

Gastric DNA-binding proteins recognize upstream sequence motifs of parietal cell-specific genes

(zinc finger/GATA family/gastrointestinal epithelium)

SHIGEHIKO TAMURA, XIAO-HUI WANG, MASATOMO MAEDA, AND MASAMITSU FUTAI

Department of Organic Chemistry and Biochemistry, The Institute of Scientific and Industrial Research, Osaka University, Ibaraki, Osaka 567, Japan

Communicated by Paul D. Boyer, July 26, 1993

ABSTRACT Polymerase chain reaction amplification of cDNA from pig gastric mucosa demonstrated the presence of zinc-finger proteins called GATA-GT1, GATA-GT2, and GATA-GT3, each having zinc-finger sequences similar to previously characterized GATA-binding proteins. Subsequently, full-length cDNAs of GATA-GT1 and GATA-GT2 were obtained from rat stomach. The zinc-finger domains of GATA-GT1 and -GT2 were 66–86% identical on the amino acid level with each other and with other GATA-binding proteins. Potential protein kinase phosphorylation sites were present in the zinc-finger region. In contrast, regions outside the zinc fingers shared significantly lower similarities. GATA-GT2 was found to bind to the upstream sequence of the H⁺/K⁺-ATPase β gene and to a sequence containing the GATA motif. GATA-GT1 and -GT2 were expressed predominantly in the gastric mucosa and at much lower levels in the intestine (GATA-GT2, also in testis), their tissue distributions being distinct from those of GATA-1, -2, or -3. These results clearly suggest that GATA-GT1 and GATA-GT2 are involved in gene regulation specifically in the gastric epithelium and represent two additional members of the GATA-binding protein family.

The highly differentiated parietal cells in the stomach have a characteristic morphology and the specialized function of secreting hydrochloric acid into the gastric lumen (1). The major enzyme in this system is an ATP-driven proton pump, the H⁺/K⁺-ATPase, which is responsible for translocation of protons across the apical membrane (2). The gene structures of the ATPase α and β subunits have been reported by us (3–5) and others (6–9). The sequences revealed that the H⁺/K⁺-ATPase is closely related to Na⁺/K⁺-ATPase. In addition, the intron/exon organizations of the genes are very similar, indicating that the two enzymes diverged from a common ancestor. On the other hand, as expected from its specific expression in the stomach (1, 5, 10–12), the 5' upstream sequences of the H⁺/K⁺-ATPase genes are different from those of the sodium pump (3–5). We have reported that a DNA sequence motif, (G or C)RR(G or C)NGAT(A or T)RY in which R and Y are unspecified purine and pyrimidine bases, respectively, is present in the human and rat α and β subunit genes (4, 13). This motif was recognized by a DNA-binding protein(s) from gastric tissue but not by those from other tissues. Interestingly, the motif includes the (T or A)GATA(G or A) sequence recognized by the GATA-binding proteins (14–24), a group of transcriptional factors implicated in the regulation of erythroid-specific genes, endothelial determinants, and others.

From the overlap of the DNA sequences recognized by the gastric-specific and GATA-binding proteins (4, 13–24), we wondered whether the gastric proteins had similar zinc-finger domains. cDNA was probed based on this idea, and three

gastric DNA-binding proteins were identified. Two of these proteins, GATA-GT1 and GATA-GT2, were each found to have tandem zinc fingers very similar to those of GATA-binding proteins.* We found that the GATA-GT1 and GATA-GT2 proteins were predominantly expressed in the gastric mucosa, not in the underlying muscular layer and were expressed in lower amounts in the intestine (GATA-GT2, also in testis) as well. These gastric-specific DNA-binding proteins are likely to play important roles in transcriptional regulation in the gastrointestinal tract.

MATERIALS AND METHODS

Amplification of Zinc-Finger Domains by Polymerase Chain Reactions (PCR). Poly(A)⁺ RNA was isolated by using Oligotex-dT30 (Dai-ichi Pure Chemicals, Tokyo). Degenerate primers corresponding to the conserved regions of the zinc-finger domains (119 amino acid residues) (23) of mammalian and chicken GATA-1, -2, and -3 (14–22, 24) were synthesized with a Pharmacia Gene Assembler Plus: primer 1, 5'-CGTCGACTACYTRTGAAAYGCCTGYGG-3' (in which Y = C or T; R = A or G; V = G, A, or C; and N = G, A, T, or C) represented protein sequence Tyr-Leu-Cys-Asn-Ala-Cys-Gly; and primer 2, 5'-CGTCGACRCAGGCRTTR-CANACVGG-3' represented protein sequence Pro-Val-Cys-Asn-Ala-Cys. A *Sal*I site was included at the 5' ends of the primers (underlined bases). These primers together with cDNA (25) were used in PCR (26). A 194-base-pair fragment was ligated into the *Hinc*II site of pUC18 (25) and was sequenced by the dideoxy chain-termination method (27).

Screening a Rat Stomach cDNA Library. PCR-amplified sequences from the zinc-finger domains of pig gastric proteins (GATA-GT1, -GT2, and -GT3) (Fig. 1) were labeled with ³²P by using a random-primed DNA labeling kit (Boehringer Mannheim) and were used for colony hybridization (25). A *Sac*I-*Eco*RI fragment of a partial clone of GATA-GT1 (base pairs 494–854, see Fig. 2) was used to probe for the remainder of the coding sequence. The resultant positive clones were subcloned into pUC18, and both strands were sequenced.

