Internal duplication and sequence homology in factors V and VIII
(coagulation/cofactors)

DAVID N. FASS*, RODNEY M. HEWICK†, GAYLOR J. KNUTSON*, MICHAEL E. NESHEIM*,
AND KENNETH G. MANN*

*Section of Hematology Research, Mayo Clinic/Foundation, Rochester, MN 55905; and †Genetics Institute, Boston, MA 02115
Communicated by Oscar D. Ratnoff. November 26, 1984

ABSTRACT Blood coagulation factors V and VIII each serve cofactor functions with different vitamin K-dependent serine proteases of the coagulation cascade. Physical, physiologic, and kinetic data suggest analogous structures and functions for these two proteins. Proteolytically activated factor V (factor Va) is required for the efficient production of thrombin from prothrombin by factor Xa. Similarly, activated factor VIII (factor VIIIa) performs its cofactor activity with factor IXa to produce the activated form of factor X (factor Xa). The studies reported here on the sequences from the thrombin-activated and unactivated cofactors provide evidence that factor V and factor VIII are chemically related and that the structures of both cofactors involve some tandem duplication.

Factor V and factor VIII have a formal similarity with respect to their position in the coagulation cascade. Kinetic studies indicate that in complexes of Ca²⁺ phospholipid and vitamin K-dependent coagulant proteases, the cofactors significantly enhance the Vₘₐₓ of the proteolytic activation of either prothrombin, in the case of factor V (1), or of factor X in the presence of factor VIII (2, 3). It has been pointed out that, in addition to their analogous roles as cofactors to two vitamin K-dependent serine proteases, factor V and factor VIII share several other characteristics (4, 5). They both are activated proteolytically by the enzyme thrombin; they are inactivated by the vitamin K-dependent serine protease, activated protein C (6, 7); and both appear to have a divalent cation requirement to maintain a productive association between different portions of the molecule (8–11). As more studies are reported dealing with the structure of factor VIII, comparisons with structural data on factor V further reinforce the similarities between these two proteins. The products of thrombin activation, factor Va and factor VIIIa, are similar in size (5, 11, 12), and (from consideration of Stokes radii, mass, and sedimentation coefficient) both cofactors are apparently highly asymmetrical (4, 5). In our studies of porcine factor VIII and bovine factor V, we isolated proteolytic fragments of these molecules and subjected the peptides to NH₂-terminal amino acid sequence analysis. Our results (this paper and in ref. 13 with colleagues) suggest that the analogies enumerated above are the result of some true molecular homology.

EXPERIMENTAL

Bovine factor V (Fig. 1) was isolated as a single 330-kDa chain by procedures developed in these laboratories (14). Activation with bovine thrombin produces intermediates of 205 and 150 kDa, which associate in an EDTA-sensitive manner (12). Further thrombin cleavage produces a 94-kDa (D) peptide from the 150-kDa intermediate and a 74-kDa chain from the 205-kDa intermediate. These factor Va products of 94 and 74 kDa remain linked together through divalent cation binding (9, 10). The bovine factor V thrombin-generated peptides were isolated by ion-exchange procedures carried out in 0.005 M EDTA (15).

The porcine factor VIII was isolated as three chains from an immobilized anti-porcine factor VIII mouse monoclonal antibody column (11). The constituent chains consisted of 166-, 130- and 76-kDa polypeptides as determined by NaDodSO₄/polyacrylamide gel electrophoresis (Fig. 1). When the anti-factor VIII agarose gel was loaded with porcine factor VIII, the 166- and 130-kDa chains were released by treatment of the immobilized immune complex with EDTA under conditions previously described (11). The 76-kDa chain was retained and was obtained by subsequent elution with 50% ethylene glycol (11). The treatment of porcine factor VIII with thrombin results in the loss of the 166-, 130-, and 76-kDa peptides with the appearance of new chains at 82 and 69 kDa (11, 16). With longer incubation in the presence of thrombin, there appear, in the digestion mixture, new polypeptide chains at 44 and 35 kDa concurrent with the loss of the 82-kDa chain (16). The 166-, 130-, 82-, 76-, and 69-kDa peptides were isolated from radioactive bands cut from NaDodSO₄/polyacrylamide gel slab electrophoresis gels that were used to analyze mixtures containing 99% unlabeled and 1% ¹²⁵I-labeled samples (17) of either porcine factor VIII or thrombin-activated porcine factor VIII. The acrylamide slices containing ¹²⁵I-labeled protein were subjected to electroelution and concentration (18), thereby providing samples suitable (0.1–1 nmol) for gas-phase amino acid sequence analysis.

