Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 7363–7367, December 1984
Biochemistry

Vaccinia virus 19-kilodalton protein: Relationship to several mammalian proteins, including two growth factors

(transforming growth factor type 1/epidermal growth factor/blood coagulation factors/evolution/conserved sequence patterns)

MARGARET C. BLOMQUIST, LOIS T. HUNT*, AND WINONA C. BARKER

National Biomedical Research Foundation, Georgetown University Medical Center, Washington, DC 20007

Communicated by Russell F. Doolittle, August 9, 1984

ABSTRACT **We observed two unusual patterns of cysteine and glycine residues in transforming growth factor type 1 (TGF1); our computer search found only one other protein, the 19-kilodalton early protein of vaccinia virus, having both patterns. The sequences of epidermal growth factor (EGF) and of the light chains from several components of the blood coagulation system also have one of the patterns, but gaps are required to adjust conserved cysteine and glycine residues in the second pattern. We used several computer analyses to confirm these relationships; the 19-kilodalton protein appears to be related to TGF1 and EGF to the same degree that they are related to each other; all three are more distantly related to the coagulation factors. An evolutionary scheme is presented for these proteins. We suggest that the conservation of cysteine residues, which form the disulfide bonds present in the active EGF molecule, may extend to conservation of disulfide bonds in these other proteins. We also suggest that the structural similarities may be correlated with a protein-binding capability.**

The transforming growth factor type 1 (TGF1), secreted by rat and mouse retrovirus-transformed fibroblasts and by human melanoma cells, has been found by Marquardt and coworkers (1, 2) to be structurally and functionally related to mouse (3, 4) and human (5) epidermal growth factors (EGFs). While preparing to add the rat TGF1 sequence to the Protein Sequence Database, one of us observed in this sequence two unusual patterns, involving cysteine and glycine residues, that were shared by the two (TGF1 and EGF) growth factor sequences. We searched the database for any other occurrences of these two patterns.

The sequence pattern in the segment of TGF1 containing residues 32–43 occurred in seven kinds of proteins in the database, as is illustrated in Fig. 1. Two of the proteins having the pattern were EGF and vaccinia virus 19-kDa protein. Four more proteins having the pattern were components of the blood coagulation system (termed clotting factors in this paper). Three of these components that were already known to have a domain related to EGF (7, 8) include human (9) and bovine (10) factor IX (EC 3.4.21.22), human (11) and bovine (12, 13) factor X (EC 3.4.21.6), and bovine protein C (EC 3.4.21.-) (14, 15); the fourth component having an EGF-related domain (this paper) is human PLA (EC 3.4.21.31) (16). The pattern also occurred in sequenced fragments of human (17, 18) and bovine (19) fibronectins once and twice, respectively.

The other sequence pattern, in the segment of TGF1 containing residues 8–21, is C-x-x-x-x-x-x-C-x-x-G-x-C; it was found in only one other sequence, the vaccinia virus 19-kDa protein. If an additional residue position is inserted in the sequence pattern, as is done to obtain maximum homology between TGF1 and EGF (1), then the sequence pattern also matches that of EGF; no exact matches are found with any other sequences in the database. These observations prompted us to undertake a detailed analysis of the possible relationships between these sequences.

		C x C x x G x x G x x C
19K protein, Vaccinia	69	C R C S H G Y T G I R C
EGF, Mouse	1007	C N C V I G Y S G D R C
EGF, Human	31	C N C V V G Y I G E R C
TGF1, Rat	32	C V C H S G Y V G V R C
PLA, Human	108	C Q C P E G F A G K C C
Factor X, Human	70	C T C L E G F E G K N C
Bovine	70	C T C A E G F E G K N C
Factor IX, Human	117	C W C P F G F E G K N C
Bovine	71	C W C Q A G F E G T N C
Protein C, Bovine	78	C D C A E G W E G R F C
Fibronectin, Human	45	C T C Y G G S R G F N C
Bovine	45	C T C Y G G S R G F N C
Bovine	844	C T C F G G Q R G W R C

FIG. 1. Segments found in the search of a sequence pattern conserved in TGF1 and EGF. Conserved residues are boxed. The number refers to the sequence position of the first cysteine residue in the conserved segment. Mouse EGF is residues 977–1029 in the EGF precursor sequence (3, 4). The standard one-letter amino acid abbreviations are used (6). PLA, plasminogen activator, tissue-type; 19K protein, 19-kDa protein.

