Cloning of human and bovine homologs of SNF2/SWI2: a global activator of transcription in yeast *S.cerevisiae*

Ichiro Okabe^{1,2+}, L.Charles Bailey², Olivier Attree^{2,§}, Sudha Srinivasan², Jeffrey M.Perkel², Brehon C.Laurent³, Marian Carlson³, David L.Nelson⁴ and Robert L.Nussbaum^{1,2,*}

¹Howard Hughes Medical Institute and ²Department of Genetics, University of Pennsylvania School of Medicine, Philadelphia, PA 19104-6145, ³Department of Genetics and Development, College of Physicians And Surgeons, Columbia University, New York, NY 10032 and ⁴Institute for Molecular Genetics, Baylor College of Medicine, Houston, TX 77030, USA

Received March 24, 1992; Revised and Accepted July 29, 1992

GenBank accession nos M88163, M89907

ABSTRACT

We performed positional cloning of genes carried on yeast artificial chromosomes that span a human translocation breakpoint associated with a human disease and isolated by chance human and bovine genes with strong homology to the *S.cerevisiae* genes, *SNF2/SWI2* and *STH1*, and the *D.melanogaster* gene *brahma*. We report here sequence analysis, expression data, and functional studies for this human SNF2-like gene (hSNF2L) and its bovine homolog (bovSNF2L). Despite strong homology at the amino acid level, hSNF2L is not capable of complementing the yeast mutations *snf2* or *sth1* in *S.cerevisiae*. Furthermore, in contrast to SNF2 itself, a fusion protein consisting of the DNA binding domain of LexA and hSNF2L did not transactivate a reporter gene downstream of LexA binding sites in a yeast expression system. The strong similarity between hSNF2L and these yeast and drosophila genes suggest that the mammalian genes are part of an evolutionarily conserved family that has been implicated as global activators of transcription in yeast and fruitflies but whose function in mammals remains unknown.

INTRODUCTION

We have been isolating genes contained on yeast artificial chromosomes that span a human t(X;3) breakpoint in a female with the Lowe Oculocerebro-renal Syndrome (OCRL; McK309000) in order to find the gene responsible for this disease(1,2). As a byproduct of this project, we found, entirely by chance, a gene that, although not involved in OCRL, does appear to be a mammalian homolog of *SNF2/SWI2* and *STH1* from *S.cerevisiae* (3,4) and *brahma* from *D.melanogaster* (5), which are representatives of a class of genes termed global activators of transcription (5). We report here the isolation and

sequencing of the human hSNF2L gene and its bovine homolog, bovSNF2L, and provide data on tissue expression, map location, sequence homologies, and functional analysis in yeast.

MATERIALS AND METHODS

cDNA library screening

Three cDNA libraries were screened: a bovine lens cDNA library in λ gt 11 and two human oligo dT primed human cDNA libraries (fibroblast and kidney) in λ gt 10 (Clontech). 1×10^6 recombinants were screened in each library. Two replica plaque lifts on Colony/Plaque Screen membrane (Dupont/NEN) were prepared from each plate according to the manufacturer's instructions.

Two YAC clones (RS88 and RS41(1), each comprising ~60–70 kb of genomic DNA) were recovered in agarose from a CHEF (clamped homogenous electric field) gel, labeled with ³²P dCTP, preannealed with sonicated human placental DNA to block repetitive sequence hybridization, and used as hybridization probes as described previously (2). When cDNA was used to screen libraries, the cDNA was labeled with ³²P dCTP by the random priming method (Bethesda Research Laboratories) and used to screen phage plaque lifts according to manufacturer's instructions (Dupont/NEN).

Positive clones derived from cDNA library screening were plaque-purified and bacteriophage DNA was extracted from plate lysates (6). Each DNA was digested with EcoRI and purified via electrophoresis. The separated DNAs were then subcloned into pBluescript-SK(–) or pGEM-3Zf(–) by standard techniques (7).

Southern and Northern analysis

DNA used for Southern blotting were obtained from: a normal male, a normal female, a 48,XXXX female (Human Genetic Mutant Cell Repository), a female with oculocerebrorenal syndrome (OCRL) and a balanced translocation between X and

* To whom correspondence should be addressed

Present addresses: ⁺Department of Pediatrics, Jichi Medical School, Minamikawachi-Machi, Kawachi-Gun, Tochigi-Ken, 329-04, Japan and [§]CRSSA, Unite de Biologie Moleculaire, 38702 La Tronche Cedex, Grenoble, France