Other Procedures. Total RNA was extracted from the tissues of 5-week-old Sprague-Dawley rats by the guanidine thiocyanate-CsCl method (25). Pig stomach was obtained from a local slaughter house, and RNA was prepared from the mucosal or muscle layer. They were subjected to RNA blot (Northern) hybridization (25) with DNA fragments from cDNA clones of GATA-GT1 or -GT2. For gel mobility-retardation assay (13), a cDNA fragment from GATA-GT2 (base pairs 380–1104, amino acids 128–368) was inserted in the *Sma*I site of the T7 expression vector, pT7-7 (28), and introduced into *Escherichia coli* BL21(DE3) (29). T7 polymerase was induced, and cells were incubated for 2 hr at 37°C to express the peptide fragment. Cell extracts were prepared

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.

*The sequences reported in this paper have been deposited in the GenBank data base (accession nos. L22760 for GATA-GT1 and L22761 for GATA-GT2).

```
GATA-1  LYHKMNGQNRPLIRPKRLIVSKRAGTQCTNCQTTTTLWRRNASGD
GATA-2  F---K-----K--R--SAAR---C-A-----N--
GATA-3  -----K--R--SAAR---S-A-----N--

GATA-GT1  --S---LS---K-QK-VPSR-L-LS-A--H-----E-E
GATA-GT2  -----I---K-QR--SASR-V-LS-A-----E-E
GATA-GT3  --S--RLS---K-QR--SASR-V-LS-A-----E-E
```

FIG. 1. Alignment of partial zinc-finger domains determined from PCR-amplified cDNA sequences. cDNA of pig (gastric mucosa) or rat (whole stomach) was added to the PCR reaction mixture with degenerate primers corresponding to conserved amino acid sequences of the zinc-finger domains of the GATA-binding proteins (14–22, 24). The deduced amino acid sequences of clones of the amplified DNA are shown (except for the sequences covered by the primers) together with the corresponding sequences from human GATA-1 (16, 17), GATA-2 (22, 24), and GATA-3 (20, 21). Human GATA-2 from ref. 24 has phenylalanine and lysine as the first and sixth residues, respectively, in the above sequence possibly because of polymorphism. Only different residues from GATA-1 are indicated.

as described (30). Protein was measured with a Bio-Rad assay kit (31) with bovine serum albumin as a standard.

Chemicals. Restriction enzymes, T4 DNA ligase, and the Klenow fragment were obtained from Takara Shuzo (Kyoto), Toyobo (Osaka), or Nippon Gene (Toyama, Japan). Sequ-

nase was from United States Biochemical. [α - 32 P]dCTP (3000 Ci/mmol; 1 Ci = 37 GBq) was from the Radiochemical Center, Amersham. All other chemicals used were of the highest grade commercially available.