Vaccinia virus, the cause of cowpox, is a member of the Poxviridae; it has double-stranded DNA and replicates within the cytoplasm of infected cells (20, 21). The 187,000 base-pair genome has a 10,000 base-pair inverted terminal repetition; the gene for the 19-kDa protein from the WR strain (American Type Culture Collection) is one of at least three lying within the inverted terminal repetition, at map position 6.54–7.16 (22), and is transcribed early in infection (20, 21). The actual molecular mass of the unmodified protein chain is ≈15.5 kDa (20). The function of this protein is unknown (20).

COMPUTER METHODS

To compare proteins for possible similarities in secondary structure, we used program PRPLOT, which is based on the algorithm of Hopp and Woods (23), to plot hydrophilicity and regions of potential α-helices, β-sheets, and β-turns. In this program, user-specified amino acid scoring data are averaged over each segment of a user-specified length (window size), and the average values are plotted versus sequence position. The average scores are plotted at the position corresponding to the middle of each segment. The hydrophilicity data of Hopp and Woods (23) and the hydropathy index of Kyte and Doolittle (24) were used. For indications of possible α-helices, β-sheets, and β-turns, the data of Chou and Fasman were used (25).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: EGF, epidermal growth factor; TGF1, transforming growth factor type 1; PLA, plasminogen activator, tissue-type; PAM, accepted point mutation per 100 amino acid residues.

*To whom all correspondence and reprint requests should be addressed.

Program SEARCH is designed to compare test protein segments of a specified length with all segments of the same length in the protein sequence database (26). The program allows for amino acid substitutions but not insertions or deletions. Segment scores are computed by using a pairwise amino acid scoring matrix. We used the Mutation Data Matrix (MDM250) for 250 PAMs, or accepted point mutations per 100 residues (26–28). The 100 top-scoring segments were selected for further examination.

To evaluate the possibility of distant relationships, we used several programs. Program DOTMATRIX (29) is a graphic matrix display program that uses an algorithm similar to that of program RELATE (30) to locate regions of similarity between two sequences. Segments of a specified length from a sequence represented on the horizontal axis are compared with segments (of the same length) from a sequence on the vertical axis, and similarities are scored with the MDM250; segment scores equal to or greater than a specified minimum score are plotted (31). Segment lengths of 20 residues and a minimum score of 20 were used; a dot occurs at the appropriate position when the MDM250 score for a 20-residue comparison is 20 or greater.

In program RELATE (30), all segments of a given length in one sequence are compared with all segments in the second sequence. In program ALIGN (30), the best alignment of two sequences is made. For both methods, a scoring matrix (the MDM250) is required. ALIGN also requires a matrix bias factor and a gap penalty; we used a bias of 6 and a gap penalty of 6 or 10. A numerical property of the real comparison is calculated; this same property is also calculated for a large number of pairs of randomly permuted sequences (with the same amino acid compositions as the real sequences). With ALIGN, we used 3000 pairs of permutations, and with RELATE, we used 100. The mean and standard deviation of the property are estimated from the distribution of scores of the permuted sequences. The ALIGN scores and the RELATE (segment comparison) scores are expressed in standard deviation (SD) units.

From the alignment that was constructed, a matrix of percent differences was calculated and corrected for inferred parallel and superimposed mutations to yield a matrix of evolutionary distances between sequences in PAMs (32). The matrix of PAM values was then used to derive the evolutionary tree; our tree program (32, 33) is a modified version of the matrix method first used by Fitch and Margoliash (34).

RESULTS AND DISCUSSION

We constructed an alignment (Fig. 2) of extended portions of the 10 sequences containing both sequence patterns. (The fibronectin sequences were not included, as they do not have the complete first sequence pattern in the alignment, and, although they have some similarity to the other sequences over the first half of the alignment, it is very weak.) The alignment length (55 positions) was based on the alignment of the entire sequences of TGF1 and human EGF; corresponding regions from the sequences of the other proteins were used. The sequences and sequence segments in this alignment vary slightly in number of residues (49–55) when the necessary gaps are inserted. The insertion of one gap area (alignment positions 13–15) in the sequences from three of the clotting factors allows these sequences to match the conserved cysteine and glycine residues of the other sequence pattern. However, in this same area, protein C has 8 more residues than the other clotting factors and 5 more than the two growth factors; these 5 residues (P-C-D-L-P) are omitted from the alignment. Rat TGF1 and the two EGF sequences have identical residues at 16 positions and the 19-kDa protein shares 13 of these, including the 6 cysteine residues. EGF and the 19-kDa protein share identical residues at 18 positions and highly similar residues at 6 more positions. Furthermore, human EGF and the 19-kDa protein share the hexapeptide sequence D-G-Y-C-L-H (the only occurrences in the database).

