

Full Length Research Paper

Amino Acid Sequence of α -Subunit in Hen Egg White
Ovomucin Deduced from Cloned cDNA*KENJI WATANABE^{a,†}, MAKOTO SHIMOYAMADA^b, TAKUO ONIZUKA^c, HIDEO AKIYAMA^c, MOTOHIRO NIWA^b,
TOMOE IDO^b and YOJI TSUGE^b^aFaculty of Agriculture, Tokyo University of Agriculture, Funako 1737, Atsugishi, Kanagawa 243-0034, Japan; ^bLaboratory of Food Material Engineering, Faculty of Agriculture, Gifu University, 1-1 Yanagido, Gifu 501-1193, Japan; ^cBiological Science Laboratories, Toray Research Center Inc, 1111 Teihiro, Kamakurashi, Kanagawa 248-8555, Japan

(Received 2 March 2004)

The primary amino acid sequence of α -subunit in ovomucin (OVM) from hen thick egg white was determined. The 2087 amino acid residues with a relative molecular mass of 230.9 kDa along the full length of the α -subunit were represented. The α -subunit contains domains, arranged from the N- to C-terminals in the following order: D1-D2-D'-D3-R (central region)-D4-C1-CK (Cystine-knot), in a manner similar to the arrangement of D, C and CK domains in human pre-pro-von Willebrand factor (hpp-vWF) and hMUC2. The α -subunit showed identities on amino acid sequences with hpp-vWF and hMUC2 at 33 and 41% in the N-terminal region and 30 and 38% in the C-terminal region, respectively. The numbers and positions of cysteine residues were highly conserved among α -subunit, hpp-vWF and hMUC2. However, R showed no virtual sequence homology with the corresponding regions in two proteins. It was estimated that α -subunit was not part of a large peptide of OVM, but was independently synthesized from β -subunit.

Keywords: Egg white ovomucin; α -Subunit; Cloning; Primary amino acid sequence

INTRODUCTION

Mucins are a heterogeneous family of high-molecular mass proteins broadly defined by their high content of carbohydrates (50–90%), which are mainly O-linked, but in some cases also N-linked. Mucins are known to be expressed in various epithelia (Forstner and Forstner, 1994). A vast

amount of structural information has been obtained on human mucins and those from other vertebrates. To date, at least 18 mucin genes have been identified: MUC1, -2, -3A, -3B, -4, -5AC, -5B, -6, -7, -8 (see Gendler and Spicer (1995) and Shankar *et al.*, 1997 for reviews), -9 (Lapensee *et al.*, 1997), -10 (Melnick *et al.*, 2001), -11, -12 (Williams *et al.*, 1999), -13 (Williams *et al.*, 2001), -14 (Melnick *et al.*, 2001), -15 (Pallesen *et al.*, 2002) and -16 (Yin and Lloyd, 2001). Mucins can be divided into at least two structurally distinct classes, the secreted (gel-forming or nongel-forming) type and the membrane-associated type.

Although the studies on the mucin genes from various sources have been reported, those from avian egg white have not been previously carried out. Ovomucin (OVM) is an egg white protein representing about 2–4% of the total egg albumen protein responsible for the gel-like properties of thick egg albumen, and is characterized by its high molecular weight and high carbohydrate content. OVM has been described as probably composed of extended glycoprotein bound by disulfide bonds between two subunits, that is, an α -subunit (apparent molecular mass ~220 kDa) containing 9–15% carbohydrate, and a β -subunit (apparent molecular mass ~400 kDa) containing 50–65% carbohydrate (Donovan *et al.*, 1970; Kato *et al.*, 1970; Hayakawa and Sato, 1977; Tsuge *et al.*, 1997a; Hiidenhovi *et al.*, 1999; Yokota *et al.*, 1999).

*The nucleotide and amino acid sequences described in this paper have been submitted to the DDBJ/EMBL/GenBank databases under accession number AB046524.

[†]Corresponding author. E-mail: watak@nodai.ac.jp

It is generally acknowledged that α -subunit contains a high proportion of *N*-linked oligosaccharides, while β -subunit contains a high proportion of *O*-linked oligosaccharides on a repeated domain rich in threonine and serine. Thus, if criteria such as model of secretion, rheological behaviour, biochemical composition and structure are considered, OVM gel in avian egg white may be considered as a member of the mucus family.

In laying hens, the oviduct consists of five portions: infundibulum, magnum (albumen-secreting portion), isthmus, uterus and vagina. Tubular glands in the magnum are responsible for the production of the majority of egg white proteins including OVM under the stimulatory control of estrogen.

We are now studying hen egg white OVM from the viewpoint of relationships between its molecular structure and function as a biologically and food-physiologically active substance (Tsuge *et al.*, 1997a,b; Watanabe *et al.*, 1998a,b). In the present study, we describe the cloning and sequencing of the cDNA on α -subunit in OVM, the complete amino acid sequence deduced from the cDNA, and the homologies with hpp-vWF, hMUC2 and the like.

RESULTS

The α -subunit N-terminal sequence was analyzed as Lys Glu Pro Val Gln Ile Val Gln (amino acids 1–8 in Fig. 1, which showed the nucleotide and amino acid sequences in full length determined as described below).

Southern blot analysis of a magnum cDNA library showed the presence of cDNA clones coding α -subunit at lengths of around 8.0 and 7.4 kb with a vector size of 4.1 kb in the Not-I-digested cDNA library (data not shown). Therefore, the molecular sizes of the cDNA clones obtained for α -subunit appeared to be around 3.9 and 3.3 kb, respectively. To determine the molecular size of the total coding α -subunit, northern blot analysis of mRNA prepared from magnum was carried out. The result showed a single band with a molecular size of around 7 kb (Fig. 2). Thus, the involved cDNA clones in the cDNA library for α -subunit were not complete and lacked part of the cDNA sequence.

The selected 3.9 kb cDNA contains a putative polyadenylation signal consensus, AATAAA, located 176 base pairs downstream of the translation stop codon, TGA and 15 base pairs upstream of the poly(A) (Fig. 1). These results indicated that the nucleotide and deduced amino acid sequences in the obtained 3.9 kb cDNA corresponded to the parts of 3'-terminal side from the nucleotide residue number 2666 (amino acid residue number 863) in Fig. 1. The deduced protein sequence in 3.9 kb cDNA contains 1225 amino acids.