RESULTS AND DISCUSSION

Amplification of cDNA for Gastric-Specific Zinc-Finger Proteins. Degenerate primers corresponding to the completely conserved sequences in the zinc-finger domains were combined with cDNA synthesized from pig gastric mucosa poly(A)⁺ RNA and subjected to PCR. Products of expected length (194 base pairs) were amplified, cloned, and sequenced. Three zinc-finger segments were identified; the translated sequence of each was strongly similar to previously known GATA-binding proteins (14–24) and were named GATA-GT1, GATA-GT2, and GATA-GT3 (Fig. 1). Despite their similarities, these sequences were not identical with any of the three GATA-binding protein groups that are highly conserved in different species. For example, the human and mouse segments of GATA-1 or GATA-3 shown in Fig. 1 are at least 98% identical, whereas the identities of the GATA-GT1, -GT2, and -GT3 sequences showed a maximum identity of 85% with GATA-1, -2, or -3 (14, 16, 17, 19–22, 24). Moreover, clones exactly matching GATA-2 and GATA-3

```
-268  CGACTTGGGAGGACCTGTGTCTTCACTGACCTCGATCAGGCCGCGACCGCCAGCAAGCTGCTGTGGTCCAGCCGGGGCGCCAAAGCT
-180  GAGTCCCTTCGCGGCCGAGCAGCCGGAGGAAATGTACCAGACGCTCGCCGCTCTGTCCAGCCAGGGGCCCGCCGCTTACGACGGCGCTCC
-190  CGGCGGCTTCGTACACTCCGACGCTGCAGCGCGCGCGCGCGCCGCGGAGCCAGCTCTCCGGTCTACGTGCCACACACGCGTGGGTCC
  1  ATGCTGCCCGCCCTGCCTTACCTGCAGGGGGCGGGCAGCGGGCCAGCAATACGCGGGCGCGCGGGCCACCCGGGCTGGCCCCAG
  M L P G L P Y L Q G A G S G P S N H A G G A G A H P G W P Q  30
  91  GCCTCCGCCGACAGCCCCCGTATGGCGGGGTGGTGCAGCGGGCGCGCGCGCCGACCCGGAGGTGCGGGATCAGCCACAGCCAC
  A S A D S P P Y G G G G A A G G G A A G P G G A G S A T A H  60
  181  GCCTGTCACGCTTCCCTACTCGCCAGCCCGCCATGGCCAAACGGCGCCGCGGGACCCGGGGGTACGTGGCTGGGGCGGGCGG
  A S A R F P Y S P S P P M A N G A A R D P G G Y V A A G G A  90
  271  GCGCGGGTAGCGTGAGTGGCGGTGGCGGCAGCCTGGCGCCATGGTGGCGGGAGCACCAGTACAGCTCGCTGTCCGACCCAGGCCG
  G A G S V S G G G G S L A A M G G R E H Q Y S S L S A A R P  120
  361  CTGAACGGGACGTACCACCACCACCATCATCACCACCCGACCTACTCGCCCTACATGGGCGCTCCTCTGACTCCCGCTGGCCAGCGGGA
  L N G T Y H H H H H H H P T Y S P Y M G A P L T P A W P A G  150
  451  CCCTTCGAAACGCGCGGTGCTCCACAGTTTACAGAGCCGCGGGAGCTCCACTCCCGGTGCCACGAGGCCACGCGCAGACCTGTGGAG
  P F E T P V L H S L Q S R A G A P L P V P R G P S A D L L E  180
  541  GACCTGTCGGAGACCCGCGAGTGCCTGAACTGTGGTCCATCCAGACGCCACTGTGGAGACGAGATGGCACCAGGTCATTACCTGTGCAAC
  D L S E S R E [C] V N [C] G S I O T P L W R R D G T G H Y L [C] N  210
  631  GCATGCGGTCTTACAGTAAGATGAACGGCTCAGCAGGCCCTCATCAAGCCACAGAAGCGCGTGCCTTCATCAGCAGCGCTTGGACTG
  A [C] G L Y S K M N G L S R P L I K P O K R V P S S R R L G L  240
  721  TCCTGTGCCAACTGTACACCACAACCACTACCTTATGGCGTAGAAACGCTGAGGGTGAGCCGCTGTGCAATGCTTGGGLTYM
  S [C] A N [C] H T T T T T L W R R N A A E G E P V [C] N A [C] G L Y M  270
  811  AAACCTCCATGGGTCCTCGACCACTTGCTATGAAAAAGAAGGAATTCAAACCAGAAACGAAACCTAAAAATATAAATAAGTCGAAA
  K L H G L V P R P L A M K K E G I O T R K R K P K N I N K S K  300
  901  GCTTGCTCCGGTAACAGCTCTGTTCCTATGACTCCAACCTCTCTCTCTCTAATTCAGATGACTGCACAAAAATACTTCTCTCCACAA
  A C S G N S S V P M T P T S S S S N S D D C T K N T S P P T  330
  991  CAGTCGACTGCCTCAGGGGTGGGTGCATCAGTGATGTCTGCGAGTGGGAGAAAGTGCCAACCTGAGAACAGTGACCTCAAGTATTCAGGT
  Q S T A S G V G A S V M S A V G E S A N P E N S D L K Y S G  360
  1081  CAAGATGGCCTTACATAGGTGTGCTGCTTGTCTCCCTCCCTGCGGAAAGTTACATCCCTCGGTGCGACAGGATCTTGGTGTGCTCTGGCCCTG
  Q D G L Y I G V S L S S P A E V T S S V R Q D S W C A L A L  390
  1171  GCCTGAGCTGGTGTGCCAAGGGCTCTGAAGGCCCTCATACCACTTGTGTCTGATTTTGTCCAGCAGTCCAGATGGCGGCAAAAAATGCAG
  A *  391
  1261  ACATAACATTCCTTCGATGCGTGATTTCTGTGCCCTTGTGTTTGAAGAGATATATTTCCAAGAAGCTTATTGAAGTAAGAAGAGATGGG
  1351  TCCTTGTAGGAAGGGGCGAGCACAGTGGCGCTGTGGCCTATTCCTGTGAGCTGGGTCTGCTTCCAGCCAGCTTGCCTCTGGCTCTGCTCC
  1441  ACACCAGTGGGAGAACTGTGACAATGACTGGGCCTTGTCTGCTAAGGAAGATTGAGAGATTAAAAAAAATAGTTTGTGTATTGCTC
  1531  CAAATCATGTGCTTCTTGTGATCAACCTTGGTTATCCCAAGATTC
```

FIG. 2. Nucleotide sequence of cDNA for rat GATA-GT1 with its deduced amino acid sequence. Nucleotides are numbered on the left of each line from the first letter of the initiation codon, and numbers on the right are those of amino acid residues starting from Met-1. The potential sites for phosphorylation (32) by cAMP-dependent protein kinase A [Arg-Arg-Xaa-(Ser or Thr) or Lys-Arg-Xaa-Xaa-(Ser or Thr)], protein kinase C [Arg-Xaa-Xaa-(Ser or Thr)-Xaa-Arg], and multifunctional calmodulin-dependent protein kinase II [Arg-Xaa-Xaa-(Ser or Thr)] are indicated by straight, dotted, and wavy bars, respectively. The arrows below the nucleotide sequence correspond to those of the PCR primers used. The zinc-finger domain is underlined. Cysteine residues in the zinc fingers are boxed.

