Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 2478–2482, March 1994
Pharmacology

Molecular cloning of bullfrog saxiphilin: A unique relative of the transferrin family that binds saxitoxin

(neurotoxin/sodium channel/thyroglobulin domain/amphibian/molecular evolution)

MARIA A. MORABITO* AND EDWARD MOCZYDLOWSKI*†

Departments of *Pharmacology and †Cellular and Molecular Physiology, Yale University School of Medicine, New Haven, CT 06510

Communicated by John H. Law, December 15, 1993 (received for review October 27, 1993)

ABSTRACT **Plasma and tissues of certain vertebrates contain a protein called saxiphilin that specifically binds the neurotoxin saxitoxin with nanomolar affinity. We describe the isolation of a cDNA clone of saxiphilin from liver of the North American bullfrog (*Rana catesbeiana*). The cDNA sequence encodes a protein that is evolutionarily related to members of the transferrin family of Fe^{3+}-binding proteins. Pairwise sequence alignment of saxiphilin with various transferrins reveals amino acid identity as high as 51% and predicts 14 disulfide bonds that are highly conserved. The larger size of saxiphilin (91 kDa) versus serum transferrin (≈78 kDa) is primarily due to a unique insertion of 144 residues. This insertion contains a 49-residue domain classified as a type 1 repetitive element of thyroglobulin, which is shared by a variety of membrane, secreted, and extracellular matrix proteins. Saxiphilin also differs from transferrins in 9 of 10 highly conserved amino acids in the two homologous Fe^{3+}/HCO_3-binding sites of transferrin. Identification of saxiphilin implies that transferrin-like proteins comprise a diverse superfamily with functions other than iron binding.**

Saxitoxin (STX) is tricyclic organic molecule that is produced by various dinoflagellates and cyanobacteria (1). STX ranks among the most potent paralytic neurotoxins by virtue of its nanomolar blocking effect on voltage-sensitive Na^+ channels of neurons and skeletal muscle. In previous studies using [^{3}H]STX to measure STX-binding sites of Na^+ channels, an unusual high-affinity binding site ($K_d \approx 0.2$ nM) was found in soluble extracts of frog heart (2) and skeletal muscle (3). Further work showed that this soluble STX-binding site is associated with an ≈90-kDa monomeric protein (named saxiphilin) that is present in bullfrog plasma at a concentration of ≈300 nM (4, 5). Partial amino acid sequences of purified saxiphilin were found to exhibit similarity to members of the transferrin family (5). However, saxiphilin is biochemically and functionally distinct from bullfrog serum transferrin (6). Transferrins are a family of ≈80-kDa proteins noted for their exceptionally high affinity for Fe^{3+} with a K_d in the range of 10^{-20} M (7). By transporting Fe^{3+} into eukaryotic cells through binding to the transferrin receptor and subsequent endocytosis, serum transferrin functions as an important growth factor required for synthesis of Fe^{3+}-containing proteins.

This paper describes the isolation of a cDNA clone encoding saxiphilin from bullfrog liver.‡ Sequence analysis indicates that saxiphilin is an evolutionary relative of the transferrin family but differs in two major respects. Saxiphilin has substitutions of 9 of the 10 highly conserved residues that form the two Fe^{3+}/HCO_3-binding sites of transferrin. It also has a unique insertion of 144 residues that contains a type 1 thyroglobulin domain (Thyr-1). These findings lead to the conclusion that saxiphilin originated from an ancestor of the transferrin family but diverged to perform a different function. The unique ability of saxiphilin to bind STX and the similarity of its tissue distribution with that of transferrin suggest that its physiological role may be to transport or sequester an endogenous organic molecule rather than Fe^{3+}

MATERIALS AND METHODS

Isolation of a cDNA Corresponding to Saxiphilin. Adult bullfrogs (*Rana catesbeiana*) were purchased from Connecticut Valley Biological Supply (Southampton, MA). Total RNA from bullfrog liver was prepared (8) and further purified on a CsCl step gradient (9). First strand cDNA was synthesized using $(dT)_{15}$ primer and murine leukemia virus reverse transcriptase according to recommendations of the manufacturer (GIBCO/BRL). The following degenerate oligonucleotides, X, Y, and Z, were designed from tryptic peptides of saxiphilin (5), Sax-133 (X and Y) and Sax-101 (Z), and they were synthesized by the Yale Medical School Protein and Nucleic Acid Facility: X sense (CAA/GTAT/CATGTAT/CGAA/GGCIC/TTIATGTGT/CGG), Y sense (GAA/GTAT/CCAT/CAAT/CAAA/GGAT/CGAT/CTTT/CGGICC), and Z antisense (CCA/GTCIGTA/GTTT/CTCA/GAAIACIACIGTG/ATCIGG). X and Z were first used as primers for PCR using the oligo(dT)-tailed cDNA as a template. The reaction was run for 30 cycles with 5 μM each of X and Z. The cycle was 1 min of denaturation at 94°C, 1 min of annealing at 47°C, and 2 min of extension at 72°C. The final cycle included an extension of 7 min at 72°C. A PCR product of ≈450 bp was obtained by reamplification of the latter reaction mixture using Y and Z by nested PCR run under the same conditions. All PCR reactions used Amplitaq DNA polymerase (Perkin–Elmer/Cetus). The ≈450-bp product was purified on a 2% agarose gel and cloned in the plasmid vector pCR1000 (Invitrogen). A number of clones were isolated and sequenced (10) using Sequenase (United States Biochemical).

