Computational comparison of antigenic sites of C2-V3-C3 regions of gp-120 of the Iranian HIV-1 with the homologous regions of different subtypes of the virus

H. Mohabatkar: Shiraz University

Abstract

In this study the predicted epitopes of C2-V3-C3 domains of gp120 of HIV-1, present in Iran, were compared to the epitopes of the homologous domains in subtypes A, B, C, D, E, F, G, H and I of this virus. Since epitopes are regions between the sequences with secondary structure which are hydrophilic and accessible, these parameters were used to predict the epitopes. The number of predicted helix (one) and sheet (five) regions in the Iranian isolate was equal to these numbers in subtypes A and F. Hubbard method recognized seven potential glycosylation sites on the Iranian isolate. In all other HIV-1 viruses, the number of putative glycosylation sites was less. In all subtypes, including the Iranian one, an epitope in the same region was predicted. In all analyzed sequences (excluding the Iranian one and subtypes D and H) a single long epitope was predicted in another region. In subtype H no epitope was predicted in that region. Similar to subtype D in the Iranian subtype B, two short epitopes were predicted in the same part. The computational analysis predicted similarities and dissimilarities between the locations of epitopes of the Iranian and other HIV-1 viruses. Although the primary structure of gp120 of the Iranian HIV-1 is highly related to subtype B, some differences were even predicted between the secondary and tertiary structure of the Iranian and consensus subtype B.

Introduction

More than two decades after the first report of AIDS, the pandemic continues to spread and the number of people who are suffering from AIDS is increasing in Asia [1].

Key words: HIV-1, Secondary structure, Computational prediction, gp120, Genomic diversity

One of the most important properties of HIV is its variability [2],[3]. Variation in the HIV-1 envelope nucleotide sequence has been used to classify HIV-1 isolates into different subtypes [4-7]. Within a subtype, nucleotide divergence in the envelope gene is <15%, whereas intersubtype variation ranges from 20% to 30% with no distinction of sample time [8]. The variability of HIV plays a major role in the natural history of HIV infection and AIDS [9]. Genomic variations of HIV seem to be manifested mainly in the envelope glycoproteins [9],[10].

Comparison of the gp120 sequences of different HIV-1 viruses keveals five variable regions (V1-V5) and five conserved regions (C1-C5) on this glycoprotein [11]. Sugars comprise approximately 50% of the molecular weight of gp120 the major glycoprotein of envelope of HIV-1. The presence of glycans seems to be necessary for the functional disposition of this glycoprotein and for protecting the polypeptide from denaturation [12].

Knowledge of the secondary structure of gp120 is very useful for predicting its antigenicity [13], binding to CD4 [11] and interaction with gp41 [14]. By the time of this study, just the sequence of C2-V3-C3 domain of gp120 of the Iranian isolate of HIV-1 has been identified [15]. In the last few decades, highly predictive patterns have been occurred, suggesting that local primary amino acid sequences can predict the secondary structure of a protein [16]. It should be mentioned here that predictions cannot be taken as definitive conclusions but they give information helping to interpret results or to design new experiments [17], [18].

It is thought that the entire surface of a protein can be considered as an antigen. Antigenic sites (epitopes) are defined as those residues of a native protein that are bound by antibodies raised to a native protein, native protein fragments, or synthetic peptides. By definition, since antigenic sites are those parts recognized by antibodies, it is most likely that these sites are accessible or on the surface of a protein and these regions are probably more mobile than interior regions. Since these sites are on the surface, they are

probably hydrophilic [19]. In addition, as epitopes are often loops between the regular secondary structures, a guide to their location would be to predict from the sequence which parts of the chain are not α-helix or β-sheets [20]. In the present work, the numbers of N-glycosylation sites and secondary structure in combination with the other values for predicting the epitopes of C2-V3-C3 regions of gp120 of the Iranian isolate of HIV-1 have been compared to the homologous regions of different subtypes of the virus by computational analysis.

Methods

Gp120 amino acid sequences:

To study the epitopes of C2-V3-C3 regions of gp120 of the Iranian HIV-1, the amino acid sequence of this protein from the Swissprot databank was fetched. The sequence of this glycoprotein (accession No. AAK21312) was compared to consensus gp120 amino acid sequences of nine different subtypes of the virus [7].

