Volume 333, number 3, 211–213 FEBS 13186 November 1993

Hypothesis

Does the HIV Nef protein mimic the MHC ?

Uwe Hobohm*, Chris Sander

European Molecular Biology Laboratory, 69012 Heidelberg, Germany

Received 2 September 1993; revised version received 15 September 1993

The sequence of the HIV Nef protein has no significant homology to other proteins in the SwissProt database, and experimental data concerning its function are sparse and contradictory. Using a novel protein sequence comparison method, we find similarities between different Nef sequences and the α chain of human MHC class I proteins. The possible biological implications of this finding are discussed.

HIV; Nef; MHC; Protein mimicry

1. INTRODUCTION

The function of the Nef-protein of HIV, one of the nine gene products of the virus, is still a mystery. While virus lacking the nef gene infects cultured cells and replicates in vitro, no aquired immunodeficiency syndrome in monkeys seems to develop without Nef [1]. Besides the putative function as an element downregulating growth, which gave Nef its name ('negative regulatory factor') and which has been refuted since [2,3], other functions for Nef have been suggested. In the literature, many 'significant homologies' between Nef sequences and other proteins in the database have been reported. For instance, Nef has been proposed to be related with MHC class II β-chains [4], with membrane-associated G-proteins [5], with scorpion neurotoxins [6], with leucine zipper transcriptional activators [7], with superantigens [8], to name only few. The significance of those sequence relationships, however, appears to be based on the experience of the authors, but not on a quantitative estimation. It is obvious, for instance, that short alignments need a higher residue identity to be considered as 'significant'. Recently, a database-derived significance threshold has been derived, which assesses the probability of structural (and possibly functional) homology between two proteins as a function of the alignment length and the sequence identity. According to the significance threshold, a sequence identity of 25% in alignments of length 80 or more indicates structural homology of the aligned portions of the sequence; a higher sequence identity is needed to draw the same conclusion for a shorter alignment [9]. Unfortunately, we were not able to confirm the significance of any relationship reported for Nef by searching Nef sequences against the SwissProt protein sequence database [10] with 28154 entries using the database-derived significance threshold.

*Corresponding author. Fax: (49) (6221) 387 517.

2. PROPSEARCH, A NEW PROTEIN SEQUENCE SEARCHING METHOD

Conventional database searching tools compare the sequential order of amino acid residues between two sequences. We have developed another searching tool (PROPSEARCH), which instead compares the amino acid composition of two sequences while neglecting any sequential order of residues [11]. This approach tries to avoid some of the difficulties of conventional sequence comparison, such as different connectivity of secondary structure elements in similar tertiary structures or compensatory point mutations (charge flips, hydrophobic flips) and represents an attempt to discover generic similarities between protein sequences not dependent on the precise order of residues. In PROPSEARCH, a sequence is represented as a vector of 144 compositional properties (property vector), where properties are simple values such as the content of single or paired amino acids. Two sequences are compared by calculating the euclidian distance between their respective property vectors. A database search is performed by comparing the properties of a query sequence with all sequences in the SwissProt database [9] and sorting distances, with the smallest distances indicating potential structural or functional similarities. A typical database search takes about 4 seconds on a common workstation.

Prior to database searches, the weights for individual properties have been trained using a genetic algorithm and a selection of structurally distinct protein families as training set [11], resulting in a vector of 144 weights.

This property search approach is in some cases capa-

ble of identifying protein pairs with similar structure, but without detectable sequence relationship. For instance, searching for yeast hexokinase-B, which is known to have the same overall structure as actin and some members of the Hsc70 heat shock protein family, but no sequence similarity [12], we found 25 actins and 18 heat shock proteins among the first 200 hits. On the other hand, in some cases the program fails to detect known remote structural homologues and may report false positives. Therefore, the detected relationships merely represent hypotheses that need to be scrutinized by other means.

3. RESULTS

Applying the property search program to 26 HIV Nef sequences from the SwissProt database (sequences with an obvious deletion were not considered), the most abundant protein species found besides HIV Nef itself are α chains of the MHC class I proteins. HIV Nef sequences were from different isolates of HIV1 and HIV2 and are mutualy diverse among themselves, with a sequence identity between 27% and 98%. Class I sequences were from ten different species with a sequence identity between 51% and 99%. 26 HIV Nef sequences find, on average, 24 HLA class I sequences among the first 200 hits. No other protein family shows comparable similarity to Nef with respect to amino acid properties.

For comparison, we did a property search using Nef sequences from SIV. The average distance between SIV Nef and the MHC class I family is greater: 12 Nef sequences from SIV find, on average, only 6 MHC class I α chains among the first 200 hits, and the MHC class I α chain is not the most abundant protein family found by SIV Nef.

4. DISCUSSION

Viruses can 'steal' genes from their hosts and use them for their own purposes. Particularly, several important immune system proteins are used by viruses to elude the immune response (for review, see [13]). No sequence similarity between Nef and HLA class I α chains can be detected. Proteins can, however, exchange all amino acids and retain the structure. The globins, for instance, have diverged down to the level of random sequence identity (< 15%). The high mutation rate of retroviruses may have permitted a complete separation of the two protein families in sequence space.