The three disulfide bonds known for EGF (35) are shown (Fig. 2); it has been proposed that TGF1 may have the same arrangement (2). These bonds presumably maintain the molecular conformation necessary for binding of EGF (and TGF1) to the EGF receptor, as reduction of the bonds prevents binding (2). Komoriya *et al.* (36) proposed that the loop between Cys-3 and Cys-4 in EGF (alignment positions 23–32) forms a major receptor-binding region. Disruption of this loop, by cyanogen bromide cleavage at the single methionine residue of EGF, produced a molecule having 2 segments still joined by the three disulfide bonds (37, 38); however, the shape of the molecule that these bonds normally maintained could have been altered. Greatly reduced receptor-binding capacity was found as one of the effects of this treatment (37, 38). The 19-kDa protein has 8 cysteine residues, but it is not known if any disulfide bonds are present (20). If all 8 cysteine residues are free to form disulfide bonds, and do so, then the pairing cannot be the same as in EGF, because the 2 cysteine residues not shown in the alignment are at adjacent positions within a long hydrophobic region in the protein and would not be able to bond with each other. However, if this hydrophobic region lies within a membrane (see below), the 2 adjacent cysteine residues may not be able to interact with others in the protein; in that case, the 6 cysteine residues in the central region of the 19-kDa protein may be able to form the same disulfide bonds as EGF. However, even if the pairing is found not to be the

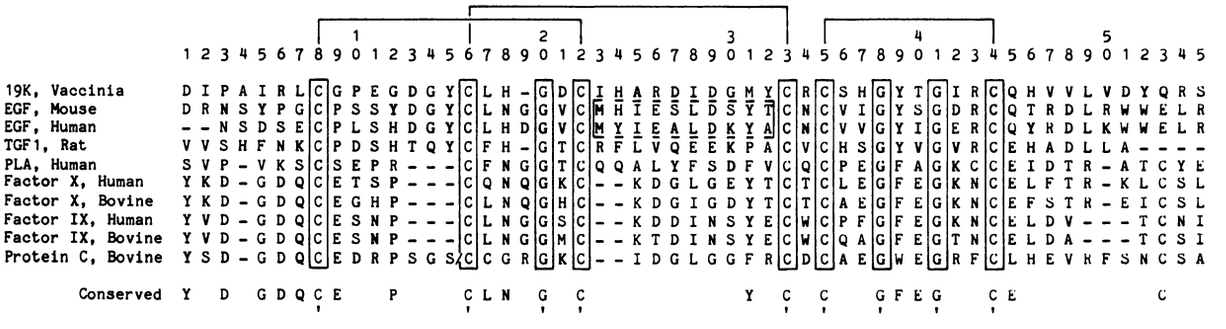

FIG. 2. Alignment of vaccinia virus 19-kDa protein with EGF, TGF1, and the first EGF-like domain from the light chains of four clotting factors. Residues conserved in at least half of the sequences appear below the aligned sequences. Boxes enclose the 9 invariant residues; the first 4 invariant residues compose the first pattern and the last 5, the second pattern. The three disulfide bonds of EGF are indicated (35). A proposed major receptor-binding region of EGF (36), which includes residues at positions 25–32, is enclosed by a dashed line. Sequence residue numbers for the segments, in order shown, are 1, 38–91; 2, 975–1029; 3, 1–53; 4, 1–50; 5, 80–129; 6, 44–91; 7, 44–91; 8, 91–136; 9, 45–90; 10, 44–57 and 63–100. A slash indicates the position of 5 residues omitted from sequence 10.

same, the cysteine residues are at matching positions in the 19-kDa protein and the two growth factors.