On the basis of the determined nucleotide sequence on 3.9 kb cDNA, the cloning strategy for the α -subunit N-terminal region was arranged schematically as shown in Fig. 3. The specific primer (5' CTCATCC-CTTGAGTTGTTTCAATCAC) used for preparing cDNA in the N-terminal region was synthesized. cDNA synthesis was performed using its primer with the magnum mRNA followed by subsequent

```

1  TACCACCTCATGGCAAGATGGAATAAAAAAGGAGAGGTCAATTTGGATCTTCGCTCATTTGGAGTTTTGCAAGGGGAAAGAACAGTTTCAGATAGTCCAGG
-26  P P H G K M E I K K E R S P W I F C L I W S F C K G K E P V Q I V Q V
106 TCTCCACTGTTGGTAGGAGCGAGTGTACCACTGGGGCAATTTCCACATTTTCATACATTTGACCACGTTAAGTTCACCTTTCCCGAACCTGCACATACGTCTTTG
10  S T V G R S E C T T W G N F H F H T F D H V K F T F P G T C T Y V F A
      -> D1 domain
211 CTTCTCACTGCAATGACAGCTACAGGACTTCAACATCAAAATCAGACGAGTACACAAAACAGTCACCTGATCTACTTCACTGTCACTACTGATGGAGTGATCT
45  S H C N D S Y Q D F N I K I R R S D K N S H L I Y F T V T T D G V I L
316 TAGAGGTGAAGGAAACAGGCATCACAGTGAATGGGAACAGATCCCTCTGCCCTTCAGCTTGAAAGCATCCTGATTGAGGACACCTGTGCATACCTTCCAAGTCA
80  E V K E T G I T V N G N Q I P L P F S L K S I L I E D T C A Y F Q V T
421 CCTCCAAGCTGGGTCTGACGCTCAAGTGAACCTGGGCTGACACTCTCCTACTGGACCTGGAAGAAACATATAAGAGAAAAATATGTGGTCTGTGTGGGAACCTATG
115 S K L G L T L K W N W A D T L L L D L E E T Y K E K I C G L C G N Y D
526 ATGGCAATAAGAGAATGATTTAATTTTGGACGGATATAAAATGCACCCAAAGGAGCTTTGGGAACCTTTCACAAAGTTGAGGATCCATCAGAAAAGTGTCTGTATG
150 G N K K N D L I L D G Y K M H P R Q F G N F H K V E D P S E K C P D V
631 TGAGGCCAGATGACCACACTGGAAGACATCCTACAGAGGATGACAATAGATGCAGTAAATATAAAAAATGTGCAAGAACTTCTGTCCGATTGGAAACTCC
185 R P D D H T G R H P T E D D N R C S K Y K K M C K K L L S R F G N C P
736 CCAAAGTGGTGTGCTTTTIGATGATTATGTAGCTACCTGCAGTGGATATGTGCAACTGTGTGGTTAATTCTTCTCAATCTGATTGGTTTCCAGCTGCATATGCA
220 K V V A F D D Y V A T C T E D M C N C V V N S S Q S D L V S S C I C S
841 GCACTTTAAACCAATACTCTCGAGACTGTGTCTCAGCAAGGAGACCTGGGGAATGGAGAACCAAGAGCTTTGCTATCAGSAATGCCAAGCAACATGGAGT
255 T L N Q Y S R D C V L S K G D P G E W R T K E L C Y Q E C P S N M E Y
946 ACATGGAATGTGGAACTCCTGTGCTGACACGTGTGCTGACCCAGAAAGGAGTAAGATATGCAAGCCCCATGTACAGATGTTGCTTCTGTCTCTCTGGAACCTA
290 M E C G N S C A D T C A D P E R S K I C K A P C T D G C F C P P G T I
1051 TCCTTGATGACCTTGGTGGCAAAAAATGTGTTCAGAGACAGCTGTCCCTGTATGTTCAGGCAAGCTACTCATCTGGAGGGACTTACTCCACTCCTTGCC
325 L D D L G G K K C V P R D S C P C M F Q G K V Y S S G G T Y S T P C Q

```

1156 AAAACTGTACTTGCAGGAGGCCACTGGAGCTGTATCTCTGCTTGTCTGGAAGCTGCAGTATAGATGGAGGATTCCATATAAAAAACCTTTGATAACAAGA
 360 N C T C K G G E W S C I S L P C S G S C S I D G G F H I K T F D N K K
 ♦
 1261 AGTTCAATTTTCATGGTAACTGCCATTATGTCTTAGCTAAGAACACCCGATGACACATTTGTGGTGATTGGAGAGATCATCCAGTGTGGACATCAAGACAAATGA
 395 F N F E G N C H Y V L A K N T D D T F V V I G E I I Q C G T S K T M T
 1366 CTTGCTTGAAGATGTCTTAGCTACACTAGGAAGAACAACATAAAAACTGTCTCTGTGGAAGTATCTACATGAACAATTTTCATTGTGAAGCTCCAGTAAGCA
 430 C L K N V L V T L G R T T I K I C S C G S I Y M N N F I V K L P V S K
 1471 AAGATGGCATACCATCTTTAGACCTCAACATTTTTCATCAAAATCTTGAGCTCAGCTGGAGTACAGATTTCGAGTGCAAAATGAAACCTGTAATGCAGCTCTCCA
 465 D G I T I F R P S T F F I K I L S S A G V Q I R V Q M K P V M Q L S I
 1576 TAACAGTTGATCATCTTATCAGAATCGCACTTCTGGTCTTTGTGGAAATTTCAATAACATCCAGACTGACGATTTTCAAGACGCCACTGGAGCTGTGGAAGATT
 500 T V D E S Y Q N R T S G L C G N F N N I Q T D D F R T A T G A V E D S
 ♦
 1681 CAGCTGCTGCTTTTGGTAACTCTGGAAAACCCAGGCCACCTGTTTTCATCTCGAGGACACCTTTGAAGATCCTTGTCTCAACACTGTGGATAAAGAAAAATTTG
 535 A A A F G N S W K T R A S C F D V E D S F E D P C S N S V D K E K F A
 1786 CCCAACACTGGTGTCTCTTCTGTCTAATACAAGCAGTACGTTTGGCGCTGCCATTCTGTTGTAGACCCTTCTGTGTACATCAAGAGATGTATGTACGACACCT
 570 Q H W C A L L S N T S S T F A A C H S V V D P S V Y I K R C M Y D T C
 ♦
 1891 GTAATGCTGAGAAGAGTGAAGTTGCAGTGTGTAGCGTGTCTCTACTTACTCCAGAGACTGTGCTGCAGCAGGCATGACCTTAAAGGCTGGAGGCAGGGGATCT
 605 N A E K S E V A L C S V L S T Y S R D C A A A G M T L K G W R Q G I C
 1996 GTGATCCCTCAGAGGAGTGTCTGAGACTATGGTTTATAACTATAGCGTGAAGTACTGCAACCACTTGTCCGATCGCTGGATGAACCTGATCCCTGTGTAAAG
 640 D P S E E C P E T M V Y N Y S V K Y C N Q S C R S L D E P D P L C K V
 ♦
 2101 TTCAGATTGCTCCTATGAGGGGTGCGGTTGCCCGAAGGGACCTACCTCAATGACGAGGAGGAATGTGTGACTCCAGATGACTGTCCCTGCTACTACAAGGCA
 675 Q I A P M E G C G C P E G T Y L N D E E E C V T P D D C P C Y Y K G K
 2206 AGATTGTTTCAAGCTTTCCTTCCAAAGAACAACTGTGTGCAAAATGCATTCAAGGAAGATTAGACTGCATTGAGAACTGTCTGTTAAAGATTGCC
 710 I V Q P G N S F Q E D K L L C K C I Q G R L D C I G E T V L V K D C P
 ♦
 2311 CAGCTCCTATGTACTACTTCAACTGCAGCTCAGCGGGTCTGGTGCCTATTGGTTCGAGTGTGAGAAGAGCTGTAAGACACAGGATATGCACTGCTATGTCACAG
 745 A P M Y Y F N C S S A G P F A I G S E C Q K S C K T Q D M H C Y V T E
 ♦
 -> D' domain
 2416 AATGTGCTCTGGCTGCATGTGCTCCTGATGGGCTGGTGTGGATGGCAGTGGGGATGTATTCCCAAGACCAATGCCCGTGTGCTCATGGAGACACTTTTATA
 780 C V S G C M C P D G L V L D G S G G C I P K D Q C P C V H G G H F Y K
 2521 AACCTGGTGAACCATCAGAGTCGACTGCAACACATGCACCTGCAACAAAAGACAGTGGAACTGCACAGACAATCCCTGCAAGGGAACCTGCATGTGTACGGCA
 815 P G E T I R V D C N T C T C N K R Q W N C T D N P C K G T C T V Y G N
 ♦
 2626 ATCGGCACTACATCAGCTTCGATGGGAGAAAGTTTGAATTCCTGGGAGACTGTGACTATATCCTAGCTCAGGATTTTGTCCCAATAACATGGATGCTGGCACCT
 850 G H Y M S F D G E K F D F L G D C D Y I L A Q D F C P N N M D A G T F
 ♦
 2731 TCCGAATTTGAATTCAGAACAATGCCCTGTGGGAAGTCACCTCTCCATCTGCTCCCTAAAGATCACATTGATTTTGTAGAGCAGTGAATCAGGCTTTTGAAGGAA
 885 R I V I Q N N A C G K S I S I C S L K I T L I F E S S E I R I L I F G R
 2836 GAATTCAGGAGATAGCCACTGACCCAGGAGCTGAGAAGAATTATAAGTGGATCTCAGAGGAGGTTATATTGTGATTGAAACAACTCAAGGGATGAGCTTCATGT
 920 I Q E I A T D P G A E K N Y K V D L R G G Y I V I E I T T Q G M S F M W
 ♦
 2941 GGGACCAAGACGACCGTTGTGTTCATGTGACACCATCTTTCCAGGAAAAGCTGTGGCCTTTGTGGGACTTCGATGGCCGTTCAAGGAATGATTTCACAA
 955 D Q K T T V V V H V T P S F Q G K V C G L C G D F D G R S R N D F T T
 ♦
 3046 CCAGAGGGCAATCTGTGGAGATGAGCATCCAGGAGTTTGGAAACAGCTGGAATAACAAGCACTTGTCTCAATATCAACATGACAGACCTGTGTGCTGATCAGC
 990 R G Q S V E M S I Q E F G N S W K I T S T C S N I N M T D L C A D Q P
 ♦
 3151 CTTTCAAACTGCTTGGGACAGAAGCACTGCAGCATCATTAAAGCAGTGTCTTTGAAGCCTGCCACAGTAAGGTGAATCCAATACCATATTATGAATCTTGTG
 1025 F K S A L G Q K H C S I I K S S V F E A C H S K V N P I P Y Y E S C V
 3256 TTTCTGACTTCTGTGGCTGTGACAGTGTGGGAGACTGTGAGTGTCTTTGTACCTCGGTGGCTGTATGCAAGAAGCTGTAGCACAGCGGTGTCTGTATCAACT
 1060 S D F C G C D S V G D C E C F C T S V A A Y A R S C S T A G V C I N W
 3361 GGAGAAGCCAGCCATCTGCCCTGTGTTCTGTGATTACTACAATCCACCTGACAAACATGAGTGGTTCATAAACCCTGTGGAGCACCTGTCTAAAGACCTGCA
 1095 R T P A I C P V F C D Y Y N P P D K H E W F Y K P C G A P C L K I C R
 3466 GGAACCCACAAGGAAATGTGGAACATACIGTACAGCTTAGAAGGCTGTTACCCAGAGTGTAGCCCTGACAAGCGTATTTTGTAGAGGAGAGGAGGAGTGGC
 1130 N P Q G K C G N I L Y S L E G C Y P E C S P D K P Y F D E E R R E C V
 3571 TCTCCCTCCCTGACTGCATTCATGTAATCCCTGAAGAGAAGCTGTGTACTGAAGACTCCAAAGATTGCCTTTGTTGCTACAATGGGAAGACCTACCCATTAAATG
 1165 S L P D C T S C N P E E K L C T E D S K D C L C C Y N G K T Y P L N E
 ♦
 3676 AAATATATACAGCCAGACCGAAGGACGAAGTGTGGTAACTGCTTCTGTGGCCCTAAATGGAATGATCATTGAAACATTCATCCCTGCAGTACACTATCCGTGC
 1200 T I Y S Q T E G T K C G N A F C G P N G M I I E T F I P C S T L S V P
 ♦
 3781 CAGCACAAGAACAACCTTATGCAACCTGTAAACATCCGACCACTGTTATCAACAGAGCCACTCCCTGTTTCTGCACTGACAAATGGACAATTGATACAAATGGGAG
 1235 A Q E Q L M Q P V T S A P L L S T E A T P C F C T D N G Q L I Q M G E
 ♦
 3886 AAAATTTTCTCTGCCAATGAATATATCAGGTCATTGTGCTTACTCCATTTGCAATGCATCATGCCAGATTGAATTAATCTGGGAGAGTGAAGTTGTTCAAA
 1270 N V S L P M N I S G H C A Y S I C N A S C Q I E L I W A E C K V V Q T
 ♦
 3991 CTGAGGCCCTTGAGACCTGTGAACCAAACTCTGAAGCCTGTCCACCAACAGCTGCTCCCAATGCAACAAGCCTAGTTCTCTGCTACTGCGCTGGCTCCTATGAGCG
 1305 E A L E T C E P N S E A C P P T A A P N A T S L V P A T A L A P M S D
 ♦
 4096 ATTCTCTTGGGCTCATCTCTCCAGAAAGTTTAAATGAATCATGGGACTTTCGAAACTGCCAGATTGCCACTTCCCTGGGAGAAGAAAATAATTAAGCTCTCCA
 1340 C L G L I P P R K F N E S W D F G N C Q I A T C L G E E N N I K L S S
 ♦