-562 CCCGAGGCTCCGCTGTGCTG
 -541 CAGCTCCGCGGACTCACGGAGATCGCGCCGGCTCTCTGGGAACTGGAGCTGGCCAGGACTGTCGCTTCAAGGGAACGGGCCCTCTTCG
 -451 TCCTTCGCTGGAGCCGACTGGAGCCACCAATGCGCCTGGGTGTGCAGCAGGAGAAAGCAAGGACTAGGCACCTTAGCCCGTGGGTG
 -361 ATCCGAAGGCTGCGCAGGGTGTTCGAGACCAGTCTTGACTGAGTCTGGGCATCTCCAGCCTCTGGGCCCTGGAGTAGAGTACAGCCCTC
 -271 CTGCAGTTCTGGAGCAACCAAAATCCAAATTTGGGATTTTGGTTTCTTCTGGAGCAAAACCAGAGCCTAAAGATTTGCTGTGATGCT
 -181 GGATTTAATTCGTATATATTTTGGAGGAGTTGGGCTGTCTCATTTGTTTTTGTATCTCCGTTTTTCGCGACAGTTCTGCACACCTGTATT
 -91 CCAATTCCTGTGCTTTGGTCTGCCGCGCCCGCTGGGCTCTTACTTGCAAATAAAGGGGAGCACACCGAAGCTCAGAGCTTGGGGCG
 1 ATGTACCAAAGCCTGGCTATGGCCGCAACACGGGCCCGCCGCGGGCGCTACGAAGCGGGTGGACCTGGCGCCTTCATGCACAGTGGC
 M Y Q S L H G P P P G A Y E A G G P G A F M H S A M A A N A 30
 91 GCGCCGCGTCTCGCCCGTCTACGTGCCCACTCCGCGGGTCCGCTCTGTGCTGGGCTGTCTACCTGCAGGGCGGAGGAGTGGT
 G A A S S P V Y V P T P R V P S S V L G L S Y L Q G G G S G 60
 181 GCTGCTTCTGGAGCCACCTCGGGTGGCAGCTCCGGGGCGGCGCCGCTGGGTGCTGGGCTGGGACCCAGCAGGGTAGCCCGGGCTGGAGC
 A A S G A T S G G S S G A G P S G A G P G T Q Q G S P G W S 90
 271 CAGGCTGGAGCTGAGGGAGCCGTTACACCCACCGCCTGTGTCCTCCCGGCTTCTCCTTCCCGGGGACTACTGGGTCCCTGGCGGGCCG
 Q A G A E G A A Y T P P P V S P R F S F P G T T G S L A A A 120
 361 GCCCGCGCTGCGCAGCCCGGGAAGCTGCAGCCTACAGCAGTAGCGGGCGGCGAGCCGGCGCTGGCTGGCCGAGAGCAGTACGGG
 A A A A A R E A A Y S S G A A G A G L A G R E Q Y G 150
 451 CGTCCGGCTTCGAGGCTTACTCCAGCCCTACCCAGCTACAGCAGTGGCGATGTGGGCGCATCTGGGCGCAGCCGCTGCCCTCC
 R P G F A G S Y S S P Y P A Y M A D V G A S W A A A A A A S 180
 541 GCGGCCCTTCGACAGCCAGTCTGCACAGCCTACCTGGCCGGCCAACTTCGAGACACCCCAATCTCGATATGTTTGTGACTTC
 A G P F D S P V L H S L P G R A N P A R H P N L D M F D D F 210
 631 TCAGAAGGCAGAGAGTGTGCAACTGCGGGCCATGTCCACCCACTCTGGAGGCGAGATGGGACAGGACTACCTATGCAACGCGCTGT
 S E G R E [C] V N [C] G A M S T P L W R R D G T G H Y L [C] N A [C] 240
 721 GGCTCTACCAAGATGAATGGTATCAACCGGCCCTCATCAAGCCTCAGCGCGGCTGTCTGCCTCCCGCGGGTAGCCCTCTCTGT
 G L Y H K M N G I N R P L I K P O R R L S A S R R V G L S [C] 270
 811 GCCAACTGCCAGACTACCACCACCACTGTGGCGCGTAACGCTGAGGGCGAGCCTGTTTGCATGCTGGCGCCTTACATGAAGCTC
 A N [C] O T T T T L W R R N A E G E P V [C] N A [C] G L Y M K L 300
