Positional cloning of the mouse *obese* gene and its human homologue

Yiying Zhang*†, Ricardo Proenca*†, Margherita Maffei†, Marisa Barone*†, Lori Leopold*† & Jeffrey M. Friedman*†‡

* Howard Hughes Medical Institute, † The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA

The mechanisms that balance food intake and energy expenditure determine who will be obese and who will be lean. One of the molecules that regulates energy balance in the mouse is the *obese* (*ob*) gene. Mutation of *ob* results in profound obesity and type II diabetes as part of a syndrome that resembles morbid obesity in humans. The *ob* gene product may function as part of a signalling pathway from adipose tissue that acts to regulate the size of the body fat depot.

OBESITY is the commonest nutritional disorder in Western societies. More than three in ten adult Americans weigh at least 20% in excess of their ideal body weight[1]. Increased body weight is an important public health problem because it is associated with type II diabetes, hypertension, hyperlipidaemia and certain cancers[2]. Although obesity is often considered to be a psychological problem, there is evidence that body weight is physiologically regulated[3].

The molecular pathogenesis of obesity is unknown. To identify components of the physiological system controlling body weight, we have applied positional cloning technologies in an attempt to isolate mouse obesity genes. Five single-gene mutations in mice that result in an obese phenotype have been described[3]. The first of the recessive obesity mutations, the *obese* mutation (*ob*), was identified in 1950[4]. *ob* is a single gene mutation that results in profound obesity and type II diabetes as part of a syndrome that resembles morbid obesity in humans[5]. Neither the primary defect nor the site of synthesis of the *ob* gene product is known. Cross-circulation experiments between mutant and wild-type mice suggest that *ob* mice are deficient for a blood-borne factor that regulates nutrient intake and metabolism[6], but the nature of this putative factor has not been determined.

We report the cloning and sequencing of the mouse *ob* gene and its human homologue. *ob* encodes a 4.5-kilobase (kb) adipose tissue messenger RNA with a highly conserved 167-amino-acid open reading frame. The predicted amino-acid sequence is 84% identical between human and mouse and has features of a secreted protein. A nonsense mutation in codon 105 has been found in the original congenic C57BL/6J *ob*/*ob* mouse strain, which expresses a twentyfold increase in *ob* mRNA. A second *ob* mutant, the co-isogenic SM/Ckc-+$^{Dac}ob^{2J}$/ob^{2J} strain, does not synthesize *ob* RNA. These data suggest that the *ob* gene product may function as part of a signalling pathway from adipose tissue that acts to regulate the size of the body fat depot.

For the positional cloning of mutant genes from mammals, it is necessary first to obtain genetic and physical maps, then to isolate the gene and detect the mutation. Here we describe the successful use of this approach to identify the *ob* gene.

Genetic and physical mapping of *ob*

The first *ob* mutation (carried on the congenic C57Bl/6J *ob*/*ob* strain) was found proximal to the *Microphthalmia* (*Mi*) and *waved-1 (wav-1)* loci on proximal mouse chromosome 6 (ref. 7); a second co-isogenic allele of *ob* has been identified in the SM/Ckc-Dac mouse strain (S. Lane, personal communication). We previously positioned *ob* relative to a series of molecular markers on mouse chromosome 6 and mapped the *ob* gene close to a restriction-fragment length polymorphism (RFLP) marker, D6Rck 13, derived from chromosome microdissection[5,8]; we also found that *Pax4* in the proximal region of mouse chromosome 6 is tightly linked to *ob* (ref. 9). Both loci were initially used to type a total of 835 informative meioses derived from both interspecific and intersubspecific mouse crosses that were segregating *ob*. *Pax4* was mapped proximal to *ob* and was recombinant in two animals (111 and 420 in Fig. 1); no recombination between D6Rck13 and *ob* was detected among the first 835 meioses scored[8].

To isolate the *ob* gene we cloned the DNA in the region of *Pax4* and D6Rck13 (Fig. 1), using both probes to start the construction of a physical map in the region of *ob*. Yeast artificial chromosomes (YACs) corresponding to *Pax4* and D6Rck13 were isolated and characterized. Centromeric and telomeric ends of each YAC were recovered, and ends mapping closer to *ob* used to screen for new YACs. YAC ends were recovered using either vectorette polymerase chain reaction (PCR) or the plasmid end-rescue technique[10,11]. One of the ends (labelled (1) in Fig. 1) of a D6Rck13 YAC, 902A0653, was recombinant in animal 257, positioning *ob* between this YAC end and *Pax4*. We were unable to recover any YACs linking the ends of YACs yB1S4A5 and 902A0653 (labelled (2) and (3) in Fig. 1). Pulsed-field gel electrophoresis (PFGE) indicated that there was a ~70-kb gap separating the two YAC ends. To bridge this gap, we used both YAC ends to isolate a set of mouse P1 clones[12,13]. Analysis of the ends of these P1 clones showed that they overlapped and that the gap in the YAC contig was ~70 kb. The size of the contig spanning the *ob* locus was 2.2 megabases (Mb) and *ob* was localized to the 900 kb between the distal end of YAC 903E1016 (labelled (5) in Fig. 1) and the distal end of YAC 902A0653 (labelled (1) in Fig. 1).

To position *ob* more precisely, we genotyped an additional 771 meioses derived from both a C57BL/6J *ob*/*ob* × DBA/2J intercross and backcross[5]. The typing of the intraspecific crosses required the development of informative single-strand length polymorphisms (SSLPs) for both D6Rck13 and Pax4. Sequencing of the *Pax4* gene itself revealed a microsatellite sequence, and an SSLP near D6Rck13, D6Rck39, was identified by sequencing cosmid subclones from YAC y902A0653 (a YAC isolated with D6Rck13). PCR amplification of genomic DNA with primers flanking these microsatellites revealed polymorphisms among the various progenitor strains of the genetic crosses. No additional recombinants between *ob* and *Pax4* were identified after genotyping the obese backcross and intercross progeny from the crosses to DBA mice. The genetic results indicated that D6Rck39 was distal to *ob* and recombined with *ob* in a single obese animal

‡ To whom correspondence should be addressed.