One of the PCR-derived clones containing saxiphilin sequence was used as a hybridization probe to screen a bullfrog liver cDNA library. Poly(A)$^+$ RNA was isolated by oligo(dT)-cellulose chromatography of total RNA extracted from liver of adult bullfrog and used to synthesize double-stranded cDNA using the ZAP-cDNA kit from Stratagene. The cDNA was cloned in the Lambda ZAPII vector following the Stratagene protocol. Five micrograms of poly(A)$^+$ RNA yielded about 2.4×10^6 recombinant phages prior to amplification. Five hundred thousand recombinant phages were screened using a *Bam*HI–*Eco*RI fragment isolated from the PCR-derived cDNA clone according to the map in Fig. 1. The screening of the library was performed by standard proce-

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

Abbreviations: STX, saxitoxin; Thyr-1, type 1 thyroglobulin domain.

‡The sequence reported in this paper has been deposited in the GenBank data base (accession no. U05246).

dures (9). One of the positive clones isolated from the screen contained the DNA sequence shown in Fig. 2.

Analysis of Saxiphilin mRNA. Five micrograms of poly(A)$^+$ RNA from bullfrog liver was fractionated by electrophoresis on a 1% agarose gel containing formaldehyde (8) and transferred to a Hybond-N$^+$ nylon membrane (Amersham). The filter was first hybridized to probe 1 (Fig. 5*a*, lane 1) and then stripped and hybridized to probe 2 (Fig. 5*a*, lane 2). Probe 1 was a 2.25-kb *Sma* I fragment derived from the saxiphilin cDNA clone (Fig. 1). Probe 2 was a 0.43-kb fragment specific to the unique insertion sequence in saxiphilin (Fig. 1) that was prepared by PCR amplification of nucleotides 303–738 in Fig. 2. The probes were labeled with [^{32}P]dCTP using a randomly primed labeling kit (Boehringer Mannheim). Alkali fixation, prehybridization, and removal of the first probe were performed according to Amersham. Autoradiography of the filters was performed by exposure to film for 1 hr with an intensifying screen. Probes 1 and 2 were similarly used to detect saxiphilin message in total RNA extracted from various bullfrog tissues in a slot blot hybridization assay. Hybridization was performed at 65°C and the final wash was in 1× SSC/0.1% SDS for 1 hr at 50°C (Fig. 5*a*) or 0.2× SSC (Fig. 5*b*).

RESULTS AND DISCUSSION

Identification of Saxiphilin cDNA. Cloning of saxiphilin was accomplished by combining PCR amplification of an 0.45-kb fragment of cDNA with screening of a bullfrog liver cDNA library. One of the positive clones from the library contained a 2681-bp cDNA with an open reading frame of 845 amino acids. Fig. 1 shows a restriction map of the cDNA clone and the location of the PCR-amplified probe used to screen the library. The nucleotide sequence of the cDNA clone and the deduced amino acid sequence of saxiphilin are shown in Fig. 2. The clone contains a 5′ untranslated region of 23 bp followed by an ATG codon for Met-1 and a TAA termination codon following Cys-845. The N terminus of the mature protein begins at Ala-20 as recognized by a 25-residue sequence previously obtained by Edman degradation of the intact native protein (5). The 19-residue sequence preceding Ala-20 corresponds to a secretory signal sequence as found for all known transferrins (7). The 3′ end of the clone contains a consensus sequence (AATAAA) for polyadenylylation that

