International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-2 Issue-3, February 2013

Signature Patterns for Major Virulent Proteins of HIV1

Shashank Mittal, Reshu Nautiyal, Swati Mamgain, Kumud Pant,Tribhuvan Chandra

Abstract- Realizing the importance of understanding major virulent proteins of HIV we are attempting to unravel various amino acid signatures that exist for two major proteins in HIV i.e. Gag and Env. The results have been obtained through freely available software VESPA available at HIV database project funded by the Division of AIDS of the National Institute of Allergy and Infectious Diseases (NIAID), a part of the National Institutes of Health (NIH). The results obtained help us to understand the sites in proteins which can be hot spot of mutations or sites where propensity of finding different amino acid exists. This is a novel work where signature patterns for various HIV proteins have been deduced.

Keywords- HIV, AIDS, (NIAID), (NIH).

I. INTRODUCTION

HIV1, one of the members of Lentivirus genus is a retrovirus that encodes following genes from its approximately 9749 nucleotide long genome. Gag (coding for the viral capsid proteins), Pol (coding for reverse transcriptase), Env (coding for HIV's envelope-associated proteins), and the regulatory genes: tat, rev, nef, vif, vpr and vpu. Understanding the functioning of these genes has been a major challenge of all the major laboratories of the world. Analysis and identification of signature pattern can help in revealing specific sites in alignments of different amino acid or nucleic acid sequences that can represent query set of sequences in comparison with background set. The signature patterns also help us to understand the polymorphisms that exist between various viral subtypes. Once signature pattern has been obtained it can be used to examine sequences from other organisms and in turn find out genetic relatedness that exists between them. Signature patterns can also help to understand the amino acids or nucleotides which are most representative of the group and also to understand the origin of disease or transmission of disease from infected individual as has been done in case of HIV transmission in a dental practice (1).

Manuscript published on 30 February 2013.
***** Correspondence Author (s)
Shashank Mittal, Department of Biotechnology, Graphic Era University, Dehradun, India.
Reshu Nautiyal, Department of Biotechnology, Graphic Era University, Dehradun, India.
Swati Mamgain, Department of Biotechnology, Graphic Era University, Dehradun, India.
Kumud Pant, Department of Biotechnology, Graphic Era University, Dehradun, India.
Tribhuvan Chandra, Department of Biotechnology, Graphic Era University, Dehradun, India.

Various epidemiological markers have been generated through molecular biological techniques like oligonucleotide fingerprinting of RNA genomes with ribonuclease, mapping of DNA genomes with restriction endonucleases, and genomic sequencing, and have been used to study viral transmissions from person to person, within communities, and between countries. For the above studies the genetic variation between viruses can help to differentiate between strains of viruses (1, 2). One of the biggest problems faced in vaccine design is high mutation rate of virus because of which viruses from different individuals are found to be distinct (3, 4). The same property can be exploited to generate signature patterns for each of the strain obtained experimentally from different organisms. For our current studies we have taken major virulent proteins of HIV available in UniProt database and have used a freely available epidemiological tool VESPA available at http://www.hiv.lanl.gov.

II. MATERIALS AND METHOD

1. The dataset- It comprised of 68 proteins belonging to Env or envelop proteins, 38 proteins for Gag, 62 for Nef, 55 for rev, 54 for tat, 60 for vif and vpr. All the sequences were carefully chosen so as to remove any fragments or sequences belonging to any other data set. The dataset was obtained from uniprot databank of Expasy (7).
2. Vespa- Vespa is an on-line tool maintained and developed by Korbes and Meyers at www.hiv.lanl.gov. This program can be used to quickly detect amino acids or nucleotides which characterize differences between two groups of sequences. It compares two groups of sequences and looks for a "signature" pattern or the set of amino acids or nucleotide that is conserved among each set, but differing between the sets. It will pick out those distinguishing characters, and calculate their frequencies in each set. In this it gives signature pattern to all our query sequences (5, 6).

