Frog integumentary mucin B.1 (FIM-B.1) contains various cysteine-rich modules. In the past, a COOH-terminal "cystine knot" motif has been found that is similar to von Willebrand factor; this region is generally known to be responsible for dimerization processes. Furthermore, a "complement control protein" motif is present as an internal cysteine-rich domain in FIM-B.1. We characterize here the missing 75% toward the NH2 terminus of the FIM-B.1 precursor by molecular cloning. Analogous to prepro-von Willebrand factor, four elements with considerable similarity to D-domains are present (i.e. D1-D2-D'-D3). These domains have been described as essential for the multimerization of von Willebrand factor. Thus, the general structure of FIM-B.1 resembles that of the human mucin MUC2 as well as prepro-von Willebrand factor; these three molecules at least seem to share common structural elements allowing similar multimerization mechanisms.

During phylogeny, mucus gels have been conserved as the essential extracellular matrices that protect delicate epithelial surfaces in many ways (1, 2). Mucins have been established as the molecules that primarily determine the defined rheological and viscoelastic properties of these gels. The key step in the formation of such a three-dimensional complex network is the ordered aggregation of linear rodlike monomeric mucins (3). The stiff and extended conformation of the monomers is the result of highly O-glycosylated repetitive serine/threonine-rich regions (4). In contrast, aggregation to multimers is achieved via cysteine-rich modules. Two models would describe such a network: (i) a cross-linked network model and (ii) an entangled network model. However, only the latter fulfills the physicochemical criteria that define a dynamic mucus gel (5).

Due to technical problems, complete molecular structures of mucins are still rare. For example, more than seven human MUC genes (6), bovine and porcine salivary mucins (7, 8) as well as three frog integumentary mucins (FIM-A.1, FIM-B.1, FIM-C.1) have been at least partially characterized (9). The latter represent typical extracellular mosaic proteins with astonishing structural similarities to other peptides and proteins (10). FIM-B.1 from Xenopus laevis certainly shows the most interesting molecular architecture. However, only about 25% from the COOH-terminal portion of the sequence has been reported thus far. In addition to a variable number of O-glycosylated type B repeats responsible for polydispersities (11), FIM-B.1 contains at least two different cysteine-rich modules: (i) internally, the "complement control protein motif" (CP, also known as "Shushi structure" or "short consensus repeat" (SCR) (12)) and (ii) a COOH-terminal region with homology to von Willebrand factor (vWF) (13). In vWF, this part is responsible for dimerization (14). A motif spanning 11 cysteine residues named "cystine knot" has been proposed as the active site of the latter (15) and is responsible for dimerization of certain cytokines as well (16). Subsequently, the cystine knot motif has also been found in a variety of other mucins, e.g. bovine salivary mucin (7), porcine salivary mucin (8), MUC2 (17), rMUC2/rMLP (18), and MUC5 (19) as well as the human sublingual gland mucin MG1 (20). Recently, dimerization of MUC2 has been reported (21), and for porcine salivary mucin it has been clearly shown that it forms dimers via its COOH-terminal domain (22). Thus, this motif is now considered to trigger homodimerization as an early event in the biosynthesis of many mucins.

We report here the full-length sequence of the FIM-B.1 precursor starting with the signal sequence as deduced from cDNA cloning.

EXPERIMENTAL PROCEDURES

Isolation of mRNA from the skin of a single adult X. laevis (purchased from the Herpetological Institute, Dr. W. de Rover, Belgium), cyclic thermal amplification via the polymerase chain reaction (PCR), and purification and sequencing of plasmid DNA as well as computerized analysis and homology searches have been described previously (12).