4201 GCATCACTTGCCTCTCAGCAGCTGAAACTCTGTGTCATGGTTTCCCATTCATGAAACACCATGATGAAACTGGTTGCTGTGAAGTCTTTGAGTGCCAAATGTA
1375 I T C P P Q Q L K L C V N G F P F M K H H D E T G C C E V F E C Q C I
R domain <- -> D4 doma:
4306 TTTGCACTGGGTGGGGAATGAGCATTATGTGACTTTCGATGGAACCTATTATCACTTCAAAGAAAAGTGCACATATGCTCTTGTGGAGTTAATTCAGCCAGCT
1410 C S G W G N E H Y V T F D G T Y Y H F K E N C T Y V L V E L I Q P S S
4411 CTGAAAAGTTCTGGATTCATATAGACAACCTACTACTGTGGTGGCCGATGGAGCGATTGTTCATATGCTCTACTCATTTTCTACTCCAATTCTCTTGTATTTC
1445 E K F W I H I D N Y Y C G A A D G A I C S M S L L I F H S N S L V I L
4516 TGACACAAGCGAAGGAACATGGAAGGAACAACTGGTTTGTGTTAAAGATAAGAAAGTTGTTCCAGACATTTCTAAGAAATGGCATCAGGATAACAGCTCTG
1480 T Q A K E H G K G T N L V L F N D K K V V P D I S K N G I R I T S S G
4621 GCCTTTACATTCATAGTTGAAATACCTGAGTTAGAAGTTTATGTCTCTTACAGTCGACTTGCATTTTACATAAACTCCCTTTTGAAAGTATTATAACAATACCA
1515 L Y I I V E I P E L E V Y V S Y S R L A F Y I K L P F G K Y Y N N T M
4726 TGGGACTGTGGTACATGCACCAACCAGAAGTCTGATGATGCTAGGAAGAGGAATGGTGAGGTCACTGACTCCTTCAAAGAGATGGCATGGAAATGGAAAGGCC
1550 G L C G T C T N Q K S D D A R K R N G E V T D S F K E M A L D W K A P
4831 CGGTTTCTACCAACAGATACCTGTAATCCAGGCATCTCAGAACCACTTAAAGATTGAAATTTATCAGCACTGTGAGCCTTCTGAGCTGTGTAAATCATCTGGAACC
1585 V S T N R Y C N P G I S E P V K I E N Y Q H C E P S E L C K I I W N L
4936 TGACTGAGTGCCACCGTGTGGTCCCCCGCAGCCCTACTATGAGGCACTGCTAGCCAGCAGATGCTCGCAGCAGACCCAGCACTGAGTGCCAGAGCATGCAGA
1620 T E C H R V V P P Q P Y Y E A C V A S R C S Q Q H P S T E C Q S M Q T
5041 CATACGAGCCTTGTGTGGGCTCCACGGGATCTGTGTGGACTGGAGGGGCGAGACGAATGGGCAGTGCAGAGGCCACCTGTGCTCGGGATCAGGTCTACAAGCCAT
1655 Y A A L C G L H G I C V D W R G Q T N G Q C E A T C A R D Q V Y K P C
5146 GTGGGGAAGCAAAAAGAAACACTTGTCTCAGCAGAGAAGTCATGTGGACACCTCTCTCCCCGAAACAACACACCACTGTTTGTGAGGGATGTTACTGCCTG
1690 G E A K R N T C F S R E V I V D T L L S R N N T P V F V E G C Y C P D
5251 ATGGGAATATTCTGCTGAATGAGCATGATGGCATCTGTGTTTCTGTCTGCGGTTGCACTGCACAAGATGGATCAGTGAAAAAGCCTAGAGAGGCTTGGGAGCATG
1725 G N I L L N E H D G I C V S V C G C T A Q D G S V K K P R E A W E H D
5356 ACTGTCAATACTGTACTTGTGATGAAGAGACCTTGAATATCTCTTGCTTCCCCGCCCTTGTGCTAAATCCCCACCTATCAACTGCATAAAGAGGGCTTCGTAC
1760 C Q Y C T C D E E T L N I S C F P R P C A K S P P I N C T K E G F V R
D4 domain <-
5461 GCAAAATAAAACACGCTGAGATGATCCATGCTGCACAGAGACAGTCGTGAATGTGATATAAAACCTGCATTATCAACAAAACAGCATGCGACTTAGGTTTCC
1795 K I K P R L D D P C C T E T V C E C D I K T C I I N K T A C D L G F Q
-> C1 domain
5566 AGCCAGTGGTAGCTATATCIGAGGATGGTGTGTCTCTATCTTCTCCGTATACCAAAAGGTGTATGTGTTTCTGAAGGAGTTGAATTTAAGCCTGGGGCAGTGG
1830 P V V A I S E D G C C P I F S C I P K G V C V S E G V E F K P G A V V
5671 TCCCAAAAAGCTCTTCTGAAGACTGTCTCTGCACAGATGACCAAGATGCTGTGACAGCAACCAACCCGATCCAAATGCCCTCCAGTGAATGCCAAACCACTCC
1865 P K S S C E D C V C T D E Q D A V T G T N R I Q C V P V K C Q T T C Q
5776 AGCAGGGTTTCCGCTACGTGGAAGAGGGTCACTGCTGCAGCCAGTGTGACGAAGTGGCTTGTGTGGCAAACTTCCCATTTGGCAGTGTGACCATCGAGGTTG
1900 Q G F R Y V E K E G Q C C S Q C Q Q V A C V A N F P F G S V T I E V G
C1 domain <-
5881 GAAAGAGTTATAAAGCACCATACGATAACTGTACTCAGTACACATGCACTGAATCTGGGGGCCAGTTCTCCTTGAAGTACAGTAAAGGTCTGCCTACCATTTG
1935 K S Y K A P Y D N C T Q Y T C T E S G G Q F S L T S T V K V C L P F E
5986 AGGAGTCAAACCTGTGCTCCAGGACTGTTGATGTGACTTCAGATGGCTGTGTGAAGACATGTATCGACCTGCCACATAAATGCAACGAGCATGAAAGAGCAGT
1970 E S N C V P G T V D V T S D G C C K T C I D L P H K C K R S M K E Q Y
-> CK domain
6091 ACATTGTTACAAAGCACTGTAAGTCTGCTGCTCCGTTCCAGTGGCCTTCTGCGAAGGGACGTGCAGTACCTATCTGTGTATTCCTTTGAGAACAATGAAATGG
2005 I V H K H C K S A A P V P V P F C E G T C S T Y S V Y S F E N N E M E
6196 AACATAAATGCATTGTGCTGCATGAAAAAAGAGTCACTGAGAGAAGTGGAACTGGTTTGTCTGAGCATAAGACACTCAAATTTCTATATGTGACGTGGATG
2040 H K C I C C H E K K S H V E K V E L V C S E H K T L K F S Y V H V D E
6301 AATGCGGATGTGTGGAACCAATGTCCAATGAGGAGAACATGAGCATAAACTAAAGGAGACATTTACAGTCCCTGTTTCACGCTTCTCTAGCAAGTGGGTAGTTT
2075 C G C V E T K C P M R R T TER
CK domain <-
6406 AGAAAAAGCTGTGAGTAACCTTCAAGTGACCGTACTTCTCTCTGTCGCTTCTCTTCTGTTGTTATATATGAGTATGAGAAATTAGCTGAAGATTTTGTCTAA
6511 ATGTTTGTAGAATAAACTGCATTTCAAAGAGAAAAA