Sufficient quantities of the factor V thrombin-generated...
peptides had been available to permit fluid-phase, spinning-
cup, automated Edman amino acid sequence-analysis, and
gas-phase studies. Some factor V sequences had been deter-
dined and reported previously (15), and these present re-
results are extensions of those studies as well as the data of
Canfield et al. (19).

The factor VIII data are the result of sequence determina-
tions on four preparations of the protein containing approxi-
ately 1320, 280, 800, and 1200 units of factor VIII, on one
preparation of 30 units that was only analyzed for the se-
quence of the 69-kDa peptide, and on one preparation of 180
units that was analyzed for the sequence of the 76-kDa pe-
ptide. Factor VIII concentration initially was estimated from
absorbance at 280 nm by assuming an extinction coefficient
ε\text{\%} of 10. From these considerations and a proposed mo-
lecular mass of 240 kDa (11), a working relationship of 0.83
pmol per unit was established. Because of the scarcity of
factor VIII, only one destructive mass estimate was carried out
during an amino acid analysis; from recovery data, the
value of 1.15 pmol per unit was established. This latter value
has been adopted inasmuch as the unhydrolyzed remainder of
the amino acid analysis sample was subsequently se-
quenced and provided our first yield data. Based on a calcu-
lated value of 1.15 pmol of factor VIII per unit, one sequence
determination each for the 166-, 130-, and 82-kDa chains and
drivers on the 69-kDa thrombin-generated pep-
tide gave between 50% and 60% yield at the second sequenc-
ing cycle.

RESULTS AND DISCUSSION

Limited amino-terminal sequences had been found for the
330-kDa intact factor V, the 150-kDa intermediate, and the
94-kDa thrombin-generated product. These were identical
(15), indicating their position in the NH\text{\textsubscript{2}} termi-
ns of the intact cofactor. The 74-kDa thrombin-generated product had a
similar, but unquestionably distinct, sequence from that
found at the amino terminus of factor V (19). The 205-kDa
intermediate has an amino-terminal sequence distinct from
that found for any other polypeptide, suggesting that the 74-
kDa chain is not the amino terminus of the 205-kDa interme-
diate (15). The existence of a 120-kDa reaction product that
is liberated by thrombin treatment of the 205-kDa interme-
diate and that bears the same amino-terminal amino acid se-
quency as in the 205-kDa intermediate supports the conclu-
sion that the 74-kDa chain occupies a carboxyl-terminal lo-
cation (15) in the 205-kDa fragment (Fig. 1). In addition, the
74-kDa chain is released intact from the clot by a single cleavage catalyzed by Russell's Viper venom factor V activ-
ator, RVV-V (19).

Further gas-phase sequence analysis of 1 nmol of bovine
factor V, as a partially thrombin-activated product, extended
the number of identified residues. Two electrophoretic pep-
tide separations followed by electroelution and concentra-
tion (18) provided material for two sequence determinations
on the 94-kDa peptide and two determinations on the 74-kDa
thrombin-cleavage product. Each determination was per-
duced on 400 pmol of peptide (nominal). The bovine factor
V 94-kDa peptide yielded sequence information to residue
29, thus increasing the identified residues by 18 (two posi-
tions in parentheses were not interpretable). The amino-
ter-
inal sequence of this peptide, in the usual single-letter code
for amino acids (20), is:

\text{AKLRQFYVAASQRWNYR(PESTHL(\text{PKP})}

where the underlined sequence had been reported previously
(15).