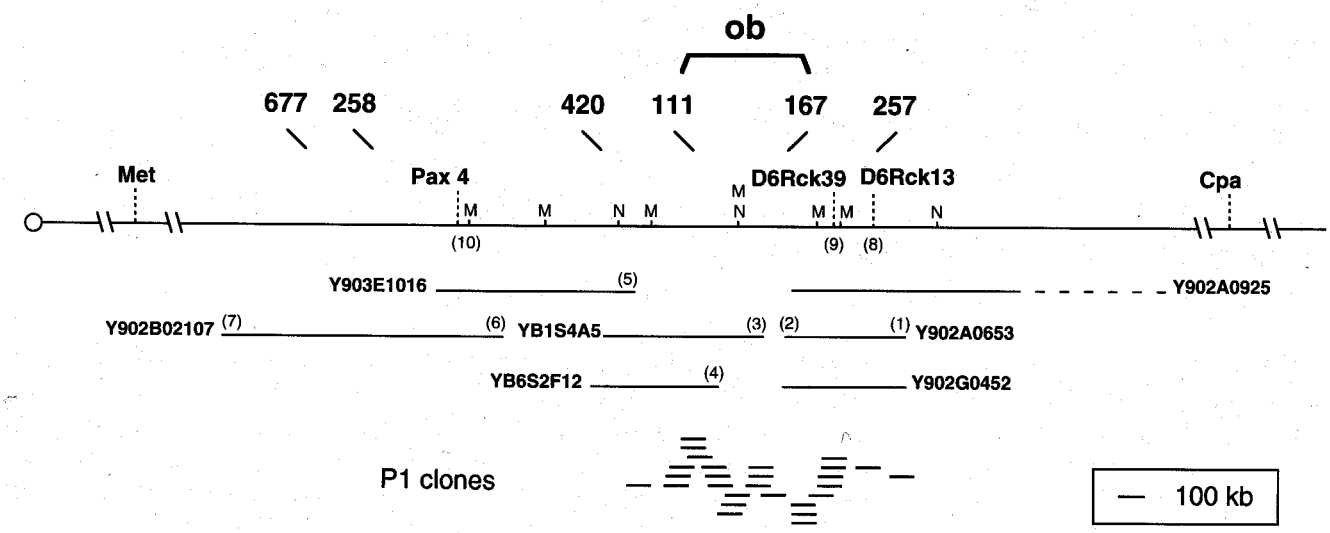


FIG. 1 Physical map of the mouse *ob* locus. A chromosome walk across the *ob* locus was completed in YACs and P1 clones using the flanking markers D6Rck13 and *Pax4* as starting points. The position of the rare-cut restriction enzymes *Mlu*1 (M) and *Not*1 (N) are indicated. Numbers in bold type indicate designation of recombinant animals in the region of *ob* among the 1,606 meioses scored. Each of the ends of the YAC clones isolated with D6Rck13 and *Pax4* were recovered using vectorette PCR and/or plasmid end rescue and used to type the recombinant animals. These ends were sequenced and used in turn to isolate new YACs. The resulting YAC contig is shown on the middle panel. One of the YACs in this contig, y902A0925, was chimaeric. Each of the probes used to genotype the recombinant animals is indicated in parentheses. On this basis, *ob* was localized to the interval between the recombination events in animals 111 and 167, (between end (5) and D6Rck39). Selected probes were also used to isolate bacteriophage P1 clones across the non-recombinant region. The resulting P1 contig is shown in the bottom panel. These P1 bacteriophage spanned a ~70-kb interval between probes (2) and (3) which could not be recovered in a YAC clone. The *ob* gene was identified in a P1 clone isolated using the distal end of YAC yB6S2F12(end 4).

METHODS. YAC clones were isolated by PCR screening or hybridization to high-density filters of available mouse YAC libraries[38–40]. The YACs from ICRF begin with a 902 prefix. YAC clones were sized on a CHEF MAPPER (Bio-Rad). Restriction enzyme digestions were done according to the manufacturers recommendations. YAC ends were recovered using vectorette PCR and plasmid end rescue[10,11]. P1 clones were isolated by sending specific pairs of PCR primers to Genome Systems Inc (St Louis, MO) who provided single picks of individual mouse P1 clones[12]. P1 ends were recovered using vectorette PCR[13]. Cosmid subclones from YAC y902A0653 were isolated as described[41]. Primer selection and PCR amplification of simple sequence repeats were performed as described: initial denaturation at 94 °C for 3 min, 25 cycles of 94° for 1 min, 55° for 2 min and 72° for 3 min. Primer sequences for *Pax4*: Pax4F, 5′-GGAGGTAGAGATGGCAGCAG-3′; *Pax4R*, 5′-ACAGA-AAGCAAGGAGGATTTC-3′, with a product size of 126 bp. Primer sequences for D6Rck39 were: 39GTF, 5′-GCACACTGACAGTGCCCTTA-3′; 39GTR, 5′-TGTAACCTGGAATTGGGAGC-3′ with a product size of 128 bp. The breeding and maintenance of the various mouse crosses have been described[8,42].

derived from the C57BL6/J *ob*/*ob* × DBA/2J backcross, animal 167. This recombination occurred between D6Rck39 and the distal end of YAC yB6S2F12 (end (4) in Fig. 1) because a B × D polymorphism defined by this end indicated that it was non-recombinant in animal 167 as well as animals 111 and 420. Thus the distal end of YAC yB6S2F12 (end (4) in Fig. 1) failed to recombine with *ob* among the 1,606 meioses scored. As recombination in animal 111 was localized between this end and the distal end of YAC y902E1016 (end (5), which was non-recombinant in animal 420), the combined data from animals 111 and 167 placed *ob* in, at most, a 650-kb interval between D6Rck39 and YAC end (5). The exact size of the critical region could not be determined until the points of recombination in these animals were precisely localized. For reasons discussed later, fine mapping of these recombination events was not necessary. There was a total of 6 recombination events in this 2.2-Mb contig among 1,606 meioses. Therefore in the region of *ob*, 1 cM of genetic distance corresponded to ~5.8 Mb, a rate of recombination nearly threefold lower than average for the entire mouse genome[14].

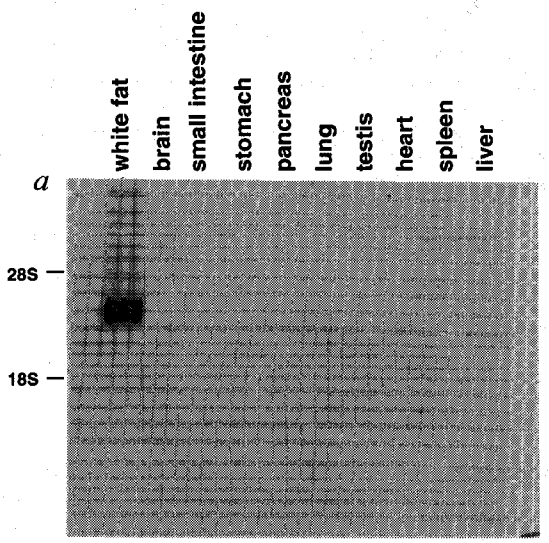
To facilitate the identification of genes in the region of *ob*, we isolated P1 clones using the ends of the YACs in this region as well as the ends derived from the initial set of P1 clones. Twenty-four P1 clones were used to construct a contig across most of the 650-kb critical region of *ob*.

Gene identification

Genes from this 650-kb interval were isolated using the method of exon trapping[15]. DNA from individual or pools of P1 clones was subcloned as *Bam*H1/*Bgl*II digests into the pSPL3 exon trapping vector. Briefly, each exon trapping product was sequenced and compared to all sequences in Genbank using the BLAST computer programme[16]. Putative exons were screened for the presence of corresponding RNA from a variety of tissues using northern blots and reverse-transcription PCR. Six genes were identified: four mapped within the 650-kb critical region of *ob* and two to outside this region.

One of the trapped exons, 2G7, was hybridized to a northern blot of mouse tissues and detected a ~4.5-kb RNA only in white adipose tissue (Fig. 2*a*). This exon was derived from a P1 isolated with the distal end of YB6S2F12 YAC (labelled (4) in Fig. 1). Even when autoradiographs were exposed for up to a week, no signal was detected in any other tissue, but we cannot exclude the possibility that this gene may be expressed elsewhere or below the level of detection in some tissues. Actin mRNA was detected in all samples (data not shown). Apparent adipose-tissue-specific expression was also seen after RT-PCR of RNA from a variety of tissues using specific primers from the 2G7 exon, with a strong signal being seen only in white fat and actin being detectable in all tissues tested (Fig. 2*b*). The high level of expression of 2G7 in adipose tissue suggested that this exon might be derived from the *ob* gene.