 901 CATGGGTTCCAGGCTCTTGAATGCGGAAGGAGGGGATTCAAACAGAAAACGGAAGCCCAAGAACTGAATAAATCAAGACACCA
 H G V P R P L A M R K E G I O T R K R K P K N L N K S K T P 330
 991 GCAGGCTCTTGGGAGAGCCTCCCTCCCTCCAGTGGTGCCTCCAGCACTCCAGCAACGCCACCAGCAGCAGCAGCAGTGAAGAG
 A G P P G E S L P P S S G A S S N S S N A T S S S S S S E E 360
 1081 ATGCGCCCATCAAGACAGAGCCCGGCTGTCTACTACTATGGGCACAGCAGCTCCATGTCCAGACATTAGTACTGTGTCGGCCAT
 M R P I K T E P G L S S H Y G H S S S M S Q T F S T V S G H 390
 1171 GGTCTCCATCCATCCAGTGTGTGGCTCTGAAGTCTCCCCAAGGCTATCCATCTCCTGTCACTCAGACATCAAGGCCAGCTCC
 G S S I H P V L S A L K L S P Q G Y P S P V T Q T S Q A S S 420
 1261 AAGCAGGACTCTTGAACAGCCTGGTCTGGCTGACAGTCAATGGGACATAATCACCGGTAATCAGCACCCCTTCCCTCTTCAAATTC
 K Q D S W N S L V L A D S H G D I I T A * 440
 1351 CTGCTCGGACTTGGGACCTGGGGCCAGCAAAGTAAAAGGCTGGGGCACCTGGCCAGCCCTCTGTGTGGGAACAACCTCTGAAGAACAA
 1441 CTGGTAGAAGCTTGAAGTTGTTGACAATCACTTAGGGATAAGGGTGTGGGTTTCTTCAAATACCTTCCAGGTGGAGCAGTGAAGAAGC
 1531 CTGCATTCTCAGAAAAGCCACCTTGGCTGCAAGCAGCAGCAGTGGGCTCTCAGTGAGCAGGAGACTTCTCCTCTCTTCTTACC
 1621 TGCCCTGACAGGACAGACATAAATCTTCCACCCAGCTCCCAAGCTCCCAAGTGGCTGTGGTGTGATGGGTTTTTCTTTGTGATCCTAGAGTGGC
 1711 TGTAGGACCGAGGCTTCAAGACACTGTCTACAGCCGGCGCAGGGTGTCTACTTGTGGACTAGACATAGCAGCCCTGCCCTTGTCCAACA
 1801 CTCCCTTGAAGCATGGAACATCTCATGTGTCCCATACCTGATCTGACTCCAAAGCGCTGGGTGCAGTGCAGATGTTACTGAGTGTCTCC
 1891 TGGGGGTTTAGGTGAGGAGAAGGCATCACCCATCACACAGATCGCAGCCCGCTCAGAGGTGCCTTCCCTTCTCCAGGAACAGCA
 1981 ACCCCTTGTCTCCAGCTGAACATCTCTACCTCTGTAATAACTAGAGCCAGTGTACAGCAGGCTAATACCCTAGTGTCTGTCTCC
 2071 CAGCATCCATTTCCACCTTCTCACACACAGGGTTCCAAGGAGGAACAAAACCTGTACTAAAGCAGGGTGGTGTGCTGTCTCATCTGC
 2161 ACCTTGGGACCTCTGGAGTTGTGTCTACAGCACAATCTGCCCCAGGACTCCAGCTTACTTGTCTGTAGCCTGCCAAGCAAGCTCT
 2251 CTCAAGTCAGACAGTAACTGGCTCTGGGACTTCTCCAGCAGATCCTTTGTCTAGAAAATACGGACTGTTTGCAAAAACAATTGAGA
 2341 GCAGAAAACAAGGGAAATTTTGAAGGACAAGGGTGTGGATGGAAGAAGATGTCCCGAGGCTGATGGGACAGCCATGCTTGCAG
 2431 TTGTGTGTAAGTGGCAACCCCTCCCTCCCTCAATGAGGGGCTCCCAAAGATTCTTCTTGTCTTCATACCCACAGAGCTGTAGCC
 2521 AACTGTGGCATTACCTTGTTTTGCCCCAAATTTCCAGCCCGCCCTAAACCTTACTGGCGTAGCAGAGAATAGCTTTGAACCAAGATT
 2611 CTGTTGTAATCATTTACGCTGTTTCTCCCTCAAGCCGCTTCTATGCTCCCTCCCAACCCGTTAACATTGCTTAAAGGTGA
 2701 AATGGCTGTAATAATCAGTATTAACTAATAAATTTATCTGTATTACTCTTAA

