PyMol to structurally align the proteins and find the RMSD for which the 'align structure' function can be used. The two structures can then be submitted to CASTp which finds the active sites of the protein and highlights the residues involved, area and volume of the active site pockets.

RESULTS AND DISCUSSION

Sequence and structure Retrieval

The CCR5-D32 sequence was retrieved from UniProt. The X-ray crystal structure (at 2.71 Å resolution) of the protein CCR5 was retrieved from RCSB Protein DataBank which had a PDB ID 4MBS (Figure 1).[12]

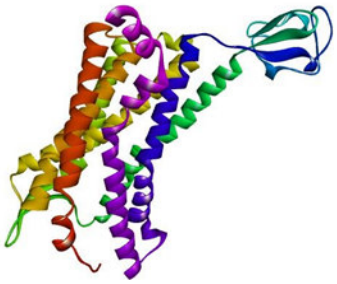


Figure 1
Crystal structure of the protein CCR5 having the PDB ID 4MBS

Homology Modeling

The sequences of CCR5-D32 and CCR5 were aligned and the crystal structure of the CCR5 receptor was taken as the template for Homology Modelling using Modeller 9.16. Hundred models of the protein CCR5-D32 were created and the model with the best stability energy was chosen for further experimentation (Figure 2). The stability energy of the generated models ranged from 3400.34937 to 2985.46094 Kcal/mol and the stability energy of the chosen model was 2985.46094 Kcal/mol. As the stability energy of the predicted structure was too high, this cannot be considered a stable structure unless the energy is minimized. Hence, energy minimization and molecular dynamic simulations were used to validate the protein structure.

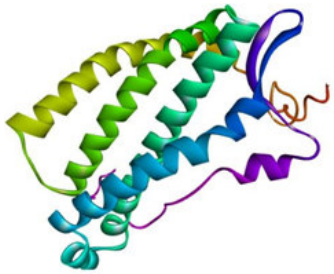


Figure 2
Predicted structure of the protein CCR5-D32 predicted by Modeller 9.16

Protein Structure validation

The chosen CCR5-D32 model was then submitted to PDB-sum. PROCHECK analysis showed that only two of the 215 amino acids were in the disallowed regions of the Ramachandran plot. The amino acids in the outlier region were Cys 20 and Leu 204. These amino acids were then looped into the allowed regions using loop modelling function of SPDBV but, other residues like glycine and proline in the outlier region were neglected as they don't contribute towards steric clashes and can be let in the disallowed region. The protein was then re-submitted to PDB-sum to verify loop modelling. The PROCHECK analysis showed that all the amino acids lied in the allowed regions (Figure 1 and Table 1). This shows that the protein is free of steric hindrances and can be used for further experimentation.

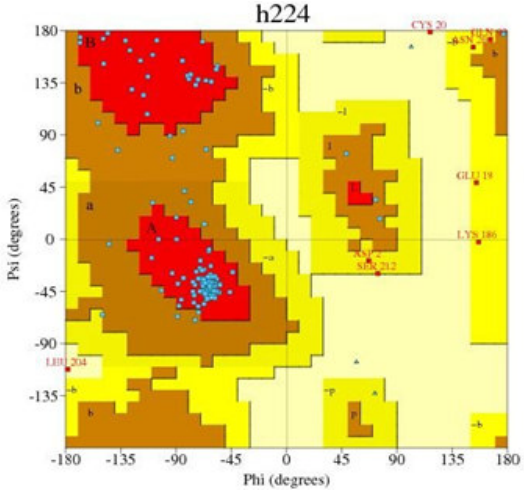


Figure 3A
Ramachandran plot generated by PDB-sum, Before loop modelling using SPDBV