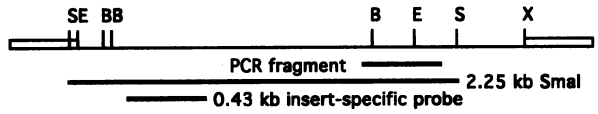

FIG. 1. Restriction map of saxiphilin cDNA. Solid bars indicate a PCR fragment (0.45 kb) used to isolate a full-length cDNA clone and probes corresponding to a *Sma* I restriction fragment and a saxiphilin-specific insertion used in mRNA hybridization experiments. Restriction sites are *Bam*HI (B), *Eco*RI (E), *Sma* I (S), and *Xho* I (X).

```
                                                                -23 GGCACGAGGAGGCACTGAGAAGG
    ATGGCTCCGACTTTCCAAACAGCTCTGTTTTTCACCATCATTAGCTTGAGCTTTGCGGCACCAAATGCAAAACAAGTCCGGTGGTGTGCGATATCAGATCTTGAGCAGAAG 111
  1 M  A  P  T  F  Q  T  A  L  F  F  T  I  I  S  L  S  F  A  A  P  N  A  K  Q  V  R  W  C  A  I  S  D  L  E  Q  K
    AAGTGCAATGATCTAGTGGGATCCTGCAATGTCCCGGACATTACCCTTGTGTGTGTCTTAGATCCTCCACTGAGGACTGTATGACAAGCTATAAAGGATGGACAAGCAGAT 222
 38 K  C  N  D  L  V  G  S  C  N  V  P  D  I  T  L  V  C  V  L  R  S  S  T  E  D  C  M  T  A  I  K  D  G  Q  A  D
    GCCATGTTCTTAGATAGTGGAGAGGTTTATGAAGCTTCAAAGGATCCTTATAACCTAAAGCCAATTATTGCAGAACCATACAGTTCCAATAGAGACTTGCAAAAATGCTTA 333
 75 A  M  F  L  D  S  G  E  V  Y  E  A  S  K  D  P  Y  N  L  K  P  I  I  A  E  P  Y  S  S  N  R  D  L  Q  K  C  L
    AAGGAGCGTCAACAAGCTCTGGCAAAGAAGATGCATCTGGTCATTATATTCCACAATGTGATGAAAAGGGCAATTACCAACCACAGCAGTGCCATGGCAGCACTGGGCACT 444
112 K  E  R  Q  Q  A  L  A  K  K  M  H  L  V  I  I  F  H  N  V  M  K  R  A  I  T  N  H  S  S  A  M  A  A  L  G  T
    GCTGGTGTGTTAACGTCAATGGGTGAAAAGATTTCTGGAACAAATACTCCACCTGGACAAACACGAGCTACATGTGAAAGGCATGAATTGCCAAAATGCTTGAAGGAACGC 555
149 A  G  V  L  T  S  M  G  E  K  I  S  G  T  N  T  P  P  G  Q  T  R  A  T  C  E  R  H  E  L  P  K  C  L  K  E  R
    CAGGTAGCTCTGGGAGGAGACGAGAAAGTGTTAGGTCGTTTTGTTCCTCAGTGTGATGAAAAGGGCAATTATGAACCACAACAGTTCCATGGCAGCACTGGGTACAGCTGG 666
186 Q  V  A  L  G  G  D  E  K  V  L  G  R  F  V  P  Q  C  D  E  K  G  N  Y  E  P  Q  Q  F  H  G  S  T  G  Y  S  W
    TGTGTTAATGCAATCGGTGAAGAGATTGCTGGAACAAAAACTCCACCTGGAAAAATACCAGCTACATGTCAAAAACACGATTTGGTTACTACATGTCATTACACAGTTGCT 777
223 C  V  N  A  I  G  E  E  I  A  G  T  K  T  P  P  G  K  I  P  A  T  C  Q  K  H  D  L  V  T  T  C  H  Y  T  V  A
    ATGGTGAAAAAGTCCAGCGCGTTCCAGTTTAATCAGCTGAAAGGCAAGAGGTCCTGTCACTCAGGTGTATCAAAGACAGATGGCTGGAAAGCTCTTGTGACTGTTCTTGTT 888
260 M  V  K  K  S  S  A  F  Q  F  N  Q  L  K  G  K  R  S  C  H  S  G  V  S  K  T  D  G  W  K  A  L  V  T  V  L  V
    GAAAAGAAGTTGTTGTCCTGGGATGGACCTGCCAAAGAATCCATTCAACGAGCCATGTCAAAGTTCTTCTCAGTCAGCTGTATCCCTGGAGCAACCCAAACCAACCTCTGC 999
297 E  K  K  L  L  S  W  D  G  P  A  K  E  S  I  Q  R  A  M  S  K  F  F  S  V  S  C  I  P  G  A  T  Q  T  N  L  C
    AAGCAATGTAAGGGAGAGGAAGGGAAGAACTGTAAGAACAGCCATGATGAGCCATACTATGGCAACTATGGAGCATTCAGATGTTTGAAAGAAGATATGGGAGATGTAGCT 1110
334 K  Q  C  K  G  E  E  G  K  N  C  K  N  S  H  D  E  P  Y  Y  G  N  Y  G  A  F  R  C  L  K  E  D  M  G  D  V  A
    TTTCTGAGAAGCACAGCTTTGTCAGATGAACATTCAGAGGTGTATGAGCTTCTGTGTCCAGACAATACTAGGAAGCCCCTTAACAAATACAAGGAGTGTAACCTAGGAACA 1221
371 F  L  R  S  T  A  L  S  D  E  H  S  E  V  Y  E  L  L  C  P  D  N  T  R  K  P  L  N  K  Y  K  E  C  N  L  G  T
    GTTCCAGCAGGTACAGTGGTGACCAGAAAAATCAGTGACAAGACTGAGGACATCAATAACTTTCTCATGGAGGCTCAGAAAAGGCAATGCAAATTGTTCAGTTCTGCACAT 1332
408 V  P  A  G  T  V  V  T  R  K  I  S  D  K  T  E  D  I  N  N  F  L  M  E  A  Q  K  R  Q  C  K  L  F  S  S  A  H
    GGGAAAGATTTGATGTTTGATGATTCTACCTTACAATTAGCTCTTCTCTCATCTGAAGTGGATGCGTTCCTCTATCTCGGAGTAAAATTGTTCCATGCCATGAAAGCACTG 1443
445 G  K  D  L  M  F  D  D  S  T  L  Q  L  A  L  L  S  S  E  V  D  A  F  L  Y  L  G  V  K  L  F  H  A  M  K  A  L
    ACTGGAGATGCACATCTTCCATCCAAAAATAAAGTGCGGTGGTGTACAATAAATAAACTGGAAAAGATGAAGTGTGATGATTGGTCAGCTGTAAGCGGTGGGGCCATTGCA 1554
482 T  G  D  A  H  L  P  S  K  N  K  V  R  W  C  T  I  N  K  L  E  K  M  K  C  D  D  W  S  A  V  S  G  G  A  I  A
    TGCACAGAGGCATCCTGTCCAAAGGGTTGTGTTAAACAGATTCTGAAAGGTGAAGCTGATGCAGTGAAACTTGAGGTACAATACATGTATGAGGCTTTGATGTGCGGACTG 1665
519 C  T  E  A  S  C  P  K  G  C  V  K  Q  I  L  K  G  E  A  D  A  V  K  L  E  V  Q  Y  M  Y  E  A  L  M  C  G  L
    CTGCCAGCAGTAGAAGAATACCACAATAAAGATGATTTTGGGCCCTGTAAAACCCCTGGATCCCCATACACAGATTTTGGGACACTGCGTGCTGTAGCCTTGGTAAAAAAA 1776
556 L  P  A  V  E  E  Y  H  N  K  D  D  F  G  P  C  K  T  P  G  S  P  Y  T  D  F  G  T  L  R  A  V  A  L  V  K  K
    AGCAACAAAGACATCAACTGGAACAACATTAAAGGCAAGAAGTCCTGCCATACTGGTGTTGGTGATATTGCTGGCTGGGTTATCCCTGTTAGTCTTATTAGAAGGCAGAAT 1887
593 S  N  K  D  I  N  W  N  N  I  K  G  K  K  S  C  H  T  G  V  G  D  I  A  G  W  V  I  P  V  S  L  I  R  R  Q  N
    GACAATTCTGATATTGATTCCTTCTTCGGTGAGAGCTGTGCTCCTGGATCAGATACTAAATCCAACCTCTGTAAACTGTGCATTGGTGACCCCAAGAATTCCGCGGCCAAT 1998
630 D  N  S  D  I  D  S  F  F  G  E  S  C  A  P  G  S  D  T  K  S  N  L  C  K  L  C  I  G  D  P  K  N  S  A  A  N
    ACAAAATGTTCTCTCAGTGATAAGGAGGCCTATTATGGAAACCAAGGTGCCTTTAGATGTCTGGTGGAAAAAGGAGATGTGGCATTTGTGCCTCACACTGTTGTATTTGAA 2109
667 T  K  C  S  L  S  D  K  E  A  Y  Y  G  N  Q  G  A  F  R  C  L  V  E  K  G  D  V  A  F  V  P  H  T  V  V  F  E
    AACACAGATGGTAAAAATCCAGCAGTGTGGGCAAAAAATTTGAAATCAGAAGATTTTGAACTATTATGTTTGGATGGCTCACGTGCCCCTGTCAGTAATTACAAGAGCTGC 2220
704 N  T  D  G  K  N  P  A  V  W  A  K  N  L  K  S  E  D  F  E  L  L  C  L  D  G  S  R  A  P  V  S  N  Y  K  S  C
    AAACTTTCAGGCATCCCACCTCCCGCCATTGTCACCCGGGAAGAGAGCATCAGTGATGTTGTAAGAATTGTTGCTAATCAACAGTCACTGTATGGACGGAAGGGATTTGAA 2331
741 K  L  S  G  I  P  P  P  A  I  V  T  R  E  E  S  I  S  D  V  V  R  I  V  A  N  Q  Q  S  L  Y  G  R  K  G  F  E
    AAAGATATGTTCCAGTTGTTTTCTTCAAATAAAGGCAACAACCTCCTCTTCAATGACAACACTCAGTGCCTGATTACATTTGATAGACAACCAAAAGACATTATGGAGGAT 2442
778 K  D  M  F  Q  L  F  S  S  N  K  G  N  N  L  L  F  N  D  N  T  Q  C  L  I  T  F  D  R  Q  P  K  D  I  M  E  D
    TACTTTGGGAAACCCTACTACACTACAGTGTATGGTGCGAGCAGATCTGCCATGTCATCAGAACTGATATCTGCCTGCACAATTAAACACTGTTAAGAAGATCTAGCAAAC 2553
815 Y  F  G  K  P  Y  Y  T  T  V  Y  G  A  S  R  S  A  M  S  S  E  L  I  S  A  C  T  I  K  H  C  *
    ATTTGTGCTATAACTTTCTGCCTGGTTGGCACGTTGATCTTCTTTTTAATCCATATATCACTTCTGTTTTCTTTTTGTTTTTAAATAAATATTTAGCCTACCATG(A)18 2676
```