1	CAAGACTGGAAGCAGAGAGAGAGCAAGAGTGAGAGAGAGCGGACGAGCTAGTCAGGAGATGGTGTGTTGATTCCAGGAGAAAACGTTTCTTCATTCTCAACTCAAACCTGCTGCTAA	120
121	GCGCCTAAATCTGAAAAGGAATGGACCCGAATATGAAGAGAAAATGAAAGCCGACCGAGCAAGAGATTGAAATTTTACTGAAGCAGACAGAACTTTTGCACATTTTCATTAGCCT	240
1	M D P E Y E E K M K A D R A K R F E F L L K Q T E L F A H F I Q P	33
241	TCAGCACAGAAATCTCCAACATCTCCACTGAACATGAAATTTGGGACGTCCTCCGAATAAAGAAAGATGAAAAGCAGAGCTTAATTTCTGCTGGAGACTACCGCCATAGGCGCACAGACAA	360
34	S A Q K S P T S P L N M K L G R P R I K K D E K Q S L I S A G D Y R H R R T E Q	73
361	GAAGAAGATGAAGAGCTACTGTCTGAGAGTCGGAACATCTAATGTGTGATTAGATTGAGGTGTACCTTCATATGTGAAAGGGGGCCACTGAGAGATTATCAGATTTCGAGGACTG	480
74	E E D E E L L S E S R K T S N V C I R F E V S P S Y V K G G P L R D Y Q I R G L	113
481	AATGGTGTGATCTCTTTATATGAAAATGGAGTCAATGGCATTGTGGCTGATGAAATGGGCCTTGGGAAAACCTTACAAACAATTCGTTTGTCTGGTTTACCTGAACACTACCGAAATATT	600
114	N W L I S L Y E N G V N G I L A D E M G L G K T L Q T I A L L G Y L K H Y R N I	153
601	CCTGGACCTCAGATGGTTTTAGTTCCAAAGTCTACTTTACACAACCTGGATGAATGAATTTAAACGATGGGTCCCATCTCTCCGTGTCATTGTGTTTGTGCGGAGACAAGGATGCCAGAGCT	720
154	P G P H M V L V P K S T L H N W M N E F K R W V P S L R V I C F V G D K D A R A	193
721	GCTTTTATTCGTGATGAAATGATGCCAGAGAGTGGGATGTTTGGCTTACTTCTTATGAGATGGTAATTAAGAAAAATCTGTATTCAAAAAGTTTCTACTGGCGATACCTGGTCATTGAT	840
194	A F I R D E M M P G E W D V C V T S Y E M V I K E K S V F K K F H W R Y L V I D	233
841	GAAGCTCAGAGAAATAAGAAATGAAAAATCTAAGCTTTCAGAGATTGTCGTGAGTTCAAGTCGACTAACCGCTTGCTCCTAACTGGAACACCTTTCAGAGATAACCTGCATGAACGTGG	960
234	E A H R I K N E K S K L S E I V R E F K S T N R L L L T G T P L Q N N L H E L W	273
961	GCCTTACTCAACTTTTTATTGCCTGATGTCTTTAATCTCGAGATGACTTTGATTCTTGGTTTGACACTAAAAATGTCTTGGTGATCAAAAACCTCGTGGAAAGACTTCATGCAGTTTAA	1080
274	A L L N F L L P D V F N S A D D F D S W F D T K N C L G D Q K L V E R L H A V L	313
1081	AAACCTTTTGTGTACGCGTATAAAAACTGATGTAGAGAAGAGTCTGCCACCTAAAAAGGAAATAAGATTACTTGGGGCTGAGTAAGATGCAACGAGAAATGGTATACAAAAATCTCTG	1200
314	K P F L L R R I K T D V E K S L P P K K E I K I Y L G L S K M Q R E W Y T K I L	353
1201	ATGAAGATATTGATGTTTTAAACTCTTCTGGCAAGATGGACAAGATGGGACTCTTAAACATTTGATGCAGCTTCGAAAGTGTGTAATCATCCATATCTGTTTGTAGGTGCTGAACCT	1320
354	M K D I D V L N S S G K M D K M R L L N I L M Q L R K C C N H P Y L F D G A E P	393
1321	GGTCCACCTTATACCCTGATGAGCATATTGTGACAGAACAGTGGTAAATGGTAGTTCTTGATATAACTTTGGCCAAACTCAAAGAACAGGGTTCAAGGGTTCTCATTTTCAGCCAGATG	1440
394	G P P Y T T D E H I V S N S G K M V V L D K L L A K L K E Q G S R V L I F S Q M	433
1441	ACTCGTCTGCTGGATATTTTGAAGATTATTGCATGTGGCTGGTTATGAGTATTGTCGACTGGATGGACAAACCCCGCATGAAGAAAGAGAGGATAAATTCCTAGAAGTGAATTTCG	1560
434	T R L L D I L E D Y C M W R G Y E Y C R L D G Q T P H E E R E D K F L E V E F L	473
1561	GGTCAAAGGGAAGCAATAGAGGCTTTTAATGCTCCTAATAGTAGCAATTCATCTTTATGCTAAGTACCAGGCTGGAGGTCTCGGAATTAACCTGGCAAGTCTGATGTGGTTATACTA	1680
474	G Q R E A I E A F N A P N S S K F I F M L S T R A G G L G I N L A S A D V V I L	513
1681	TATGATTCAGACTGGAACCCACAGGTTGATCTACAAGCTATGATCGAGCACATCGTATTGGTCAGAAGAACAGTACGTGTATTCCGTCTCATCACTGACAACACTGTTGAAGAGAGG	1800
514	Y D S D W N P Q V D L Q A M D R A H R I G Q K K P V R V F R L I T D N T V E E R	553
1801	ATTGTAGAAAGAGCTGAGATAAACTGAGACTCGATTCAATTTGTATACAACAAGGAAGACTCATTGACCAACGGTCTAACAAGCTGGCAAAAGAGGAAATGTTACAATGATACGGCAT	1920
554	I V E R A E I K L R L D S I V I Q Q G R L I D Q R S N K L A K E E M L Q M I R H	593
1921	GGAGCCACCCATGTTTTTCTTAAAGAGAGTGAGTTGACAGATGAAGACATTTACAACATTTCTGGAAAGAGGGGAAAAGAACTGCAGAGATGAATGAACGCCTGCAAAAAATGGGA	2040
594	G A T H V F A S K E S E L T D E D I T T I L E R G E K K T A E M N E R L Q K M G	633
2041	GAGTCTTCTTAAGAAATTTTGAATGGACATTGAACAAGTTTATACAAATTTGAGGGAGAAGATTATAGAGAAAAACAAGAGCTTGGCATGGTGAATGGATTGAACCTCCTAAACGA	2160
634	E S S L R N F R M D I E Q S L Y K F E G E D Y R E K Q K L G M V E W I E P P K A R	673
2161	GAACGCAAGCAAACTACCGAGTGGATGCTACTTTAGAGAGGCTTTGGCGTGCAGCGAGCCAAAGATTCCAAGGCTCCACGGCTCCAAAACAGCCAAATGTTTCAGGATTTTCAATT	2280
674	E R K A N Y A V D A Y F R E A L R V S E P K I P K A P R P P K Q P N V Q D F Q F	713
2281	TTCCACCACGCTTATTGAGCTCCTGGAAAAGGAAATCTTTATTATCGGAAGACAATAGGCTATAAGGTTCCAAGGAATCCTGATATCCAAATCCAGCTCTGGCTCAAAGAGAAGAG	2400
714	F P P R L F E L L E K E I L Y Y R K T I G Y K V P R N P D I P N P A L A Q R E E	753
2401	CAAAAAAGATTGATGGAAGCTGAACCTCTTACACAGAGAGACTGAAGAAAAGGAAAACTTCTCACACAAGGTTTCAACAACTGGACTAAACAGAGATTTTAAACAGTTTATTAAGCT	2520
754	Q K K I D G A E P L T P E E T E E K E K L L T Q G F T N W T K R D F N Q F I K A	793
2521	AATGAGAAATATGGAAGAGATGACATTGATATACATAGCTCGAGAGGTAGAGGGCAATCCCTCGAGGAGGTTCAGGATGATTTCAGCTGATTTTGGGAACGTTGCAATGAATTACAGGAC	2640
794	N E K Y G R D D I D N I A R E V E G K S P E E V M E Y S A V F W E R L A Q D G	833
2641	ATTGAGAAATATTGGCTCAAATTTGAACGTTGGAGAAGCAAGAAATCAACGAAGGATCAGTATCAAGAAGCCCTGGATGCCAAAATTCGAAGATACAAGGCTCCATTTTCATCAGTTGCGC	2760
834	I E K I A M T G Q I E R G E A R I Q R R I S I K K A L D A K I A R Y K A P F H Q L R	873
2761	ATTCAGTATGGAACCGACGAAGGAAGAACTACTGAGGAAGAAGATAGATTCTTGATTGTTATGTTACAAAATGGGCTTGATAGAGAAATGTATATGAAGAATTAAAGACAGTG	2880
874	I Q Y G T S K G K N Y T E E E D R F L I C M L H K M G C T T G D R E N V Y E E L R Q C	913
2881	GTACGAAATGCTCCCAAGTTTATAGTTTGAATGCTGTTTATCAAGTCTAGGACTGCCATGGAATTCAGAGACGCTGTAACTCTGATTTTCATTGATTGAGAAAGAAAATATGGAATTTGAG	3000
914	V R N A P Q F R F D W F I K S R T A M E F Q R R C N T L I S L I E K E N M E I E	953
3001	GAAAGAGAGAGAGCAGAAAAGAGAAACGGGCACTAAAACCTCCAATGGTAAAAATTTTTCAGCATTTTCTTAACCTTTAGATTTAACATTGTTGGGCCATTTAAATGTGCATATTGGAGC	3120
954	E R E R A E K K K R A T K T P M V K F S A F S	976
3121	AGAACATTAAATCTGTTTCCATTTTACTCACAGAAAAGAAAAGCAGAGTCAAGTCTGAGAGCTCTGGAAGAAGGATGTCAGAAGGTTGAAATCTTAAGCGCTAGAAATAAGATTTTAA	3240
3241	ATGGGAAACTGCTATTTTCTGTTCCCACTCTCAAAATGCTAATTCAGTCTCCAGTGTATTCAATGCTACTCAAGAAAATCTCTTTGGTTTGATTCTTCGATATTTATATATTTTA	3