III. RESULTS AND DISCUSSION

1. Envelop protein- The 54 proteins belonging to envelop category were divided into two groups of 27 each one as query and one as background. Beginning 7 amino acids were found to be same and at 8th position lysine (K) and Methionine (M) was found. Similarly at 9th position Glutamic acid (E) and Glycine (G) was found in query and background sequences respectively. Rest of the positions with variable amino acids is depicted below in the results. With these results following signature patterns were deduced for **env** proteins. Out of the 926 nucleotide long env protein 87 amino acids were found to be significantly differing in these two sets of Env proteins

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

Env Query Signature

K	E	Y	R	T	K	D	L	K	D	0.630	0.556
Pos 8	Pos 9	Pos 14	Pos 19	Pos 43	Pos 45	Pos 145	Pos 146	Pos 147	Pos 149		
T N	S	S	G	R	M	I	I	S			
0.556	0.519	0.519	0.481	0.333	0.222	0.296	0.259	0.630	0.556		
Pos 150	Pos 151	Pos 162	Pos 163	Pos 164	Pos 165	Pos 166	Pos 167	Pos 173	Pos 183		
V	Q	E	K	I	N	D	S	S	N		
0.481	0.630	0.741	0.704	0.556	0.778	0.370	0.407	0.519	0.593		
Pos 188	Pos 189	Pos 191	Pos 197	Pos 200	Pos 205	Pos 219	Pos 222	Pos 227	Pos 262		
T	R	V	R	R	T	G	K	K	I		
0.630	0.630	0.556	0.407	0.741	0.704	0.370	0.519	0.481	0.630		
Pos 264	Pos 284	Pos 303	Pos 342	Pos 349	Pos 353	Pos 359	Pos 379	Pos 385	Pos 387		
S	Q	G	V	T	Q	F	S	T	G		
0.333	0.593	0.370	0.593	0.593	0.556	0.259	0.296	0.333	0.333		
Pos 389	Pos 394	Pos 396	Pos 418	Pos 434	Pos 435	Pos	Pos 442	Pos 444	Pos 445		
S	N	T	E	G	D	I	E	S	R		
0.444	0.481	0.667	0.481	0.556	0.519	0.778	0.481	0.296	0.519		
Pos 454	Pos 456	Pos 457	Pos 458	Pos 459	Pos 461	Pos 473	Pos 478	Pos 489	Pos 493		
E	S	I	D	Q	I	L	M	A	A		
0.370	0.333	0.556	0.889	0.704	0.556	0.556	0.407	0.593	0.481		
Pos 518	Pos 519	Pos 521	Pos 528	Pos 566	Pos 576	Pos 579	Pos 596	Pos 678	Pos 683		
E	N	E	N	S	N	T	R	S	V		
0.556	0.556	0.778	0.333	0.444	0.370	0.222	0.630	0.778	0.593		
Pos 692	Pos 695	Pos 701	Pos 707	Pos 711	Pos 745	Pos 794	Pos 813	Pos884	Pos 892		
T W	S V	A	V Y								
0.593	0.556	0.667	0.593	0.630	0.519	0.333					
Pos 852	Pos 876	Pos 884	Pos 892	Pos 897	Pos 913	Pos 917					

Env Background Signature

M	G	C	K	A	N	N	I	N	N
0.296	0.593	0.407	0.556	0.481	0.444	0.481	0.185	0.370	0.407
S	T	-	-	N	T	T	T	M	V
0.185	0.222	0.407	0.333	0.259	0.185	0.111	0.222	0.481	0.259
K	K V	R	V	G	N	E	N	D	
0.370	0.333	0.519	0.481	0.519	0.333	0.444	0.259	0.815	0.778
K	K	I H	Q	A	E	D	Q	V	
0.407	0.556	0.519	0.296	0.519	0.556	0.296	0.222	0.222	0.667
T	H	P	T S	G	N	T	-	-	
0.333	0.407	0.222	0.593	0.704	0.407	0.481	0.370	0.333	0.333
N	T	N	S	T	G	V	G	A	Q
0.259	0.259	0.296	0.185	0.407	0.296	0.630	0.370	0.296	0.259
N	T	T	N	E	L M	I	T	S	
0.222	0.481	0.556	0.519	0.444	0.481	0.370	0.556	0.630	0.407
N	D	Q	D	G	D S	Q	N	A	
0.370	0.444	0.481	0.481	0.370	0.444	0.259	0.519	0.556	0.593
A	L	G	I	T	A	G			
0.593	0.667	0.444	0.444	0.630	0.370	0.370			

The number below the amino acid symbols represents the frequency of it in the group comprising query or background.