Theoretical analysis:

All prediction calculations were based on propensity scales for each of the 20 amino acids. Nine scales of inverted hydrophobicity and two scales of hydrophilicity which were mostly derived from the study of partition coefficient of amino acids in two non-interacting isotropic phases were taken into consideration. Four scales of accessibility, which were constructed by measuring the accessible surface of all the residences in a number of proteins, were considered. Three scales of secondary structure which were based on the prediction of turns and troops obtained from statistical analysis of proteins of known structure were considered for secondary structure prediction. The 19 scales were grouped as follows [13]: Nine scales of inverted hydrophobicity (Doolittle, Heijne, Manavala, Prils, Rose, Sweet, Totls, Ges, Zimmermann); two scales of hydrophilicity (Hopp, Parker); Four scales of accessibility (Janin, Chothia, Chothia 8, Acrophil); one scale of antigenicity (Welling) and three scales of secondary structure (Chouf 3, Garnier

3, Levitt). Applied programs and the ways to use the programs are available at **http://au.expasy.org**.

The amino acid sequences of the glycoproteins were read as a moving window of seven residues and their values corresponding to each of the 19 scales taken here into consideration and the mean was plotted against the fourth residue of the window. In order to compare the profiles obtained by different methods, various scales were normalized where the original values of each scale were set between +3 and -3.

N-glycosylation sites are searched as Asn-X-Thr or Asn-X-Ser sequences, where X is any residue [21],[22],[23].

Results and discussion

The results of the computer-assisted comparison of the number of N-glycosylation sites and secondary structure of C2-V3-C3 domains of gp120 glycoproteins are shown in Table 1.

The result of this analysis for the Iranian isolate indicates that one region predicted to be α-helix and five regions predicted to be β-sheet. According to this analysis 14.4%, 25.8% and 30.9% of this protein was in the α-helix, β-sheet and turn forms respectively. The same analysis was performed to study the homologous region of gp120 of nine different subtypes of HIV-1. Although the HIV-1 isolate reported from Iran belongs to subtype B of the virus, there was a 29.9% amino acid sequence difference between the Iranian HIV-1 and the other subtype B isolates of the virus (29 out of 97 amino acids). One of the differences between these proteins is in the residue number 17. In this location, asparagine has replaced lysine in the Iranian isolate. Since the sequence of this part of gp120 in all other HIV-1 viruses is Lys-Glu-Ser, this change makes an Asn-Glu-Ser sequence. As it has already been mentioned in the methods section, this sequence is a potential glycosylation site. All other glycosylation sites remain the same in the Iranian isolate (glycosylation of residues number 4, 23, 29, 61, 82, and 88).

The core of proteins usually contains a combination of helices and sheets, which are hydrophobic. Since the core is mostly devoid of water molecules, formation of intramolecular hydrogen bonds is favored. In contrast, turns are situated on the surface of the protein in contact with solvent atoms. Thus, turns are accessible and hydrophilic [13]. Similar to subtypes D and I in the Iranian isolate, the percentage of turns is more than 30% in the analyzed sequence. The number of α-helices and β-sheets in the Iranian isolate is equal to the number of these structures in subtypes A and F. Similarities and dissimilarities shown in table1 can have influence, on the overall structure of gp120 of the Iranian isolate.

Table1: Prediction of secondary structure of C2-V3-C3 regions of gp120 HIV-1

Subtype	No. of helices	No. of sheets	Percentage of helices	Percentage of sheets	Percentage of turns	No. of N-glycosylation sites
A	1	5	21.6	25.8	24.7	5
B	2	4	22.7	17.5	23.7	6
C	2	5	22.7	23.7	28.9	5
D	1	4	8.2	30.9	34.0	5
E	2	4	23.7	23.7	21.6	4
F	1	5	16.5	27.8	26.8	5
G	1	3	13.4	27.8	24.7	5
H	1	6	10.3	35.1	21.6	5
I	1	4	10.0	31.1	35.6	6
Iranian B	1	5	14.4	25.8	30.9	7

The overall results of prediction of antigenic domains using different structural parameters have been shown in figure 1. Despite the differences some similarities can be seen between different subtypes. The figure indicates that in all subtypes including the Iranian one, one epitope in an identical region (amino acids number 27 to 33) has

been predicted. In all analyzed sequences (excluding the Iranian one and subtypes D and H) a single long epitope was predicted in another region (amino acids number 57 to 65). In subtype H no epitope was predicted in this region. In this aspect there was a similarity between the Iranian subtype B and subtype D because in both of them two short eitopes were predicted in the same location.