The activated clotting factors are composed of an NH_2-terminal light chain and COOH-terminal heavy chain, linked by at least one disulfide bond (39). The heavy chain contains the active site residues of a serine protease. In factor IX, factor X, and protein C the light chain contains two tandem, growth factor-related regions (7, 8), but the two cysteine and glycine patterns are not conserved in the second region. In human factor IX these two regions are coded by the fourth and fifth exons of the gene; the fourth exon corresponds to residues at alignment positions 3–46 in Fig. 2 (40). PLA and urokinase, the plasminogen activator from urine (41), have only one growth factor-related region; in urokinase, however, the two cysteine and glycine patterns are not conserved in this region. The light chains of the clotting factors contain at least 15 cysteine residues, but the arrangement of possible disulfide bonds is not known for any of them. The possibility that the EGF-like region in the 19-kDa protein and the clotting factors could have the same disulfide bond arrangement as EGF and/or could bind to a protein receptor remains to be experimentally investigated.

The hydrophilicity profiles for the same sequence regions shown in the alignment (Fig. 2) indicate that these sequences have short alternating hydrophobic and hydrophilic segments. The predicted locations for β-turns are quite similar in all of the sequences examined: just before Cys-1, between Cys-1 and Cys-2, between Cys-3 and Cys-4 [in the center of a proposed receptor-binding region (36) of EGF], and midway between Cys-5 and Cys-6. These locations of predicted β-turns agree with those proposed by Holladay *et al.* (42) and by Doolittle *et al.* (8) and also with those that might be expected on the basis of the loops formed by the disulfide bonds in EGF (35). We obtained no clear indications of α-helical or β-sheet structures between Cys-1 and Cys-6 in these sequences; however, there are experimental data for some β-sheet structure in EGF (42). The hydrophilicity profiles for the complete 19-kDa protein and residues 930–1070 of mouse EGF precursor were also compared and found to be similar. Aside from the central conserved region, the most remarkable resemblance is a region containing >25 hydrophobic residues close to the COOH end in each sequence. Doolittle *et al.* (8) suggest that this region in the EGF precursor could be a membrane-spanning segment, allowing most of the molecule to be exposed outside the cell in which it was synthesized. The presence of a similar long hydrophobic segment in the 19-kDa protein might allow it to interact with some membrane.

This previously unsuspected relationship of vaccinia virus 19-kDa protein to EGF and TGF1 was investigated by several other computer methods. First, various 30-residue segments covering the entire 19-kDa protein were searched against the Protein Sequence Database, which contained >450,000 30-residue segments. If at least half the residues in a segment were derived from the middle portion (residues 46–90) of the 19-kDa protein, then TGF1 was the highest-scoring segment found, and both EGF sequences were next. These search scores were high enough to indicate relationship of the segments. The 20 top-scoring segments also included the EGF-related regions from most of the clotting factors. Searches with the NH_2-terminal and COOH-terminal regions of the 19-kDa protein did not reveal any obviously related sequences. Similar searches were performed with 2 overlapping segments each from TGF1 and mouse EGF precursor. In three of these four searches, the 19-kDa protein was among the 5 top-scoring segments and several of the clotting factors also occurred in the top 10 segments of one search (residues 1000–1029 of EGF precursor).

Although the EGF precursor has several other sequence regions related to the functional EGF sequence (residues

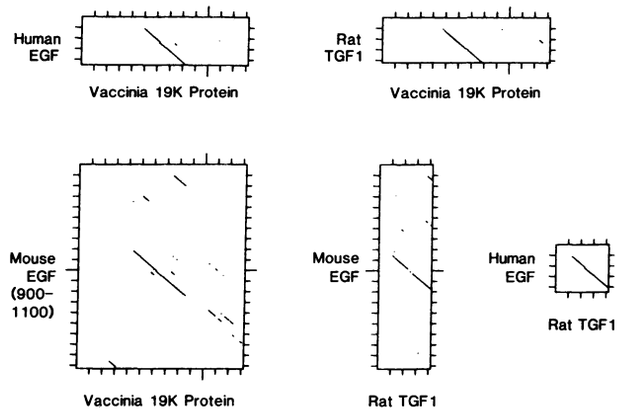

FIG. 3. Graphic matrix plots of the 19-kDa protein, TGF1, human EGF, and mouse EGF precursor (residues 900–1100). A minimum score of 20 and a segment length of 20 residues were used in all plots. Units on the axes of the plots correspond to 10-residue intervals. Each 20-residue comparison with a score of 20 or greater is indicated by a dot, so that regions of similarity between two sequences appear as diagonal lines.