In order to elongate the incompletely known nucleotide sequence encoding the COOH-terminal portion of FIM-B.1 toward the 5'-end, a multistep amplification procedure (RACE protocol) has been employed (23). Starting from the region encoding the CP motif (12), the oligonucleotide SCR1 d(CACAGCTTGGTGTATTTC) was used as a specific primer for cDNA synthesis. After dC tailing, amplification occurred with Taq polymerase and a combination of oligonucleotides REP7 d(C- CCTCGAGAATTCGGATCCTGCTACCGTTCCGTTT) and PCR5' d(CGGATCCCTCGAGAATTCTAGAGG(6)GA). The underlined region is complementary to part of the CP motif in FIM-B.1 (12). After subcloning the products into the BamHI/EcoRI sites of pBluescript-II KS-(Stratagene), clone p5Sr7-2 was obtained. Further cDNA clones were generated in a similar way by a multistep amplification procedure using a set of specific primers toward the 5'-end (Fig. 1C).

Based on this sequence information obtained from relatively short cDNA clones, long overlapping cDNA clones were generated by PCR and subsequently analyzed (Fig. 1B).

RESULTS

Fig. 2 represents the cDNA sequence obtained from a set of overlapping clones using the RACE protocol toward the very 5'-end. The deduced amino acid sequence encodes the aminoterminal portion of the FIM-B.1 precursor, starting with the
signal sequence until it reaches the CP motif (which served as the anchor for the first specific oligonucleotide used). However, the CP motif cloned here does not show the identical sequence as characterized previously (12). The two CP motifs differ in precisely two point mutations also changing two amino acid residues: K to E and A to G. These two mutations have been confirmed to be highly specific by the analysis of a series of independent cDNA clones. In order to distinguish between these two CP motifs, we designated them SCR (12) and SCR* (sequence from Fig. 2). Thus, we assumed that the FIM-B.1 precursor could theoretically contain at least two CP motifs that differ slightly.

To test this hypothesis, oligo(dT)-primed cDNA from X. laevis skin was amplified with Taq polymerase using the oligonucleotides FIM8 d(CCCGGATCCTCGAGAATTCAAATCAAGCTATAACAG) and SCR4 d(CCCGGATCCGCACAACCTCTT). The underlined part in FIM8 represents positions 3979–3995 from Fig. 2, and SCR4 is complementary to the SCR motif (12) and does not recognize SCR*. After subcloning the PCR products into the BamHI/EcoRI sites of pBluescript- II/SK, clones pF8S4.2–5, -7, and -8 were characterized (Fig. 3). All three clones indeed contained two different CP motifs (i.e. SCR* and SCR). However, the clones were not identical but differed in their repetitive parts by specific insertions and deletions. Such polydispersities are typical of FIM-B.1 and have been shown to result from alternative splicing of repetitive cassettes (11).

**DISCUSSION**

The combined amino acid sequences deduced in Figs. 2 and 3 now complete the missing amino-terminal portion of FIM-B.1. Together with the published COOH-terminal part (12, 13), the FIM-B.1 precursor consists of at least about 2700 amino acid residues (Fig. 4) encoded by a polydisperse mRNA population with a length of more than 8.3 kilobases. This is in fairly good agreement with Northern blot analysis which revealed a smear of up to 10 kilobases (24). As indicated in Fig. 4, the difference is probably due to the existence of a polydisperse cluster of additional CP motifs and repetitive highly O-glycosylated regions (25). In particular, multiple CP motifs could represent potential anchor points that non-covalently cross-link mucin subunits.

The amino-terminal portion of the FIM-B.1 precursor presented here can be clearly divided into separated domains. As is typical of secretory proteins, the sequence starts with a hydrophobic signal sequence that is probably cleaved off after alanine 19 (Fig. 2). Then a mainly basic repetitive region follows with the motif PAKGG. This glycine-rich (until gly-
cine-77) sequence a β-turn structure can primarily be expected. Similar terminal sequences have been detected in cytokeratins (26) and synapsins (27). Starting with proline 78, the pattern changes drastically to a threonine-rich sequence also containing proline and alanine. Such a composition is typical of mucins (2, 4); however, the acid residues flanking some threonine residues at positions 211 probably diminish their potential to become O-glycosylated (28). Similarly, as shown previously for type B repeats (12), analysis of further cDNA clones revealed polydispersities by insertion of a variable number of tandem repeats with the motif PAATDSET after amino acid 122 (25). Thus, the sequence given in Fig. 2 represents a minimal length variant within a polydisperse population.