It has already been shown that in gp120, carbohydrates have a powerful effect on the antigenicity of this protein [24]. The Iranian isolate contains the maximum number of potential glycosylation sites (7sites) and this feature of the protein may affect its pathogenicity.

Investigations have demonstrated that antigenic determinants are surface features of proteins and indicate that they are frequently found on regions of a molecule that have an unusually high degree of exposure to solvent (regions which project into the medium). This together with the fact that charged, hydrophilic amino acid side chains are common features of antigenic determinants, led scientists to investigate the possibility that antigenic determinants might be associated with stretches of amino acid sequence, that contain a large number of charged and polar residues and are lacking large hydrophobic residues [25]. For predicting epitopes, different parameters should be considered because no single parameter contains enough information to allow a transmission from primary structure data to a tertiary structure entirely [13].

The viral glycoproteins displayed on the surface of enveloped viruses are responsible for interacting with the receptor on the target cell and are also exposed to humoral selection by circulating antibodies [26]. Genomic diversity of the surface glycoprotein among independent HIV-1 isolates is a well-characterized feature of this virus. Although this sequence heterogeneity is distributed throughout the genome, most of the heterogeneity is located on the *env* gene [27].

Completely perfect predictions with accuracy of 100% are not achievable, since proteins with essentially identical 3-D fold differ in secondary structure [28]. The goal

of structure prediction can therefore be redefined to reach 85% three-stage per-residue accuracy [29]. Like any other protein prediction of structure of gp120 can provide us important information about the interactions and functions of this protein. For example, prediction of a β-strand contact map may provide information on long range interactions within gp120 [28]. For specific identification, corresponding synthetic peptides of the consensus sequence of the gp120 of the Iranian HIV-1 can be tested by ELISA. Candidates for peptide synthesis are selected from the protein sequence on the basis of prediction algorithms derived from the observed correlation between the location of continuous epitopes and structural parameters such as the secondary structure, hydrofilicity, accessibility, and mobility [30]. Understanding the similarities and the differences between the antigenic determinants of the Iranian HIV-1 and other subtypes of the virus may help us in more specific identification of the virus [31].