FIG. 2. Nucleotide sequence and deduced amino acid sequence of a cDNA clone of saxiphilin from bullfrog liver. Solid underlines mark the amino acid sequences of six fragments of native saxiphilin identified previously (5).

is 22 bp upstream of the poly(A) tail. Identification of the cloned sequence as saxiphilin was confirmed by finding the sequences of all five tryptic fragments (underlined in Fig. 2) previously reported (5) as well as the N-terminal sequence of the native protein. The predicted molecular weight of the 826-residue mature protein is 90,818, which is in good agreement with that of the native protein (90,000 ± 3000) as estimated by polyacrylamide gel electrophoresis (5, 6).

Homology between Saxiphilin and Transferrins. A sequence comparison search (11, 12) (SwissProt and GenBank data base) using the deduced amino acid sequence of saxiphilin established an evolutionary relationship to the transferrin family of Fe^{3+}-binding proteins. This is illustrated in Fig. 3 by an alignment of saxiphilin, serum transferrin from the African clawed frog (*Xenopus laevis*) (13), and human serum transferrin (14). Considerable sequence similarity is found throughout the whole alignment except for a 144-residue insertion that occurs after Gln-89 of saxiphilin. If this large insertion is considered as a gap, pairwise sequence alignments (17) between saxiphilin and various transferrins yield values of 28%, 44%, and 51% identity with transferrin from tobacco hornworm (*Manduca sexta*) (18), human serum (14), and *X. laevis* (13), respectively.

Another key feature identifying saxiphilin as an evolutionary relative of transferrin is the presence of internal duplication between residues 20–487 and 488–845 (Met-1 numbering) as detected by dot plot analysis (19) (not shown). The transferrin protein family is characterized by similarity between the first ≈350 N-terminal residues and the last ≈350 C-terminal residues, which indicates that the protein arose from an intragenic duplication (7). This internal duplication is the basis of the bilobal tertiary structure and similar secondary structure folding pattern of the N-terminal half (N lobe) and C-terminal half (C lobe) of human lactoferrin (15) and rabbit serum transferrin (16) as deduced by x-ray crystallography. Excluding the 144-residue insertion, saxiphilin contains the same percentage of internal similarity as *Xenopus* transferrin (38% identity), but somewhat less than that exhibited by various human transferrins (44–48% identity). The transferrin family can also be recognized by a large number of conserved disulfide bonds (7, 16, 20). In human lactoferrin six disulfide bonds appear at homologous positions in the N lobe and C lobe. All 12 of these disulfide bonds appear to be conserved in saxiphilin as identified by sequence alignment. In Fig. 3 these are labeled *a–f* and *a′–f′* in the putative N and C lobe regions of the sequence, respectively. Human lactoferrin and several other transferrins also contain four additional disulfides in the C lobe that are not present in the N lobe. Two of these can be identified in saxiphilin, labeled as *g′* and *h′* in Fig. 3.

