Vol. 179, No. 2, 1991 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
September 16, 1991 Pages 1056-1062
Reprinted from Biochemical and Biophysical Research Communications 179, 1056–1062 (1991)

Protein and DNA-sequence homologies between the V3- loop of human immunodeficiency virus type I envelope protein gp120 and immunoglobulin variable regions

Radmilla Metlas[1], Veljko Veljkovic[1], Rudy D. Paladini[2] and Sàndor Pongor[2,3]

[1]Boris Kidric Institute, P.O. Box 522, 11001 Beograd, Yugoslavia

[2]International Centre for Genetic Engineering and Biotechnology
Padriciano 99, 34012 Trieste, Italy

[3]ABC Institute of Biochemistry and Protein Research
Szent-Györgyi A. u. 4, 2103 Gödöllö, Hungary

Received July 11, 1991

We found that a common amino acid sequence motif exists between the *V3*-loop region of the human immunodeficiency virus type I envelope protein *HIV gp120* and the human immunoglobulin heavy chain variable regions of subclass III (*Ig V_H-III*). In the *Ig V_H-III* sequences, the common motif overlaps with framework-1, complementarity-determining-region-1 and framework-2. In the homologous regions, the two groups of sequences also have a similar distribution of residue variability. On the DNA sequence level, the homology includes the conserved rearrangement signals of the *V_H-III* genes, which lends support to the speculation that the *V3* region of *gp120* also may be involved in rearrangement processes.

One of the striking aspects of *HIV-1* is the large degree of genetic variation between different viral strains. This variability is not uniformly distributed along the sequence but is confined to certain "hypervariable" regions that alternate with domains that are well conserved among different isolates (1). Sequence similarities have been detected by several authors between different *gp120* proteins and various members of the immunoglubulin superfamily of proteins (2,3). The motifs detected are distributed in various parts of the *gp120* sequence excluding however the *V3*-loop which is thought to be the principal neutralizing determinant that elicits a type but not group specific neutralization response (4). Identification of motifs

Abbreviations: *HIV-1*: human immunodeficiency virus type I; *gp120*: human immunodefficiency virus type I envelope protein. *Ig V_H-III*: human immunoglobulin heavy chain variable regions of subclass III ; *V3-loop*: third variable region of *gp120*. FW: framework, respectively; CDR: complementarity-determining-region.

0006-291X/91 $1.50

common between two groups of sequences is not a straightforward task, especially if the groups compared are as variable as the *gp120* and and the *Ig-V* sequences. With more new *HIV env gp120* sequences available (5, 6) we often find that some of the similarities detected by conventional pairwise alignment are not found in all HIV isolates.

With this in mind we decided to reanalyze the conserved similarities between *gp120* and the *Ig* variable regions using an exhaustive search for local similarities. This analysis revealed a previously undetected homology region between the *V3*-loop of *gp120* and the *Ig* V_H-*III* family of genes. We also found that the recombination signals that flank the region of homology in the V_H-III genes are also well conserved within the gp120 sequences.

METHODS

Sequences of *HIV gp120* and immunoglobulins were taken from the the following databases: Genbank release 65 (7), EMBL, release 24. (8) PIR, release 25 (23) and Swiss-Prot release 14 (24). Variability of *gp120 V3* was calculated as described by Kabat et al. (9) using the sequences taken from the databases as well as from references (5,6). For the detection of local homologies we used *HIV* -1 *gp120* sequences from the following isolates Swiss-Prot database identification code given in parenthesis): ARV2/SF2 (env$hiv1a);BH8 (env$hiv18); BH10 (env$hiv10); BRAIN (env$hiv1i); BRU (env$hiv1b); HXB2 (env$hiv1x); HXB3 (env$hiv1y); MN (env$hiv1L); PV22 (env$hiv1p); RF/HAT (env$hiv1r); SC (env$hiv1s); New-York-5 (env$hiv1n); OYI (env$hiv10); WMJ2 (env$hiv1w); Zaire-6 (env$hiv1z); Z2/CDC-Z34 (env$hiv16). All 10 residue long segments of these sequences were compared with a collection of Ig variable region sequences from the PIR and Swiss-Prot databases, according to the algorithm of Brutlag et al. (10). Average score values were calculated for each position of the aligned gp120 sequences and plotted as a function of the residue position, using the BH10 numbering of the BH10 gp120 sequence (22). The motifs underlying the local homologies were then identified from multiple sequence alignment. The pattern search program QUEST of Intelligenetics was used to identify sequences that contain the motifs identified by the above procedures.