FIGURE 1 Nucleotide and deduced amino acid sequences of cDNA encoding α -subunit. The N-terminal amino acid (Lys) is numbered one. The amino acid sequences upstream the N-terminal Lys are in italic type. The first amino acid (Leu, number 863) from cDNA clone at the length of around 3.9kb is marked with an arrowhead ($\blacktriangle$). The amino acid sequences determined by protein sequencing of trypsin digested α -subunit and used for preparation of probes are underscored (T2-d and T2), and the nucleotide sequences determined from cDNA clones of the C-terminal region and used to synthesize the specific cDNA for the N-terminal region are underscored (specific primer). Potential Asn-X-Thr/Ser glycosylation sites are indicated by cloned triangles ($\blacktriangle$). The location of glycosylated Asn in the sequence Asn-Ser-Cys is shown by a closed circle ($\bullet$). The vicinal cysteine motif C-G-X-C is indicated by dashes (...). Protein family domains are indicated below the alignments.

insertion into pUC18 plasmid to prepare a specific cDNA library for the N-terminal portion. A probe for the N-terminal portion was prepared from 3.9kb cDNA by PCR amplification using the following sense 5'CTGGGAGACTGTGACTATATCCTA and

antisense 5'GGCTATCTCCTGAATTCTTCCTTCCA oligonucleotide primers (probe B in Fig. 3), and screening of the specific cDNA library for the N-terminal region was performed by the method described for the C-terminal region.

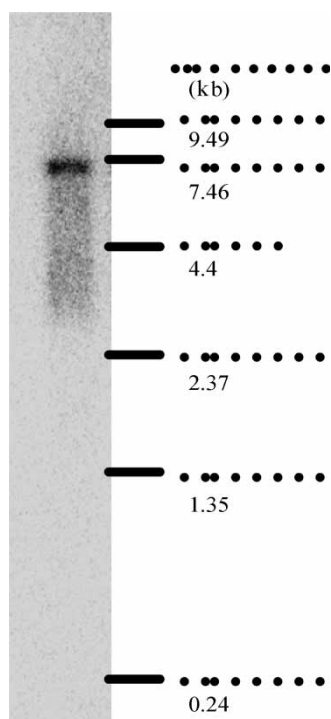


FIGURE 2 Northern blot analysis of α -subunit mRNA from magnum. The experimental methods are as described in the text. The nucleotide length was determined by using RNA marker as standard.

The complete nucleotide and deduced amino acid sequences of α -subunit using the two prepared cDNA clones with an overlap (the overlapping location: nucleotides 2666–2932) are shown in Fig. 1. The full sequence is 6558 bp, and it codes for 2087 amino acid residues with a predicted molecular mass of 230.9 kDa and at least 21 residues of a signal sequence upstream of the N-terminal Lys. The sequence contains a high amount of cysteine (9.5%) as well as asparagine (5.0%), glutamine (3.7%), proline (5.6%), serine (7.7%) and threonine (7.0%).

Our earlier study (Tsuge *et al.*, 1997a) indicated that α -subunit consisted of 91% protein and 9% carbohydrate, and the apparent molecular size determined by SDS-PAGE for α -subunit was

220 kDa. The predicted size for α -subunit in this study was calculated to be around 254 kDa on the basis of molecular mass of 230.9 kDa for protein portion described above and carbohydrate content of 9%. This is higher than the apparent molecular size of the 220 kDa.

DISCUSSION

Mucin cDNAs are difficult to clone full length because of many other repetitive and/or large cDNA sequences. This is also true for egg white OVM that is highly expressed in the magnum of hen oviduct. However, by using the two strategically prepared clones, the primary amino acid sequence could be determined.