In most known transferrins, both the N lobe and C lobe domains contain a high-affinity ($K_d \approx 10^{-20}$ M) binding site for Fe^{3+}. X-ray crystallography (15, 16) and sequence analysis have previously shown that ligand residues in these two Fe^{3+}-binding sites are highly conserved (7, 20). In each lobe, Fe^{3+} is coordinated by the same four residues: Asp-63 (Asp-392), Tyr-95 (Tyr-426), Tyr-188 (Tyr-517), and His-249 (His-585), with sequence numbers corresponding to human serum transferrin in the N lobe (C lobe), respectively. Physiological binding of Fe^{3+} to these two sites in transferrin is also known to require bicarbonate anion (HCO_3^-), which appears to bridge (15) between Fe^{3+} and the highly conserved

```
      -19                  1         a         b                                 b         a          *
Sax   MAPTFQTALF FTIISLSFAA PNAKQVRWCA ISDLEQKKC- ---NDLVGSC NVPDITLVCV LRSSTEDCMT AIKDGQADAM FLDSGEVYEA SKDPYNLKPI   77
Xltf  MDFSLRVALC LSMLALCLAI QKEKQVRWCV KSNSELKKC- ---KDLVDTC KNKEIKLSCV EKSNTDECSL LFRKTMQMQF VWTGGDVYKG SLQPYNLKPI   77
Hstf  MRLAVGALLV CAVLGLCLAV P-DKTVRWCA VSEHEATKCQ SFRDHMKSVI PSDGPSVACV KKASYLDCIR AIAANEADAV TLDAGLVYDA YLAPNNLKPV   80

                                            !
Sax   IAEPYSSNRD LQKCLKERQQ ALAKKMHLVI IFHNVMKRAI TNHSSAMAAL GTAGVLTSMG EKISGTNTPP GQTRATCERH ELPKCLKERQ VALGGDEKVL  177
Xltf  MAENYGSHTE TD-------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------   89
Hstf  VAEFYGSKED PQ-------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------   92

                                                             *                    c     *
Sax   GRFVPQCDEK GNYEPQQFHG STGYSWCVNA IGEEIAGTKT PPGKIPATCQ KHDLVTTCHY TVAMVKKSSA FQFNQLKGKR SCHSGVSKTD GWKALVTVLV  277
Xltf  ---------- ---------- ---------- ---------- ---------- ------TCYY AVAVVKKSSK FTFDELKDKK SCHTGIGKTA GWNIIIGLLL  133
Hstf  ---------- ---------- ---------- ---------- ---------- ------TFYY AVAVVKKDSG FQMNQLRGKK SCHTGLGRSA GWNIPIGLL-  135

                                  d          e   d        e           *   c
Sax   EKKLLSWDGP AKESIQRAMS KFFSVSCIPG ATQ---TNLC KQCKGEEGKN CKNSHDEPYY GNYGAFRCLK EDMGDVAFLR STALSDEHS- ----EVYELL  369
Xltf  ERKLLKWAGP DSETWRNAVS KFFKASCVPG AKE---PKLS QLCAGIKEHK CSRSNNEPYY NYAGAFKCLQ DDQGDVAFVK QSTVPEEFH- ----KDYELL  225
Hstf  ---YCDLPEP -RKPLEKAVA NFFSGSCAPC ADGTDFPQLC QLCPG----- CGCSTLNQYF GYSGAFKCLK DGAGDVAFVK HSTIFENLAN KADRDQYELL  226

      f             f       *
Sax   CPDNTRKPLN KYKECNLGTV PAGTVVTRKI SDKTEDINNF LMEAQ----- -KRQ-CKLFS SAHG-KDLMF DDSTLQLALL SSEVDAFLY- -LGV---KLF  456
Xltf  CPDNTRKSIK EYKNCNLAKV PAHAVLTRGR DDKSKDIIEF LQEAQ----- -KTQECKLFR LPG-----MG KGSNFQGQRS ESYSPPIFYG QFSVPRSRLF  314
Hstf  CLDNTRKPVD EYKDCHLAQV PSHTVVARSM GGKEDLIWEL LNQAQEHFGK DKSKEFQLFS SPHG-KDLLF KDSAHGFLKV PPRMDAKMYL GYE-----YV  320

                            a'         b'          b'         a'                  *          g'               h'
Sax   HAMKALTG-- -DAHLPSKNK -VRWCTINKL EKMKCDDWSA VSGGAIACTE ASCPKGCVKQ ILKGEADAVK LEVQYMYEAL MCGLLPAVEE YHNKDDFGPC  552
Xltf  QCIQALKEGV KEDDSAAQVK -VRWCTQSKA EKTKCDDWTT ISGGAIECTE ASTAEECIVQ ILKGDADAVT LDGGYMYTAG LCGLVPVMGE YYDQDDLTPC  413
Hstf  TAIRNLREGT CPEAPTDECK PVKWCALSHH ERLKCDEWSV NSVGKIECVS AETTEDCIAK IMNGEADAMS LDGGFVYIAG KCGLVPVLAE NYNKSD--NC  418

                 *                      c'     *                                 d'          e' d'            e'
Sax   KTPGSPYTDF GTLRAVALVK KSNKDINWNN IKGKKSCHTG VGDIAGWVIP VSLIRRQNDN SDIDSFFGES CAPGSDTKSN LCKLCIGDP- KNSAANTKCS  651
Xltf  QRSCSQAK-- GVYYAVAIVK KG-TQVSWSN LRGVKTCHTA VGRTAGWNIP VGLITSETAN CDFASYVGES CAPGSDVKSN LCALCIGDPE KLSEREKKCS  510
Hstf  ED-----TPE AGYFAVAVVK KSASDLTWDN LKGKKSCHTA VGRTAGWNIP MGLLYNKINH CRFDEFFSEG CAPGSKKDSS LCKLCMGS-- ----GLNLCE  507

         *       c'                                       f'             f'     *
Sax   LSDKEAYYGN QGAFRCLVEK -GDVAFVPHT VVFENTDGKN PAVWAKNLKS EDFELLCLDG SRAPVSNYKS CKLSGIPPPA IVTREESISD VVRIVANQQS  750
Xltf  PSASEAYYGY SGAFRCLVEK -GQVGFAKHT TVFENTDGKN PAGWAKDLKS EDFELLCPDG SRAPVTDYKR CNLAEVPAHA VVTLPDKREQ VAKIVVNQQS  609
Hstf  PNNKEGYYGY TGAFRCLVEK -GDVAFVKHQ TVPQNTGGKN PDPWAKNLNE KDYELLCLDG TRKPVEEYAN CHLARAPNHA VVTRKDKEAC VHKILRQQQH  606

                                h'                                          g'
Sax   LYGRKGFE-K DMFQLFSSNK GNNLLFNDNT QCLITFDRQP KDIMEDYFGK PYYTTVYGAS RSAMSSELIS ACTIKHC                          826
Xltf  LYGRKGFQ-K DIFQMFQSTG GKDLLFKDST QCL--LEIPS KTTMQEFLGD KYHTAVTSLN KCSTSNEASW LPAQFHSCMK IYMIVDCPLS II       698
Hstf  LFGSNVTDCS GNFCLFRS-E TKDLLFRDDT VCLAKL--HD RNTYEKYLGE EYVKAVGNLR KCSTSS-LLE ACTFRRP                          679
```