B

```

hSNF2L-2   1   ctgctccatctagctctttccagtgcagccactgccgcccaggagccctcgctccct
hSNF2L-2   61  gccttgccccctactcgttcccgcctccacggcatggagcaggacactgccgcagtgcc
hSNF2L-2  121  agccacgctggcagccgcggatgcgaccgccactatcgtggatagaggacgagcagcc

hSNF2L-2  181  cgggccgtccacctctcaggaggaggagcgccgcccggccaccgaagccaccgcggc
hSNF2L      1   caagactggaagcagagagagagagcaagagtgagagagagcgagcgagcgta
consensus  -----

hSNF2L-2  241  cacggagaagggcgagaagaagaAGGAGAAAAACGTTTCTTCATTTCAACTCAAACTTGC
hSNF2L      54  gtcaggagatggtgtgtattccAGGAGAAAAACGTTTCTTCATTTCAACTCAAACTTGC
consensus  -----AGGAGAAAAACGTTTCTTCATTTCAACTCAAACTTGC

hSNF2L-2  301  TGCTAAAGCGCCTAAATCTGAAAAGGAAATGGACCCAGAATATGAAGAGAAAAATGAAAGC
hSNF2L     114  TGCTAAAGCGCCTAAATCTGAAAAGGAAATGGACCCAGAATATGAAGAGAAAAATGAAAGC
consensus  TGCTAAAGCGCCTAAATCTGAAAAGGAAATGGACCCAGAATATGAAGAGAAAAATGAAAGC