A searching of the databases indicated that the amino acid sequence of α -subunit was similar to those of hpp-vWF, hMUC2 and other mucins such as human gastric mucin MUC5AC, human secretory mucin MUC6, and human tracheobronchial mucin MUC5, although their similarities were not as high in every case. Sequence similarities among the deduced amino acid sequences of α -subunit, hpp-vWF, hMUC2 and the C-terminal region (rMLP, discussed below) of a rat intestinal mucin-like peptide are presented in Fig. 4.

pp-vWF is a blood multimeric glycoprotein required for normal hemostasis, and contains four types of repeated domains designated A–D and arranged from the N- to C- terminals in the following order; D1-D2-D'-D3-A1-A2-A3-D4-B1-B2-B3-C1-C2-CK (cystine knot). The sequence of ppvWF is remarkable for the clustering of cysteine residues in the N-terminal and C-terminal regions. Many of the binding functions of vWF to collagen, blood coagulation factor VIII and so on have been localized to discrete regions of its subunit (Sadler *et al.*, 1985; Bonthron *et al.*, 1986; Shelton-Inloes *et al.*, 1986; Verweij *et al.*, 1986). On the other hand, MUC2 is expressed in the small intestine and colon, and also contains cysteine-rich regions on either side of the tandem repeat domain which resemble the D, C1 and CK domains of hpp-vWF (Toribara *et al.*, 1991; Gum *et al.*, 1992; Gum *et al.*, 1994). Analogous to pp-vWF and MUC2, the α -subunit contains several domains arranged from the N- to C-terminals in the following order; D1-D2-D'-D3 (N-terminal region)-R (central region)-D4-C1-CK (C-terminal region) (Fig. 1). Data presented here also make it clear that the complete amino acid sequence of α -subunit lacks the A1, A2, A3, B1, B2, B3 and C2 domains in hpp-vWF, as well as the domain of tandem repeats in hMUC2. The lack of tandem repeats is also seen in the mucin-like glycoproteins of mouse MUC14 (Melnick *et al.*, 2001) and human breast cancer tissue MUC1/Y (Zrihan-Licht *et al.*, 1994).

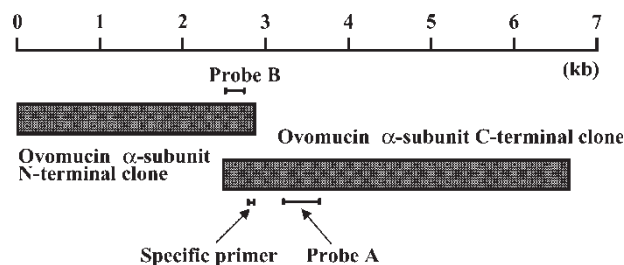


FIGURE 3 Strategy to isolate the cDNA clones covering the whole region of α -subunit. Specific primer was used to synthesize the specific cDNA for the N-terminal region. Probe A: DNA fragment for cloning the C-terminal region. Probe B: DNA fragment for cloning the N-terminal region.

A

-N-terminal D domain->

	*	20	*	40	*	60	*	80	
α -sub :	-----	KEPVQIVQVSTVGR	SECTTGNFHFHFDHVK	TFPCTCTVVFASHND	SYQDN	NIKIR	-----		60
hnp-vWF :	--MIPARF	CVLALATILPGT	CAECTGRSSTARCS	LEGSDFVMAFDC	SMYSFACYSMLLAG	CCOKRSFSIIIGDFQ	-----		78
hMUC2 :	MGLPLARL	AAVCLAISSAGGSE	QTECRTRYHGRNV	STWGNFHYKTFD	DVREFFCLDYNFASD	RGSYKEFAVHLK	GPQG		84
consensus	ar a v lal l	el eg	gr	CstwGnfh	TFDg	f FpG C Y fAs C sy f	r		
	*	100	*	120	*	140	*	160	
α -sub :	SDKNSHLI	FTVTDGVILEVKETGI	VNCGNIPLEFSL	SILLIEDTCAYFO	TSKLG	LTIDKTNWADT	LLDTEETK	KIKC	143
hnp-vWF :	GKRVSLSV	LGEFFD	IHLFVNGTV	QGDQVSMVYASK	GLYLETEAGY	YLSGEYGFVARID	SGSNFOVL	SDRFNKTCC	160
hMUC2 :	AEAPAGVESILL	IKDDTIYLRHLAVL	NCVSTPHYS	PGLLLEKSDAT	KVYSRAG	LTIMNREDA	MMEDTKFR	HTCC	167
consensus	s y t d i	t ng vs P	skgliiE	aY kv s ag ltl wn d l l l	y nktCG				
	*	180	*	200	*	220	*	240	
α -sub :	LCCNYDNK	KNDLITGYKMH	PROISGNFHV	VEDSEKCP	DVREDDHTGRHPT	EDDNRCSKYKMK	CKKLIS	REGNCPKVVA	223
hnp-vWF :	LCCNFNIFA	EDDMFTQEGT	TSDEYDANSWAL	SSGEQWCERAS	PPSSSCN	--ISSGEMOKGL	WEQCOLK	STSVFRCHEPLVD	242
hMUC2 :	LCCDYNGLQSYSE	FLSFCVFFS	GLEENMQRINO	DVVCDE	PEEVAP	-----	ASCSEHRAE	CERLTAEPAD	242
consensus	LCCgnyng	dG l s P	FgN k p	Ced p		cs	C Lls	Fa C lV	
	*	260	*	280	*	300	*	320	
α -sub :	FDDVVA	TCTEDMNCVNVSSQ	SDLVSSCISTINOY	SRDVLKSD	PCENRTKEL	CYQECESNMEY	ECNSCADT	CADPERSK	307
hnp-vWF :	PEEFV	LCEKTLCEAGGL	-----	ECACPA	LEARTCAQ	EMVLYGTDHSAC	SPVCHAG	MEYRQCVSPC	319
hMUC2 :	LEPYLRACQ	DRCRCPGGD	-----	TCVCS	VAFESRCSHAGGR	PGNWRATL	CPKTCFG	LVNLES	319
consensus	epyva C	d C C gg		C Cstl eysR C	gg pg Wrt	lC	CP nmeY	ecgspCadTC le s	
	*	340	*	360	*	380	*	400	
α -sub :	ICKAP	NDGCFCPFGIT	LDLGGKKCVPRDS	COMFOG	YVSSCGT	YSTPQNTCKG	CHWSCLIS	PCSCSCSIDG	391
hnp-vWF :	MCQER	VDGSCSCEPGL	DEG	-----	LCVEST	ECVHSGRP	PTSTLSR	ONTICRNS	400
hMUC2 :	GKTYTF	FCDCVYVLA	GDHNSYALL	CELPCC	STDKQ	CKNTVLLADKKK	AVFKD	CS	403
consensus	C e c	DGCFcPeGt	ldd g	Cvp	CpC	Gk Y pG	s dC	C C g W C	
	*	440	*	460	*	480	*	500	
α -sub :	NKKENF	CHNVLA	NTDDT	VVI	CEIL	CCSTKMT	CKNVLTG	RTTIKIC	471
hnp-vWF :	NRYTE	SGICOMLARD	COHSE	SIVIT	ETV	CADDDRAV	CTRSVT	R	482
hMUC2 :	GKTYTF	FCDCVYVLA	GDHNSYALL	CELPCC	STDKQ	CKNTVLLADKKK	AVFKD	CS	486
consensus	nk ftFhG	C YvLak	ddsf	gE	qCg	tClk V v l	n v s	Gs v mn	
	*	520	*	540	*	560	*	580	
α -sub :	PSTFF	KILSSAGVQR	VMKPVMSI	VHSHON	RTSG	LCGNFN	NIOTDDERT	TGAV	555
hnp-vWF :	QHVTAS	VRISYED	OMDWDGR	LLVK	SPVMA	KTCGLCGN	YNENQDD	ELTP	566
hMUC2 :	PSSYH	VMMAIGVR	LOVLA	PVMOLF	VTLDQAS	QCQVGL	LCGNFN	LEGDDEK	570
consensus	pst i v s	Gv lqvq	pvmqL	vtld	yqg t	GLCGNFng	qgDDF	TasGlvE	
	*	600	*	620	*	640	*	660	
α -sub :	EDPCS	NSVDKEK	FAHWCAL	LSNTSS	FA	CHSVVD	SVTKRCMYD	TCAEKS	639
hnp-vWF :	SDPCAL	NPRMTR	SEEA	CVLT	--SP	FEACH	RVSELP	PLNRNRYD	647
hMUC2 :	DDPCS	NTFESANY	AEHWC	LSLKK	ETEP	GRCSA	VDFAEY	KRCKYD	654
consensus	DPCSln	faehwCaLl	ts tF	aCHsaVdP	Y krC	YDtCnc	e cLC	aLssYaraCaa	
	*	680	*	700	*	720	*	740	
α -sub :	DPSE	ECPE	TMV	N	SVKY	CNOSCR	STDEPD	PLCKVQIA	722
hnp-vWF :	CELN	--CPK	GQVIL	CCGTP	CNLC	RSLSY	DEEON	--EAC	727
hMUC2 :	NKDVG	SCPN	SQVRL	NLT	TCQ	CRSL	SEAD	SHCLE	738
consensus	CP	qVlyl	t Cnqt	CRSLsepD	C	Ap	eGCGP	gtylde g	
	*	760	*	780	*	800	*	820	
α -sub :	LCKIC	QGRD	CIGET	VLV	-----	KDPA	PMYF	FNCS	787
hnp-vWF :	MCYCE	GFMC	CTMSG	PGSLL	PDVAVL	SSPLSHRS	KRSL	SCRPMV	811
hMUC2 :	RCVCR	GRRLH	CRQIRL	IG	-----	QSCT	AKIHMD	CSNL	803
consensus	C C	dGrlhC	v g			sC aPm	Cs	A g ec ksCqt d	
	*	860	*	880	*	900	*	920	
α -sub :	DGLVL	DGSG	GCIPK	QCP	VEG	CHFYK	EGETH	RVDCNT	871
hnp-vWF :	PGM	RHEN	RCVAL	ERC	PC	HO	KE	MAP	887
hMUC2 :	DGLMD	GRGGC	VEKE	CPV	HNDL	SSCAK	IKVDC	NTCK	887
consensus	dGlv dg	ggCv	CPCvH g	Y p	Getikvd	CNTCtC	r WnCTd	vC gTCs	
	*	940	*	960	*	980	*	1000	
α -sub :	QDFCP	NMDAG	TRF	IVION	NACCK	SLSIC	SLKTL	IFESSE	955
hnp-vWF :	QDYCES	NP	--G	FRIL	VGNK	GCSHPS	KCKR	VT	975
hMUC2 :	QDYCES	NP	--G	FRIL	VGNK	GCSHPS	KCKR	VT	971
consensus	QDYCg	N	GtFrI	N	Cg	v Csk	iti e	Ei L dg	
	*	1020	*	1040	*	1060	*	1080	
α -sub :	QKRTV	VHVTP	ES	FGK	VCG	LCCD	FDGR	SND	1033
hnp-vWF :	RHLIS	SV	KQ	TOE	VCG	LCCD	FDGR	SND	1059
hMUC2 :	KRTV	FIK	APSY	K	TVCG	LCCD	FDGR	SND	1050
consensus	ttv v l	psyqgk	VCG	LCCD	FDGR	SND	FTtr	Ve s	