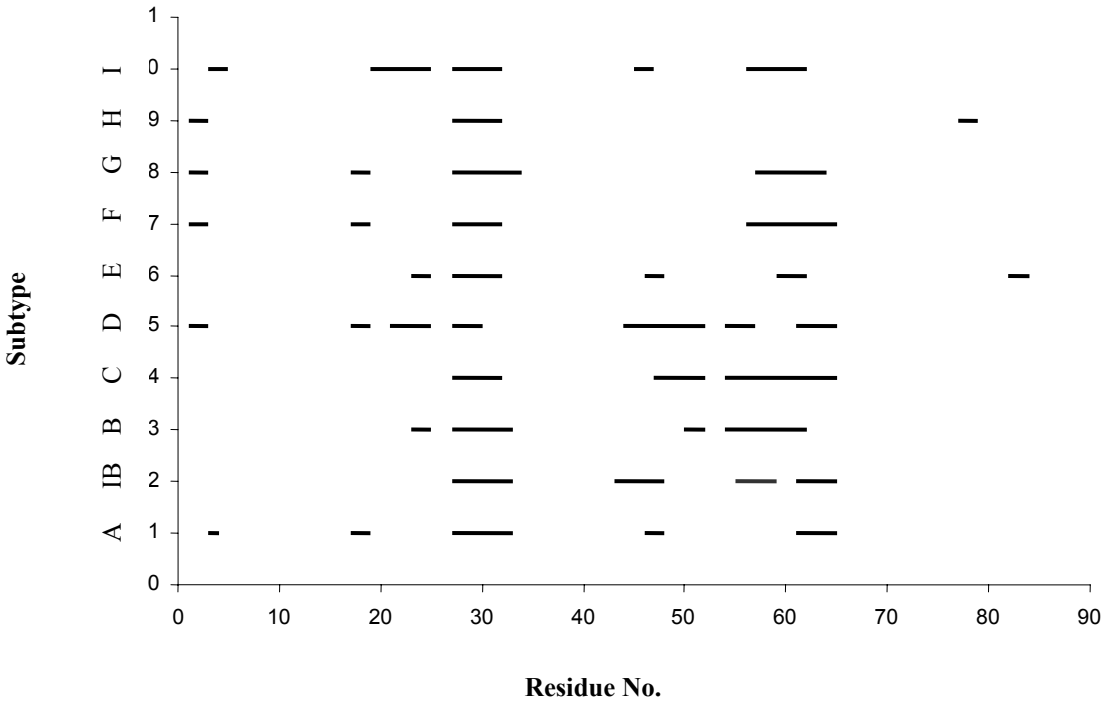

Figure1: Predicted antigenic determinants of C2-V3-C3 regions of different subtypes of HIV-1 (IB stands for the Iranian Subtype B)

Acknowledgement

Here by we acknowledg the support of this study by Shiraz University (project number 82-SC-1654-C253).

References

1. L. Garbus, The HIV/AIDS drug pipeline. A status report. BETA, 15 (2002) 29-40.
2. A.M. Klevytska, M.R. Mracna, L. Guay, G. Becker-Pergola, M. Furtado, L. Zhang, J.B. Jackson, and S.H. Eshleman, Analysis of length variation in the V1-V2 region of env in nonsubtype B HIV type 1 from Uganda. AIDS Res. Hum. Retroviruses, 18 (2002) 791-796.
3. H. Tian, C. Lan, and Y.H. Chen, Sequence variation and consensus sequence of V3 loop on HIV-1 gp120. Immun. Lett., 83 (2002) 231-233.
4. R. Downing, D. Pieniazek, D.J. Hu, B. Biryahwaho, C. Fridlund, M.A. Rayfield, S.D.K Sempala, and R.B. Lal, Genetic characterization and phylogenetic analysis of HIV-1 sub type C from Uganda. AIDS Res. Hum. Retroviruses, 16 (2000) 815-819.
5. S.G. Tisminetzky, E.A. Scodeller, P. Evangelisti, Y. Chen, M. Schiappacassi, F. Porro, F. Bizik, T. Zacchi, G. Lunazzi, S. Miertius, and F.E. Baralle, Immunoreactivity of chimeric proteins carrying the HIV-1 epitope IGPGRAF. Correlation between predicted conformation and antigenicity. FEBS Lett., 353 (1994) 1-4.
6. S. Tripathy, B. Renjifu, W.K Wang, and M. F. McLane, Envelope glycoprotein 120 sequences of primary HIV type 1 isolates from Pune and New Delhi, India. AIDS Res Hum Retroviruses, 12 (1996) 1199-1202.
7. F. Gao, S.G. Morrison, D.L. Robertson and C. Thornton, Molecular cloning and analysis of functional envelope genes from human immunodeficiency virus type 1 sequence subtypes A through G. The WHO and NIAID Networks for HIV Isolation and Characterization. J. Virol., 70 (1996) 1651-1667.
8. L.E. Soto-Ramirez, S. Tripathy, B. Renjifo, and M Essex, HIV-1 pol sequences from India fit distinct subtype pattern. J. Acquir. Immune. Defic. Syndr. Hum. Retrovirol., 13 (1996) 299-307.