977–1029) (3, 8), none of these has either of the two patterns shared with TGF1 and the 19-kDa protein nor were these related sequence regions found among the highest-scoring segments on the various searches.

Graphic matrix plots of the 19-kDa protein with TGF1 and both EGF sequences indicated the same extent of sequence similarity as that between TGF1 and both EGFs (see Fig. 3). In these comparisons, residues 900–1100 of the mouse EGF precursor were used; residues 977–1029 correspond to the human EGF sequence. In each plot the longest diagonal indicates a region of similarity between the 2 sequences that extends over ≈40 residues at least and over 50 residues for mouse EGF precursor and the 19-kDa protein.

As the alignment indicated that the sequence of the 19-kDa protein was >60% different from those of the two growth factors and as it was almost three times as long, we used our programs ALIGN and RELATE to evaluate the relationships among these proteins (30). Comparison of the entire 19-kDa protein with TGF1, using program RELATE and a 20-residue segment length, obtained a segment comparison score of 11.6 SD. The 19-kDa protein was also compared with residues 950–1050 of mouse EGF; segment lengths of 20, 15, and 10 residues obtained scores of 8.1, 7.3, and 6.2 SD, respectively. All of these scores are above 5.0 SD and the sequences are therefore considered to be related (26, 30).

Table 1 presents ALIGN scores for comparisons among 19-kDa protein, TGF1, and EGF; both EGFs are included to

Table 1. ALIGN scores of 19-kDa protein and growth factors

	19-kDa protein, vaccinia virus		TGF1, rat	EGF, human
	(residues 1–140)	(residues 38–91)	(residues 1–50)	(residues 1–53)
TGF1, rat (residues 1–50)	7.5	8.5		
EGF, human (residues 1–53)	6.1	7.7	8.8	
EGF precursor, mouse (residues 930–1070)	6.0	6.5	5.3	15.7
EGF precursor, mouse (residues 975–1029)	6.6	8.9	8.0	17.2

A gap penalty of 10 and 3000 random permutations were used in deriving these scores.

provide scores from two closely related proteins. The sequences of the active growth factors are just over 50 residues in length, whereas the 19-kDa protein sequence is 140 residues long, with residues 38–91 being the region that corresponds to the sequences of the two growth factors. It is better to compare sequences of approximately the same length, but, on the other hand, shortening a sequence involves preselecting the "best match." Therefore, in the comparisons we used both residues 1–140 and 38–91 of the 19-kDa protein and residues 930–1070 and 975–1029 of mouse EGF precursor. Even using the longer sequences, all except one of the ALIGN scores were 6.0 SD or higher, indicating that the sequences are related (26).

ALIGN scores for comparisons between the sequence segments in the alignment (Fig. 2) from the various clotting factors indicated, as expected, that all are related to one another. Factors IX and X (bovine and human) are fairly closely related and the scores fall between 11 and 18 SD, whereas protein C and plasminogen activator are more distant, with scores ranging between 4 and 9 SD. Comparisons of EGF with these proteins, again using the sequence segments in the alignment, produced ALIGN scores between 5 and 7 SD for factor X, but slightly lower scores (4–5 SD) for factor IX, plasminogen activator, and protein C. No scores >3.2 SD were obtained in comparisons of these proteins with TGF1. The comparisons of the 19-kDa protein with plasminogen activator, factor IX, and protein C produced scores between 4 and 5 SD, but those with factor X were <4 SD. However, EGF is related to the four clotting factors on the one hand and to TGF1 and 19-kDa protein on the other, so it forms a link between the two sets of protein families. Except for 1 cysteine residue in the first pattern region that is sometimes unmatched, all invariant cysteine and glycine residues shown in the Fig. 2 alignment were matched in all the ALIGN comparisons.

To interpret these scores, one can consider them in relation to another group of proteins whose members are known to be distantly related to one another, such as the group that includes insulin, insulin-like growth factors (ILGFs), and relaxin. This group of proteins may also be appropriate, as most have some effects on growth, all have certain cysteine residues (and the disulfide bonds they form) conserved, and the active molecules contain about 50–60 residues. Comparisons (using the same parameters) of human insulin (43), rat insulin 2 (44), and the corresponding regions of human ILGF I (45), rat ILGF II (46), and rat relaxin (47) produce these ALIGN scores (in SD units): insulins, 20.3; ILGF I and ILGF II, 16.5; insulin and ILGF I, 13.6; insulin and ILGF II, 12.1; insulin and relaxin, 5.3; and ILGFs and relaxin, 5.2 and 3.7.

