Brief Definitive Report

VACCINIA VIRUS HEMAGGLUTININ

A Novel Member of the Immunoglobulin Superfamily

By DONGYAN JIN, ZHILIANG LI, QI JIN, HAO YUWEN, AND YUNDE HOU

From the Laboratory of Virology and Genetic Engineering, Institute of Virology, Chinese Academy of Preventive Medicine, Beijing 100052, People's Republic of China

Vaccinia virus hemagglutinin (VVHA) has long been recognized as a lipid-linked glycoprotein that can not only agglutinate chicken as well as human erythrocytes, but can also play an important role in the intercellular dissemination of the virus (1). The binding element on the surface of the erythrocytes and the mechanism responsible for the virus spreading remain elusive for >40 yr. Being a nonvirion protein unnecessary for the virus replication and expressed on the surface of the infected cells, VVHA has been used in recent years as a selection marker in constructing the recombinant vaccinia virus (2), which has the potential to develop polyvalent vaccines. It is of vital importance to elucidate its structure and its influence on the cytopathology and virulence of the virus. The VVHA gene from a nonvaccine strain (WR strain) of vaccinia virus was mapped within the Sal I/Hind III region at the right edge of the Hind III A fragment of the genome and was sequenced recently (3). For safety's sake, the better choice of vaccinia virus in developing vaccines for human use should be a less virulent vaccine strain (4). To provide a molecular basis for understanding the function of VVHA and its interaction with the erythrocytes and the infected cells, we now sequenced and analyzed the VVHA gene from the Chinese vaccine strain (Tiantan Strain) of vaccinia virus.

Materials and Methods

Bacterial Strains, Bacteriophages, and Plasmids. *Escherichia coli* K12 JM103 was originally obtained from Boehringer Mannheim Biochemicals, Penzberg, FRG, and was grown on M9 minimal-medium plates, as recommended by the supplier. Bacteriophage M13 mp18/mp19 RF and plasmids pAT153 and pUC19 were purchased from Pharmacia Fine Chemicals, Uppsala, Sweden. Plasmid pV630, containing the Sal I G fragment of the Tiantan strain of vaccinia virus, was constructed in our laboratory. The sequence coding for VVHA was confirmed by the restriction map compared with that of the WR strain, and also by the loss of HA trait in the recombinant viruses with a galactosidase gene inserted into the region (Hao Yuwen, manuscript in preparation).

Enzymes and Chemicals. Restriction enzymes were mainly purchased from New England Biolabs, Beverly, MA. Deoxyribonucleoside triphosphates (dNTPs, ddNTPs, α-[^{32}P]dATP, and α-[^{35}S]dATP-α-S) were obtained from Amersham International, Amersham, UK. A modified T7 DNA polymerase (Sequenase) and all the other chemicals used in nucleotide

This work was supported in part by High Technology Research Grant 863-102-10 from Chinese National Committee of Science and Technology.

sequencing were from United States Biochemical Corp., Cleveland, OH. All enzymes and chemicals were used according to the manufacturer's recommendations.

Nucleotide Sequencing. The 1.8-kb Sal I/Hind III fragment of plasmid pV630 was subcloned into plasmid pUC19. Overlapping fragments were then taken from the subclone and inserted into M13 mp18 and mp19 in both orientations. The nucleotide sequence was determined on both strands by the M13 dideoxy chain termination method using a modified T7 DNA polymerase (5). The nucleotide and peptide sequences were analyzed on an IBM PS-2 computer with the CalTech software package developed by Dr. Alan Goldin from the California Institute of Technology, Pasadena, CA, and kindly provided by the Molecular Biology Computer Research Resources (MBCRR), Boston, MA. The routine to perform dot-matrix analysis was written by G. Gutman and B. Ward from the University of California, Irvine, CA. The FASTA program (6), developed and kindly provided by Dr. W. R. Pearson from the University of Virginia, Charlottesville, VA, was used to search the National Biomedical Research Foundation (NBRF) protein database release 12.0 (7) obtained from MBCRR.

Results

The VVHA gene from the Tiantan strain is located at the right edge of the Sal I G fragment of the virus genome. The 1,458-bp nucleotide sequence starting from the right terminus of the Sal I G region reveals a single open reading frame with 315 amino acids (Fig. 1). Of them, 11 nucleotides and eight deduced amino acids were found to be different from those in the WR strain (3).