RESULTS

The local similarities between *gp120* and *Ig* variable domain sequences are shown in Figure 1. In this representation, peaks correspond to a motif conserved among *gp120* sequences and homologous with at least one variable region sequences. In order to see the conservation within the Ig sequences, we carried out multiple alignments for the 6 highest peaks in the plot. Four of the highest peaks corresponded to known structural motifs (refs. 11-14). In addition, we found 3 previously undected homology regions designated by *H,I* and *K* in Figure 1. H and I (Figure 2) are short motifs characteristic to light chain sequences; *H* maps onto the V3 region of gp120. A third new homology (*K*) was found between the *V3*-loop *gp120* and the heavy chain V_H-*III* regions (Figure 3). Conserved residues of this

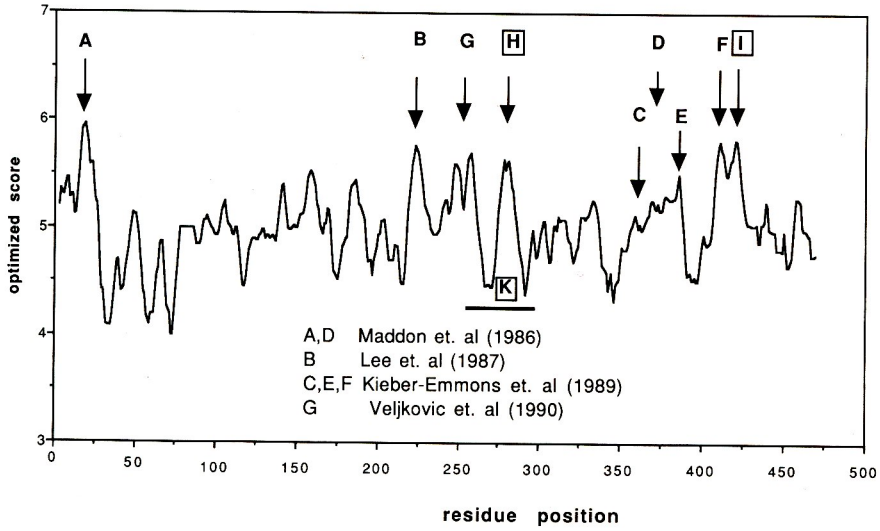


Figure 1. Local similarities between gp120 and immunoglobulin variable regions. *gp120* sequences are listed under Methods. Ig variable region sequences included all the 197 sequences in Swiss-Prot and PIR. Local similarity was calculated as the maximal value of the optimized score obtained when a 10 residue segment of one *gp120* sequence was compared to the *Ig-V* sequences. The scores were averaged for the 10 gp120 sequences analyzed and plotted as a function of *gp120* residue position. Numbering corresponds to the mature *gp120* sequence of Ratner et al. (22).

```
Motif  H
  280   QRGPGRAFYVTI    291   gp120 (BH10 isolate)

   37   QQKPGKAPKVLI     48   kv1d$human
   37   QQKPGKAPKVLI     48   kv1e$human
   37   QQKPGKAPQVLI     48   kv1h$human
   37   QQKPGKAPKFLI     48   kv1c$human     Light chain
   36   QQKPGRAPVMVI     47   lv4c$human     variable regions
   38   QQRPGRAPTTVI     49   lv6c$human
   39   QQHPGRAPKLVI     50   lv2g$human
   39   QQYPGKAPKVLI     50   lv2a$human

        Q--PG[R/K]A----I

Motif  I
  423   LLTRDGGNSNN     433   gp120 (BH10 isolate)

   29   LLYKDGKTYLN      40   kv2a$mouse
   49   LLYKDGKTYLN      60   kv2b$mouse
   29   LLYKDGKTYLN      40   kv2c$mouse
   29   LLHSDGFDYLN      40   kv2d$human     Light chain
   46   LL-IYGAS--N      54   kv5s$mouse     variable regions
   66   LLIRDASSRAN      77   kv3k$human
   66   LLIRDASSRAN      77   k3hu41
   46   LLIYDAS---N      54   kv1b$human

        LL---G----N
```