The level of expression of the 2G7 exon was then assayed in the two available *ob* strains by hybridization to northern blots as well as by RT-PCR. Northern blots showed that 2G7 RNA was absent in SM/Ckc-+$^{Dac}ob^{2J}$/ob^{2J} adipose tissue (Fig. 3*a*). This lack of 2G7 RNA in these animals was demonstrated by RT-PCR of the fat cell RNA, as shown by the absence of a signal after thirty cycles of amplification (Fig. 3*b*). As the ob^{2J} mutation is relatively recent and is maintained as a co-isogenic



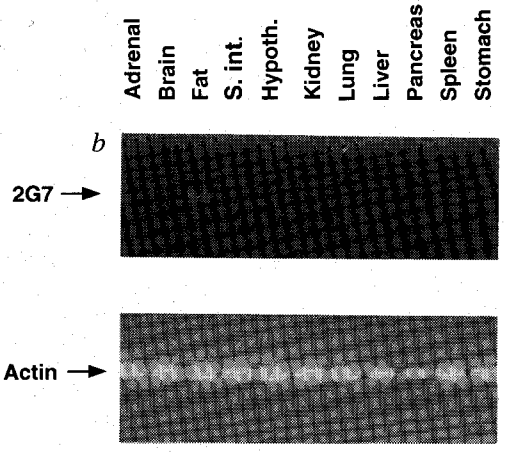
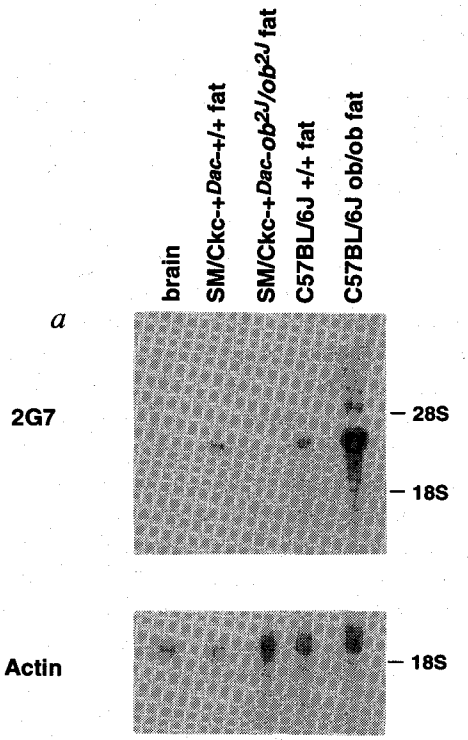


FIG. 2 Tissue distribution of the 2G7 transcript. *a*, Northern blot of total RNA (10 μg) from various tissues probed with labelled 2G7 exon. The 2G7 exon was identified using exon trapping with DNA from a pool of P1 clones in the region of *ob*. This probe hybridized specifically to RNA from white adipose tissue. Autoradiograph signals appeared after 1-h exposure (24-h exposure shown here). The transcript migrated between 28S and 18S ribosomal RNA markers and is estimated to be ~4.5 kb. *b*, Reverse transcription-PCR (RT-PCR) was performed with RNA from each of the tissue samples shown using primers specific for the 2G7 exon or actin. A positive signal was detectable only in white adipose tissue, even when PCR amplification was continued for 30 cycles.
METHODS. *a*, Exon trapping was done by ligating *Bgl*II/*Bam*H1 digestion products of a pool of P1 clones into the pSPL III vector as described[15]. Total RNA was prepared and electrophoresed in formaldehyde gels as in ref. 43. Northern blots were transferred to Immobilon and hybridized to radiolabelled probes as described[44]. *b*, Reverse-transcription PCR reactions were performed using 100 ng total RNA[45]. First-strand cDNA, prepared using random hexamers as primers, was PCR-amplified using primers derived from the 2G7 exon as well as mouse actin. The primers used to detect 2G7 were selected using the Primer program and were: 2G7F(5′CCAGGGCAGGAAAATGTG3′): 2G7R(5′CATCCTGGACTTTCTGGATAGG3′). The mouse actin primers were purchased from Clonetech. PCR amplification was performed for 30 cycles with 94° denaturation for 1 min 55° hybridization for 1 min, and 72 extensions for 2 min with a 1-s autoextension per cycle. RT-PCR products were resolved in a 1.5% low-melting-point agarose (Gibco/BRL) gel run in 0.5 × TBE buffer.



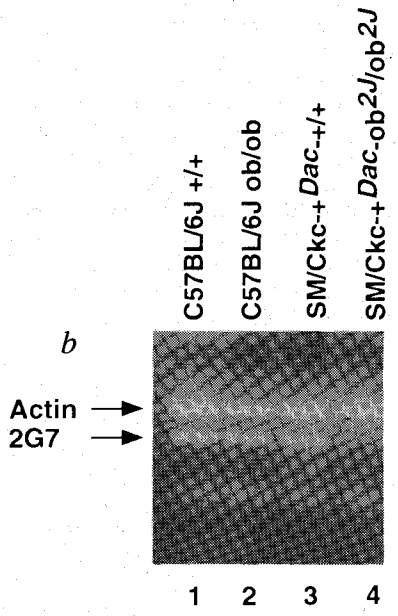


FIG. 3 2G7 expression in mutant mice. *a*, Northern blot of fat cell RNA isolated from *obese* and lean mice. The 2G7 exon was hybridized to northern blots with 10 μg total RNA from white adipose tissue from each of the strains indicated. An approximately 20-fold increase in the level of 2G7 RNA was apparent in white fat RNA from the C57BL/6J *ob/ob* strain relative to lean littermates. There was no detectable signal in RNA from the SM/Ckc-+$^{Dac}ob^{2J}/ob^{2J}$ mice even after a 2-week exposure. A 24-h autoradiographic exposure is shown. The same filter was hybridized to an actin probe (bottom panel). SM/Ckc-+^{Dac} and SM/Ckc-+$^{Dac}ob^{2J}/ob^{2J}$ mice were provided by S. Lane and B. Paigen of the Jackson Laboratory. C57Bl/6J *ob/ob* mice were purchased from the Jackson Laboratory. *b*, RT-PCR of mutant RNA. Reverse-transcription PCR was performed for 30 cycles of amplification using 100 ng total fat RNA from each of the samples shown. In this experiment both the actin and 2G7 primer pairs were included in the same PCR reaction. The complete absence of 2G7 RNA was demonstrated in the SM/Ckc-+$^{Dac}ob^{2J}/ob^{2J}$ adipose tissue. For RT-PCR reactions, 100 ng total RNA was used[45].

strain, these data indicate that 2G7 encodes an exon from the *ob* gene. This was confirmed by characterization of the mutation in C57L/6J *ob/ob* mice. Northern blots of adipose tissue RNA from C57BL/6J *ob/ob* mice showed a ~20-fold increase in the level of *ob* RNA (Fig. 3*a*), suggesting that the original *ob* allele (C57BL/6J *ob/ob*) was associated with a non-functional gene product and that the mRNA was increased as part of a possible feedback loop. This turned out to be the case (Fig. 4*b*) as the C57BL/6J *ob/ob* phenotype is the result of a nonsense mutation.

Sequence of the *ob* gene

The 2G7 exon was used to isolate a total of 22 complementary DNA clones from a mouse adipose tissue cDNA library. None of the cDNA clones extended more than 97 base pairs upstream of 2G7, whereas each extended a variable distance downstream. Sequencing of the cDNA clones revealed a methionine initiation codon in the 2G7 exon, with a 167-amino-acid open reading frame followed by a long 3′-untranslated sequence (Fig. 4). A potential Kozak translation initiation consensus sequence was present with an adenosine residue three bases upstream of the ATG[17]. One of the cDNA clones extended to the 5′ end of the mRNA because its sequence was identical to that of the 5′ RACE (rapid amplification of cloned ends) products of adipose tissue RNA (data not shown)[18]. A total of 2.9 kb of cDNA sequence is shown, most of which is 3′-untranslated sequence. A search for internal homology within the cDNA revealed a 50-base-pair (bp) direct repeat in both the 5′ and 3′ untranslated sequence; this sequence was not found in Genbank and there were no additional segments of internal homology.