FIG. 3. Homology relationships of saxiphilin. The amino acid sequence of saxiphilin (Sax) is aligned with *X. laevis* transferrin (Xltf) (13) and human serum transferrin (Hstf) (14). The alignment was produced by using the PILEUP program of the Genetics Computer Group analysis package (12). Gaps in the alignment are shown as a hyphen (–). Residues that are identical in two of three proteins at any position are shown in boldface type. Position 1 is the N terminus and residues −19 to −1 correspond to the signal sequence. A single consensus site for N-linked glycosylation in saxiphilin is noted by an exclamation point (!). Asterisks (*) denote the positions of 10 highly conserved residues in the two Fe^{3+}/HCO_3^- sites of transferrins (15, 16). The locations of 14 probable disulfide bonds in saxiphilin were identified by homology to human lactoferrin (15) and are labeled as *a–f* in the N lobe and *a′–h′* in the C lobe above underlined pairs of Cys (C) residues.

```
Sax (90-160)   KCLKERQQAL-AKKMHLVIIFHNVMKRAITNHSSAMAALGTAGVLTSMGEKISGTNTPPGQTRATCERHELP-
                |||||||  ||  :    |  :      ::  :     :  |    :  :||  |  ||  ||||   |||::|:|
Sax (161-233)  KCLKERQVALGGDEKVLGRFVPQCDEKGNYEPQQFHGSTGYSWCVNAIGEEIAGTKTPPGKIPATCQKHDLVT

Sax (178-226)                  GRFVPQCDEKGNYEPQQFHGSTGYSWCVNAIGEEIAGTKTPPGKIPATC
nidogen (842-889)              GMFVPQCDEYGHYVPTQCHHSTGYCWCVDRDGRELEGSRTPPGMRPP-C
invariant chain (210-248)      GAFRPKCDENGNYMPLQCHGSTGYCWCVFPNGTEVPHTK
EGP (93-123)                   GLYDPDCDESGLFKAKQCNG-TSMCWCVNTAG
thyroglogulin 1.1 (29-73)        YVPQCAEDGSFQTVQCQNDGRSCWCVGANGSEVLGSRQP-GR-PVAC
thyroglobulin 1.2 (97-141)       YLPQCQDSGDYAPVQCDVQHVQCWCVDAEGMEVYGTRQL-GR-PKRC
thyroglobulin 1.5 (597-639)      FVPSCTTEGSYEDVQCFSGE--CWCVNSWGKELPGSRVRDGQ-PR-C
thyroglobulin 1.6 (664-707)      FVPACTSEGHFLPVQCFNSE--CYCVDAEGQAIPGTRSAIGK-PKKC
```

FIG. 4. Homology relationships of the 144-residue insertion unique to saxiphilin. The upper two sequences are a pairwise alignment of saxiphilin residues 90–160 and 161–233 showing significant two-fold internal homology within the 144-residue insertion. A vertical line marks an identity and a colon marks a conservative substitution. The lower eight sequences illustrate homology between residues 178–226 within the saxiphilin insertion and a domain observed in other proteins that is known as a type 1 repeat of thyroglobulin (22). The comparison sequences are mouse nidogen (23), rat invariant chain (24), human epithelial glycoprotein (EPG) (25), and human thyroglobulin (22). Residues in boldface type are identical to those in saxiphilin and are present in at least three of the eight sequences.

residue, Arg-124 (Arg-456). In Fig. 3 the positions of these 10 critical residues are identified with asterisks; only 1 of these residues is found to be absolutely conserved in saxiphilin at Asp-60. Two of these residues in the N lobe have also apparently diverged in *X. laevis* transferrin (Asp-63 → Thr, Arg-124 → Lys); however, direct measurement of the Fe^{3+}-binding capacity of transferrin from *R. catesbeiana* indicates the presence of two functional Fe^{3+}-binding sites (6). Two other members of the transferrin family are known to contain only one functional Fe^{3+}-binding site [i.e., *M. sexta* transferrin (18) and human melanotransferrin (21)], but saxiphilin is the only example of a related protein in which nearly all of the Fe^{3+}-binding site residues have diverged.

Homology Relationships of the Insertion Sequence. The 144-residue insertion found in saxiphilin has several interesting features. Dot plot analysis (19) of the insertion sequence reveals that this domain also appears to have arisen from a duplication event. Fig. 4 shows an alignment of the two halves of the insertion sequence (saxiphilin residues 90–160 and 161–233) indicating 35% identity that is mainly clustered at the ends of the alignment. Assuming that saxiphilin/transferrin homology corresponds to a similar tertiary structure, then the insertion would be predicted to occur at a location between the previously identified d and e strands of β-sheet in the crystal structure of lactoferrin (15). This location is at the hinge region separating domains 1 and 2 that form the Fe^{3+}-binding pocket in the N lobe of lactoferrin and suggests that the saxiphilin insertion could form an autonomous globular domain that would not disrupt the basic folding pattern of lactoferrin. The 144-residue insertion also contains the only consensus site (N-X-T/S) for N-linked glycosylation in saxiphilin at Asn-119 (Fig. 3), suggesting that part of this region forms a surface domain.