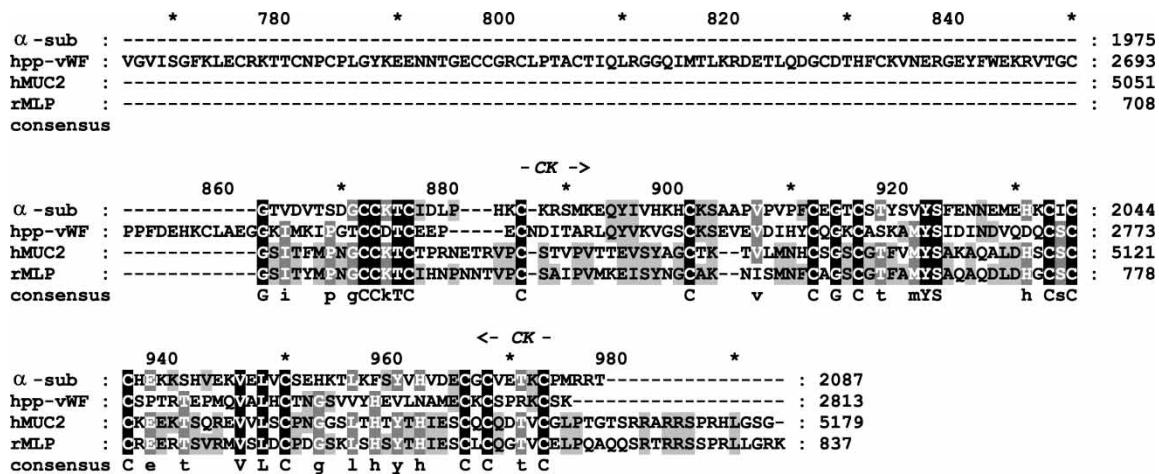


FIGURE 4 Comparison of amino acid sequence of α -subunit with hpp-vWF and hMUC2 (A, N-terminal region), and that of α -subunit with hpp-vWF, hMUC2 and rMLP (B, C-terminal region). Identical amino acids between three (A) and four sequences (B) are indicated by black background, and similar amino acids between two (A) and three sequences (B) are indicated by gray background. Gaps indicated by dots have been introduced to optimize alignments. Numbering indicates amino acid positions within the respective sequences. The consensus sequence indicates amino acids that are present in three and four sequences.

It was suggested that the cysteine-rich motifs in various mucins are critical for determining the tertiary structure, aggregation and physical properties of mucins via intra- and/or intermolecular disulfide bond formation (Voorberg *et al.*, 1990). The D3 domain, which contains cysteines that participate in interdimer disulfide-bond formation (Keates *et al.*, 1997), and the adjacent short D' domain in which no interchain disulfide bonds have been located are required for multimer formation (Marti *et al.*, 1987). Both the D1 and D2 domains are also crucial in vWF multimerization. It was reported that the vicinal cysteine motif C-G-X-C seen in vWF is required for the higher order multimerization of protein (Mayadas and Wagner, 1992). Its motif is similar to vicinal cysteines at the active site of disulfide isomerases that catalyze thiol protein disulfide interchange. The α -subunit also contains that motif (Fig. 1, amino acids 142–145 in D1, 973–976 in D3 and 1553–1556 in D4).

The C-terminal CK motif (a motif spanning 11 Cys residues) in α -subunit comprises 90 amino acids including six cysteines arranged in a knot-like topology, as generally seen in the CK domain of various mucins (Meitinger *et al.*, 1993). Based on the arrangement of disulfide bonds found in the CK family [Cys1-Cys4, Cys2-Cys5 and Cys3-Cys6], it is possible that the CK disulfide bonds in α -subunit are arranged: Cys1996-Cys2045, Cys2021-Cys2075 and Cys2025-Cys2077. Moreover, dimerization of α -subunit may be mediated by additional cysteines as described on the CK domain of hpp-vWF (Katsumi *et al.*, 2000).

On the basis of the above description and the view that the CK domain is a region responsible for growth factors (McDonald and Hendrickson, 1993),

one might expect that the several domains and/or subunits containing cysteine residues in α -subunit also fulfill the biological functions involved in the process of hatching and so on.

A comparison of the amino acid sequences of α -subunit with hpp-vWF and hMUC2 showed somewhat less sequence similarity both in the N- and C-terminal regions as shown in Fig. 5. However, the cysteine residues in α -subunit, hpp-vWF and hMUC2 could be aligned as described above. The central region (R; 191 amino acid residues) contained relatively large amounts of threonine, serine and proline (7.3, 8.9 and 9.3%, respectively), as well as cysteine and leucine (8.9 and 8.3%, respectively). However, the threonine, serine and proline residues did not specifically form the clustered segment in R of α -subunit, as was seen in the Ser-Thr-Pro-rich repeat domains of hMUC2 and the other mucins (Toribara *et al.*, 1991; Gum *et al.*, 1992; Gum *et al.*, 1994).

The predicted sequence of α -subunit includes 21 potential N-glycosylation sites with the sequence of Asn-X-Thr (or Ser) (Fig. 1). In addition, there is a potential carbohydrate-binding site with the sequence of Asn-X-Cys. There are also threonine and serine residues in α -subunit that may provide an abundance of sites available for O-linked glycosylation.