Two classes of cDNA were found which differed by inclusion or exclusion of a single glutamine codon (underlined in Fig. 4). This residue is found in a position immediately 3′ to the splice acceptor of the 2G7 exon. As the CAG codon of glutamine includes a possible AG splice-acceptor sequence, there appears to be slippage at the splice-acceptor site, with an apparent 3-bp deletion in a subset of the cDNAs[19]. This glutamine residue is located in a highly conserved region of the molecule but its significance is unknown.

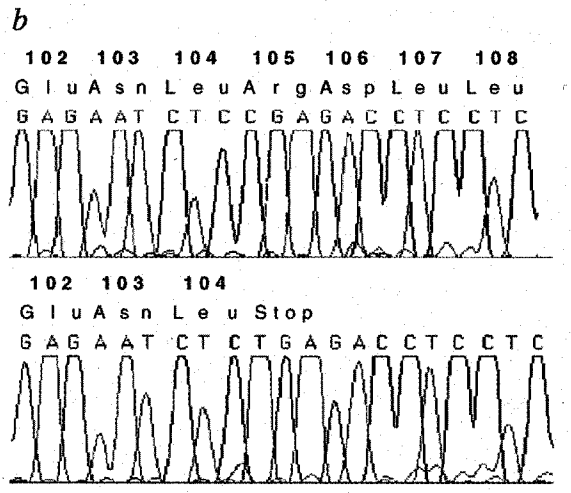
To identify the mutation in C57B6/J *ob/ob* mice, RT-PCR products from the entire open reading frame (ORF) were prepared from adipose tissue RNA from this mutant and both strands sequenced. The coding sequences were identical apart from a C→T mutation in C57BL/6J *ob/ob* mice that results in a change of an arginine at position 105 to a stop condon (Fig. 4*b*). This amino acid is also underlined in Fig. 4*a*. This base change did not occur in genomic DNA from v/LE or C57/BL10 mice, the strains in which this mutation was carried before it was transferred to C57BL/6J (S. Lane, personal communication). DNA sequence changes from CGA to TGA are not uncommon as a result of the high mutation rate of methyl cytosine to thymidine[20]. This *ob* mutation is presumed to inactivate the protein as the phenotype in both C57BL/6J and SM/Ckc-$+^{Dac}ob^{2J}/ob^{2J}$ mice is identical (S. Lane, personal communication; data not shown).

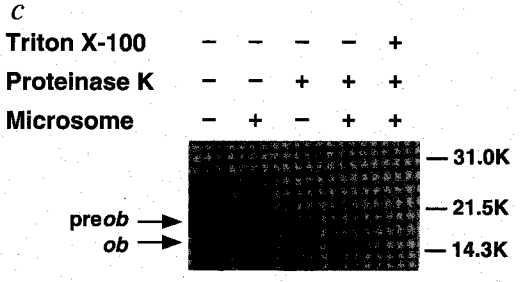
A database search of the *ob* protein using the BLAST program[16] identified no significant homology to any sequences in Genbank. The predicted polypeptide was largely hydrophilic and had a putative N-terminal signal sequence[22]. Amino-acid sequence analysis using the *sigseq* computer program indicated a 98% likelihood that a signal sequence was present at the N terminus[21] (underlined in Fig. 4*a*). The predicted signal sequence cleavage site is C terminal to an alanine at position 21. Computer analysis of the human protein (Fig. 6*b*) also suggests that it has an N-terminal signal sequence, so human and mouse *ob* may both encode secreted proteins. There is some divergence in the mouse and human signal sequences, but both give identical scores when analysed using *segseq*. To confirm the presence of a functional signal sequence, a human cDNA that included the entire ORF was subcloned into a pGEM vector (suitable mouse subclones were not recovered). Positive-strand human *ob* RNA