```

Figure 1. A. Sequence of human fibroblast hSNF2L gene. This sequence has been deposited in GenBank under accession # M88163. **B.** Comparison of 5' sequence of kidney (hSNF2L-2) and fibroblast cDNA (hSNF2L). Regions of identity are shown as capital letters connected by vertical lines. The kidney cDNA sequence has been deposited in GenBank under accession # M89907.

3 [46,X,+t(X;3)(q25;q27),t(14;17)(q24;q23)] (8); a human-hamster cell hybrid F649-5 which contains the der3 chromosome (Xq25-qter) from the t(X;3) female OCRL patient but lacks the derX and the normal X (RLN, unpublished data); a human-mouse cell hybrid Lo1-6TG containing the derX chromosome from the same patient (Xpter-q25) without the derivative 3 or normal X (9); a human-mouse cell hybrid, RJK734, retaining Xq26-qter (10); a human-hamster cell hybrid X3000-11 containing the region Xq24-qter as its only human component (11); Chinese hamster V79 cells; mouse A9 cells.

10 μ g of each genomic DNA was digested with BglII overnight using 10 units enzyme/ μ g of DNA, phenol and chloroform-extracted, ethanol precipitated and redigested using 3 units of enzyme/ μ g of DNA for several hours to ensure completion of the reaction. Digested genomic DNA was resolved by electrophoresis, transferred onto Hybond N (Amersham), hybridized and washed according to instructions of the manufacturer.

A human multiple tissue Northern blot filter (Clontech) was used for analysis of steady-state RNA levels in various human tissues. In addition, total RNA from fibroblast and lymphoblastoid cell cultures were prepared by the GuSCN-CsCl₂ method and 15 μ g of each RNA was separated on a 1.2% agarose-formaldehyde gel (7) and transferred to a Zetabind (Cuno) membrane. Hybridization was performed under standard conditions (50% formamide, 3 $\times$ SSPE, 1%SDS, 5 $\times$ Denhardt's, 8.5% dextran sulfate) at 42°C for 24 hours and washed to 0.5 $\times$ SSC, 0.1% SDS at 65°C, followed by a five days exposure with intensifying screens. The Northern filters were also probed with the cDNA for the ribosomal protein rPL32(6) to control for integrity of the RNAs and loading in each lane.

Probes were prepared by random priming (Bethesda Research

Laboratories) of plasmid inserts isolated from low-melting point agarose gels (7). The 5' probe specific to human kidney cDNA was prepared by polymerase chain reaction amplification using oligonucleotides 5'CCTTGTCCTCCCTACTCGTTCC3' and 5'TCTTCTTCTCGCCCTTCTCCG3' (which bracket the region of HKcL15 not present in λ HK; see below), the cloned HKcL15 as template, and manufacturer's reagents (Perkin-Elmer).

DNA sequencing and analysis

Nucleotide sequence was determined directly on recombinant plasmid purified on Qiagen columns according to manufacturer's instructions. Dideoxy-terminator sequencing reactions were performed with fluorescent dye-labeled terminators with manufacturer's protocols and reagents (Applied Biosystems). Samples were purified with Sephadex G-50 columns (Quick Spin Columns; Boehringer Mannheim) and stored at -20°C until ready for sequencing on a 373A DNA Sequencer (Applied Biosystems). Both strands were sequenced completely to eliminate all ambiguities in sequence.