The 74-kDa bovine factor V peptide produced sequence
information to residue 25, which is 16 residues beyond those
that had been reported previously. The sequence determined
was:

\text{SN\text{\texttt{G}}\text{\texttt{N}}\text{\texttt{R}}\text{\texttt{K}}\text{\texttt{Y}}\text{\texttt{Y}}\text{\texttt{I}}\text{\texttt{A}}\text{\texttt{E}}\text{\texttt{E}}\text{\texttt{I}}\text{\texttt{S}}\text{\texttt{W}}\text{\texttt{D}}\text{\texttt{Y}}\text{\texttt{S}}\text{\texttt{K}}\text{\texttt{F}}\text{\texttt{V}}\text{\texttt{Q}}},

where the underlined sequence had been reported previously
(19).

Sequence analysis of the 166-, 130-, and 82-kDa peptides of
porcine factor VIII was consistent with the three chains
being identical, implying that the 130- and 82-kDa chains are
derived from the amino terminus of the 166-kDa chain (Fig.
1). The sequence obtained as a consensus of the three deter-
minations was:

\text{(S)IRRYGAVELSWD YRQS ELLR ELHVDTRFP A},

with the parentheses indicating tentative identification. The
thrombin-cleavage product of the 76-kDa chain, which mi-
grates as a 69-kDa band in NaDodSO\text{\textsubscript{4}}/polyacrylamide gel
electrophoresis, has a sequence similar to, but distinct from,
that found for the 166-, 130-, and 82-kDa peptides. The se-
quence obtained for the 69-kDa peptide was:

\text{( FQKRTRHYFIAAVEQLWDYG MSES PRAL RNR AQNGE VPR}.

The 76-kDa peptide was obtained by electroelution of pro-
tein bands from cylindrical NaDodSO\text{\textsubscript{4}} electrophoresis gels.
The peptide was prepared for electrophoresis by lyophiliza-
tion from 0.02 M acetic acid. On three occasions no se-
quence was obtained for the 76-kDa peptide obtained in this
manner. One sample of the 76-kDa peptide was collected from
the immunoaffinity column after the heavy chains were
stripped with 0.020 M EDTA. The peptide was concentrated
and prepared for electrophoresis without exposure to acidic
conditions by lyophilization from 0.01% NaDodSO\text{\textsubscript{4}} in
water. The sequence obtained for the 76-kDa peptide was:

\text{(ISSL P(A)FQPEEDKMDY(D)D I F}.

Based on the initial quantity of factor VIII in this analysis,
the yield calculated at residue 2 (isoleucine) was 70.5%. The
76-kDa peptide is not blocked at the amino terminus, and
our original supposition, based on the acid-treated material,
that the 76-kDa chain represented the amino terminus of a puta-
tive single-chain factor VIII is, therefore, neither supported
nor contradicted. The similarities noted between the amin-
terminal sequences of the 94- and 74-kDa peptides of bovine
factor V\text{a}\text{\textsubscript{a}} and the similarities between the amino-terminal se-
quences of the 82- and 69-kDa peptides of the porcine factor
VIII can be seen in Fig. 2, which portrays these four chains.
In addition to these evidences of internal gene duplication, it
can be seen that, within a group of some 14 residues (num-
bered 3–16) in all four chains, there exists a strong homology
not only within the factor V and factor VIII structures but
also between the two cofactors. The amino acid sequences
within this region are identical at 11 positions in at least one
factor VIII chain and one factor V chain. The extent of this
homology is all the more striking in that there is a species
disparity in this analysis. In addition to the outright homolo-
gy, those residues not identical within this 14-residue region
are generally conservative substitutions that may have ar-
sen as single-base changes of other codons resident to those
positions. The characteristics of the factor V and factor VIII
regions of homology include a pair of aromatic residues at
positions indicated in Fig. 2 as 5 and 6 and a single aromatic