Certainly one of the most interesting domains in FIM-B.1 is the cysteine-rich region between positions 172 and 1330 (Fig. 2) because it reveals pronounced similarities with pro-vWF (29). In particular, three subdomains with internal homology (named D1, D2, and D3) as well as a truncated version located between D2 and D3 (designated as D9) can be recognized. This set of D-domains has been reported to be obligatory for multimer assembly of pro-vWF (30). This biosynthetic event occurs unusually late in trans-Golgi and post-Golgi acidic compartments (30) and seems to be independent of dimerization in the endoplasmic reticulum (31). Furthermore, multimerization via the D1 and D2 domains plays an important role in storage granule formation (32). Small vWF multimers are secreted constitutively, whereas large multimers are packed into Weibel-Palade bodies and then released via the regulated pathway (30). An analogous domain structure (as in vWF and FIM-B.1) has also been reported for the amino-terminal part of MUC2.

---

**FIG. 2.** Nucleotide sequence and deduced amino acid sequence of the amino-terminal portion of the FIM-B.1 precursor. This sequence was compiled from the DNA cloned shown in Fig. 1B, which were all obtained from a single individual. Two point mutations were observed in clone pF18F38–3 in its overlapping region with pF30F56–4. The T at position 1950 is changed to C (Ile→Thr mutation on amino acid level) and at position 2147 the G is changed to A (Asp→Asn mutation). Restriction sites are marked. The presumed signal sequence of the precursor, as well as potential N-glycosylation sites, is underlined. Also denoted are the cysteine-rich domains homologous to vWF (D1-D2-D9-D3) and the CP motif.
which also forms multimers (34)). As shown in Fig. 5, nearly all cysteine residues are conserved in these three molecules. However, the general similarity of the sequences is not particularly pronounced. The two most conserved continuous stretches of amino acid residues are regions in the D1 and the D3 domain with the sequences TCGLCG and VCGLCGN, respectively. Remarkably, the vicinal cysteine residues in the underlined CGLCG motifs are similar to those at the active site of disulfide isomerase, and they have been proposed to play a role in multimerization of pro-vWF (35). In the mature vWF (after cleavage of its pro-sequence (i.e. at the D2/D3 junction; see Fig. 5)), homophilic intersubunit disulfide bonds have been determined within the D3 domain at positions Cys-379 (36), Cys-459, Cys-462, and Cys-464 (37). The homologous cysteine residues are conserved in FIM-B.1 and MUC2 (indicated by triangles in Fig. 5). Whether proteolytic processing of the FIM-B.1 precursor occurs similarly to pro-vWF is not known yet. A potential cleavage site would be next to the D2/D3 junction between positions 888 and 889 of the FIM-B.1 precursor (sequence SRKR\_T; Fig. 5) liberating a polydisperse mucin-like

Fig. 3. Molecular analysis of cDNA clones containing two CP motifs in FIM-B.1. PCR-generated clones pFS8 4.2–5, -7, and -8 were obtained from a single individual. These clones were compared with the corresponding sequence from pSS26–3 shown in Fig. 2. The two CP motifs detected are denoted as SCR and SCR*. Furthermore, polydispersities in the repetitive part are indicated by stars (representing gaps). Bars denote identical nucleic acid residues.
Pro-peptide. This sequence is close to the equivalent position in pro-vWF and also remarkably resembles the known processing site in the vWF precursor (sequence RSKR | S; Fig. 5). It is noteworthy that proteolytic cleavage of pro-vWF is not essential for multimer formation (38). Taken together, many mucins seem to mimic the covalent stepwise aggregation of vWF to linear clusters. Molecular structures supporting such a model are now available for MUC2 (33), rMUC2 (39), FIM-B.1, and obviously also porcine salivary mucin (22). Furthermore, partial sequences of MUC5 (19), bovine salivary mucin (7), and MG1 (20) indicate that these mucins could follow the same common hypothetical scheme. Also, the sperm membrane protein zonadhesin (40) containing a mucin-like domain and a cluster of D-domains would be a candidate for a similar molecular mechanism. However, based on the observation that vWF D-domains bind heparin (41), non-covalent interactions of mucin D-domains with sulfated carbohydrates should also be taken into consideration. Such a lectin bond-mediated polymerization model has already been proposed in the past for mucus gels (42).