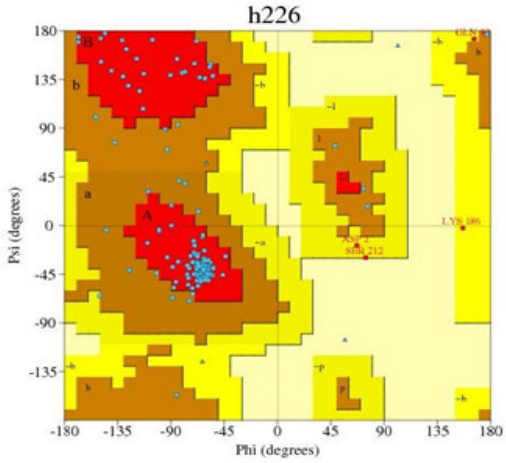
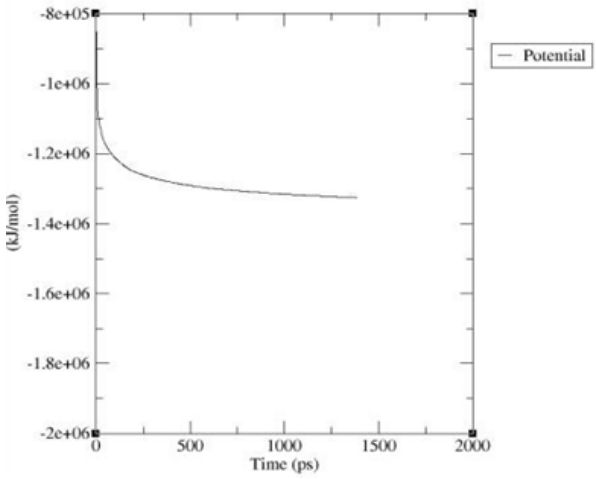


Figure 3B
Ramachandran plot generated by PDB-sum, after loop modelling using SPDBV

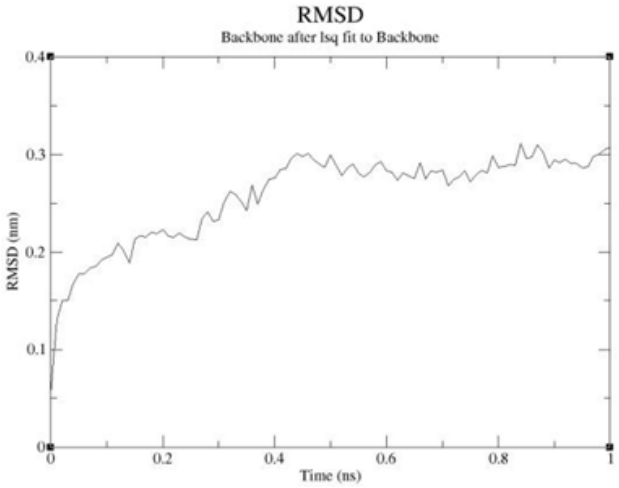
Table 1
statistics of the amino acid residues generated by PDB-sum before (A) and after (B) loop modelling using SPDBV

Regions (Colour code)	No. of residues		% of residues	
	A	B	A	B
Most favoured Region (red)	162	165	84.8%	86.4%
Additionally allowed regions (brown)	21	22	11%	11.5%
Generously allowed regions (yellow)	6	4	3.1%	2.1%
Disallowed regions (white)	2	0	1%	0%
Total non-glycine and non-proline residues	191	191	100%	100%
Glycine and proline residues	22	22	-	-
End residues	2	2	-	-
Total residues	215	215	-	-

Then the protein was submitted to GROMACS to perform Energy Minimization and Molecular Dynamic Simulations. The process of energy minimization decreased the energy of the protein and Molecular Dynamic Simulations validated the structure of the protein. After the energy minimization operation, the stability energy of the protein reduced to '-1.3e+06 Kcal/mol'. As the stability energy of the protein has reached a constant value (from the graph 1), it is assumes that the protein has reached its minimum energy. Hence the structure was used for further experimentation. The energy was brought down to '-1.3e+06 Kcal/mol' in the simulation time span of 1472 picoseconds (ps) (Graph 1). The protein was then subjected to molecular dynamic simulations for a span of 1 nanosecond (ns). The simulation produced a protein backbone RMSD graph. The RMSD of the protein globular structure fluctuated between 0.05 and 0.3 nm in the span of 1 ns (Graph 2). The low RMSD value stated that protein has a proper fold, as the dynamic structures of the protein were very similar to each other.



Graph 1
T*he energy minimization graph (potential against time) produced by GROMACS*.