Biochemistry: Fass et al.

residue at 16. This last residue is a tyrosine characteristically found in a Trp-Asx-Tyr triplet at aligned residues 14, 15, and 16. This raises the question of whether the homology is more related to the nature of thrombin-cleavage sites than to an overall evolutionary homology of factors V and VIII. Inspection of fibrinogen α chains, as a thrombin substrate, shows a Trp-Asn-Tyr triplet at the same approximate position carboxyl to a thrombin-cleavage site. Prothrombin contains a site at the junction of fragment 1 and fragment 2 that is also thrombin sensitive (21, 22); however, no similar sequence is obvious in the first 30 residues of fragment 2 (carboxyl to the thrombin-sensitive site).

Comparison of the sequences numbered 3–16 in Fig. 2 shows that the factor VIII 166-, 130-, and 82-kDa sequences are 57% identical with the factor VIII 69-kDa sequence. The sequence of factor V 74-kDa chain peptide is 57% identical with that of the factor VIII 166-kDa amino terminus in this limited region. The factor V 94-kDa peptide sequence is only 36–43% identical with the sequence of each of the other three chains when compared individually but is 50% identical when measured against the aggregate sequences of the other three peptides (Fig. 2, shaded areas). Inspection of the sequences in the text determined for residues beyond those in Fig. 2 shows that the similarities among the chains disappear after residue 17 in Fig. 2. Because this study only compares approximately 2% of factor V sequences with 2% of factor VIII sequences, the data can only suggest a common precursor of both cofactors. Within the regions shown in Fig. 2, however, portions of factor VIII show a greater percentage of identity to portions of factor V than that seen between teleost and mammalian hemoglobin α chains (23). On the same basis within the limited areas of identity, factors V and VIII measured in pairwise combinations suggest that these small regions of the two molecules are at least as related as human hemoglobin α chain is to human hemoglobin β chain. The same can be said of the putative repeats within the respective cofactors (factor VIII, 166:69 kDa; factor V, 94:74 kDa).

To evaluate the probability of these similarities occurring by chance, random 14-residue amino acid sequences were generated without regard to natural amino acid abundance. These sequences were compared to the sequences of the residues 3 through 16 in Fig. 2. The composition of the 74-kDa factor V sequence matched the composition of the 166-kDa factor VIII sequence better than 96% of random compositions. When the sequences from factor V and factor VIII were compared residue by residue for identity, with no gaps involved, it was found that the number of identities between the cofactor peptides (8) was 7–11 standard deviations higher than the mean number of identities between the cofactor sequences and the random sequences. The statistical significance of the similar sequences occurring in two proteins that appear to perform analogous cofactor roles in analogous serine protease complexes and the fact that all of the similar sequences occur within 20 residues of a naturally occurring amino terminus is not readily apparent. We believe, however, that these ancillary observations contribute to the strong evidence that these similarities did not occur by chance. Clarification of evolutionary relationships awaits the sequence determination of the coagulation cofactors of additional species and more extensive sequence determination of porcine factor VIII and bovine factor V.

A structural, as well as functional, homology between factor V and factor VIII gives investigators of blood coagulation protein chemistry and protein structure the opportunity to study paired systems: the factor X-activating complex and the prothrombin-activating complex. In this pair the enzyme for the prothrombin-activating (factor Xa) complex is derived by proteolysis of the substrate (factor X) for the factor X-activating complex. Subtle differences between factor V and factor VIII may contribute to determining whether a factor X species is the enzyme or substrate for a proteolytic event. Because the primary structure of factor IX and factor X are already known, the identification of binding sites in the cofactors should be very informative. The data presented here represent a small percentage of the total sequence of the cofactors, and further work may or may not show additional homology. However, it should be enlightening to determine which regions of these very large cofactors are conserved.

This work was supported by Grants HL-17430, HL-16150, and HL-30365 from the National Institutes of Health, the Mayo Foundation, and Genetics Institute.