When subjected to reduction, purified intestinal mucins have been shown to undergo a decrease in molecular mass and to liberate a 118-kDa glycopeptide (Robertson *et al.*, 1989). The latter has been called a putative "link" component because it is assumed to be important for disulfide bond-mediated mucin polymerization. Controversy exists as to whether such a "link" is an integral mucin

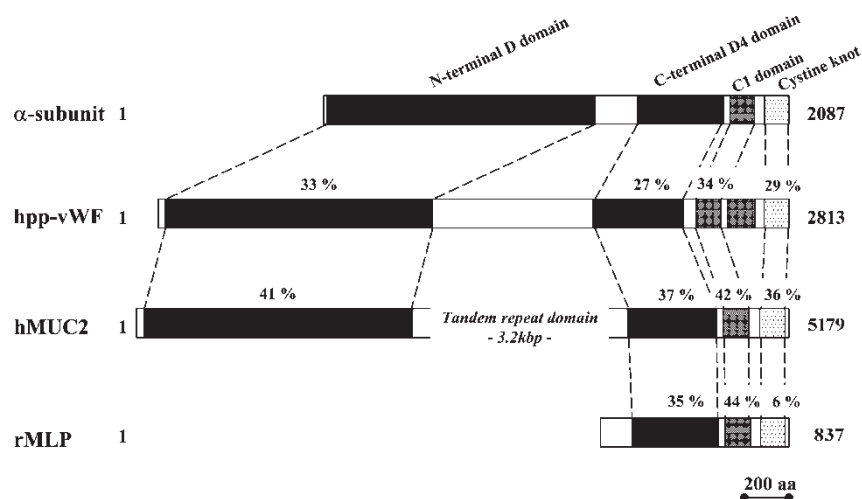


FIGURE 5 Domain comparisons of α -subunit for each of hpp-vWF, hMUC2 and rMLP. Numbering at the ends of the figures indicate the amino acid numbering used in each reference. Percentage above each domain indicates the amount of amino acid similarity (identical) compared to the corresponding region of α -subunit.

component or a separate mucin-associated glycopeptide.

The deduced amino acid sequence on a cDNA specific for the 118 kDa glycopeptide showed that rMLP corresponding to the putative "link" peptide is equivalent to the C-terminal 689 amino acids of a larger peptide (Xu *et al.*, 1992). Thus, rMLP was demonstrated to be a C-terminal region in a very much larger mucin-like peptide. A comparison of amino acid sequence between C-terminal regions of α -subunit and rMLP also showed a sequence similarity, although it was somewhat less pronounced (Fig. 5). The α -subunit in this paper was prepared by treatment with a disulfide-reducing agent in the presence of SDS; however, this subunit was different from rMLP in that the size of its molecular mass depended on the presence of the N-terminal D domain. Moreover, cDNA clones encoding α -subunit gene was independently obtained from the prepared library in this study. It may be speculated that α -subunit was not part of a large peptide of OVM, but was independently synthesized from β -subunit which is probably composed of mucin-like peptide. OVM may be made up of the homodimer and heteromultimer via intra- and intermolecular disulfide bonds of α - and β -subunits.

The similarity of domains in α -subunit with hpp-vWF and hMUC2 is surprising, since hens are phylogenetically distant from mammals. Even more significantly, the numbers and positions of cysteine residues are highly conserved among all three sequences, suggesting important structure-function relationships. Although the biological role of α -subunit is not known, the structure and function of pp-vWF and various mucins may provide important insights into those of the α -subunit. Further studies will identify the amino acid

sequences of the β -subunit to clarify in more detail the entire structural features of OVM.

MATERIALS AND METHODS

Preparation of α -Subunit, and Partial Sequence Analysis

The α -subunit was isolated from OVM in thick egg white of eggs from White Leghorn hens within 10 h after laying as described in a previous study (Tsuge *et al.*, 1997a), and the partial amino acid sequences of the N-terminus of α -subunit and its tryptic digested peptides were determined by Edman analysis using a Procise 491 protein sequencer (Applied Biosystems, Norwalk, CT). On the basis of the determined N-terminal amino acid sequences of two fragments (T2-d and T2 in Fig. 1) from the tryptic digestion of α -subunit, several primers were synthesized on the sense and antisense sequences and used for the polymerase chain reaction (PCR) amplification of probe DNA (probe A in Fig. 3).

Preparation and Screening of cDNA Library

Total RNA was prepared after freezing from magnum in the oviduct of hens (White Leghorn) in their laying period, using Isogen (Nippon Gene, Tokyo, Japan) according to the manufacturer's instructions. Poly (A)⁺ RNA was isolated from the total RNA by affinity chromatography on oligo(dT)-cellulose (Aviv and Leder, 1977), using an mRNA purification kit (Amersham Biosciences, Piscataway, NJ). cDNA synthesis was carried out with mRNA using a superscript plasmid system (Invitrogen, Carlsbad, CA), followed by subsequent ligation into pSPORT1 plasmid. The clones were selected from

a cDNA library using the colony filter hybridization technique (Maniatis *et al.*, 1989), and the nucleotide sequence was determined on both strands using a 373S DNA sequencer (Applied Biosystems).

Southern and Northern Blot Analyses

On southern blot analysis, plasmid DNA in a magnum cDNA library was completely digested with the restriction enzyme Not I (Takara, Kyoto), and 2 µg of the DNA was electrophoresed on a 1.0% agarose gel. The DNA was transferred to a nylon membrane Hybond™-N (Amersham Biosciences) and UV cross-linked. The probe A in Fig. 3 was used as a hybridization probe. The membrane was hybridized at 55°C for 16 h with an alkaline phosphate-labelled probe. A chemiluminescent signal was detected with Kodak X-ray film. On northern blot analysis, total RNA was prepared again from magnum as described above, and 5 µg of the isolated mRNA and commercially available RNA marker (Invitrogen) was electrophoresed on a 1.0% agarose-formaldehyde gel denatured by soaking it in 10 mM NaCl/50 mM NaOH for 10 min, transferred to a nylon membrane Hybond™-N, and UV cross-linked. The membrane was prehybridized for 4 h at 42°C, followed by hybridization at 42°C for 16 h with a [³²P]-labelled probe. Membranes were washed in 2XSSC/0.1% SDS and then underwent autoradiography with Kodak X-ray film.

Homology Search

A comparison of the amino acid sequence was carried out on searches using the Swiss Prot databases (<http://www.ncbi.nlm.gov/BLAST>) with the FASTA system (<http://fasta.genoma.ad.jp>). The sequence alignments were achieved with the Clustal W program (Thompson *et al.*, 1994).

References

Aviv, H. and Leder, P. (1977) "Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose", *Proc. Natl Acad. Sci. USA* **69**, 1409–1412.

Bonthron, D., Orr, E.C., Mitsock, L.M., Ginsburg, D., Handin, R.I. and Orkin, S.H. (1986) "Nucleotide sequence of pre-pro-von Willebrand factor cDNA", *Nucleic Acids Res.* **14**, 7125–7127.

Donovan, J.W., Davis, J.G. and White, L.M. (1970) "Chemical and physical characterization of ovomucin, a sulfated glycoprotein complex from chicken eggs", *Biochim. Biophys. Acta* **207**, 190–201.

Forstner, J.F. and Forstner, G.G. (1994) "Gastrointestinal mucus", In: Johnson, L.R., ed., *Physiology of the Gastrointestinal Tract*, 3rd Edn. (Raven Press, New York) **Vol. 2**, pp 1255–1283.

Gendler, S.J. and Spicer, A.P. (1995) "Epithelial mucin genes", *Ann. Rev. Physiol.* **57**, 607–634.

Gum, J.R., Jr, Hicks, J.W., Toribara, N.W., Rothe, E-M., Lagace, R.E. and Kim, Y.S. (1992) "The human MUC2 intestinal mucin has cysteine-rich subdomains located both upstream and

downstream of its central repetitive region", *J. Biol. Chem.* **267**, 21375–21383.

Gum, J.R., Jr, Hicks, J.W., Toribara, N.W., Siddikil, B. and Kim, Y.S. (1994) "Molecular cloning of human intestinal mucin (MUC2) cDNA; Identification of the amino terminus and overall sequence similarity to prepro-von Willebrand factor", *J. Biol. Chem.* **269**, 2440–2446.