A search in the NBRF protein sequence database revealed proteins belonging to the Ig superfamily with similarities to VVHA. In addition to a 22% overall identity between the first 110 residues of VVHA and the human Ig λ chain V-I region[1], consensus residues were found clustered around the two conserved cystein residues. Sequence alignment was then extented to other members of the Ig superfamily (Fig. 2). Most of the residues that are conserved among the IgV domains (8) are also relatively invariant in the VVHA molecule. The best example is that although there is only one tryptophan in the deduced 315-amino acid VVHA molecule, its position is very similar to that of the conserved tryptophan in the V region (9). Moreover, the size of the similar region (100 residues) and the distance between the two cysteins (70 residues) resemble those of the Ig V region (8). It is reasonable to suggest that VVHA contains an Ig-like domain of ~100 amino acids at its NH_2 terminus, with a three-dimensional structure characteristic of the Ig-like fold (9).

The most intriguing finding in the self-comparison of the VVHA sequence is that two tandem repeating units exist head to tail in the middle of the VVHA gene and deduced peptide (Fig. 3, *A* and *B*). They were located in a region from 170 to 240 residues at the amino acid level, just after the Ig-like domain. These two units share significant sequence homology with each other but show little similarity to proteins belonging to the Ig superfamily. They might possibly have evolved from the duplication of a gene fragment unrelated to the Ig superfamily. It is not known whether this region has a useful viral function.

The deduced protein sequence (Fig. 1) and its hydrophobicity plot (Fig. 4 *A*) demonstrate that VVHA should be a typical transmembrane glycoprotein (Fig. 4 *B*). The first 16 amino acids of VVHA comprise a hydrophobic region rich in leucine,