All computed sequence analysis was performed using GCG software (12). The GenBank (Release 70.0 + updates) and EMBL (Release 29.0 + updates) nucleic acid sequence databases were searched with the hSNF2L nucleic acid sequence, as well as translated and searched with the hSNF2L amino acid sequence, using the FASTA algorithm (13) as implemented in the GCG programs FASTA and TFASTA. The hSNF2L amino acid sequence was scanned for the presence of protein sequence motifs in the ProSite database (release 8.0), using the GCG program Motifs. Local similarities between individual sequences were identified according to the method of Smith and Waterman using the GCG program BestFit. Local alignments of multiple sequences were constructed by use of the above algorithms under

varying parameters on individual sequence pairs, followed by manual construction and optimization of the overall alignments.

Strains, media, and genetic methods

S.cerevisiae strains MCY829 (*MAT* α his3- Δ 200 ura3-52 lys2-801 SUC2), MCRY920 (*MAT* α /MAT α his3- Δ 200/his3- Δ 200 ura3-52/ura3-52 ade2-100/+ +/lys2-801 sth1-1::URA3/+ SUC2/SUC2), and MCY2156 (*MAT* α *snf2-141* his3- Δ 200 ura3-52 SUC2) were used as hosts. Transformation and genetic analysis of yeast were by standard techniques (14). Synthetic complete (SC) media (14) lacked nutrients to maintain selection for plasmids. Raffinose utilization was scored by comparing anaerobic growth on SC medium containing either 2% glucose or 2% raffinose.

Transactivation assay

The portion of the human hSNF2L gene encoding residues 2 through 977 was amplified by PCR and cloned into pSH2-1 in both the sense (pLexA-hSNF2L) and antisense (pLexA-hSNF2Lopp) orientations (15–17). Accurate, in-frame insertion of the hSNF2L gene in the sense orientation was confirmed by sequencing. Multiple independent clones were isolated and used for subsequent studies. Transactivation by pLexA-hSNF2L was assayed in yeast MCY829 using β -galactosidase as a reporter gene carried in vectors pLR1 Δ 1 and p1840 which, respectively, lack or contain LexA binding sequences upstream of β -galactosidase (16,18). β -galactosidase assays in liquid culture of yeast were by standard protocols (19).

Immunoblot analysis

Protein extracts were prepared from strain MCY829 (wild type) carrying pLexA-hSNF2L. Proteins were separated by SDS-PAGE and immunoblot analysis was carried out as described previously (4). Polyclonal LexA antibody was a gift of J.Kamens and R.Brent.

RESULTS

cDNA library screening

In our search for expressed sequences contained within YAC RS41, a bovine lens cDNA library was initially screened with labeled YAC isolated from a pulsed field gel. Over 160 positive signals were found, all but three of which were shown to be the result of plasmid contamination of the library when pYAC4 alone was used to probe the same filters. However, three clones (BLcL12, 14 and 16) were detected with RS41 but not with pYAC4 (2). The insert in BLcL12 was 2.2 kb and BLcL14 (and its identical sibling clone BLcL16) had an insert of 1.4 kb. The screening of human kidney and fibroblast cDNA libraries with BLcL12 and BLcL14 as probes resulted in the isolation of lambda clones containing a 2.7 kb kidney cDNA (λ HKcL15) and a 3.6 kb fibroblast cDNA (λ HF).

DNA sequence

Sequence analysis was performed on BLcL12, HKcL15 and HF. The complete sequence of HF, is shown in Figure 1A (hSNF2L; GenBank accession # M88163).

In clone HF, an open reading frame (ORF) was found beginning at base 142, as defined by an in-frame ATG with an in-frame stop codon upstream at position 51. The ORF is capable of encoding a polypeptide of 976 amino-acids and is followed by a 3' untranslated region of ~0.6 kb with poly-A tract.

The human fibroblast, human kidney, and bovine lens clones were found to have essentially identical sequence in the 3' direction beginning at position 77 in HF. As shown in Figure 1B, however, the fibroblast cDNA clone HF contained 76 bp of sequence at its 5' end, in the putative 5' untranslated region, that were not present in the human kidney clone HKcL15 (hSNF2L-2; Accession # M89907) or in the bovine lens clone BLcL12; furthermore, HKcL15 and BLcL12 both contained 263 bp of 5' sequence not present in the fibroblast clone. This alternative 5' sequence in HKcL15 and BLcL12 eliminated the upstream stop codon at position 51 in the fibroblast clone and provided another upstream, in-frame ATG that could encode an additional 78 amino acid residues at the amino terminus of the protein.

Northern and Southern analysis

Northern blot analysis using HF as a probe revealed a 4.4 kb transcript present in human heart, brain, placenta, lung, liver, skeletal muscle and kidney (Fig 2); in a separate experiment, a similar transcript was found in human fibroblast RNA but not in Epstein–Barr virus-immortalized lymphoblastoid cells. Control probing with a ribosomal protein gene cDNA rpL32 (6) showed the loading of RNA in figure 2 was equivalent in each lane and that the lymphoblastoid RNA which failed to give a signal with HF gave a good signal with the control probe (data not shown). A 260 bp probe prepared from the 5' end of HKcL15, comprising the region of the kidney cDNA not present in fibroblast clone HF, detected this same 4.4 kb hSNF2L transcript in fibroblasts (data not shown), indicating that this 5' sequence, although not present in our particular fibroblast cDNA clone, is present in fibroblast mRNA.