Figure 2. New common structural motifs found between *gp120* and Ig light chain variable regions (***H*** and ***I*** in Figure 1). The starting and finishing residues of the patterns are indicated at the right and left side of the sequence, respectively. Swiss-prot/PIR names of the sequence entries are listed right from the alignment. Kappa and lambda chain variable reagions have entry names starting by k and l, respectively.

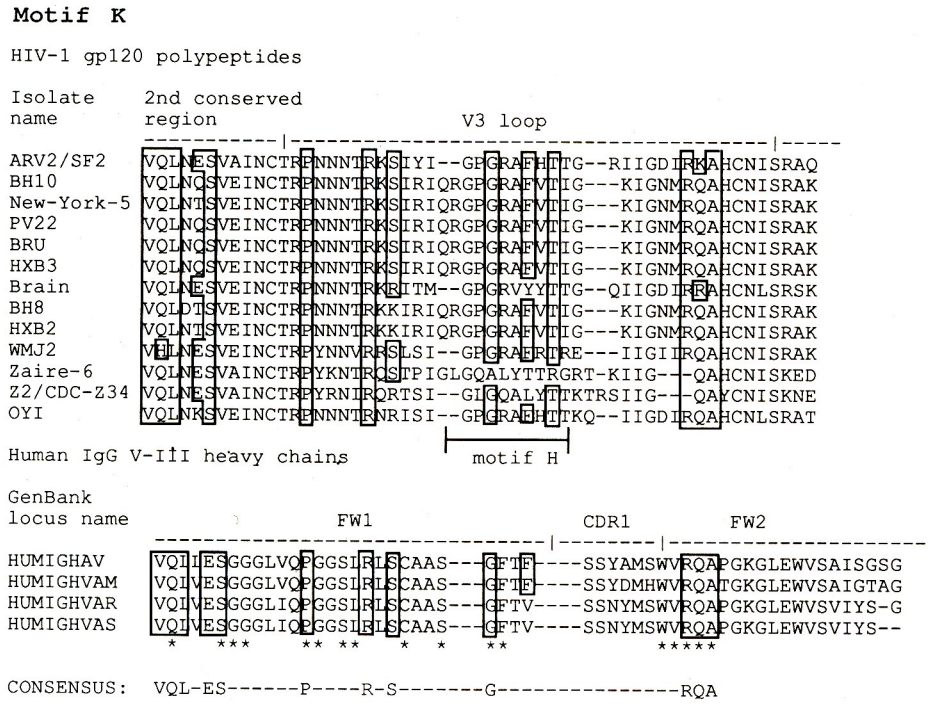


Figure 3. Extended common structural pattern found 256-299 of *gp120* sequences (***K*** in Figure 1). * designates the residues of the Kabat consensus of the Ig V_H-III gene products (9).

motif are in good agreement with the Kabat consensus *V_H-III* genes (Figure 3). of As both *V3* and *Ig V_H-III CDR1* are highly variable regions, we compared the variabilities of the homologous regions using the 245 recently published *V3* sequences (5, 6). Figure 4 shows that the distribution of variability is qualitatively quite similar in the two group of sequences, i.e. variable segments align with variable ones and vice versa.