a

```
ATATTGCTGAGCTCAGGGAGTGAGGGCCCCACATTTGAGACAGTGAGCCCCAAGAAGAGG    60
GATCCCTGCTCCAGCAGCTGCAAGGTGCAAGAAGAAGAAGATCCCAGGGAGGAAAATGTG   120
                                                      M  C        2
CTGGAGACCCCTGTGTCGGTTCCTGTGGCTTTGGTCCTATCTGTCTTATGTTCAAGCAGT   180
W  R  P  L  C  R  F  L  W  L  W  S  Y  L  S  Y  V  Q  A  V      22
GCCTATCCAGAAAGTCCAGGATGACACCAAAACCCTCATCAAGACCATTGTCACCAGGAT   240
P  I  Q  K  V  Q  D  D  T  K  T  L  I  K  T  I  V  T  R  I      42
CAATGACATTTCACACACGGCAGTCGGTATCCGCCAAGCAGAGGGTCACTGGCTTGGACTT   300
N  D  I  S  H  T  Q  S  V  S  A  K  Q  R  V  T  G  L  D  F      62
CATTCCTGGGCTTCACCCCATTCTGAGTTTGTCCAAGATGGACCAGACTCTGGCAGTCTA   360
I  P  G  L  H  P  I  L  S  L  S  K  M  D  Q  T  L  A  V  Y      82
TCAACAGGTCCTCACCAGCCTGCCTTCCCAAAATGTGCTGCAGATAGCCAATGACCTGGA   420
Q  Q  V  L  T  S  L  P  S  Q  N  V  L  Q  I  A  N  D  L  E     102
GAATCTCCGAGACCTCCTCCATCTGCTGGCCTTCTCCAAGAGCTGCTCCCTGCCTCAGAC   480
N  L  R  D  L  L  H  L  L  A  F  S  K  S  C  S  L  P  Q  T     122
CAGTGGCCTGCAGAAGCCAGAGAGCCTGGATGGCGTCCTGGAAGCCTCACTCTACTCCAC   540
S  G  L  Q  K  P  E  S  L  D  G  V  L  E  A  S  L  Y  S  T     142
AGAGGTGGTGGCTTTGAGCAGGCTGCAGGGCTCTCTGCAGGACATTCTTCAACAGTTGGA   600
E  V  V  A  L  S  R  L  Q  G  S  L  Q  D  I  L  Q  Q  L  D     162
TGTTAGCCCTGAATGCTGAAGTTTCAAAGGCCACCAGGCTCCCAAGAATCATGTAGAGGG   660
V  S  P  E  C  *                                               167
AAGAAACCTTGGCTTCCAGGGGTCTTCAGGAGAAGAGAGCCATGTGCACACATCCATCAT   720
TCATTTCTCTCCCTCCTGTAGACCACCCATCCAAAGGCATGACTCCACAATGCTTGACTC   780
AAGTTATCCACACAACTTCATGAGCACAAGGAGGGGCCAGCCTGCAGAGGGGACTCTCAC   840
CTAGTTCTTCAGCAAGTAGAGATAAGAGCCATCCCATCCCCTCCATGTCCCACCTGCTCC   900
GGGTACATGTTCCTCCGTGGGTACACGCTTCGCTGCGGCCCAGGAGAGGTGAGGTAGGGA   960
TGGGTAGAGCCTTTGGGCTGTCTCAGAGTCTTTGGGAGCACCGTGAAGGCTGCATCCACA  1020
CACAGCTGGAAACTCCCAAGCAGCACACGATGGAAGCACTTATTTATTTATTCTGCATTC  1080
TATTTTGGATGGATCTGAAGCAAGGCATCAGCTTTTTCAGGCTTTGGGGGTCAGCCAGGA  1140
TGAGGAAGGCTCCTGGGGTGCTGCTTTCAATCCTATTGATGGGTCTGCCCGAGGCAAACC  1200
TAATTTTTGAGTGACTGGAAGGAAGGTTGGGATCTTCCAAACAAGAGTCTATGCAGGTAG  1260
CGCTCAAGATTGACCTCTGGTGACTGGTTTTGTTTCTATTGTGACTGACTCTATCCAAAC  1320
ACGTTTGCAGCGGCATTGCCGGGAGCATAGGCTAGGTTATTATCAAAAGCAGATGAATTT  1380
TGTCAAGTGTAATATGTATCTATGTGCACCTGAGGGTAGAGGATGTGTTAGAGGGAGGGT  1440
GAAGGATCCGGAAGTGTTCTCTGAATTACATATGTGTGGTAGGCTTTTCTGAAAGGGTGA  1500
GGCATTTTCTTACCTCTGTGGCCACATAGTGTGGCTTTGTGAAAAGGACAAAGGAGTTGA  1560
CTCTTTCCGGAACATTTGGAGTGTACCAGGCACCCTTGGAGGGGCTAAAGCTACAGGCCT  1620
TTTGTTGGCATATTGCTGAGCTCAGGGAGTGAGGGCCCCACATTTGAGACAGTGAGCCCC  1680
AAGAAAAGGGTCCCTGGTGTAGATCTCCAAGGTTGTCCAGGGTTGATCTCACAATGCGTT  1740
TCTTAAGCAGGTAGACGTTTGCATGCCAATATGTGGTTCTCATCTGATTGGTTCATCCAA  1800
AGTAGAACCCTGTCTCCCACCCATTCTGTGGGGAGTTTTGTTCCAGTGGGAATGAGAAAT  1860
CACTTAGCAGATGGTCCTGAGCCCTGGGCCAGCACTGCTGAGGAAGTGCCAGGGCCCCAG  1920
GCCAGGCTGCCAGAATTGCCCTTCGGGCTGGAGGATGAACAAAGGGGCTTGGGTTTTTCC  1980
ATCACCCCTGCACCCTATGTCACCATCAAACTGGGGGGCAGATCAGTGAGAGGACACTTG  2040
ATGGAAAGCAATACACTTTAAGACTGAGCACAGTTTCGTGCTCAGCTCTGTCTGGTGCTG  2100
TGAGCTAGAGAAGCTCACCACATACATATAAAAATCAGAGGCTCATGTCCCTGTGGTTAG  2160
ACCCTACTCGCGGCGGTGTACTCCACCACAGCAGCACCGCACCGCTGGAAGTACAGTGCT  2220
GTCTTCAACAGGTGTGAAAGAACCTGAGCTGAGGGTGACAGTGCCCAGGGGAACCCTGCT  2280
TGCAGTCTATTGCATTTACATACCGCATTTCAGGGCACATTAGCATCCACTCCTATGGTA  2340
GCACACTGTTGACAATAGGACAAGGGATAGGGGTTGACTATCCCTTATCCAAAATGCTTG  2400
GGACTAGAAGAGTTTTGGATTTTAGAGTCTTTTCAGGCATAGGTATATTTGAGTATATAT  2460
AAAATGAGATATCTTGGGGATGGGGCCCAAGTATAAACATGAAGTTCATTTATATTTCAT  2520
AATACCGTATAGACACTGCTTGAAGTGTAGTTTTATACAGTGTTTTAAATAACGTTGTAT  2580
GCATGAAAGACGTTTTTACAGCATGAACCTGTCTACTCATGCCAGCACTCAAAAACCTTG  2640
GGGTTTTGGAGCAGTTTGGATCTTGGGTTTTCTGTTAAGAGATGGTTAGCTTATACCTAA  2700
AACCATAATGGCAAACAGGCTGCAGGACCAGACTGGATCCTCAGCCCTGAAGTGTGCCCT  2760
TCCAGCCAGGTCATACCCTGTGGAGGTGAGCGGGATCAGGTTTTGTGGTGCTAAGAGAGG  2820
AGTTGGAGGTAGATTTTGGAGGATCTGAGGGC                               2852
```

FIG. 4 *a*, Sequence of *ob* cDNA. 22 cDNA clones were isolated from a mouse white fat cDNA library and sequenced. A 97-bp 5′ leader was followed by a predicted 167-amino-acid ORF and a ~3,700-kb 3′-untranslated sequence. A total of ~2,500 base pairs of the 3′-untranslated sequence is shown. Analysis of the predicted protein sequence using the *sigseq* computer ▶





program suggests the presence of a signal sequence (underlined)[21]. Microheterogeneity of the cDNA was noted in that ~70% of the cDNAs had a glutamine codon at codon 49 and 30% did not. This amino acid is underlined as is the arginine codon that is mutated in C57BL/6J *ob/ob* mice (see *b*). After cleavage of the signal sequence, two cysteines remain at the C terminus of the molecule, possibly with formation of a disulphide bond. cDNA clones were isolated from a mouse adipose tissue library (Clonetech)[46]. Inserts were prepared directly from phage DNA or by PCR amplification of the insert using primers from the λgt10 cloning site (Clonetech). PCR products were prepared for sequencing after electrophoresis in low-melting-point agarose gels (Gibco/BRL) using agarase[47]. DNA was sequenced manually or using an ABI 373A automated sequencer[48,49]. 5′ RACE was done after dG tailing of first-strand cDNA using terminal transferase followed by PCR amplification with the reverse complement of the 2G7F primer and oligo(dC)[18]. *b*, The C57BL/6J *ob/ob* mutation. RT-PCR was carried out using white fat RNA from C57BL/6J +/+ and C57BL/6J *ob/ob* mice and primers from the 5′ and 3′ untranslated regions. PCR reaction products were gel-purified and sequenced using primers along both strands of the coding sequence. The C57BL/6J *ob/ob* mice had a C→T mutation which changed Arg 105 to a stop codon. This base change is shown as the output from the automated sequencer. The cDNA sequences of mutant and wild-type mice were otherwise identical. The wild-type cDNA sequence was also seen in genomic DNA from v/Le and C57BL/10 mice, the progenitor strains of the stock on which the C57BL/6J *ob/ob* mutation originally arose (data not shown). *c*, *In vitro* translation of *ob* RNA. A human *ob* cDNA was subcloned into the pGEM cloning vector (see Fig. 5*b*). Plus-strand RNA was synthesized using SP6 polymerase. The *in vitro* synthesized RNA was translated in the presence or absence of canine pancreatic microsomal membranes. An ~18K primary translation product was seen after *in vitro* translation. The addition of microsomal membranes (mb) to the reaction led to the appearance of a second translation product ~2K smaller than the primary translation product. The 16K product was resistant to proteinase K but was rendered protease-sensitive when the microsomal membranes were permeabilized by Triton X-100, indicating that a functional signal sequence was present. A human *ob* cDNA was cloned into the pGEM-3zf(+) plasmid. *In vitro* transcription was carried out using a Ribo MAX large-scale RNA production system with SP6 polymerase (Promega). The transcription mixture was used without further purification in a wheat-germ translation system (Promega) with or without 5 μg of a canine pancreas microsomal membrane preparation[50] at 27 °C for 2 h. Proteinase K (100 μg ml^{-1}) digests were performed on ice for 1 h with and without 0.1% Triton-X100. Translation products were analysed by SDS–PAGE.