Acknowledgments—We thank U. Schimanko and C. Cap for oligonucleotide synthesis.

Fig. 5. Homologous domains in FIM-B.1, prepro-vWF, and MUC2. The region spanning positions 172-1350 in FIM-B.1 (from Fig. 2) is compared with the D1-D2-D9-D3 domains of prepro-vWF (29) and the amino-terminal part of MUC2 (33). Gaps are introduced to maximize similarity. Identical amino acid residues in prepro-vWF and MUC2 (when compared with FIM-B.1) are enclosed in boxes. The cleavage site in prepro-vWF is indicated by an arrow, as well as a potential processing site in FIM-B.1. Triangles indicate cysteine residues probably involved in homophilic intermolecular disulfide bridges.
REFERENCES

1. Forstner, J. F., and Forstner, G. G. (1994) in Physiology of the Gastrointestinal Tract (Johnson, L. R., Alpers, D. H., Christensen, J., Jacobson, E. D., and Walsh, J. H., eds) Vol. 2, 3rd Ed., pp. 1553–1583, Raven Press, Ltd., New York
2. Strous, G. J., and Dekker, J. (1992) Crit. Rev. Biochem. Mol. Biol. 27, 57–92
3. Carlstedt, I., and Sheehan, J. K. (1984) in Ciba Foundation Symposium 109 on Mucus and Mucous (Nugent, J., and O'Connor, M., eds) pp. 157–166, Pitman, London
4. Jentoft, N. (1990) Trends Biochem. Sci. 15, 291–294
5. Verdugo, P. (1990) Annu. Rev. Biochem. 59, 157–176
6. van Klinken, B. J.-W., Dekker, J., Bu¨ller, H. A., and Einerhand, A. W. C. (1995) J. Biol. Chem. 269, 9678–9686
7. Bhargava, A. K., Woitach, J. T., Davidson, E. A., and Bhavanandan, V. P. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, 6798–6802
8. Eckhardt, A. E., Timpte, C. S., Abernethy, J. L., Zhao, Y., and Hill, R. L. (1991) J. Biol. Chem. 266, 2821–2827
9. Hoffmann, W., and Auwerx, J. (1993) Comp. Biochem. Physiol. 105B, 465–472
10. Hoffmann, W., and Joba, W. (1995) Biochem. Soc. Trans. 23, 805–810
11. Joba, W., and Hoffmann, W. (1996) Glycoconj. J. 13, 735–740
12. Probst, J. C., Hauser, F., Joba, W., and Hoffmann, W. (1992) J. Biol. Chem. 267, 6310–6316
13. Probst, J. C., Gertz, E.-M., and Hoffmann, W. (1990) Biochemistry 29, 6240–6244
14. Voorberg, J., Fontijn, R., Calafat, J., Janssen, H., van Mourik, J. A., and Pannekoek, H. (1991) J. Cell Biol. 113, 195–205
15. Meitinger, T., Meinl, A., Bork, P., Rest, B., Sander, C., Haasemann, M., and Murken, J. (1993) Nat. Genet. 5, 376–380
16. McDonald, N. Q., and Hendrickson, W. A. (1993) Cell 73, 421–424
17. Gum, J. R., Hicks, J. W., Toribara, N. W., Rothe, E.-M., Lague, R. E., and Kim, Y. S. (1992) J. Biol. Chem. 267, 21375–21383
18. Xu, G., Huan, L.-J., Khatri, I. A., Wang, D., Bennick, A., Fahim, R. E. F., Forstner, G. G., and Forstner, J. F. (1992) J. Biol. Chem. 267, 5401–5407