The alternation of well-conserved and variable regions within *V_H-III* family of genes have led other authos to the suggestion that the constant regions might carry noncoding functions involved in rearrangement processes (15,16). One of such signals, the so-called *Chi* sequence (5'-GCTGGTG-3') is homologous to a signal that promotes generalized recombination in bacteriophage λ and is located in the region coding for *FW1* of *V_H-III* genes (15). Another signal is a conserved 5'-TACTGTG-3' heptamer in the *FW3* region which was implicated as a recombination signal (16). Both of these signals were shown to take part in rearrangement processes in heterologous systems (16). Figure 5 shows that these two signals flank the *K* region of homology and that their sequence is also conserved in the *gp120* sequences. It has been shown by experiment, that a similar level of conservation (one mismatch at the 3' end) can be sufficient to elicit a recombination event

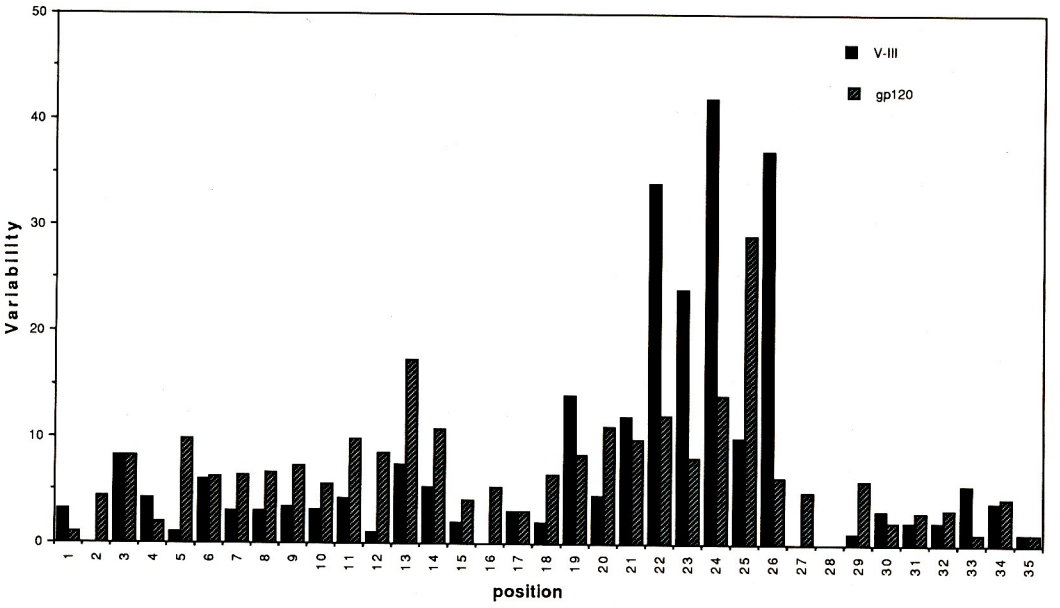
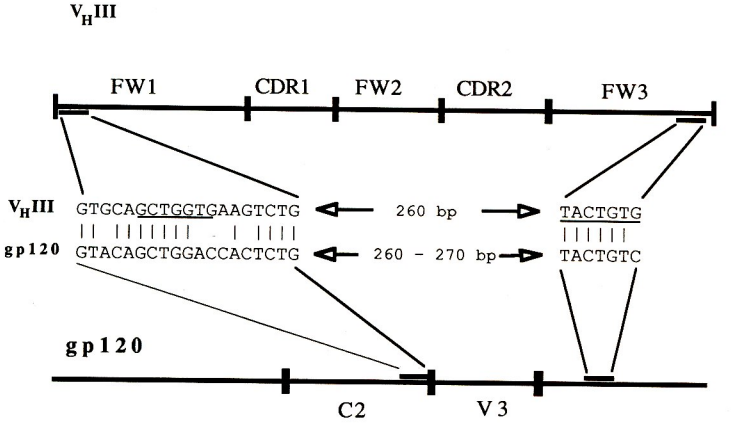