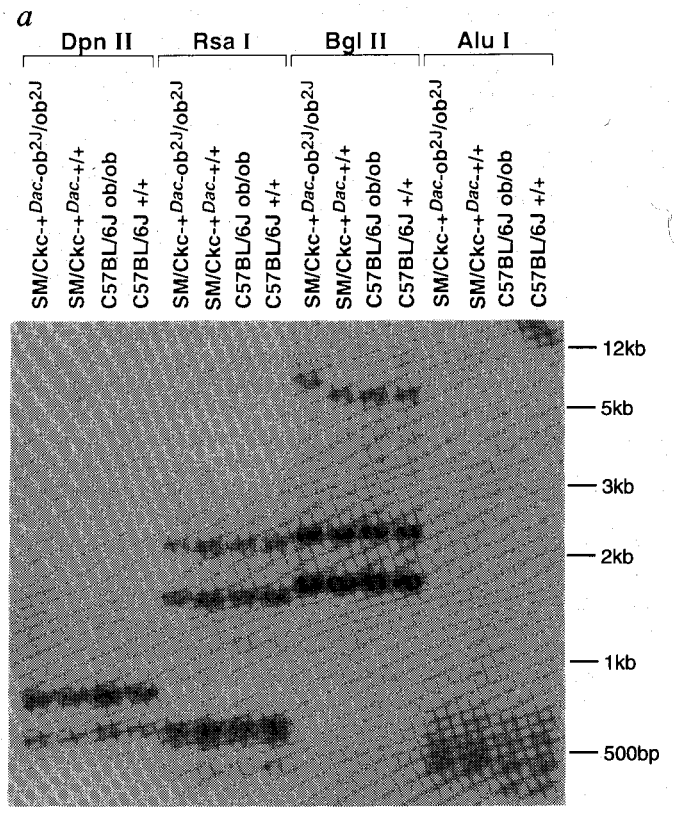
was transcribed using sp6 polymerase and translated *in vitro* with and without canine pancreatic microsomal membranes[50]. The primary translation product migrated with an apparent relative molecular mass of ~18K, consistent with that predicted from the cDNA sequence. Inclusion of microsomal membranes in the reaction inhibited translation ~5-fold, but about half of the *ob* primary translation product was truncated by ~2K in the presence of microsomal membranes, which suggests that the signal sequence is functional (Fig. 4*c*). To confirm that the *ob* protein had been translocated, we treated *in vitro* translation products with proteinase K, which caused complete proteolysis of the 18K primary translation product whereas the 16K processed form was unaffected, indicating that the translation product had been translocated into the microsomal lumen. Permeabilization of the membranes with Triton X-100 rendered the processed form protease-sensitive and is compatible with the hypothesis that *ob* is a secreted molecule. After signal-sequence cleavage, two cysteine residues would remain within the predicted protein, suggesting that the molecule may contain the disulphide bond characteristic of other secreted polypeptides. Amino-acid sequence and secondary structure prediction support a globular structure for the protein (data not shown). The largely hydrophilic amino-acid sequence had no notable structural motifs or membrane-spanning domains other than the N-terminal signal sequence. We find no consensus for *N*-linked glycosylation or dibasic amino-acid sequences indicative of protein cleavage in the predicted processed protein[22].

The ob^{2J} mutation

We next explored the molecular basis for the mutation in SM/Ckc-+$^{Dac}ob^{2J}/ob^{2J}$ mice. In these animals, the absence of 2G7 RNA was associated with an increase in the size of an ~9 kb *Bgl*II fragment of genomic DNA in affected animals (Fig. 5*a*). Although the precise nature of this polymorphism is not established, the altered *Bgl*II site maps ~7 kb upstream of the mRNA start site for the 4.5 kb *ob* RNA (data not shown), suggesting that this mutation may be the result of a structural alteration or sequence variation in the promoter. None of the other restriction fragments reach the promoter region. Nevertheless, this polymorphism is significant because it is always associated with the obese phenotype in a colony that segregates the *ob* 2J mutation (Fig. 5*b*). In this experiment, DNA from a total of six obese and five lean animals was Southern-blotted after *Bgl*II digestion and probed with 2G7. In each case homozygosity for the larger of the polymorphic *Bgl*II fragments was associated with an obese phenotype. Each of the lean animals was either homozygous for the smaller *Bgl*II allele (+/+) or heterozygous (*ob*/+). As reported, no overt phenotypic differences were apparent between +/+ and *ob*/+ mice[23].

ob sequence conservation

The coding sequence of the *ob* gene was hybridized to genomic Southern blots of vertebrate DNAs (Fig. 6*a*). Even at moderate stringency, there were detectable signals in all vertebrate DNAs tested, demonstrating that *ob* is highly conserved among vertebrates. To determine the extent of this *ob* sequence conservation, we isolated and sequenced cDNA clones hybridizing to *ob* from a human adipose tissue cDNA library. The nucleotide sequences from human and mouse were highly homologous in the predicted coding sequence, but had only 30% homology in the available 5′ and 3′ untranslated regions. Homology of the human polypeptide apparent N-terminal signal sequence to the mouse equivalent region was slightly lower than in the rest of the molecule. The signal sequence is seen to be functional from the truncation of the primary translation product in the presence of canine pancreas microsomal membranes (Fig. 4*c*). Alignment of the predicted human and mouse amino-acid sequences (Fig. 6*b*)



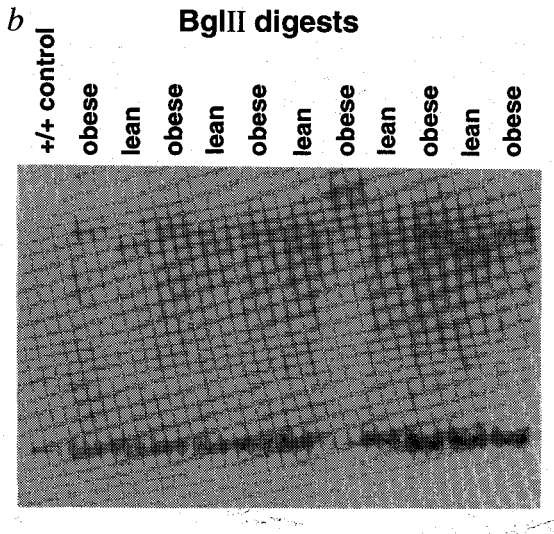
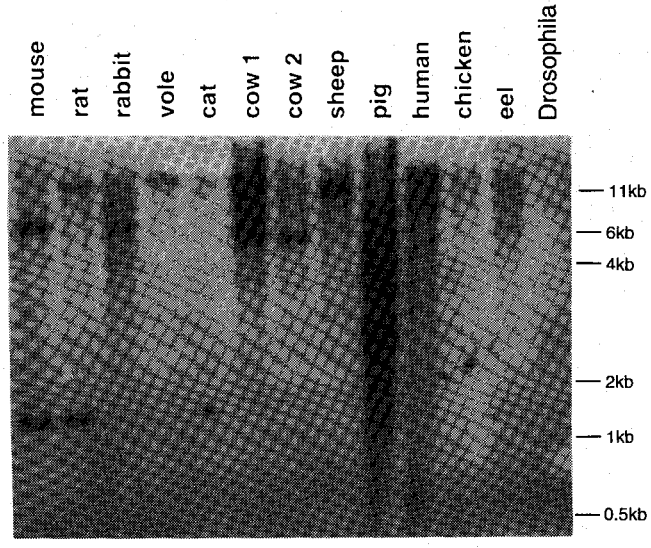