A matrix of evolutionary distances calculated from the alignment of Fig. 2 was used to derive evolutionary trees. The sequences of human factor IX and factor X were omitted in order not to exceed eight branches. For a tree of eight branches, the program derives all 10,395 possible topologies. In this case, only 6 topologies with all positive branch lengths were found; there was <3.0 PAMs overall difference in length among the 6. The shortest tree is shown in Fig. 4. The EGF-like domain of the 19-kDa protein fits very well on the growth factor side of the tree. Duplication of an ancestral gene and divergence of genes from EGF and TGF1 may have occurred before the origin of vertebrates (1). However, it is not clear whether the 19-kDa protein is more closely related to EGF or to TGF1, as the relative order of branching of the three lines varied among the six trees (indicated by the dashed oval), nor can this uncertainty be resolved by comparing ALIGN and RELATE scores. Doolittle *et al.* (8) have suggested that part of a gene for a factor X-like protein was duplicated and translocated in the genome, or that two such exchanges may have occurred; we are assuming that these

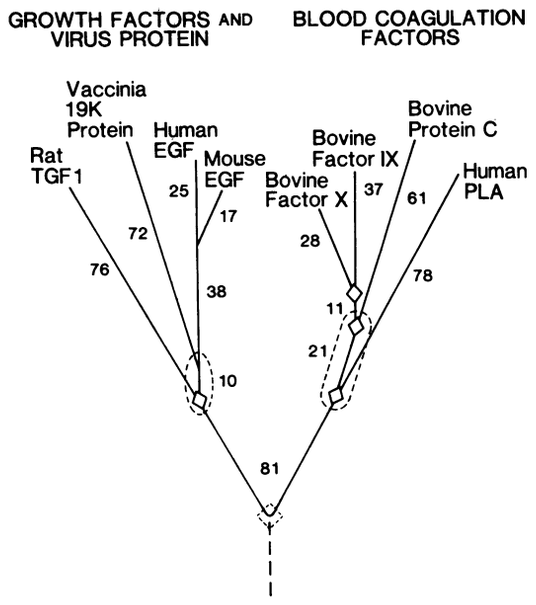

FIG. 4. Evolutionary tree for the 19-kDa protein, TGF1, EGF, and the EGF-like domain from the four clotting factors, derived from the sequence regions used in the alignment of Fig. 2; human factor IX and factor X were omitted. Branch lengths are given in PAMs. Dashed ovals indicate some uncertainty in the order of branching (see text). The trunk of the tree was arbitrarily positioned. Diamonds represent possible gene duplications.

events (represented by the dotted diamond at the "base" of the tree) would have preceded the divergence of the lines leading to the 19-kDa protein and the two growth factors. The EGF gene, which now contains 10 homologous regions, would have elongated by tandem intragenic duplications (8). The tree of Fig. 4 is in agreement with these suggestions; however, it is based on the sequence of the active EGF only. None of the nine other homologs in the EGF precursor has the complete conserved patterns, and some or all of the elongation of the EGF gene may have occurred after the divergence of the 19-kDa protein branch. Probably the active EGF has the most ancient and highly conserved sequence, correlated with its function of binding to and thus activating its receptor protein.

The above analyses are evidence that vaccinia virus has a protein related to two growth factors and, more distantly, to a region in the light chains of four clotting factors produced by mammalian cells. The sequence of residues 38–91 of the vaccinia 19-kDa protein appears to be related to those of TGF1 and EGF (active factors) to approximately the same degree that EGF and TGF1 are related to each other. However, the demonstration of these relationships raises several questions that await experimental investigation. The sequence conservation, especially of the 6 cysteine residues (which in EGF form three disulfide bonds whose presence is correlated with the activity of the factor), and the potential similarities in secondary structure and hydrophobic regions among these proteins suggest that some features of functional similarity could also still exist. Could the activity of the 19-kDa protein be dependent on a protein-binding capability, which might aid in activating a viral enzyme involved in viral replication or might affect some normal metabolic process of the host cell to the advantage of viral replication? The effects of some growth factors, including EGF and TGF1, appear to result from a synergistic sequential activity by two or more factors (48); could the 19-kDa protein act in combination with host cell factors present in the infected cell? The mechanism by which the vaccinia virus came to possess a gene (or domain) coding for an EGF-like protein is unknown but