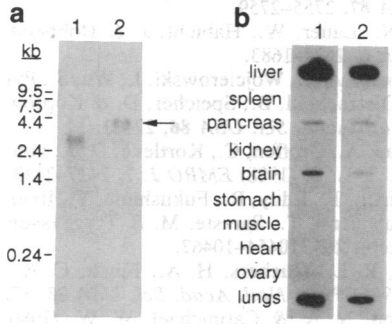

FIG. 5. Characterization and tissue distribution of saxiphilin mRNA. (*a*) Northern blot analysis of poly(A)$^+$ RNA from bullfrog liver. Five micrograms of poly(A)$^+$ RNA was fractionated on agarose, transferred to nylon, and consecutively hybridized with a 2.25-kb *Sma* I probe (lane 1) and a 0.43-kb insert-specific probe (lane 2). The arrow at the right points to the band identified as saxiphilin mRNA. (*b*) The same probes were hybridized to 5 μg of total RNA from various bullfrog tissues using a slot blot apparatus: 2.25-kb *Sma* I probe (lane 1); 0.43-kb insert-specific probe (lane 2).

A sequence comparison search (11, 12) against the 144-residue insertion identified a portion of the C-terminal half of the insertion that is 59% identical to a 48-residue fragment (Fig. 4) of nidogen (23), a ubiquitous ≈150-kDa cell matrix protein. This particular region of nidogen is itself a repetitive domain that has been previously shown to be related to a particular class of 10 repetitive domains (Thyr-1) found in thyroglobulin (22). As indicated in Fig. 4, such Thyr-1 repetitive elements have also been identified in several small integral membrane proteins: invariant chain subunit I of class II major histocompatibility complex (24, 26) and two related cell surface antigens called EPG (epithelial glycoprotein) (25) and GA733 (27). A class of insulin-like growth factor-binding proteins (28) and the B1 chain of laminin (29) also contain Thyr-1 domains. Along with saxiphilin, versions of this domain thus occur in a wide variety of membrane, secreted, and cell matrix proteins. The function of such repetitive Thyr-1 elements is unknown, but they have been suggested to play a role in intracellular protein transport and secretion (26).

Expression of Saxiphilin Message in Bullfrog Tissues. Northern blot analysis (Fig. 5*a*) was performed using the 2.25-kb *Sma* I fragment (Fig. 1) derived from the saxiphilin cDNA clone as a probe of poly(A)$^+$ RNA from bullfrog liver. Two bands were detected, an intense band at ≈3.1 kb and a fainter band at ≈4.4 kb (Fig. 5*a*, lane 1). The same filter was then stripped and hybridized to a 0.43-kb probe (Fig. 1) corresponding to the saxiphilin-specific insertion extending from nucleotide position 303 to position 738 in Fig. 2. This probe hybridized strongly to the upper ≈4.4-kb band (Fig. 5*a*, lane 2), indicating that this band corresponds to saxiphilin mRNA. The intense lower band at ≈3.1 kb in lane 1 is presumably the result of cross-hybridization with transferrin mRNA, which is known to be highly abundant in liver (30). This interpretation is consistent with the high degree of similarity between saxiphilin and transferrin mRNA and the relative abundance of the two proteins in bullfrog plasma (i.e., the concentration of transferrin is ≈100-fold greater than saxiphilin) (4–6).

The relative level of saxiphilin mRNA in various bullfrog tissues was investigated by hybridizing total RNA with the same probes used for the Northern blot analysis (Fig. 5*b*). The highest amount of saxiphilin mRNA was detected in liver, followed by lung, pancreas, and brain. In a previous study (4), we found that bullfrog kidney, heart, and ovaries have high levels of soluble [^{3}H]STX binding activity but this does not appear to correspond to a high level of mRNA. Such differences in the tissue distribution of saxiphilin mRNA versus protein activity as detected by [^{3}H]STX binding may reflect cellular uptake of plasma saxiphilin through a cell surface receptor or differences in protein turnover rates. Rat transferrin is synthesized principally in liver but also at lower

levels in brain and testis (30). Transferrin secreted from liver appears to be the source of transferrin found in plasma and other tissues such as intestine, which do not express transferrin mRNA (30). The possibility exists that saxiphilin and serum transferrin utilize similar mechanisms of gene expression, secretion, and internalization through the process of receptor-mediated endocytosis (7). Further work will be necessary to determine whether there is a receptor for saxiphilin analogous to the transferrin receptor.

Saxiphilin was originally discovered by its binding affinity for STX ($K_d \simeq 0.2$ nM), an interaction that exhibits a high degree of chemical specificity (4). Although our understanding of the functional significance of this binding interaction is incomplete, it has allowed us to identify a relative of the transferrin family that does not appear to be involved in iron metabolism. By analogy to other protein superfamilies that can be recognized on structural grounds, the case of saxiphilin and transferrin implies that transferrin-like proteins comprise a superfamily with functions more diverse than those associated with Fe^{3+} binding. Based on the high degree of sequence similarity, it appears that saxiphilin arose from a two-lobed transferrin ancestor by a process that eventually led to substitution of most of the Fe^{3+}-coordinating residues. The insertion sequence in the N lobe probably arose through a duplication event and exon shuffling of a Thyr-1 domain (27). The wide occurrence and repetition of this domain within proteins suggest that it may serve as a recognition site for a protein–protein interaction.