Figure 4. Variability of the *gp120 V3* region and of the immunoglobulin heavy chain *V-III* region. Variability of *Ig V_H-III* are from Kabat et al. (9), variability of the *V3* regions were calculated by the same method using the sequences published by Javaherian et al. (5) and LaRosa et al. (6). Numbering of *gp120*, corresponds to region 266 through 301 in the mature *gp120* sequence of Ratner et al. (122). Numbering in *Ig V_H-III* corresponds to region 2 through 42 of the Kabat consensus (9).

(15). When the Genbank database (43,903 sequences) was searched with the consensus pattern shown in Figure 5 it was found that this pattern is present only in *HIV* gp120 sequences and in immunoglobulin *V_H-III* genes.



Consensus:GT-CAGCTGG---A-GTCG (250-270 BP) TACTGT

Figure 5. Comparison of the consensus nucleotide sequences of the *Ig V_HIII* and *HIV gp120* genes. The Chi site in *FW1* and the recombination heptamer in *FW3* are underlined.

DISCUSSION

The present analysis suggests that there is a conserved framework of Ig-like sequence elements in the *V3*-loop of *gp120*. Previous work in this field revealed Ig homologies only outside the *V3*-loop (11-14). We detected two overlapping motifs, *K* and *H*, within the *V3* loop. Region *K* is similar to heavy chain V_H-*III* sequences and also displays a residue variability profile similar to that family of genes. Internal to region *K* is a shorter motif, *H*, that corresponds to light chain sequences. Experimental results of Wolfs and associates (17) provide indirect support to these findings. These authors found that, in *HIV*-infected patients, the region corresponding to motif *H* is in fact more conserved than the rest of *V3*-loop (17).

The existence of Ig-like structural motifs in the *V3*-loop region might also have implications on vaccine design. Clearly, if the structural motifs described here are in fact "seen" by the immune system then direct testing of *V3*-loop based vaccines in healthy individuals might not be as advisable as generally supposed (2,3,18,25). Tests in *HIV*-infected individuals may present even more complex problems. For example, virus-induced anti-V3 antibodies may be already present in the host due to the infection, and subsequent immunization might trigger the transformation of the disease from latent to acute state (19). As an indirect support for this mechanism we note that the V_H-*III* genes encode heavy chains that are preferentially utilized by fetal B cells. It has been suggested that perturbations in the expression of these Ig's can contribute to immunodeficiency and autoimmunitiy (20,21).

The similarity between the *V3*- loop and the V_H-*III* sequences also includes motifs of the DNA sequence that were previously implicated in rearrangements of the V_H-*III* family of genes. This finding leads us to the speculation that *gp120* might be involved in similar rearrangements, either as a mechanism through which the virus acquired this region, or as a part of the pathogenetic process itself. Since the similarity between the *V3*-loop of *gp120* and V_H-*III* genes includes three different structural aspects (i.e. the protein sequence motifs, the similar distribution of variability, and the homology of the flanking recombination signals in the DNA sequence) it seems worth while to further examine these possibilities.

Acknowledgments: This work was supported in part by the Human Health Service (USA) contract No. JF913 and by the Fund for Scientific Research of the Serbian Repulic (Yugoslavia), contract 1.42. The authors thank Prof. O. Burrone and Dr. M. Giacca, ICGEB, for useful discussions.