9. A. Petruckevitch, J. Del Amo, A. Phillips, A.M. Johnson, and J. Stephenson, Disease progression and survival following specific AIDS-defining conditions: a retrospective cohort study of 2048 HIV-infected persons in London. AIDS, 12 (1998) 1007-1013.
10. R.F. Siliciano, T. Lawton, C. Knall, R.W. Karr, P. Berman, T. Gregory, and E.L. Reinherz, Analysis of host-virus interactions in AIDS with anti-gp120 T cell clones: effect of HIV sequence variation and a mechanism for CD4+ cell depletion. Cell, 54 (1988) 561-576.
11. P.D. Kwong, R. Wyatt, J. Robinson, R.W. Sweet, J. Sodroski, and W.A. Hendrickson, Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Nature, 393 (1998) 648-659.
12. L. A. Lasky, J. E. Groopman, C.W. Fennie,. P.M. Benz, and D.J. Gapon, Neutralization of the AIDS retrovirus by antibodies to a recombinant envelope glycoprotein. Science, 233 (1986) 209-212.
13. J.L. Pellequer, E. Westhof, and M.H.V. Regenmortel, Predicting location of continuous epitopes in proteins from their primary structures. Methods Enzymol., 203 (1991) 176-201.
14. T.L. McInerney, E.I. Ahmar, B.E. Kemp, and P. Poumbourios, Mutation-directed chemical cross-linking of human immunodeficiency virus type 1 gp41 oligomers. J. Virol. 72 (1998) 1523-1533.
15. S. Kiali, E. Elahi, and A. Moghaddam, V3 loop of HIV-1 from Iran. Iran. J. Infect. Dis. Trop. Med., 3 (2001) 23-27.
16. R. Garduno-Juarez, and L.B. Morales, A genetic algorithm with conformational memories for structure prediction of polypeptides. J. Biomol. Struct. Dyn., 21 (2003) 65-87.
17. R.S. Carmenes, J.P. Freije, M.M. Molina, and J.M Martin, Predict7, a program for protein structure prediction. Biochem. Biophys. Res. Commun., 159 (1989) 687-693.
18. M.J. Rooman, and S.J Wodak Identification of predictive sequence motifs limited by protein structure data base size. Nature, 335 (1988) 45-49.
19. J.M.R. Parker, D. Guo, and R.S. Hodges, New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. Biochem. 25 (1986) 5425-5432.

20. M.J. Zvelebil, G.J. Barton, W.R. Taylor and M.J. Sternberg, Prediction of protein secondary structure and active sites using the alignment of homologous sequences. J. Mol. Biol. 195 (1987) 957-961.

21. E. Bause, Structural requirements of N-glycosylation of proteins. Studies with proline peptides as conformational probes. Biochem. J. 209 (1983) 331-336.

22. J.L. Mellquist, L. Kasturi, L. Spitalnik, and S.H. Shakin-Eshleman. The amino acid following an asn-X-Ser/Thr sequon is an important determinant of N-linked core glycosylation efficiency. Biochem. 37 (1998) 6733-6837.

23. S.C. Hubbard and R.J. Ivatt, Synthesis and processing of asparagine-linked oligosaccharides. Annu. Rev. Biochem., 50 (1981) 555-583.

24. X. Huang, J.J. Barchi, F.T. Lung, P.P. Roller, P.L. Nara, J. Muschik and R.R. Garrity. Glycosylation affects both the three-dimensional structure and antibody binding properties of the HIV-1IIIB gp120 peptide RP135. Biochem. 36 (1997) 10846-10856.

25. P. T. Hoop and K.R. Woods, Prediction of protein antigenic determinant from amino acid sequences. Proc. Natl. Acad. Sci. USA, 78 (1981) 3824-3828.

26. B.D. Freedman, Q.H. Liu, M. Del Corno, and R.G. Collman, HIV-1 gp120 chemokine receptor-mediated signaling in human macrophages. Immunol. Res., 27 (2003) 261-276.

27. L. Milich, B. Margolin, and R. Swanstrom, V3 loop of the human immunodeficiency virus type 1 Env protein: interpreting sequence variability. J. Virol., 67 (1993) 5623-5634.

28. J.E. Hansen, O. Lund, J.O. Neilsen, S. Brunak and J.E.S. Hansen, Prediction of the secondary structure of HIV-1 gp120. Proteins, 25 (1996) 1-11.

29. B. Rost, C. Sander, and S. Schneider, Redefining the goals of protein secondary structure prediction. J. Mol. Biol., 235 (1994) 13-26.

30. M.H.V. Van Regenmortel, and G.D Marcillac, An assessment of prediction methods for locating continuous epitopes in proteins. Immunol. Lett. 17 (1988) 95-108.

31. H. Mohabatkar, and S.K. Kar, Prediction of exposed domains of envelope glycoprotein in Indian HIV-1 isolates and experimental confirmation of their immunogenicity in humans. Braz. J. Med. Res. 36 (2004) 675-681.