[1] The following sequences, each with an entry name given in parentheses, are fully referenced in release 12.0 of the NBRF protein database (7): human Ig λ chain V-I region (L1HUNG), human Ig H chain V-II region (MHHUMC), rabbit poly-Ig receptor (QRRBG), human CD4 (RWHUT4), and human TCR β chain (RWMSCS).

```
                GTCGACGATTGTTCATGATGGCAAGATTTATATATCTGGAGGTT   44
   ACAACAATAGTAGTGTAGTTAATGTAATATCGAATCTAGTCCTTAGCTATAATCCGA  101
   TATATGATGAATGGACCAAATTATCATCATTAAACATTCCTAGAATTAATCCCGCTC  158
   TATGGTCAGCGCATAATAAATTATATGTAGGAGGAGGAATATCTGATGATGTTCGAA  215
   CTAATACATCTGAAACATACGATAAAGAAAAAGATTGTTGGACATTGGATAATGGTC  272
   ACGTGTTACCACGCAATTATATAATGTATAAATGCGAACCGATTAAACATAAATATC  329
   CATTGGAAAAAACACAGTACACGAATGATTTTCTAAAGTATTTGGAAAGTTTTATAG  386
   GTAGTTGATAGAACAAAATACATAATTTTGTAAAAATAAATCACTTTTTATACTAAT  443

   ATG GCA CGA TTA CCA ATA CTT TTG TTA CTA ATA TCA TTA GTA     485
 1 Met Ala Arg Leu Pro Ile Leu Leu Leu Leu Ile Ser Leu Val
   --- --- --- --- --- --- --- --- --- --- --- --- --- ---
   TAC TCT ACA CCT TCT CCT CAG ACA TCT AAA AAA ATA GGT GAT     527
15 Tyr Ser Thr Pro Ser Pro Gln Thr Ser Lys Lys Ile Gly Asp
   --- ---
   GAT GCA ACT CTA TCA TGT AAT CGA AAT AAT ACA AAT GAC TAC     569
29 Asp Ala Thr Leu Ser Cys Asn Arg Asn Asn Thr Asn Asp Tyr

   GTT GTT ATG AGT GCT TGG TAT AAG GAG CCC AAT TCC ATT ATT     611
43 Val Val Met Ser Ala Trp Tyr Lys Glu Pro Asn Ser Ile Ile

   CTT TTA GCT GCT AAA AGC GAC GTC TTG TAT TTT GAT AAT TAT     653
57 Leu Leu Ala Ala Lys Ser Asp Val Leu Tyr Phe Asp Asn Tyr

   ACC AAG GAT AAA ATA TCT TAC GAC TCT CCA TAC GAT GAT CTA     695
71 Thr Lys Asp Lys Ile Ser Tyr Asp Ser Pro Tyr Asp Asp Leu

   GTT ACA ACT ATC ACA ATT AAA TCA TTG ACT GCT AGA GAT GCC     737
85 Val Thr Thr Ile Thr Ile Lys Ser Leu Thr Ala Arg Asp Ala

   GGT ACT TAT GTA TGT GCA TTC TTT ATG ACA TCG CCT ACA AAT     779
99 Gly Thr Tyr Val Cys Ala Phe Phe Met Thr Ser Pro Thr Asn

   GAC ACT GAT AAA GTA GAT TAT GAA GAA TAC TCC ACA GAG TTG     821
113 Asp Thr Asp Lys Val Asp Tyr Glu Glu Tyr Ser Thr Glu Leu

   ATT GTA AAT ACA GAT AGT GAA TCG ACT ATA GAC ATA ATA CTA     863
127 Ile Val Asn Thr Asp Ser Glu Ser Thr Ile Asp Ile Ile Leu

   TCT GGA TCT ACA CAT TCA CCA GAA ACT AGT TCT GAG AAA CCA     905
141 Ser Gly Ser Thr His Ser Pro Glu Thr Ser Ser Glu Lys Pro

   GAG GAT ATA GAT AAT CTT AAT TGC TCG TCG GTA TTC GAA ATC     947
155 Glu Asp Ile Asp Asn Leu Asn Cys Ser Ser Val Phe Glu Ile

   GCG ACT CCG GAA CCA ATT ACT GAT AAT GTA GAA GAT CAT ACA     989
169 Ala Thr Pro Glu Pro Ile Thr Asp Asn Val Glu Asp His Thr

   GAC ACC GTC ACA TAC ACT AGT GAT AGC ATT AAT ACA GTA AGT    1031
183 Asp Thr Val Thr Tyr Thr Ser Asp Ser Ile Asn Thr Val Ser

   GCA ACA TCT GGA GAA TCC ACA ACA GAC GAG ACT CCG GAA CCA    1073
197 Ala Thr Ser Gly Glu Ser Thr Thr Asp Glu Thr Pro Glu Pro

   ATT ACT GAT AAA GAA GAA GAT CAT ACA GTC ACA GAC ACT GTC    1115
211 Ile Thr Asp Lys Glu Glu Asp His Thr Val Thr Asp Thr Val

   TCA TAC ACT ACA GTA AGT ACA TCA TCT GGA ATT GTC ACT ACT    1157
225 Ser Tyr Thr Thr Val Ser Thr Ser Ser Gly Ile Val Thr Thr

   AAA TCA ACC ACC GAT GAT GCG GAT CTT TAT GAT ACG TAC AAT    1199
239 Lys Ser Thr Thr Asp Asp Ala Asp Leu Tyr Asp Thr Tyr Asn

   GAT AAT GAT ACA GTA CCA TCA ACT ACT GTA GGA TGT AGT ACA    1241
253 Asp Asn Asp Thr Val Pro Ser Thr Thr Val Gly Cys Ser Thr

   ACC TCT ATT AGC AAT TAT AAA ACC AAG GAC TTT GTA GAA ATA    1283
267 Thr Ser Ile Ser Asn Tyr Lys Thr Lys Asp Phe Val Glu Ile
                                *** *** *** *** ***
   TTT GGT ATT ACC GCA TTA ATT ATA TTG TCG GCC GTG GCA ATT    1325
281 Phe Gly Ile Thr Ala Leu Ile Ile Leu Ser Ala Val Ala Ile
    *** *** *** *** *** *** *** *** *** *** *** *** *** ***
   TTC TGT ATT ACA TAT TAT ATA TAT AAT AAA CGT TCA CGT AAA    1367
295 Phe Cys Ile Thr Tyr Tyr Ile Tyr Asn Lys Arg Ser Arg Lys
    *** *** *** *** *** *** *** *** ***
309 TAC AAA ACA GAG AAC AAA GTC TAG   ATTTTTGACTTACATAAATGT   1412
    Tyr Lys Thr Glu Asn Lys Val End

   CTGGGATAGTAAAATCTATCATATTGAGCGGACCATCTGGTTCAGG            1458
```