Nef as a MHC class I structural homolog might interfere with the assembly and/or transport of class I molecules. The downregulation of class I molecules by viral gene products as a defense mechanism against the hosts immunological attack is known from other viruses such as adenovirus and human cytomegalovirus (HCMV). Adenovirus type 2 E3/19K protein binds to class I molecules forming a ternary complex and interferes with the transport to the cell surface [14]. The HCMV UL18 gene product binds β-2-microglobulin as a class I homologue and suppresses the assembly of mature class I molecules [15]. Both strategies most likely reduce or suppress the presentation of viral peptides on the cell surface and hide the virus from the cytotoxic immune response. Even a temporary escape from the immune response may be advantageous. It may be enough to gain some hours in the race between virus and host defense to finish synthesis and assembly of viral proteins and to start release of virions. Nef is an early gene product and is not needed later in the life cycle of HIV. The first Nef mRNA can be detected about 12 h after infection in H9 cells, virus production begins about 24 h after infection [16]. A transient down modulation of

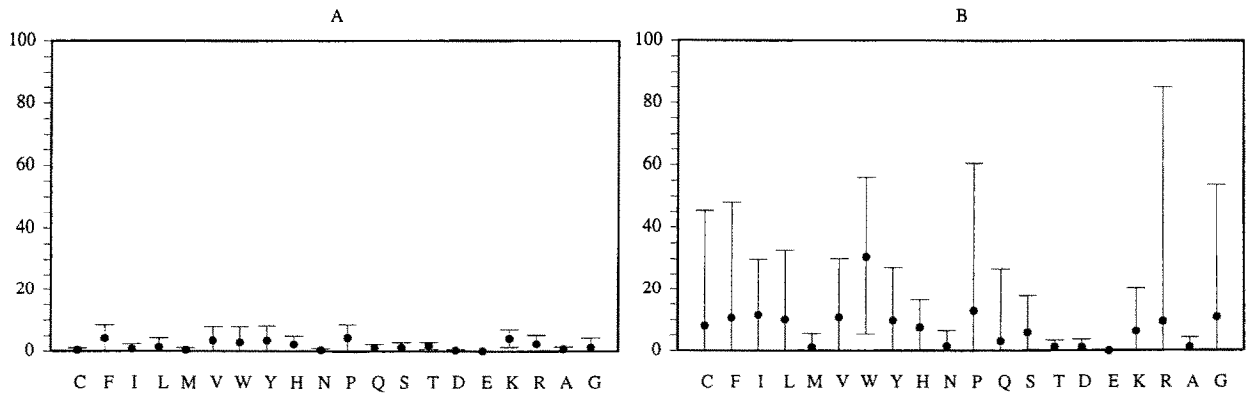

Fig. 1. Difference in amino acid content between HIV Nef and MHC class I sequences (A) and between HIV Nef and random sequences (B). (A) The amino acid content was compared between 26 HIV Nef sequences and 62 MHC class I sequences. (B) The amino acid content was compared between 26 HIV Nef sequences and 1000 sequences randomly selected from the SwissProt Database. Differences in amino acid content (in percent) were multiplied with the respective weights. This was done for all possible sequence pairs and averaged over the number of pairs. (Horizontal axis: amino acids, one letter code. Note that the weight for glutamic acid E was 0.0. Remember also that in a PROPSEARCH database search the distance between two sequences is calculated using 144 properties, of which only 20 are shown here).

MHC class I molecules in another cell line, CEM-E5, beginning 18 h post-infection and reverting to normal level 24 h post-infection has been described [17], although this temporary down-modulation was not attributed to a specific viral gene product.

The hypothesis is compatible with the finding that Nef is not important for infectivity in vitro, but for progression of AIDS in vivo [1], because there is no immune response in cell culture.

The proposal is based on a novel method of sequence comparison that has yet to be established as a reliable tool, but the similarity between HIV Nef sequences and MHC class I α chain sequences in terms of weighted amino acid residue and dipeptide-content is interesting enough to take a closer look by experiment: is there any antibody crossreactivity between Nef protein and MHC class I α chain molecules? Is the MHC class I display downregulated in Nef expressing cells?

REFERENCES

[1] Kestler, H.W., Ringler, D.J., Mori, K., Panicali, D.L., Sehgal, P.K., Daniel, M.D. and Desrosiers, R.C. (1991) Cell 65, 651–662.
[2] Kim, S., Ikeuchi, K., Byrn, R., Groopman, J. and Baltimore, D. (1989) Proc. Natl. Acad. Sci. USA 86, 9544–9548.
[3] Hammes, S.R., Dixon, E.P., Malim, M.H., Cullen, B.R. and Greene, W.C. (1989) Proc. Natl. Acad. Sci. USA 86, 9549–9553.
[4] Vega, M.A., Guido, R. and Smith, T. (1990) Nature 345, 26.
[5] Guy, B., Riviere, Y., Dott, K., Regnault, A. and Kieny, M.P. (1990) Virology 176, 413–425.
[6] Werner, T., et al. (1991) AIDS 5, 1301–1308.
[7] Samuel, K.P., Hodge, D.R., Chen, Y.M.A. and Papas, T.S. (1991) Aids Research and Human Retroviruses 7, 697–706.
[8] Haase, A.T., Jenkins, M.K., Peng, H. and Urdahl, K. (1992) Curr. Biol. 2, 130–132.
[9] Sander, C. and Schneider, R. (1991) Proteins 9, 56–68.
[10] Bairoch, A. and Boeckmann, B. (1991) Nucleic Acids Res. 19, 2247–2250.
[11] Hobohm, U. and Sander, C., manuscript in preparation
[12] Holmes, K.C., Sander, C. and Valencia, A. (1993) Trends Cell Biol. 3, 53–59.
[13] Gooding, L.R. (1992) Cell 71, 5–7.
[14] Burgert, H.-G. and Kvist, S. (1985) Cell 41, 987–997.
[15] Browne, H., Smith, G., Beck, S. and Minson, T. (1990) Nature 347, 770–772.
[16] Kim, S., Byrn, R., Groopman, J. and Baltimore, D. (1989) J. Virol. 63, 3708–3713.
[17] Scheppler, J.A., Nicholson, J.K.A., Swan, D.C., Ahmed-Ansari, A. and McDougal, S. (1989) J. Immunol. 143, 2858–2866.