19. Lesuffleur, T., Roche, F., Hill, A. S., Lacassa, M., Fox, M., Swallow, D. M., Zweibaum, A., and Real, F. X. (1995) J. Biol. Chem. 270, 13665–13673
20. Troxler, R. F., Offner, G. D., Zhang, F., Iostcheva, I., and Oppenheim, F. G. (1995) Biochem. Biophys. Res. Commun. 217, 1112–1119
21. Asker, N., Baeckström, D., Axelson, M. A. B., Carlstedt, I., and Hansson, I. (1995) Biochem. J. 308, 873–880
22. Perez-Vilar, J., Eckhardt, A. E., and Hill, R. L. (1996) J. Biol. Chem. 271, 9845–9850
23. Frohman, M. A., Dush, M. K., and Martin, G. R. (1988) Proc. Natl. Acad. Sci. U. S. A. 85, 8998–9002
24. Hoffmann, W. (1988) J. Biol. Chem. 263, 7686–7690
25. Hofe, H. (1996) Molekulare Charakterisierung des Integument-Muzins FIM-B. 1 aus Xenopus laevis. Doctoral dissertation, Ludwig-Maximilians-Universität München (Fakultät für Chemie und Pharmazie), Munich, Germany
26. Hoffmann, W., Sterrer, S., and Konigstorfer, A. (1988) FEBS Lett. 237, 178–182
27. Sudhof, T. C., Czernecki, A. J., Kao, H.-T., Takei, T., Johnston, P. A., Heriuchi, A., Kanazir, S. D., Wagner, M. A., Perin, M. S., De Camilli, P., and Green, P. (1989) Science 245, 1474–1480
28. Nehrke, K., Hagen, F. K., and Tabak, L. A. (1996) J. Biol. Chem. 271, 7061–7065
29. Verweij, C. L., Diergaarde, P. J., Hart, M., and Pannekoek, H. (1986) EMBO J. 5, 1839–1847
30. Wagner, D. D. (1990) Annu. Rev. Cell Biol. 6, 217–246
31. Voorberg, J., Fontijn, R., van Mourik, J. A., and Pannekoek, H. (1990) EMBO J. 9, 797–803
32. Wagner, D. D., Saffaripeur, S., Bonfanti, R., Sadler, J. E., Cramer, E. M., Chapman, B., and Mayadas, T. N. (1991) Cell 64, 403–413
33. Gum, J. R., Hicks, J. W., Toribara, N. W., Siddiki, B., and Kim, Y. S. (1994) J. Biol. Chem. 269, 2440–2446
34. McCool, D. J., Forstner, J. F., and Forstner, G. G. (1994) Biochem. J. 302, 111–118
35. Mayadas, T. N., and Wagner, D. D. (1992) Proc. Natl. Acad. Sci. U. S. A. 89, 3531–3535
36. Dong, Z., Thoma, R. S., Crimmins, D. L., McCourt, D. W., Tuley, E. A., and Sadler, J. E. (1994) J. Biol. Chem. 269, 6753–6758
37. Azuma, H., Hayashi, T., Dent, J. A., Ruggeri, Z. M., and Ware, J. (1993) J. Biol. Chem. 268, 2821–2827
38. Verweij, C. L., Hart, M., and Pannekoek, H. (1988) J. Biol. Chem. 263, 7921–7924
39. Ohnori, H., Doehman, A. F., Gallup, M., Tsuda, T., Kai, H., Gum, J. R., Kim, Y. S., and Basbaum, C. B. (1994) J. Biol. Chem. 269, 17833–17840
40. Hardy, D. M., and Garbers, D. L. (1995) J. Biol. Chem. 270, 26925–26928
41. Meyer, D., and Girma, J.-P. (1993) Thromb. Haemostasis 70, 99–104
42. Silberberg, A. (1987) Biorheology 24, 605–614