FIGURE 1. Nucleotide and deduced amino acid sequence starting from the Sal I G fragment of the genome of the Tiantan strain of vaccinia virus. The putative signal sequence and the probable transmembrane portion of the molecule are indicated, respectively, by dashes and asterisks below the amino acid sequence. Five potential *N*-linked glycosylation sites are also underlined. These sequence data have been submitted to the EMBL/GenBank Data Libraries.

which is probably a signal peptide to be cleaved off the mature protein. At the COOH terminus, another hydrophobic region is followed by a hydrophilic tail rich in basic residues. This unit is most likely the transmembrane-cytoplasmic portion of VVHA. Between the two hydrophobic regions are one Ig-like domain and two tandem repeating units.

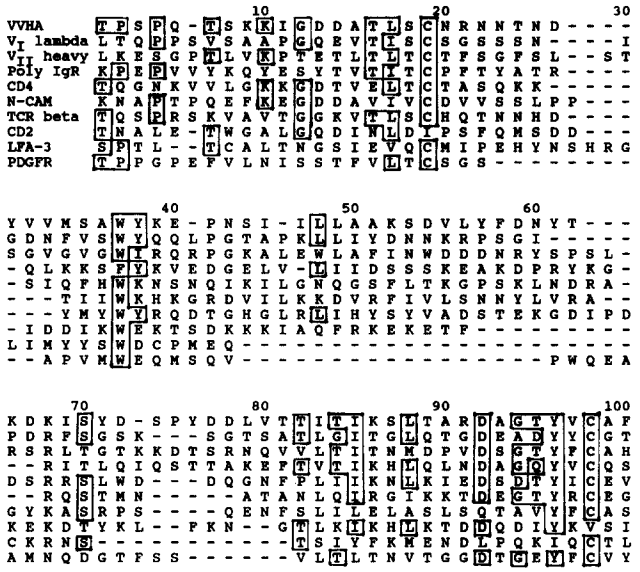

FIGURE 2. Alignment of the sequence between 100 residues at the NH_2 terminus of VVHA and the V domain sequence from different members of the Ig superfamily. Residues identical in VVHA and at least three other aligned sequence are boxed. Gaps indicated by dashes are introduced to maximize the similarities. The aligned sequences are listed as follows: V_I lambda, human Ig λ chain V-I region; V_{II} heavy, human H chain V-II region; Poly IgR, rabbit poly-Ig receptor; human CD4; chicken N-CAM; TCR beta, human TCR β chain V region; human CD2; human LFA-3; and mouse PDGFR.

Discussion

The concept of an Ig-like domain as the primordial, yet versatile structure involved in intercellular recognition in higher eucaryotes has been strongly reinforced by the sequences of many newly identified members of the Ig superfamily (10). The homophilic adhesion of the neural cell adhesion molecule (N-CAM) and the binding of CD2 to LFA-3 may represent the basic model for the interaction within the superfamily. Considering that the VVHA has an Ig-like domain exposed on the cell surface and that VVHA is responsible for the hemagglutination, the intercellular spreading, and perhaps the release of the virus (3), we believe that VVHA will be another case in support of the above model. Among the superfamily members, LFA-3 and rat OX-45, whose equivalent in humans is called Blast-1 (11), were found to be