The human and bovine cDNAs detect identical restriction fragments in somatic cell hybrids Lo1-6TG (containing the regions Xpter-q25) and X3000-11(Xq24-qter), but neither cDNA detected any human-specific fragments in hybrid F649-5 (Xq25-qter) (Fig 3). The same restriction fragments were also found in a set of overlapping YAC chromosomes covering 250 kb around the t(X;3) breakpoint (data not shown). Thus, the human gene maps to Xq24-q25, within 250 kb immediately proximal to the t(X;3) breakpoint. The 260 bp 5' probe unique to the human kidney and bovine lens cDNAs also mapped to the region around the t(X;3) indicating that it is derived from the gene and is not a ligation artifact generating during cDNA cloning.

Similarity searches using the nucleic acid and peptide sequences of hSNF2L identified significant similarities over ~600 amino acids to sequences derived from the *S.cerevisiae* transcriptional regulators SNF2/SWI2 (3, Accession # M61703) and STH1 (4, Accession # M83755), and the *D.melanogaster* homeotic regulator *brahma* (5, Accession # M85049). Within this region, particularly good conservation is seen in numerous short domains, as shown in Fig 4. Of these, the domains marked with * contain motifs associated with the nucleotide binding fold of NTP binding proteins (20), while motifs within domains marked with + are found in a number of putative helicases (4,20). Direct comparison with the predicted peptide sequence of the mouse B-cell nuclear factor NF κ B (21) also demonstrated strong similarity to hSNF2L over the entire region overlined with a dashed line.

In addition, more limited similarities were identified to sequences within the *S.cerevisiae* gene *MOT1* (22, Accession # M83224) coding for a presumed helicase, *S.cerevisiae* *RAD54* (23, Accession # M63232), which is involved in DNA repair and recombination, and the *D.melanogaster* lodestar gene (24,

Figure 2. Northern blot analysis for human tissues using a subfragment of HF as a probe. lane 1, heart; lane 2, brain; lane 3, placenta; lane 4, lung; lane 5, liver; lane 6, skeletal muscle; lane 7, kidney. Size standards are shown on the left.

Figure 3. Southern blot analysis of normal and hybrid cell line DNA using probe BLcL12 shows that BLcL12 is mapped to Xq24-q25, proximal to the t(X;3) breakpoint. A. Schematic diagram of somatic cell hybrids showing the regions of the human X chromosome each contains. B. Line 1, normal male; lane 2, normal female; lane 3, 48,XXXX; lane 4, t(X;3) patient; lane 5, the human-hamster cell hybrid F6 49-5 which contains the der3 chromosome (Xq25-qter) from the t(X;3) female OCRL patient but lacks the derX and the normal X(RLN, unpublished data); lane 6, human-mouse cell hybrid Lo1-6TG containing the derX chromosome from same patient (Xpter-q25) (6) without the der3 or normal X; lane 7, RJK734, a human-mouse cell hybrid retaining Xq26-qter (7); lane 8, human-hamster cell hybrid X3000-11 containing the region Xq24-qter as its only human component (8); lane 9, Chinese hamster V79 cell line; lane 10, mouse A9 cell line. Dots indicate fragments of human origin in somatic cell hybrid cells. Size standards are shown on the left.

Accession # X62629), which may play a role in chromosomal replication and mitosis. Scanning the hSNF2L sequence with the patterns from the ProSite database identified a potential leucine zipper motif beginning at amino acid 258 of the predicted peptide. However, the significance of this motif is unknown since this leucine zipper motif is not conserved between man, drosophila and yeast, as shown in Figure 4, and a proline residue at position 264, which is highly conserved between all three organisms, interrupts the zipper motif.

Complementation Analysis

To determine whether the hSNF2L protein provides SNF2 function in yeast, we tested whether the *hSNF2L* gene in

multicopy restores growth of a *snf2* mutant on raffinose. The *S.cerevisiae* host strain MCY2156 (*snf2*) was transformed to histidine prototrophy with pLexA-hSNF2L, which expresses the N-terminal 87 residues of LexA fused to hSNF2L from amino acid residue 2 through the entire open reading frame. As controls the strain was also transformed with the vector pSH2-1, which expresses only the DNA binding domain LexA₁₋₈₇ (17), with pLexA-hSNF2Lopp (same as pLexA-hSNF2L except that the hSNF2L sequence is cloned in the opposite orientation into pSH2-1), and with pLexA-SNF2 (expressing a bifunctional SNF2 fusion protein) (3). His⁺ transformants were then scored for raffinose utilization. Only the transformants carrying pLexA-SNF2 grew well on raffinose. We detected no differences in growth between the six transformants carrying pLexA-hSNF2L and the two negative controls, pSH2-1 and pLexA-hSNF2Lopp. These results indicate that the human hSNF2L gene does not provide SNF2 function in yeast.