FIG. 3. Nucleotide sequence of cDNA for rat GATA-GT2 with its deduced amino acid sequence. Nucleotide and amino acid sequences for rat GATA-GT2 are shown as described in Fig. 2. The polyadenylation signal is doubly underlined.

sequences were amplified from pig cDNA, suggesting that these proteins are also expressed in gastric mucosa (data not shown). PCR products corresponding to GATA-GT1 and GATA-GT2 were also amplified from a rat whole-stomach cDNA library.

Structures of GATA-GT1 and GATA-GT2. The rat whole-stomach cDNA library was probed with a mixture of labeled fragments from GATA-GT1, GATA-GT2, and GATA-GT3. Five positive clones were obtained from 4×10^5 colonies. One contained the entire coding region for GATA-GT2, while

the other four clones matched GATA-GT1 but lacked the 5' end of the message. The library was further screened by using the 5' end of one of the partial GATA-GT1 clones, and two positives were obtained carrying the entire coding region. Clones corresponding to GATA-GT3 were not found, possibly because of a lower transcriptional level than those of GATA-GT1 and -GT2.

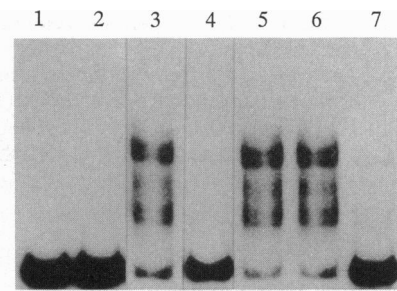
The nucleotide and deduced amino acid sequences for GATA-GT1 (391 amino acid residues) and GATA-GT2 (440

residues) are shown in Figs. 2 and 3, respectively. Like other GATA-binding proteins, the gastric proteins had two distinct zinc-finger motifs following the sequence pattern, Cys-Xaa₂-Cys-Xaa₁₇-Cys-Xaa₂-Cys (23), which were repeated near the middle of the primary structures. The zinc-finger domains of GATA-GT1 and -GT2 were 66–86% identical on the amino acid level with each other and with other GATA-binding proteins, but the remainder of the sequences had lower similarities (<34%). GATA-GT1 had a unique cluster of seven histidine residues (residues 126–132). GATA-GT1 and GATA-GT2 were easily distinguishable from GATA-1, -2, and -3 and from each other. We expect that these proteins constitute two previously unreported groups of the “GATA-binding protein” family.

Despite the differences between the sequences outside of the zinc-finger domains, all of the GATA-binding proteins have common features in their amino acid compositions. They are rich in serine and threonine residues (as high as 20% of the total) and proline residues (≈10%). This high content of α -helix breakers (proline) and β -breakers (serine and proline) suggests that the proteins are quite flexible (33). These characteristics are consistent with their protein–protein and protein–DNA interactions expected in regulation of transcription (34). In the deduced GATA-GT1 and -GT2 sequences (Figs. 2 and 3), the zinc-finger domains have potential phosphorylation sites (32) for cyclic AMP-dependent protein kinase A, protein kinase C, and multifunctional calmodulin-dependent protein kinase II, suggesting kinase-mediated regulation of GATA-binding proteins. However, the differences in the distributions of phosphorylation sites in the different GATA-binding proteins including GATA-1, GATA-2, and GATA-3 (14–24) suggest that each may be regulated differently in carrying out its specific function.

GATA-GT2 Binding to Gastric DNA Sequence Motifs. The zinc-finger region of GATA-GT2 was tested for binding to gastric-specific DNA sequences. A portion of the GATA-GT2 cDNA encoding amino acid residues 128–368 was expressed from the T7 promoter in *E. coli*, and a cell extract was prepared for binding assays with the upstream region of the H⁺/K⁺-ATPase β subunit gene (Fig. 4) (4). Migration of the β -subunit gene fragment (lane 1) was clearly shifted after preincubation with the cell extract containing the recombinant GATA-GT2 protein (lane 3) but not with a control extract (lane 2). Double-stranded fragments containing [(G or C)RR(G or C)NGAT(A or T)RY] from the rat α and β subunit genes (Fig. 4 Lower) (4, 13) also competed for binding to the GATA-GT2 protein (Fig. 4 Upper, lanes 4 and 7), while DNA fragments not containing this sequence did not (lanes 5 and 6). Interestingly, the GATA motif of the α -globin gene (35) also competed for binding (not shown). This motif (CAAC-TGATAAG) is similar to the gastric-specific sequence [(G or C)RR(G or C)NGAT(A or T)RY] (13). The truncated GATA-GT1 protein (residues 127–368) did not bind significantly to the gastric-specific sequence for some unknown reason.

Tissue Distributions of GATA-GT1 and GATA-GT2 mRNAs. Northern blot analysis was used to determine the tissue distributions of GATA-GT1 and -GT2. By using probes specific for GATA-GT1 and GATA-GT2 mRNAs, strong hybridization signals were observed with stomach RNA but not with RNAs from brain, heart, liver, kidney, spleen, and lung (Fig. 5). Weak but significant signals for GATA-GT1 and GATA-GT2 were detected with intestinal RNA. In addition, very weak signals for GATA-GT2 were detected with RNA from testis, whereas hybridization with GATA-GT1 gave no signal even on long exposures of the autoradiograph. Consistent with their possible roles in regulating expression of gastric epithelial-specific proteins such as the H⁺/K⁺-ATPase, the mRNAs for GATA-GT1 and GATA-GT2 were



R α 1 GCCAGCTGATTACAGCTGATTATGAGGCTC
 R α 2 TATACCCAGTTGGGTGAGGGTGGAGCAGAGCCAGCT
 R β 1 GACCAACTGACTTCTGGGACAGTGGAGGA
 R β 2 TGGAGGACAGATAGCACGCAAGCCCCAGCCCTCCCTTATG

FIG. 4. DNA binding of GATA-GT2 protein. The DNA binding of GATA-GT2 was tested by gel mobility-retardation assay. A bacterial lysate (4 μ g) expressing the GATA-GT2 protein was incubated with a radiolabeled DNA probe (3000 cpm) encoding the gastric-specific element of the rat H⁺/K⁺-ATPase β subunit gene [Sty I–Pst I fragment (195 base pairs); ref. 4]. Lanes: 1, radioactive probe without protein; 2, *E. coli* lysate with vector plasmid alone; 3, *E. coli* lysate with GATA-GT2; 4 and 5, rat α subunit gene sequence R α 1 (13) (lane 4) and rat β subunit gene segment R β 2 (4) (lane 5) added as competitors; 6 and 7, Unrelated DNA sequences R α 2 (13) (lane 6) and R β 1 (4) (lane 7), which did not bind GATA-GT2. The nucleotide sequences of R α 1, R α 2, R β 1, and R β 2 are shown. Gastric-specific sequences (13) are underlined. GAT sequences are doubly underlined.

found in the gastric mucosal layer but not in the underlying muscular layer.