Graph 2
The RMSD graph (RMSD against time) produced by GROMACS

Structural alignment and active site comparison
The backbone atoms of the proteins CCR5 and CCR5-D32 were then aligned using the align molecules function in PyMol (Figure 6). The RMSD was found out to be 2.202 Å. This shows that the structure of the proteins is significantly different because of the frame-shift in the protein residues. This RMSD shows that the two proteins have a significant structural and functional difference. This functional dissimilarity might be the reason because of which, 1% of the population having the homozygous CCR5-D32 mutation are resistant to HIV infection.

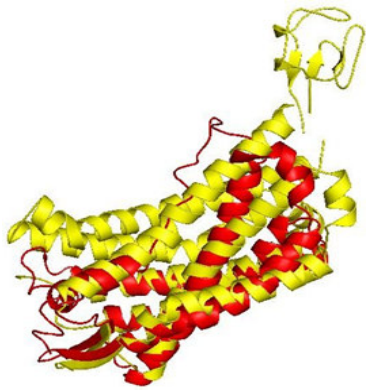


Figure 6
Structure of CCR5 (coloured yellow) aligned to the structure of CCR5-D32 (Coloured red)

To analyze the impact of the deleted 137 amino acids, I looked into the sequence shift and active site analysis of the CCR5-D32 and compared with that of CCR5. The active site analysis of the proteins using CASTp revealed that in CCR5 an active site had residues that were not deleted in CCR5-D32. When the same active site in CCR5-D32 was analyzed using CASTp, it revealed that the amino acid constitution of the active site significantly (Figure 7) due to the deletion of 137 amino acids. This also had an effect on the space filling of the active site. The active site of the CCR5 receptor has a hollow cavity, but the same site in CCR5-D32 has no such cavity. This cavity in the CCR5 has 69 amino acids, but the active site of CCR5-D32 has only 33 amino acids (the detailed list of amino acids is given in the figures 7A and 7B). The active site residues in CCR5 that are absent in CCR5-D32, might have a key role to play in the attachment of the virus to the protein followed by the viral infection. All these identified changes in CCR5-D32 compared to that of CCR5 might be responsible for its different function and hence be the reason behind the resistance of individuals possessing CCR5-D32 mutant against HIV infections.

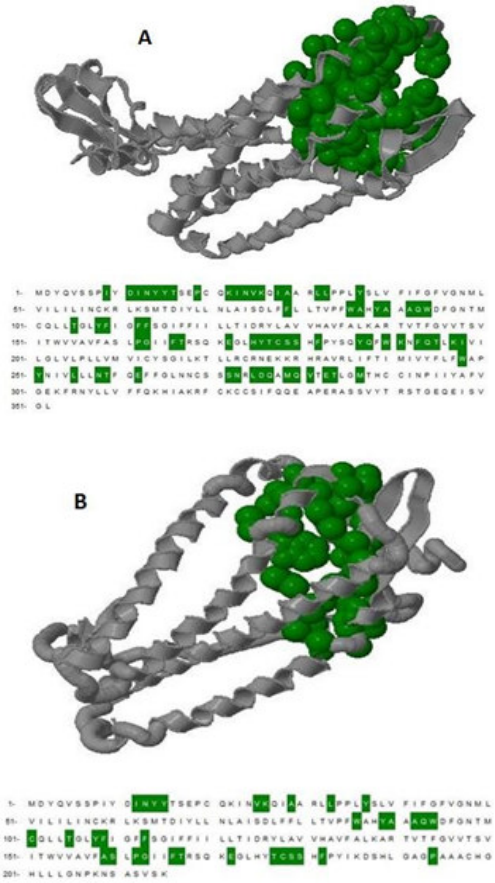


Figure 7
(***A) Active site of the protein CCR5, as predicted by CASTp.***
(B) Active site of the protein CCR5-D32 as predicted by CASTp.