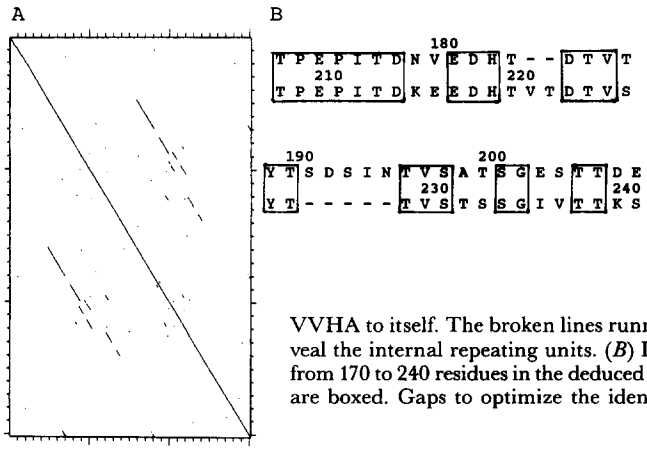

FIGURE 3. Internal tandem repeating units in the VVHA gene and deduced peptide. (*A*) Self-comparison of the partial VVHA gene using a dot-matrix program. The 900–1,200 bp of the VVHA gene are presented on both the horizontal and vertical axes. Seven matches out of eight residues are required to produce a dot on the plot. The solid diagonal bisecting the figure is the result of identity of VVHA to itself. The broken lines running parallel with the diagonal reveal the internal repeating units. (*B*) Internal homologies of the region from 170 to 240 residues in the deduced VVHA peptide. Identical residues are boxed. Gaps to optimize the identities are indicated by dashes.

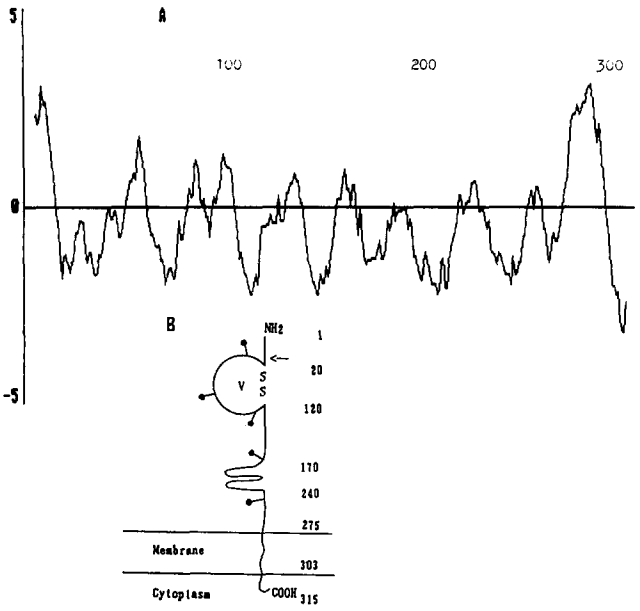

FIGURE 4. (*A*) Kyte and Doolittle hydrophobicity plot of the deduced VVHA polypeptide with a window of 11 residues. (*B*) Schematic diagram illustrating a VVHA molecule on the cell surface. The hypothetical Ig V-like domain from 20 to 120 residues is denoted by a circle marked V, and an intrachain disulphide bridge (*S-S*), which serves to stabilize the Ig-like fold, is postulated to form between cysteins 34 and 103. Also shown are two repeating units located in the 170–240 region, five potential *N*-linked glycosylation sites (*rods with solid ball at one end*), and the approximate cleavage site of the signal peptide (*arrow*).

expressed on the surface of erythrocytes. The LFA-3 antigen was also shown to mediate adhesion between T cells and erythrocytes by interacting with CD2. It would be of interest to see whether VVHA would trigger the virus-induced hemagglutination by recognizing an as yet unidentified Ig-related ligand on the erythrocytes.

It is generally accepted that all members of the Ig superfamily share a common ancestry (10). Comparisons of the amino acid sequence between VVHA and other superfamily members demonstrate that the Ig-like domain in VVHA is structurally more similar to the Ig V domain, and that the sequence flanking the Ig-like domain is perhaps dissimilar to proteins belonging to the Ig superfamily. This prompts us to consider that vaccinia virus had captured an exon encoding an Ig V domain from the eucaryotic cell when interacting with the host immune system and converted it through evolution to the VVHA molecule of its own. It is noteworthy that monoclonal autoantibodies against intermediate filaments or Thy-1.2 antigen produced by clones established after immunization with lysates from cells infected by vaccinia virus were shown to crossreact with VVHA (12). Elucidation of the influence of VVHA on the cytopathogenesis and virulence of the vaccinia virus requires further study of the potential molecular mimicry between VVHA and other members of the Ig superfamily, including the myelin-associated glycoprotein MAG and the major glycoprotein of peripheral myelin P_0 found on neural tissues (10).