would be expected to be unlike the genetic relationship found between other kinds of mammalian viruses and their hosts. The terminal inverted repetition within which the gene for the 19-kDa protein lies is somewhat unstable; deletions of segments containing protein-coding genes, reiteration of tandem repeating units, and recombination are reported to occur (49). Vaccinia virus is thought to duplicate in the cytoplasm of the host cell, using viral enzymes contained in the infectious particle (20), but many details of the replication process are not known. Vaccinia virus is still of medical interest because of its potential use as a live vaccine for other diseases (50); more information about its many proteins may lead to better understanding of their functions in relation to host cell metabolism.

This research was supported by Grants RR-01821, GM-08710, and HD-09547 from the National Institutes of Health and by Contract NASW 3954 from the National Aeronautics and Space Administration.

1. Marquardt, H., Hunkapiller, M. W., Hood, L. E., Twardzik, D. R., De Larco, J. E., Stephenson, J. R. & Todaro, G. J. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 4684–4688.
2. Marquardt, H., Hunkapiller, M. W., Hood, L. E. & Todaro, G. J. (1984) *Science* **223,** 1079–1082.
3. Scott, J., Urdea, M., Quiroga, M., Sanchez-Pescador, R., Fong, N., Selby, M., Rutter, W. J. & Bell, T. I. (1983) *Science* **221,** 236–240.
4. Savage, C. R., Jr., Inagami, T. & Cohen, S. (1972) *J. Biol. Chem.* **247,** 7612–7621.
5. Gregory, H. & Preston, B. M. (1977) *Int. J. Pept. Protein Res.* **9,** 107–118.
6. IUPAC-IUB Commission on Biochemical Nomenclature (1968) *J. Biol. Chem.* **243,** 3557–3559.
7. Young, C. L., Barker, W. C., Tomaselli, C. M. & Dayhoff, M. O. (1979) in *Atlas of Protein Sequence and Structure*, ed. Dayhoff, M. O. (National Biomedical Research Foundation, Washington, DC), Vol. 5, Suppl. 3, pp. 73–93.
8. Doolittle, R. F., Feng, D. F. & Johnson, M. S. (1984) *Nature (London)* **307,** 558–560.
9. Kurachi, K. & Davie, E. W. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 6461–6464.
10. Katayama, K., Ericsson, L. H., Enfield, D. L., Walsh, K. A., Neurath, H., Davie, E. W. & Titani, K. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 4990–4994.
11. McMullen, B. A., Fujikawa, K., Kisiel, W., Sasagawa, T., Howald, W. N., Kwa, E. Y. & Weinstein, B. (1983) *Biochemistry* **22,** 2875–2884.
12. Enfield, D. L., Ericsson, L. H., Fujikawa, K., Walsh, K. A., Neurath, H. & Titani, K. (1980) *Biochemistry* **19,** 659–667.
13. Titani, K., Fujikawa, K., Enfield, D. L., Ericsson, L. H., Walsh, K. A. & Neurath, H. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 3082–3086.
14. Fernlund, P. & Stenflo, J. (1982) *J. Biol. Chem.* **257,** 12170–12179.
15. Stenflo, J. & Fernlund, P. (1982) *J. Biol. Chem.* **257,** 12180–12190.
16. Pennica, D., Holmes, W. E., Kohr, W. J., Harkins, R. N., Vehar, G. A., Ward, C. A., Bennett, W. F., Yelverton, E., Seeburg, P. H., Heyneker, H. L., Goeddel, D. V. & Collen, D. (1983) *Nature (London)* **301,** 214–221.
17. Garcia-Pardo, A., Pearlstein, E. & Frangione, B. (1983) *J. Biol. Chem.* **258,** 12670–12674.
18. Pierschbacher, M. D., Ruoslahti, E., Sundelin, J., Lind, P. & Peterson, P. A. (1982) *J. Biol. Chem.* **257,** 9593–9597.
19. Petersen, T. E., Thogersen, H. C., Skorstengaard, K., Vibe-Pedersen, K., Sahl, P., Sottrup-Jensen, L. & Magnusson, S. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 137–141.