The results show that the protein CCR5-D32 is structurally different from CCR5. This is because of the frame-shift mutation that caused changes in the active site of the protein, in terms of residues and volume. But, it is found that, people having this mutation lead a normal life with no medical implications.[13] It is a puzzling fact that, though having a significant change in its structure, CCR5-D32 does not cause any medical implications in the individuals having a homozygous allele. This means that some of its active sites are not affected by the mutation. Future Prospects of the project would be to find the other possible active site of the protein CCR5-D32 and to find other functions of the protein using dry lab studies. This mutation can be taken advantage of as a potential therapy for HIV infections. In one such study a zinc finger nuclease was used to disrupt the gene (*CCR5*) in the same way as the natural mutation does, in $CD4^+$T cells. This will make the $CD4^+$T cells resistant to HIV infections and can also help the individual fight the infection.[14]

CONCLUSION

The aim of my study was to find out the structural and functional differences in the proteins CCR5 and CCR5-D32, using bioinformatics and computational biology. The results of the study indicate that the two proteins are significantly different as their RMSD after structural superimposition was fund out to be 2.202 Å and their active site makeup, volume and surface area is also considerably because of the frame-shift frame shift mutation. Hence, it can be concluded that the replacement of these amino acids in the active site residues in CCR5-D32 might be the key reason for the difference in the function of the two proteins. But, further wet lab studies have to be performed to verify the findings.

ACKNOWLEDGEMENT

I thank the faculty of Dept. Biotechnology, SNIST for their support and encouragement.

CONFLICT OF INTEREST

Conflict of interest declared none.

REFERENCES

1. Wain-Hobson S, Sonigo P, Danos O, Cole S, Alizon M. Nucleotide sequence of the AIDS virus, LAV. Cell. 1985 Jan ;40(1):9-17.
2. Danial C. Douek, Jason M. Brenchley, Michael R. Betts, David R. Ambrozak, Brenna J. Hill,Yukari Okamoto, et al. HIV preferentially infects HIV-specific $CD4^+$ T cells. Nature. 2002 May; 417(6884): 95-8.
3. Iribarren JA, Rubio R, Aguirrebengoa K, Arribas JR, Baraia-Etxaburu J, Gutiérrez F, et al. Executive summary: Prevention and treatment of opportunistic infections and other coinfections in HIV-infected patients. Enferm Infecc Microbiol Clin. 2016 Apr; S0213(16):30010-6.
4. Hütter G, Bodor J, Ledger S, Boyd M, Millington M, Tsie M, et al. CCR5 Targeted Cell Therapy for HIV and Prevention of Viral Escape. Viruses. 2015 Jul; 7(8):4186-203.
5. Wu 1, Deng R, Wu W. Study on CCR5 analogs and affinity peptides. Protein Eng Des Sel. 2012 Mar;25(3):97-105.
6. Lucotte G. Frequencies of 32 base pair deletion of the (Delta 32) allele of the CCR5 HIV-1 Co-receptor gene in Caucasians: a comparative analysis. Infect Genet Evol. 2002 May; 1(3):201-5
7. Webb B, Sali A.Comparative Protein Structure Modeling Using MODELLER. Curr Protoc Bioinformatics. 2014 Sep ;47(5):1-32.
8. Johansson MU, Zoete V, Michielin O, Guex N. Defining and searching for structural motifs using deep views / swiss-PDB viewer. BMC Bioinformatics, 2012 Jul; 23 (13): 173-9.
9. de Beer TA, Berka K, Thornton JM, Laskowski RA. PDBsum additions. Nucleic Acids Res. 2014 Jan;42(Database issue):D292-6.
10. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. GROMACS: fast, flexible, and free. J Comput Chem. 2005 Dec;26(16):1701-18.
11. Joe Dundas, Zheng Ouyang, Jeffery Tseng, Andrew Binkowski, Yaron Turpaz, and Jie Liang. CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. Nucleic Acids Res. 2006 Jul 1; 34(Web Server issue): W116–W118.
12. Tan Q, Zhu Y, Li J, Chen Z, Han GW, Kufareva I, Structure of the CCR5 chemokine receptor-HIV entry inhibitor maraviroc complex. Sci. 2013 Sep ;341(6152):1387-90.
13. Enrico M T, Mario T, Katleen de GD, Enrica T, Roberto C, Christina B, et al. Partial protective effect of CCR5-Delta 32 heterozygosity in a cohort of heterosexual Italian HIV-1 exposed uninfected individuals. AIDS Res Ther. 2006 Sep; 3: 22.
14. Cohen J. Building an HIV-proof immune system. Sci. 2007 Aug 3;317 (5838):612-4.