We also tested whether multiple copies of hSNF2L suppress the lethality of an *sth1* mutation (4). Plasmid pLexA-hSNF2L was used to transform diploid strain MCY920 (*sth1-1::URA3/STH1 his3/his3 ura3/ura3*), and three His⁺ transformants were sporulated with selection for the plasmid, and subjected to tetrad analysis. All 19 dissected asci yielded only two viable spores. All of the viable spore clones were Ura⁻ and His⁺ confirming that they carried the wild-type *STH1* allele and the plasmid. Thus hSNF2L on a multicopy plasmid does not compensate for loss of STH1 function.

Transactivation

The yeast protein SNF2 appears to lack intrinsic DNA binding activity but, when fused to the DNA-binding domain of LexA in a chimeric protein, is able to activate transcription of genes downstream of LexA binding sites. In order to test hSNF2L for similar transactivation activity, pLexA-hSNF2L was transformed into yeast with *lacZ* as a reporter gene carried on pLR1Δ1, which lacks LexA binding sites, and p1840, which carries LexA binding sites. For controls, yeast carrying pLR1 1 and p1840 with or without pSH2-1 alone were grown to log-phase, permeabilized, and assayed for β-galactosidase activity—no increased activity of β-galactosidase over negative controls was detected in any of the experimental conditions. Because of the negative complementation and transactivation results using pLexA-hSNF2L, an immunoblot of the pLexA-hSNF2L product was carried out and showed an ~130 kd protein, which is in agreement with the expected size of an in-frame fusion between LexA₁₋₈₇ and hSNF2L residues 2–977.

Thus, there was no evidence that hSNF2L can activate transcription of a gene downstream of LexA binding sites in yeast when it forms part of a chimeric protein containing DNA-binding domains of LexA.

DISCUSSION

We have cloned, by chance, the bovine and human homologs of a gene which bears significant peptide sequence similarity to several proteins from yeast and flies involved in transcriptional regulation. The near identity of the bovine and human clones over most of their length makes it quite likely that the functions of the encoded proteins will be similar in both mammalian species. The presence of 5' sequence in the HF clone different

Figure 4. Amino acid sequence alignment of hSNF2L family. Darkness of shading reflects the degree of conservation across all amino acids at a given position. Identity of individual sequences is detailed in the text. The overlined regions marked by * represent domains conserved among nucleotide binding proteins while the regions marked by + show similarity to several DNA helicases (14). The region overlined by a dashed line shows similarity to mouse NFXY.

from that in the BLcL12 or HKcL15 clones raises the possibility that the human gene may undergo alternative splicing or have alternative transcriptional start sites. If this is the case, though, both products must be present in fibroblast and kidney. Further study using the unique 5' sequence from clone HF will be required to confirm this.

Little is known at present about the function of hSNF2L or its bovine counterpart. Given the strong homology of hSNF2L to SNF2/SWI2 and *brahma*, it was reasonable to ask whether hSNF2L was able to complement *snf2* and *sth1* mutations and whether a LexA-hSNF2L chimeric protein will act to transactivate transcription of a gene downstream of LexA binding sites, as has been shown for SNF2. No such activity was found for hSNF2L. Its nearly ubiquitous expression in human tissues, however, indicates that a general involvement of hSNF2L in transcription regulation in mammalian cells is possible and deserves to be investigated further.

The significant degree of similarity which hSNF2L bears to the *S.cerevisiae* genes *SNF2/SWI2* and *STH1*, and the *D.melanogaster* gene *brahma* does imply that these may be a family of evolutionarily related genes or may be genes with different functions but with evolutionarily conserved domains that serve similar functions. In general, SNF2 and *brahma* appear to exert their effects as part of multifactor complexes, and may serve as bridges between specific DNA-binding proteins and the transcriptional apparatus (3,25). In addition, several motifs characteristic of DNA helicases are present. It has been theorized that these conserved helicase domains indicate that these genes may play a role in the unwinding of compact chromatin, in order to make specific regions available for transcription (4,5,25). This alternate hypothesis is supported by the observation that some suppressors of *swi2/snf2* mutants map to components of chromatin (4). However, the exact mechanisms by which the SNF2/SWI2 or *brahma* proteins effects transcriptional regulation are unclear. SNF2/SWI2 normally does not bind DNA but, if tethered by a synthetic DNA binding domain, is able to activate transcription, while STH1 lacks this property. Genetic studies have shown that *brahma* interacts with several homeotic genes, but little is known about the mechanism of its action.

There is no known phenotype in man resulting from mutations in hSNF2L. A few hereditary disorders have been mapped to the region Xq24-Xq26 for which hSNF2L might be a candidate. It is unlikely to be involved in a disorder of B-lymphocyte function, such as hyper-IgM syndrome, since it is not expressed in lymphoblasts at levels detectable on a Northern blot. The gene is present (data not shown) in a male patient with X-linked lymphoproliferative disease and an Xq25 interstitial deletion (26), making its involvement in this disorder also unlikely. A number of other defects, such as albinism-deafness syndrome (McK 300700) and X-linked mental retardation with Dandy-Walker malformation, basal ganglia disease, and seizures (McK 304340), also map broadly to this area but their phenotypes are not particularly compatible with a possible role for hSNF2L as a putative transcription co-activator.