20. Venkatesan, S., Gershowitz, A. & Moss, B. (1982) *J. Virol.* **44,** 637–646.
21. Weir, J. P., Bajszar, G. & Moss, B. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 1210–1214.
22. Wittek, R., Cooper, J. A., Barbosa, E. & Moss, B. (1980) *Cell* **21,** 487–493.
23. Hopp, T. P. & Woods, K. R. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 3824–3828.
24. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157,** 105–132.
25. Chou, P. Y. & Fasman, G. D. (1978) *Annu. Rev. Biochem.* **47,** 251–276.
26. Dayhoff, M. O., Barker, W. C. & Hunt, L. T. (1983) *Methods Enzymol.* **91,** 524–545.
27. Dayhoff, M. O., Schwartz, R. M. & Orcutt, B. C. (1979) in *Atlas of Protein Sequence and Structure*, ed. Dayhoff, M. O. (National Biomedical Research Foundation, Washington, DC), Vol. 5, Suppl. 3, pp. 345–352.
28. Schwartz, R. M. & Dayhoff, M. O. (1979) in *Atlas of Protein Sequence and Structure*, ed. Dayhoff, M. O. (National Biomedical Research Foundation, Washington, DC), Vol. 5, Suppl. 3, pp. 353–358.
29. George, D. G., Yeh, L.-S. L. & Barker, W. C. (1983) *Biochem. Biophys. Res. Commun.* **115,** 1061–1068.
30. Barker, W. C. & Dayhoff, M. O. (1983) *Proceedings of the Seventh Annual Symposium on Computer Applications in Medical Care* (IEEE Computer Soc., Silver Spring, MD), pp. 584–588.
31. Maizel, J. V., Jr., & Lenk, R. P. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 7665–7669.
32. Dayhoff, M. O. (1976) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **35,** 2132–2138.
33. Schwartz, R. M. & Dayhoff, M. O. (1978) *Science* **199,** 395–403.
34. Fitch, W. M. & Margoliash, E. (1967) *Science* **155,** 279–284.
35. Savage, C. R., Jr., Hash, J. H. & Cohen, S. (1973) *J. Biol. Chem.* **248,** 7669–7672.
36. Komoriya, A., Hortsch, M., Meyers, C., Smith, M., Kanety, H. & Schlessinger, J. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 1351–1355.
37. Schechter, Y., Hernaez, L., Schlessinger, J. & Cuatrecasas, P. (1979) *Nature (London)* **278,** 835–838.
38. Schreiber, A. B., Yarden, Y. & Schlessinger, J. (1981) *Biochem. Biophys. Res. Commun.* **101,** 517–523.
39. Jackson, C. M. & Nemerson, Y. (1980) *Annu. Rev. Biochem.* **79,** 765–811.
40. Anson, D. S., Choo, K. H., Rees, D. J. G., Giannelli, F., Gould, K., Huddleston, J. A. & Brownlee, G. G. (1984) *EMBO J.* **3,** 1053–1060.
41. Gunzler, W. A., Steffens, G. J., Otting, F., Kim, S.-M. A., Frankus, E. & Flohe, L. (1982) *Hoppe-Seyler's Z. Physiol. Chem.* **363,** 1155–1165.
42. Holladay, L. A., Savage, C. R., Jr., Cohen, S. & Puett, D. (1976) *Biochemistry* **15,** 2624–2633.
43. Nicol, D. S. H. W. & Smith, L. F. (1960) *Nature (London)* **187,** 483–485.
44. Steiner, D. F., Clark, J. L., Nolan, C., Rubenstein, A. H., Margoliash, E., Aten, B. & Oyer, P. E. (1969) *Recent Prog. Horm. Res.* **25,** 207–282.
45. Rinderknecht, E. & Humbel, R. E. (1978) *J. Biol. Chem.* **253,** 2769–2776.
46. Marquardt, H., Todaro, G. J., Henderson, L. E. & Oroszlan, S. (1981) *J. Biol. Chem.* **256,** 6859–6865.
47. Hudson, P., Haley, J., Cronk, M., Shine, J. & Niall, H. (1981) *Nature (London)* **291,** 127–131.
48. Kaplan, P. L. & Ozanne, B. (1983) *Cell* **33,** 931–938.
49. Moss, B., Winters, E. & Cooper, N. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 1614–1618.
50. Mackett, M., Smith, G. L. & Moss, B. (1984) *J. Virol.* **49,** 857–864.