Summary

Striking similarities between vaccinia virus hemagglutinin (VVHA) and proteins belonging to the Ig superfamily clearly indicate that VVHA, a 315-amino acid glycoprotein expressed on the surface of the infected cells, is a novel viral protein that can be added to the expanding list of the Ig superfamily. Its deduced amino acid sequence contains one Ig-like domain at the NH_2 terminus, followed by two tandem

repeating units and a hydrophobic region, suggestive of membrane spanning. The results offer an opportunity for the further study of the probable evolutionary and possible functional relationship between VVHA and other members of the Ig superfamily. Our observation, together with a recent finding that human CMV possibly encodes a protein similar to the MHC class I antigens (13), provides evidence supporting the fact that the viral capture of cellular Ig-related genes is more common than expected in vaccinia and other viruses, and that the usage of an Ig-like domain as recognition signals might be extended from higher animals to animal viruses.

We thank Dr. Jiming Zhu from our institute, Dr. Alan F. Williams from the University of Oxford (Oxford, UK), Dr. Don C. Willy from Harvard University (Cambridge, MA), Dr. Bernard Moss and Dr. Ronald Germain, both from the National Institutes of Health (Bethesda, MD), and Dr. Jiahuai Wang from the Institute of Biophysics, Chinese Academy of Sciences, for helpful comments and suggestions. We also thank Molecular Biology Computer Research Resources and its user coordinator Ms. Susan Russo from the Dana-Farber Cancer Institute (Boston, MA) for kindly providing the computer softwares.

Received for publication 12 April 1989.

References

1. Ichihashi, Y., and S. Dales. 1971. Biogenesis of poxvirus: interrelationship between hemagglutinin production and polykariocytosis. *Virology.* 46:533.
2. Shida, H., T. Tochikura, T. Sato, T. Konno, K. Hirayoshi, M. Seki, Y. Ito, M. Hatanaka, Y. Hinuma, M. Sugimoto, F. Takahashi-Nishimaki, T. Maruyama, K. Miki, K. Suzuki, M. Morita, H. Sashiyama, and M. Hayami. 1987. Effect of the recombinant vaccinia viruses that express HTLV-I envelop gene on HTLV-I infection. *EMBO (Eur. Mol. Biol. Organ.) J.* 6:3379.
3. Shida, H. 1986. Nucleotide sequence of the vaccinia virus hemagglutinin gene. *Virology.* 150:451.
4. Hou, Y. T., X. K. Yang, and Y. W. Hu. 1985. Variation in the Hind III restriction fragments of DNA from the Chinese Tian Tan strain of vaccinia virus. *J. Gen. Virol.* 66:1819.
5. Tabor, S., and C. C. Richardson. 1987. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci. USA.* 84:4767.
6. Pearson, W. R., and D. J. Lipman. 1988. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA.* 85:2444.
7. Sidman, K. E., D. G. George, W. C. Barker, and L. T. Hunt. 1988. The protein identification resource (PIR). *Nucleic Acids Res.* 16:1869.
8. Taylor, W. R. 1986. The classification of amino acid conservation. *J. Theor. Biol.* 119:205.
9. Amzel, L. M., and R. J. Poljak. 1979. Three dimensional structure of immunoglobulins. *Annu. Rev. Biochem.* 48:961.
10. Williams, A. F. 1987. A year in the life of the immunoglobulin superfamily. *Immunol. Today.* 8:298.
11. Killeen, N., R. Moessner, J. Arvieux, A. Willis, and A. F. Williams. 1988. The MRC OX-45 antigen of rat leukocytes and endothelium is in a subset of the immunoglobulin superfamily with CD2, LFA-3 and carcinoembryonic antigens. *EMBO (Eur. Mol. Biol. Organ.) J.* 7:3087.
12. Dales, S., R. S. Fujinami, and M. B. A. Oldstone. 1983. Infection with vaccinia favors the selection of hybridomas synthesizing autoantibodies against intermediate filaments, one of them cross-reacting with the virus hemagglutinin. *J. Immunol.* 131:1546.
13. Beck, S., and B. G. Barrell. 1988. Human cytomegalovirus encodes a glycoprotein homologous to MHC class-I antigens. *Nature (Lond.).* 331:269.