Hayakawa, S. and Sato, Y. (1977) "Physicochemical identity of α-ovomucins or β-ovomucins obtained from the sonicated insoluble and soluble ovomucins", *Agric. Biol. Chem.* **41**, 1185–1191.

Hiidenhovi, J., Aro, H.S. and Kankare, V. (1999) "Separation of ovomucin subunits by gel filtration: Enhanced resolution of subunits by using a dual-column system", *J. Agric. Food Chem.* **47**, 1004–1008.

Kato, A., Nakamura, R. and Sato, Y. (1970) "Changes in the chemical composition of ovomucin", *Agric. Biol. Chem.* **34**, 1009–1013.

Katsumi, A., Tuley, E.A., Bondo, I. and Sadler, J.E. (2000) "Localization of disulfide bonds in the cystine knot domain of human von Willebrand factor", *J. Biol. Chem.* **275**, 25585–25594.

Keates, A.C., Nunes, D.P., Afdhal, N.H., Troxler, R.F. and Offner, G.D. (1997) "Molecular cloning of a major human gall bladder mucin: complete C-terminal sequence and genomic organization of MUC5B", *Biochem. J.* **324**, 295–303.

Lapensee, L., Paquette, Y. and Bleau, G. (1997) "Allelic polymorphism and chromosomal localization of the human oviductin gene (MUC9)", *Fertil. Steril.* **68**, 702–708.

Maniatis, T., Fritsch, E.F. and Sambrook, J. (1989) *Molecular Cloning: a Laboratory Manual*, 2nd Edn. (Cold Spring Harbor Laboratory Press, New York), pp 190–1104.

Marti, T., Rosselet, S.J., Titani, K. and Walsh, K.A. (1987) "Identification of disulfide-bridged substructures within human von Willebrand factor", *Biochemistry* **26**, 8099–8109.

Mayadas, T. and Wagner, D. (1992) "Vicinal cysteines in the prosequence play a role in von Willebrand factor multimer assembly", *Proc. Natl Acad. Sci. USA* **89**, 3531–3535.

McDonald, N.Q. and Hendrickson, W.A. (1993) "A structural superfamily of growth factors containing a cystine knot motif", *Cell* **73**, 421–424.

Meitinger, T., Meindl, A., Bork, P., Rost, B., Sander, C., Haasemann, M. and Murken, J. (1993) "Molecular modeling of the Norrie disease protein predicts a cystine knot growth factor tertiary structure", *Nat. Genet.* **5**, 376–380.

Melnick, M., Chen, H., Zhou, Y. and Jaskoll, T. (2001) "An alternatively spiced MUC10 glycoprotein ligand for putative L-selectin binding during mouse embryonic submandibular gland morphogenesis", *Arch. Oral. Biol.* **46**, 745–757.

Pallesen, L.T., Berglund, L., Rasmussen, L.K., Petersen, T.E. and Rasmussen, J.T. (2002) "Isolation and characterization of MUC 15, a novel cell membrane-associated mucin", *Eur. J. Biochem.* **269**, 2755–2763.

Robertson, A.M., Mantle, M., Fahim, R.E.F., Specian, R.D., Bennick, A., Kawagishi, S., Sherman, P. and Forstner, J.F. (1989) "The putative "link" glycopeptide associated with mucus glycoproteins: composition and properties of preparations from the gastrointestinal tracts of several mammals", *Biochem. J.* **261**, 637–647.

Sadler, J.E., Shelton-Inloes, B.B., Sorace, J.M., Harlan, J.M., Titani, K. and Davie, E.W. (1985) "Cloning and characterization of two cDNAs coding for human von Willebrand factor", *Proc. Natl Acad. Sci. USA* **82**, 6394–6398.

Shankar, V., *et al.* (1997) "Chromosomal localization of a human mucin gene (MUC8) and cloning of the cDNA corresponding to the carboxyl terminus", *Am. J. Respir. Cell Mol. Biol.* **16**, 232–241.

Shelton-Inloes, B.B., Titani, K. and Sadler, J.E. (1986) "cDNA sequences for human von Willebrand factor reveal five types of repeated domains and five possible protein sequence polymorphisms", *Biochemistry* **25**, 3164–3171.

Thompson, J.D., Higgins, D.G., Gibson, T.J. and Clustal, W. (1994) "Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice", *Nucleic Acids Res.* **22**, 4673–4680.

- Toribara, N.W., Gum, J.R., Jr., Culhane, P.J., Lagace, R.E., Hicks, J.W., Petersen, G.M. and Kim, Y.S. (1991) "MUC-2 human intestinal mucin gene structure; repeated arrays and polymorphism", *J. Clin. Investig.* **88**, 1005–1013.
- Tsuge, Y., Shimoyamada, M. and Watanabe, K. (1997a) "Structural features of Newcastle disease virus- and anti-ovomucin antibody-binding glycopeptides from pronase-treated ovomucin", *J. Agric. Food Chem.* **45**, 2393–2398.
- Tsuge, Y., Shimoyamada, M. and Watanabe, K. (1997b) "Bindings of ovomucin to Newcastle disease virus and anti-ovomucin antibodies and its heat stability based on binding abilities", *J. Agric. Food Chem.* **45**, 4629–4634.
- Verweij, C.L., Diergaarde, P.J., Hart, M. and Pannekoek, H. (1986) "Full-length von Willebrand factor (vWF) cDNA encodes a highly repetitive protein considerably larger than the mature vWF subunit", *EMBO J.* **5**, 1839–1847.
- Voorberg, J., Fontijn, R., van Mourik, J.A. and Pannekoek, H. (1990) "Domains involved in multimer assembly of von Willebrand factor (vWF): multimerization is independent of dimerization", *EMBO J.* **9**, 797–803.
- Watanabe, K., Tsuge, Y., Shimoyamada, M., Ogama, N. and Ebina, T. (1998a) "Antitumor effects of pronase-treated fragments, glycopeptides, from ovomucin in hen egg white in a double grafted tumor system", *J. Agric. Food Chem.* **46**, 3033–3038.
- Watanabe, K., Tsuge, Y. and Shimoyamada, M. (1998b) "Binding activities of pronase-treated fragments from egg white ovomucin with anti-ovomucin antibodies and Newcastle disease virus", *J. Agric. Food Chem.* **46**, 4501–4506.
- Williams, S.J., McGuckin, M.A., Gotley, D.C., Eyre, H.J., Sutherland, G.R. and Antalis, T.M. (1999) "Two-novel mucin genes down-regulated in colorectal cancer identified by differential display", *Cancer Res.* **59**, 4083–4089.
- Williams, S.J., Wreschner, D.H., Tran, M., Eyre, H.J., Sutherland, G.R. and McGuckin, M.A. (2001) "MUC13, a novel human cell surface mucin expressed by epithelial and hemopoietic cells", *J. Biol. Chem.* **276**, 18327–18336.
- Xu, G., Huan, L.-J., Khatri, I.A., Wang, D., Bennick, A., Fahim, R.E.F., Forstner, G.G. and Forstner, J.F. (1992) "cDNA for the carboxyl-terminal region of a rat intestinal mucin-like peptide", *J. Biol. Chem.* **267**, 5401–5407.
- Yin, B.W. and Lloyd, K.O. (2001) "Molecular cloning of the CA125 ovarian cancer antigen: identification as a new mucin, MUC16", *J. Biol. Chem.* **276**, 27371–27375.
- Yokota, T., Ohishi, H. and Watanabe, K. (1999) "In vitro studies of cytotoxic effect on sarcoma-180 cells of β -subunit derived from egg white ovomucin", *Food Sci. Technol. Res.* **5**, 273–278.
- Zrihan-Licht, S., Vos, H.L., Baruchi, A., Elroy-Stein, O., Sagiv, D., Keydar, I., Hilken, J. and Wreschner, D.H. (1994) "Characterization and molecular cloning of a novel MUC1 protein, devoid of tandem repeats, expressed in human breast cancer tissue", *Eur. J. Biochem.* **224**, 787–795.