Significance of GATA Family Proteins. The predominant expressions of GATA-GT1 and GATA-GT2 in the gastric mucosa suggest that these DNA-binding proteins function in regulating gene expression in the gastric epithelium and may also have roles in development and differentiation of the stomach, intestine, and testis. These roles are clearly different from those of known GATA-binding proteins (14–24). GATA-1 plays a central role in development of erythroid lineages (36, 37) and appears to contribute to the initial

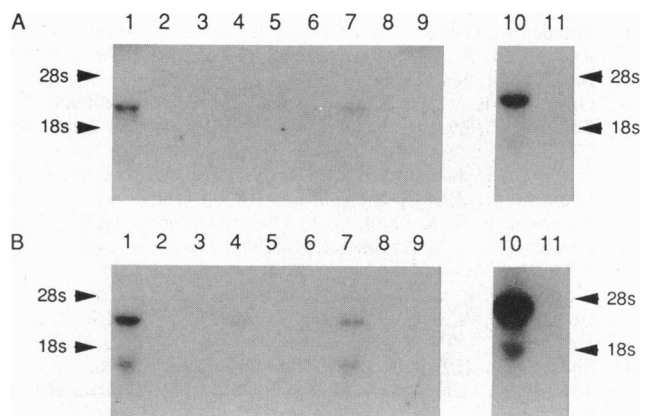


FIG. 5. Tissue distributions of GATA-GT1 and GATA-GT2 mRNAs. Total RNAs (20 μ g) from various rat tissues [stomach, brain, kidney, testis, lung, liver, intestine, spleen, and heart (lanes 1–9, respectively)] and pig stomach [mucosa and muscle layers (lanes 10 and 11, respectively)] were analyzed by Northern blot hybridization. Specific probes for GATA-GT1 (A) (Pst I–EcoRI segment, bases 1028–1570) and GATA-GT2 (B) (Pst I–Pst I segment, bases 160–386) were used. The filter was exposed to x-ray film for 12 hr at -80°C . Longer exposure for 36 hr gave essentially the same results. The positions of ribosomal RNAs (28S and 18S) are shown.

development of progenitor cells as well as regulation of maturation of erythroid cells, mast cells, and megakaryocytes (14, 38). GATA-2 has so far been found in numerous tissues including endothelial, brain, and liver cells (18, 22, 24), and GATA-3 has been found in T-cells, mast cells, brain, and liver (18–21). It is expected that more GATA-binding proteins will be found distributed in various tissues and shown to play roles in tissue-specific transcriptional regulation.

An important question is how a family of DNA-binding proteins that recognize essentially the same sequence regulate transcription in a gene-specific manner. In the gastric mucosa, for example, we found as many as five different GATA-binding proteins. Although the zinc-finger domains are strongly conserved in all GATA-binding proteins, their amino- and carboxyl-terminal regions have little if any shared sequences. Certainly, these parts of the proteins will direct specific protein-protein interactions and perhaps also protein-DNA interactions. A second level of specificity may be mediated through phosphorylation by different protein kinases as described above.

During preparation of this manuscript, the sequence of another GATA-binding protein named GATA-4 was reported (39). GATA-4 is certainly the mouse homolog of the rat GATA-GT2 reported in this paper. The two proteins were 100% identical in the zinc-finger domain and 77% identical in other regions. Mouse GATA-4 was reported to be expressed in heart, intestine, and testis, but its expression in stomach was not examined. As described above, rat GATA-GT2 (as well as GATA-GT1) was expressed in the gastric mucosa at much higher levels than in the intestine or testis (Fig. 5). These results indicate that GATA-GT2 and GATA-GT1 regulate transcription specifically in the gastric mucosa.

We thank Dr. R. K. Nakamoto for helpful discussion and critical reading of the manuscript. This research was supported by grants from the Ministry of Education, Science and Culture of Japan and a Human Frontier Science Program.