FIG. 5 The ob^{2J} mutation. *a*, Southern blots of genomic DNA from mutant animals. The 2G7 probe was hybridized to 5 μg of restriction-enzyme-digested genomic DNA from each strain. Restriction digestion with *Bgl*II revealed an increase in the size of a ~9 kb (the largest) *Bgl*II fragment in SM/Ckc-+$^{Dac}ob^{2J}/ob^{2J}$ DNA. RFLPs were not detectable with any other restriction enzymes. Preliminary restriction mapping of genomic DNA indicated that the polymorphic *Bgl*II site is ~7 kb upstream of the transcription start site (data not shown). None of the other enzymes tested extend past the mRNA start site. Genomic DNA preparation and Southern blotting have been described[8]. *b*, Segregation of a *Bgl*II polymorphism in the SM/Ckc-+$^{Dac}ob^{2J}/ob^{2J}$. Six obese and five lean progeny from the same generation of the coisogenic SM/Ckc-+$^{Dac}ob^{2J}/ob^{2J}$ colony were genotyped by scoring the *Bgl*II polymorphism in *a*. All of the phenotypically obese animals were homozygous for the larger allele of the polymorphic *Bgl*II fragment. The DNA in the control lane was prepared from an unrelated SM/Ckc-Dac$^{+/+}$ mouse, bred separately from the SM/Ckc-+$^{Dac}ob^{2J}/ob^{2J}$ colony.



b

```
Mouse   MCWRPLCRFL WLWSYLSYVQ AVPIQKVQDD TKTLIKTIVT RINDISHTQS    50
Human   MHWGTLCGFL WLWPYLFYVQ AVPIQKVQDD TKTLIKTIVT RINDISHTQS

Mouse   VSAKQRVTGL DFIPGLHPIL SLSKMDQTLA VYQQVLTSLP SQNVLQIAND   100
Human   VSSKQKVTGL DFIPGLHPIL TLSKMDQTLA VYQQILTSMP SRNVIQISND

Mouse   LENLRDLLHL LAFSKSCSLP QTSGLQKPES LDGVLEASLY STEVVALSRL   150
Human   LENLRDLLHV LAFSKSCHLP WASGLETLDS LGGVLEASGY STEVVALSRL

Mouse   QGSLQDILQQ LDVSPEC                                      167
Human   QGSLQDMLWQ LDLSPGC
```

FIG. 6 Evolutionary conservation of *ob*. *a*, Cross-species hybridization of *ob*. Genomic DNA from each of the species shown was Southern blotted after *Eco*RI digestion and probed with *ob* cDNA. Hybridization signals were detectable in every vertebrate sample even after moderate-stringency hybridization. The cat DNA used in this experiment was slightly degraded. DNA was prepared from tissues or cell lines from the organisms shown. Southern blot hybridization was at 65 °C, with two washes in 2 × SSC/0.2% SDS at 65 °C for 20 min, and autoradiograph exposure was for 3 d using Kodak X-OMAT film. *b*, Sequence of human *ob* protein. The mouse *ob* gene was used as a probe to isolate human adipose tissue cDNA clones from a λgtII library. The sequence of human *ob* cDNA was highly homologous to that of the mouse cDNA in the predicted 167-amino-acid coding sequence. There was 30% homology in the 5′ and 3′ untranslated sequences in this tissue. Predicted amino-acid sequences of the human and mouse proteins have been aligned. Conservative changes are noted by a dash and non-conservative changes by an asterisk. The variable glutamine codon is underlined, as is the position of the nonsense mutation in C57BL/6J *ob/ob* mice. Overall there is 84% identity at the amino-acid level, although only seven substitutions were found between the valine at codon 22 (immediately downstream of the signal sequence overage) and the cysteine at position 117. cDNA clones were isolated from a human adipose tissue cDNA library (Clonetech) as described for Fig. 4, except that PCR amplification of the phage insert was carried out using primers adjacent to the cloning site of λgt11. PCR products were sequenced directly.

shows that there is 84% overall identity and more extensive identity in the N terminus of the mature protein, with only four conservative and three non-conservative substitutions among the residues between the signal sequence cleavage site and the conserved cysteine at position 117. As in mouse, 30% of the clones were missing a glutamine at codon 49.

Discussion

Obesity has been a focus of discussion[24], with a line of research dating back to Lavoisier and Laplace (1783) implying that energy balance—food intake versus energy output—is physiologically regulated[25–27]. The site of this regulation has been the subject of nearly one hundred years of debate[28–30]. Of the brain regions implicated in the regulation of feeding behaviour, the ventromedial nucleus of the hypothalamus (VMH) is considered to be the most important satiety centre in the central nervous system (CNS). The increase in body weight associated with VMH lesions is a result of both increased food intake and decreased energy expenditure[29].

The energy balance in mammals was therefore postulated to be controlled by a feedback loop in which the amount of stored energy is sensed by the hypothalamus, which adjusts food intake and energy expenditure to maintain a constant body weight[31,32]. The nature of the inputs to the hypothalamus was unclear[27]. According to the lipostasis theory[32], the size of the body fat depot is regulated by the CNS, with a product of fat metabolism circulating in plasma and affecting energy balance by interacting with the hypothalamus. The glucostasis theory[33] suggested that the plasma glucose level is a key signal regulating energy stores. A third possibility is that body temperature is an important input to the CNS centres controlling food intake[31]. The inability to identify the putative signal from fat has hindered the validation of the lipostasis theory. Moreover, neither the glucostasis theory nor theories on the thermal regulation of food intake fully account for the precision with which energy balance is regulated *in vivo*. The possibility that at least one component of the signalling system circulates in the bloodstream was first suggested by Hervey[34], who showed that the transfer of blood from an animal with a VMH lesion across a vascular graft to an untreated animal (a parabiosis experiment) resulted in a reduction of food intake in the intact animal. The biochemical nature of the putative signal has remained elusive, although it has been suggested that *ob* is responsible for the generation of this blood-borne factor[6].

Our results, particularly the evidence that the *ob* protein product is secreted, suggest that *ob* may encode this circulating factor. Evidence that the gene described here is *ob* includes the identification of a nonsense mutation in the C57BL/6J *ob/ob* strain and a genomic alteration in the SM/Ckc-+$^{Dac}ob^{2J}/ob^{2J}$ strain that is associated with an absence of RNA. The data showing that one strain overproduces *ob* RNA while the other fails to express this gene, make it unlikely that these changes are secondary. This conclusion can be confirmed using transgenic mice to complement the obese mutation. The nature of the mutation in the co-isogenic SM/Ckc-+$^{Dac}ob^{2J}/ob^{2J}$ mice is as yet unknown. Preliminary analysis of the gene structure has indicated that the polymorphic *Bgl*II site maps ~7 kb upstream of the promoter. As this mutation is co-isogenic, the data strongly suggest that this polymorphism is associated with the mutation, although it is not yet clear whether there is a structural alteration (insertion or deletion) or base change. Further studies will be required to identify the molecular basis for the absence of *ob* RNA expression in SM/Ckc-+$^{Dac}ob^{2J}/ob^{2J}$ mice, as well as the molecular mechanisms that result in the increased expression of *ob* RNA in C57BL/6J *ob/ob* mice.

The increased level of *ob* RNA in the C57BL/6J mutant, which has greatly increased fat cell mass, raises the possibility that the level of expression of this gene signals the size of the adipose depot. This hypothesis is consistent with the lipostasis theory, implying that an increase in the level of the *ob* signal (as might occur after a prolonged binge of eating) may act directly or indirectly on the CNS to inhibit food intake and/or regulate energy expenditure as part of a homeostatic mechanism to maintain constancy of the adipose mass. Such effects could be explained if the *ob* protein activated the sympathetic nervous system via interactions with the VMH. Direct effects of *ob* on the CNS would require that mechanisms exist to allow its passage across the blood–brain barrier. The means by which the CNS effects changes in body weight remain to be elucidated and may include mechanisms independent of effects on food intake and energy expenditure. In addition, *ob* may have endocrine, paracrine and autocrine effects on different tissues. These and other possibilities will be testable if the active form of the *ob* protein can be shown to circulate in plasma.