Although saxiphilin and transferrins bind different ligands (6), the pH dependence of STX dissociation from saxiphilin (L. Llewellyn and E.M., unpublished data) is similar to that of Fe^{3+} dissociation from transferrin, which is important in the delivery of iron to cells (7). This suggests that the mechanism of ligand binding and release in saxiphilin and transferrin are functionally analogous. The crystal structure of lactoferrin indicates that the Fe^{3+}-binding cavity (diameter = 10 Å) is potentially large enough to accommodate an organic molecule (15). On this basis, we hypothesize that saxiphilin may function in delivering or removing an endogenous ligand. Although it is known that STX is widely distributed in various marine invertebrates in association with plankton blooms (1), there is scant information on the chemical ecology of STX in freshwater ecosystems. However, at least one species of freshwater cyanobacteria has been found to synthesize STX (31). Thus, in frogs it is possible that saxiphilin may participate in a detoxification mechanism for neutralizing a microbial toxin. In broader terms, the recognition of a transferrin-like protein that binds an organic molecule is suggestive of a physiological system for transport and sequestration of small molecules that might ultimately be exploited for antidote therapy or drug delivery.

We are indebted to B. Rossier and N. Birnberg for helpful suggestions in cloning strategy and H.-P. Gaeggeler for assistance with RNA preparation. We thank E. Ullu and our laboratory colleagues for critically reading the manuscript. This work was supported by the National Institutes of Health and the U.S. Army Medical Research and Development Command.

1. Hall, S., Strichartz, G., Moczydlowski, E., Ravindran, A. & Reichardt, P. B. (1990) in *Marine Toxins: Origin, Structure and Molecular Pharmacology*, eds. Hall, S. & Strichartz, G. (Am. Chem. Soc., Washington, DC), pp. 29–65.
2. Doyle, D. D., Wong, M., Tanaka, J. & Barr, L. (1982) *Science* **215**, 1117–1119.
3. Moczydlowski, E., Mahar, J. & Ravindran, A. (1988) *Mol. Pharmacol.* **33**, 202–211.
4. Mahar, J., Lukacs, G. L., Li, Y., Hall, S. & Moczydlowski, E. (1991) *Toxicon* **29**, 53–77.
5. Li, Y. & Moczydlowski, E. (1991) *J. Biol. Chem.* **266**, 15481–15487.
6. Li, Y., Llewellyn, L. & Moczydlowski, E. (1993) *Mol. Pharmacol.* **44**, 742–748.
7. Crichton, R. R. (1991) *Inorganic Biochemistry of Iron Metabolism* (Horwood, New York), pp. 1–263.
8. Chomczynski, P. & Sacchi, N. (1987) *Anal. Biochem.* **162**, 156–159.
9. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY).
10. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
11. Pearson, W. R. & Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2444–2448.
12. Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395.
13. Moskaitis, J. E., Pastori, R. L. & Schoenberg, D. R. (1990) *Nucleic Acids Res.* **18**, 6153.
14. Yang, F., Lum, J. B., McGill, J. R., Moore, C. M., Naylor, S. L., vanBragt, P. H., Baldwin, W. D. & Bowman, B. H. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 2752–2756.
15. Anderson, B. F., Baker, H. M., Norris, G. E., Rice, D. W. & Baker, E. N. (1989) *J. Mol. Biol.* **209**, 711–734.
16. Bailey, S., Evans, R. W., Garratt, R. C., Gorinsky, B., Hasnain, S., Horsburgh, C., Jhoti, H., Lindley, P. F., Mydin, A., Sarra, R. & Watson, J. L. (1988) *Biochemistry* **27**, 5804–5812.
17. Needleman, S. B. & Wunsch, C. D. (1970) *J. Mol. Biol.* **48**, 443–453.
18. Bartfeld, N. S. & Law, J. H. (1990) *J. Biol. Chem.* **265**, 21648–21691.
19. Maizel, J. V. & Lenk, R. P. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 7665–7669.
20. Baldwin, G. S. (1993) *Comp. Biochem. Physiol. B* **106**, 203–218.
21. Rose, T. M., Plowman, G. D., Teplow, D. B., Dreyer, W. J., Hellstrom, K. E. & Brown, J. P. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 1261–1265.
22. Malthiery, Y. & Lissitzky, S. (1987) *Eur. J. Biochem.* **165**, 491–498.
23. Mann, K., Deutzmann, R., Aumailley, M., Timpl, R., Raimondi, L., Yamada, Y., Pan, T., Conway, D. & Chu, M.-L. (1989) *EMBO J.* **8**, 65–72.
24. McKnight, A. J., Mason, D. W. & Barclay, A. N. (1989) *Nucleic Acids Res.* **17**, 3983–3984.
25. Simon, B., Podolsky, D. K., Moldenhauer, G., Isselbacher, K. J., Gattoni-Celli, S. & Brand, S. J. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 2755–2759.
26. Koch, N., Lauer, W., Habicht, J. & Dobberstein, B. (1987) *EMBO J.* **6**, 1677–1683.
27. Linnenbach, A. J., Wojcierowski, J., Wu, S., Pyrc, J. J., Ross, A. H., Dietzschold, B., Speicher, D. & Koprowski, H. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 27–31.
28. Brinkman, A., Groffen, C., Kortleve, D. J., van Kessel, A. G. & Drop, S. L. S. (1988) *EMBO J.* **7**, 2417–2423.
29. Pikkarainen, T., Eddy, R., Fukushima, Y., Byers, M., Shows, T., Pihlajaniemi, T., Saraste, M. & Tryggvason, K. (1987) *J. Biol. Chem.* **262**, 10454–10462.
30. Idzerda, R. L., Huebers, H. A., Finch, C. A. & McKnight, G. S. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 3723–3727.
31. Mahmood, N. A. & Carmichael, W. W. (1986) *Toxicon* **24**, 175–186.