- Saccomani, G., Helander, H. F., Crago, S., Chang, H. H., Daily, D. W. & Sachs, G. (1979) *J. Cell Biol.* **83**, 271–283.
- Faller, L., Jackson, R., Malinowska, D., Mukidjam, E., Rabon, E., Saccomani, G., Sachs, G. & Smolka, A. (1982) *Ann. N.Y. Acad. Sci.* **402**, 146–163.
- Maeda, M., Oshiman, K., Tamura, S. & Futai, M. (1990) *J. Biol. Chem.* **265**, 9027–9032.
- Maeda, M., Oshiman, K., Tamura, S., Kaya, S., Mahmood, S., Reuben, M. A., Lasater, L. S., Sachs, G. & Futai, M. (1991) *J. Biol. Chem.* **266**, 21584–21588.
- Oshiman, K., Motojima, K., Mahmood, S., Shimada, A., Tamura, S., Maeda, M. & Futai, M. (1991) *FEBS Lett.* **281**, 250–254.
- Newman, P. R., Greb, J., Keeton, T. P., Reyes, A. A. & Shull, G. E. (1990) *DNA Cell Biol.* **9**, 749–762.
- Newman, P. R. & Shull, G. E. (1991) *Genomics* **11**, 252–262.
- Canfield, V. A. & Levenson, R. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 8247–8251.
- Morley, G. P., Callaghan, J. M., Rose, J. B., Toh, B. H., Gleeson, P. A. & van Driel, I. R. (1992) *J. Biol. Chem.* **267**, 1165–1174.
- Shull, G. E. (1990) *J. Biol. Chem.* **265**, 12123–12126.
- Toh, B.-H., Gleeson, P. A., Simpson, R. J., Moritz, R. L., Callaghan, J. M., Goldkorn, I., Jones, C. M., Martinelli, T. M., Mu, F.-T., Humphris, D. C., Pettitt, J. M., Mori, Y., Masuda, T., Sobieszczuk, P., Weinstock, J., Mantamadiotis, T. & Baldwin, G. S. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 6418–6422.
- Canfield, V. A., Okamoto, C. T., Chow, D., Dorfman, J., Gros, P., Forte, J. G. & Levenson, R. (1990) *J. Biol. Chem.* **265**, 19878–19884.
- Tamura, S., Oshiman, K., Nishi, T., Mori, M., Maeda, M. & Futai, M. (1992) *FEBS Lett.* **298**, 137–141.
- Tsai, S.-F., Martin, D. I. K., Zon, L. I., D'Andrea, A. D., Wong, G. G. & Orkin, S. H. (1989) *Nature (London)* **339**, 446–451.
- Evans, T. & Felsenfeld, G. (1989) *Cell* **58**, 877–885.
- Zon, L. I., Tsai, S.-F., Burgess, S., Matsudaira, P., Bruns, G. A. P. & Orkin, S. H. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 668–672.
- Trainor, C. D., Evans, T., Felsenfeld, G. & Boguski, M. S. (1990) *Nature (London)* **343**, 92–96.
- Yamamoto, M., Ko, L. J., Leonard, M. W., Beug, H., Orkin, S. H. & Engel, J. D. (1990) *Genes Dev.* **4**, 1650–1662.
- Ko, L. J., Yamamoto, M., Leonard, M. W., George, K. M., Ting, P. & Engel, J. D. (1991) *Mol. Cell. Biol.* **11**, 2778–2784.
- Ho, I.-C., Vorhees, P., Marin, N., Oakley, B. K., Tsai, S.-F., Orkin, S. H. & Leiden, J. M. (1991) *EMBO J.* **10**, 1187–1192.
- Joulin, V., Bories, D., Eléouet, J.-F., Labastie, M.-C., Chrétiens, S., Mattéi, M.-G. & Roméo, P.-H. (1991) *EMBO J.* **10**, 1809–1816.
- Lee, M.-E., Temizer, D. H., Clifford, J. A. & Quertermous, T. (1991) *J. Biol. Chem.* **266**, 16188–16192.
- Zon, L. I., Mather, C., Burgess, S., Bolce, M. E., Harland, R. M. & Orkin, S. H. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 10642–10646.
- Dorfman, D. M., Wilson, D. B., Bruns, G. A. P. & Orkin, S. H. (1992) *J. Biol. Chem.* **267**, 1279–1285.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed.
- Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. & Erlich, H. A. (1988) *Science* **239**, 487–491.
- Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. & Roe, B. A. (1980) *J. Mol. Biol.* **143**, 161–178.
- Tabor, S. & Richardson, C. C. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 1074–1078.
- Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. (1990) *Methods Enzymol.* **185**, 60–89.
- Singh, H., Lebowitz, J. H., Baldwin, A. S. & Sharp, P. A. (1988) *Cell* **52**, 415–423.
- Bradford, M. M. (1976) *Anal. Biochem.* **72**, 248–254.
- Kemp, B. E. & Pearson, R. B. (1990) *Trends Biochem. Sci.* **15**, 342–346.
- Chou, P. Y. & Fasman, G. D. (1978) *Adv. Enzymol.* **47**, 45–148.
- Pugh, B. F. & Tjian, R. (1992) *J. Biol. Chem.* **267**, 679–682.
- Brady, H. J. M., Sowden, J. C., Edwards, M., Lowe, N. & Butterworth, P. H. W. (1989) *FEBS Lett.* **257**, 451–456.
- Pevny, L., Simon, M. C., Robertson, E., Klein, W. H., Tsai, S.-F., D'Agati, V., Orkin, S. H. & Costantini, F. (1991) *Nature (London)* **349**, 257–260.
- Simon, M. C., Pevny, L., Wiles, M. V., Keller, G., Costantini, F. & Orkin, S. H. (1992) *Nat. Genet.* **1**, 92–98.
- Martin, D. I. K., Zon, L. I., Mutter, G. & Orkin, S. H. (1990) *Nature (London)* **344**, 444–447.
- Arceci, R. J., King, A. A. J., Simon, M. C., Orkin, S. H. & Wilson, D. B. (1993) *Mol. Cell. Biol.* **13**, 2235–2246.