The *ob* mutation is associated with a myriad of hormonal and metabolic alterations[35]. These include effects on thermoregulation, fertility, adrenal and thyroid function as well as a wide range of biochemical abnormalities. Although we cannot exclude the possibility that many of these pleiotypic effects are the result of the expression of *ob* in sites not included in our initial survey of mouse tissues, the apparent expression of *ob* in only adipose tissue suggests that many of these changes are secondary. It has been suggested that the complex *ob* phenotype reflects an imbalance in the activity of the autonomic nervous system with low sympathetic and high parasympathetic tone[35]. If *ob* is a signal molecule that regulates body weight, it may act by interacting with CNS (and other) receptors to modulate food intake and the activity of the autonomic nervous system. Parabiosis experiments suggest that the *ob* receptor is encoded by the mouse *db* (diabetes) gene[6], a possibility that can be tested by positional cloning of *db* or by expression cloning and mapping of the *ob* receptor[36].

The extensive homology of the *ob* gene product among vertebrates suggests that its function is highly conserved. Now that the human homologue of *ob* is cloned, it will be possible to test for mutations in the human *ob* gene. As *ob* heterozygotes have an enhanced ability to survive a prolonged fast[23], heterozygous mutations at *ob* may provide a selective advantage in human populations subjected to caloric deprivation[37]. Identification of *ob* now offers an entry point into the pathways that regulate adiposity and body weight and should provide a fuller understanding of the pathogenesis of obesity. □

Received 17 October; accepted 4 November 1994.

1. Burros, M. *Despite awareness of risks, more in US are getting fat* in *The New York Times* 17 July 1994.
2. Grundy, S. M. & Barnett, J. P. *Disease-a-Month* **36,** 645–696 (1990).
3. Friedman, J. M. & Leibel, R. L. *Cell* **69,** 217–220 (1990).
4. Ingalls, A. M., Dickie, M. M. & Snell, G. D. *J. Hered.* **41,** 317–318 (1950).
5. Friedman, J. M. *et al. Genomics* **11,** 1054–1062 (1991).
6. Coleman, D. L. *Diabetologia* **14,** 141–148 (1978).
7. Dickie, M. M. & Lane, P. W. *Mouse News Lett.* **17,** 52 (1957).
8. Bahary, N. *et al. Mamm. Genome* **4,** 511–515 (1993).
9. Walther, C. *et al. Genomics* **11,** 424–434 (1991).
10. Riley, J. *et al. Nucleic Acids Res.* **18,** 2887–2890 (1990).
11. Hermanson, G. G. *et al. Nucleic Acids Res.* **19,** 4943–4948 (1991).
12. Sternberg, N. *Trends Genet.* **8,** 11–16 (1992).
13. Hartl, D. L. & Nurminsky, D. I. *BioTechniques* **15,** 201–208 (1993).
14. Dietrich, W. *et al. Genetics* **131,** 423–447 (1992).
15. Church, D. M. *et al. Nature Genet.* **6,** 98–105 (1994).
16. Gish, W. & States, D. J. *Nature Genet.* **3,** 266–272 (1993).
17. Kozak, M. *Cell* **44,** 283–292 (1986).
18. Frohman, M. A., Dush, M. K. & Martin, G. R. *Proc. natn. Acad. Sci. U.S.A.* **85,** 8998–9002 (1988).
19. Padgett, R. A. *et al. A. Rev. Biochem.* **55,** 1119–1150 (1986).
20. Cooper, D. N. & Youssoufian, H. *Hum. Genet.* **78,** 151–155 (1988).
21. Heijne, G. v. *Nucleic Acids Res.* **14,** 4683–4690 (1986).
22. Sabatini, D. D. & Adesnick, M. B. *The Metabolic Basis of Inherited Disease* (eds Scriver, C. R. *et al.*) 177–223 (McGraw-Hill, New York, 1989).
23. Coleman, D. L. *Science* **203,** 663–665 (1979).
24. Bray, G. A. *Int. J. Obesity* **14,** 909–926 (1990).
25. Rubner, M. *Ztschr. Biol.* **30,** 73–142 (1894).
26. Adolph, E. F. *Am. J. Physiol.* **151,** 110–125 (1947).
27. Hervey, G. R. *Nature* **222,** 629–631 (1969).
28. Cannon, W. B. & Washburn, A. L. *Am. J. Physiol.* **29,** 441–454 (1912).
29. Brobeck, J. R. *Physiol. Rev.* **25,** 541–559 (1946).
30. Hetherington, A. W. & Ranson, S. W. *Am. J. Physiol.* **136,** 609–617 (1942).
31. Brobeck, J. R. *Yale J. Biol. Med.* **20,** 545–552 (1948).

32. Kennedy, G. C. *Proc. R. Soc.* B **140,** 578–592 (1953).
33. Mayer, J. *Ann. N.Y. Acad. Sci.* **63,** 15–43 (1955).
34. Hervey, G. R. *J. Physiol.* **145,** 336–352 (1959).
35. Bray, G. A. *J. Nutrition* **121,** 1146–1162 (1991).
36. Bahary, N. *et al. Proc. natn. Acad. Sci. U.S.A.* **87,** 8642–8646 (1990).
37. Neel, J. V. *Am. J. hum. Genet.* **14,** 353–362 (1962).
38. Larin, Z., Monaco, A. P. & Lehrach, H. *Proc. natn. Acad. Sci. U.S.A.* **88,** 4123–4127 (1991).
39. Green, E. D. & Olson, M. V. *Proc. natn. Acad. Sci. U.S.A.* **87,** 1213–1217 (1990).
40. Kusumi, K. *et al. Mamm. Genome* **4,** 391–392 (1993).
41. Vidal, S. M. *et al. Cell* **73,** 469–485 (1993).
42. Friedman, J. M. *et al. Genomics* **11,** 1054–1062 (1991).
43. Chirgwin, J. M. *et al. Biochemistry* **18,** 5294–5299 (1979).
44. Friedman, J. M., Cheng, E. & Darnell, J. E. *J. molec Biol.* **179,** 37–53 (1984).
45. Wang, A. M., Doyle, M. V. & Mark, D. F. *Proc. natn. Acad. Sci. U.S.A.* **86,** 9717–9721 (1989).
46. Benton, W. D. & Davis, R. W. *Science* **195,** 180–182 (1977).
47. Burmeister, M. & Lehrach, H. *Trends. Genet.* **5,** 41 (1989).
48. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74,** 5463–5467 (1977).
49. Venter, C. J. *Automated DNA Sequencing and Analysis* (Academic, New York, 1993).
50. Yu, Y. *et al. Proc. natn. Acad. Sci. U.S.A.* **86,** 9931–9935 (1989).

ACKNOWLEDGEMENTS. We thank D. Coleman for prior genetic and physiological analysis; R. Leibel and N. Bahary for their important contributions to the early phases of this work; D. Koos and R. Cox for help in isolating the first set of YAC clones; A. Buckler, J. Rutter and C. Stotler for instruction in exon trapping; Y. Yu and B. Lauring for advice on *in vitro* translation and for canine pancreatic microsomal membranes; P. Gruss for providing the *Pax4* probe before publication and communicating its map position; J. E. Darnell, N. Heintz, R. Kucherlapati, S. Burley, J. Froude, D. Luck, C. Winestock and L. Safani for comments; S. Korres for preparing the manuscript; J. Sholtis for photography; F. Ilchert for help in preparing the figures; A. Popowicz, P. Dash and the staff of computing services for assistance; D. Sabatini for expert advice on the characterization of the *ob* polypeptide; and J. E. Darnell Jr, who provided the environment to initiate this work. J.M.F. is an investigator of the Howard Hughes Medical Institute. This work was supported by a grant from NIH/NIDDK.