As the Human Genome Project progresses, many new genes will be identified from their chromosomal location without knowing their normal function or the phenotype resulting from mutations in these genes. Sequence homologies to genes, identified in model organisms through genetic and biochemical studies, will be important first clues as to the possible functions of these human genes.

ACKNOWLEDGEMENTS

We thank Dr. David Purtilo for making the cell line from a patient with XLP and an Xq25 deletion available, Dr. Robert Perry for consultations and for sharing unpublished sequence from the NFκY gene, and Dr. Tom Kadesch for critical reading of the manuscript. We gratefully acknowledge the expert help of L.Johnson, N.Blanchard, D.Sosnoski and M.F.Victoria. RLN is Associate Investigator and IO is an Associate of the Howard Hughes Medical Institute. This work was supported in part by NIH HD-23245, by a French Armed Forces Scholarship to OA, by a fellowship from the Life and Health Insurance Medical Research Fund to LCB, and by grants from the Dept. of Energy (DE FG05 88ER60692) and the Hereditary Disease Foundation to DLN.

REFERENCES

1. Nelson,D.L., Ballabio,A., Victoria,M.F., Pieretti,M., Bies,R.D., Gibbs,R.A., Maley,J.A., Chinault,A.C., Webster,T.D., and Caskey,C.T. (1991) Proc. Natl. Acad. Sci.USA **88**, 6157-6161.
2. Okabe,I., Attree,O., Bailey,L.C., Nelson,D.L., Nussbaum,R.L. (1992) J.Inherited Metabolic Dis., in press.
3. Laurent,B.C., Treitel,M.A., and Carlson,M. (1991) Proc. Natl. Acad. Sci. U.S.A. **88**, 2687-2691.
4. Laurent,B.C., Yang,X., and Carlson, M. (1992) Mol. Cell. Biol. **12**, 1893-1902.
5. Tamkun,J.W., Deuring,R., Scott,M.P., Kissinger,M., Pattatucci,A.M., Kaufman,T.C., and Kennison,J.A. (1992) Cell **68**, 561-572.
6. Meyuhas,O. and Perry,R.P. (1980) Gene **10**, 113-129.
7. Maniatis,T., Fritsch,E.F., and Sambrook,J. (1989) Molecular cloning: a laboratory manual (second edition). Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
8. Hodgson,S.V., Heckmatt,J.Z., Hughes,E., Crolla,J.A., Dubowitz,V., and Bobrow, M. (1986) Am. J. Med. Genet. **23**, 837-847.
9. Reilly,D.S., Lewis,R.A., Ledbetter,D.H.,and Nussbaum,R.L. (1988) Am. J. Hum. Genet. **42**, 748-755.
10. Scott,A.F., Phillips,J.A.,III, and Migeon,B. (1979) Proc. Natl. Acad. Sci. U.S.A. **76**, 4563-4565.
11. Nussbaum,R.L., Airhart,S.D., and Ledbetter,D.H. (1986) Am. J. Med. Genet. **23**, 457-466.
12. Devereux,J., Haerberli,P. and Smithies,O. (1984). Nucleic Acids Res. **12**, 387-395.
13. Pearson,W.R. and Lipman,D.J. (1988) Proc. Natl. Acad. Sci. U.S.A. **85**, 2444-2448.
14. Rose, M.D., Winston, F., and Hieter, P. (1990) Methods in yeast genetics, a laboratory course manual. (Cold Spring Harbor, NY: Cold Spring Harbor Press).
15. Ma,J. and Ptashne,M. (1987) Cell **51**, 113-119.
16. Brent,R. and Ptashne,M. (1985) Cell **43**, 729-736.
17. Hanes,S.D. and Brent,M.R. (1989). Cell **57**, 1275-1283.
18. West,R.W.,Jr., Yocum, R.R. and Ptashne,M. (1984) Mol. Cell. Biol. **4**, 2467-2478.
19. Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J.A. and Struhl,K. (1989) Current protocols in molecular biology. (New York: John Wiley & Sons) pp 13.6.1-13.6.4.
20. Gorbatenya,A.E., Koonin, E.V., Donchenko,A.P., and Blinov,V.M. (1989) Nucleic Acids Res. **12**, 4713-4730.
21. Atchison,M.L., Delmas,V., and Perry,R.P. (1990) EMBO J. **9**, 3109-3117.
22. Davis,J.L., Kunisawa,R. and Thorner,J. (1992) Mol. Cell. Biol., **12**, 1879-1892.
23. Emery,H.S., Schild,D., Kellogg,D.E. and Mortimer,R.K. (1991) Gene **104**, 103-106.
24. Girdham,C.G. and Glover, D.M. (1991) Genes Dev. **5**, 1786-1799.
25. Peterson,C.L. and Herskowitz, I. (1992) Cell **68**, 573-583.
26. Sanger,W.G., Grierson,H.L., Skare,J., Wyandt,H., Pirruccello,S., Fordyce, R., and Purtilo,D.T. (1990) Cancer Genet. Cytogenet. **47**, 163-169.