2. Gag- For 38 proteins of Gag 19 were included in background and 19 in query. Of the 545 amino acid long gag protein only 30 amino acids formed a signature pattern at various positions.

Query Signature

R	K	I	V	S	R	R	K	A	
0.474	0.526	0.632	0.526	0.737	0.684	0.526	0.579	0.579	
Pos 15	Pos 30	Pos 34	Pos 46	Pos 54	Pos 58	Pos 76	Pos 114	Pos 116	
T	V	A	N	S	E	N	T	R	
	0.579	0.684	0.684	0.526	0.526	0.737	0.632	0.579	0.526
Pos 129	Pos 169	Pos 234	Pos 263	Pos 321	Pos 323	Pos 383	Pos 391	Pos 400	
N	K	R	Y	E	R	T	T	P	
0.474	0.684	0.789	0.368	0.579	0.684	0.579	0.579	0.632	
Pos 401	Pos 404	Pos 419	Pos 459	Pos 492	Pos 496	Pos 501	Pos 503	Pos 517	
I	T	R							
0.474	0.684	0.579							
Pos 518	Pos 532	Pos 535							

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Background Signatures

A	R	L	L	A	Q	K	Q	T	K
0.368	0.632	0.684	0.579	0.474	0.368	0.526	0.421	0.368	0.421
I	P	G	T	D	-	A	K	G	R
0.526	0.526	0.421	0.526	0.526	0.368	0.474	0.579	0.579	0.474
K	H	A	G	I	-	Q	K	A	K
0.579	0.368	0.632	0.737	0.526	0.789	0.474	0.526	0.421	0.632

The similar amino acids and their positions highlight the fact they are essentially conserved in these proteins. One of the properties of HIV is that they constantly change their amino acid composition. By careful analysis of the above data we can say that new proteins are generally formed by changing amino acids at the signature region. Moreover the essentially conserved amino acids at specific positions can be hot drug targets since their presence is required for proper functioning of the enzyme.

Since both Env and Gag proteins are very important in pathology of virus therefore our study has been done for the above two proteins. Because the mutation rate of Env proteins is tremendous therefore understanding frequency of occurrence of amino acid and change can aid research and drug design.

IV.CONCLUSION

The epidemiological markers like above can be deciphered for other significant proteins of other viruses and can help to understand the positions where mutations are most likely to occur. They also provide us with a means to understand the relationship that exist between a group of proteins with other and can help in genome comparison studies.

REFERENCES

1. Molecular Epidemiology of HIV Transmission in a Dental Practice, Chin-Yih Ou et.al. SCIENCE * VOL. 256 * 22 MAY 1992.
2. J. S. Smith etal., N. Engi. J. Med. 324, 205 (1991); H. J. Un etal., J. Infect. Dis. 164, 284 (1991); G. P. Holmes et al., Ann. Intem. Med. 112, 833 (1990); R. Rico-Hesse, M. A. Pallansch, B. K. Nottay, 0. M. Kew, Virology 160, 311 (1987).
3. P. Balfe et al., J. Wrol. 64, 6221 (1990).
4. G. Myers et al., Eds., "Human retroviruses and AIDS 1991: A compilation and analysis of nucleic acid and amino acid sequences," (Theoretical Division, T10, Los Alamos National Laboratory, Los Alamos, NM, 1991).
5. Ou C-Y, Cielsielski CA, Myers G, Bandea CI, Luo C-C, Korber BTM, Mullins JI, Shochetman G, Berkelman RL, Nikki Economou AN, Witte J, Furman LJ, Satten GA, MacInnes KA, Curran JW, and Jaffe HW: Molecular epidemiology of HIV transmission in a dental practice. Science, 1992 May; 256(5060):1165-71.
6. B. Korber and G. Myers: Signature Pattern Analysis: A Method for Assessing Viral Sequence Relatedness. AIDS Research and Hum. Retroviruses, 1992 Sep; 8(9):1549-60.
7. The UniProt Consortium, Reorganizing the protein space at the Universal Protein Resource (UniProt) Nucleic Acids Res. 40: D71-D75 (2012).

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)