REFERENCES

1. Willey L.R., Rutledge A.R., Dias S., Folks T., Theodore T., Buckler E.C., and Martin A.M. (1986) Proc. Natl. Acad. Sci. USA, 83, 5038-5042
2. Bolognesi, D.P. (1990) Trends in Biotechnology 8, 40-45
3. Bolognesi P.D. (1989) AIDS 3 (suppl. 1), S111-S118
4. Rusche R.J., Javaherian K., McDanal C., Petro J., Lynn L.D., Grimaila R., Langlois A., Gallo R.C., Arthur O.L., Fischinger J.P., Bolognesi P.D., Putney D.S., and Matthews J.T. (1988) Proc. Natl. Acad. Sci. USA. 85, 3198-3202
5. Javaherian K., Langlois J.A., LaRosa J.G., Profy T.A., Bolognesi P.D., Herlihy C.W., Putney D.S., and Matthews J.T. (1990) Science 250, 1590-1593
6. LaRosa G.J., Davide J.P., Weinhold K., Waterbury J.A., Profy A.T., Lewis J.A., Langlois A.J., Dreesman G.R., Boswell R.N., Shadduck P., Holley L.H., Karplus M., Bolognesi D.P., Matthews T.J., Emini E.A., Putney S.D. (1990) Science 249, 932-935
7. GenBank (1990): Genetic Sequence Data Bank, National Institute of Health, release 65, 15 September, 1990
8. EMBL (1990): EMBL Data Library, European Molecular Biology Laboratory,Heidelberg release 24, August, 1990
9. Kabat A.R., Wu T.T., Reid-Miller M., Perry M.H., and Gottesman S.K. (1987) "Sequences of proteins of immunological interest" U.S. Department of Health and Human Services, Bethesda Md.
10. Brutlag, D. L. , Dautricourt, J-P., Maulik, S. and Relph, J.(1990) Comput. Applic. Biosci. 6, 237-245
11. Kieber-Emmons, T., Jameson, B.A., and Morrow, W.J.W. (1989) Biochemica et Biophisica Acta 989, 281-300
12. Veljkovic, V. and Metlas, R.(1990) Immunology Letters 26, 193-196
13. Lee, M.R., Ho, D.D. and Gurney, M.E. (1987) Science 237, 1047-1051
14. Maddon, P.J., Dalgleish, A.G., McDougal, J.S., Clapham, P.R., Weiss, R.A. and Axel, R. (1986) Cell 47, 333-348
15. Kenter, A. and Birshtein, B.K. (1981) Nature 293, 402-404.
16. Tuter, A. and Riblet, R. (1989) Proc. Natl. Acad. Sci. USA. 86. 7460-7964
17. Wolfs, T.F. W., De Jong, J.-J., Van den Berg, H., Tijnagel, J.M.G.H., Krone, W.J.A. and Goudsmit, J. (1990) Proc. Natl. Acad. Sci. USA. 87, 9938-9942
18 Sonigo P., Girard M., and Dormont D. (1990) Design and trials of AIDS vaccines. Immunol. Today, 11, 465-471
19. Ascher M.S. and Sheppard H.W. (1990). Journal Acq. Immun. Def. Synd. 3, 177-191
20. Schroeder Jr., H.W. and Wang, J.Y. (1990) Proc. Natl. Acad. Sci. USA. 87, 6146-6150
21. Yancopoulos, G.D., Desiderio, S.V., Paskin, M., Kearney, J.F., Baltimore, D. and Alt, F.W. (1984) Nature 311, 727-733
22. Ratner L., Haseltine, W., Patarca, R., Livak, K.J., Sarcich, B.R., Josephs, S.F., Doran, E.R., Rafalski, J.A., Whitehorn, E.A., Baumeister, K., Ianoff, L., Pettaway, S.R. Jr., Pearson, M.L., Lautenberger, J.A., Papas, T.S., Ghrayeb, J., Chang N.T., Gallo R.C., Wong-Staal F. (1985): Nature 313, 277-284
23. PIR (1990): Protein Identification Resource, National Biomedical Research Foundation, release 25, 30 June , 1990
24. Bairoch, A. (1990): Swiss-Prot Protein Sequence Database, EMBL Data Library, European Molecular Biology Laboratory, Heidelberg release 14, April, 1990
25. Koff W.C. and Hot D.F. (1988). Development and testing of AIDS vaccines. Science 241, 426-432.