*Proc. Natl. Acad. Sci. USA*
Vol. 85, pp. 8963–8967, December 1988
Biochemistry

# Multiple genes encode nuclear factor 1-like proteins that bind to the promoter for 3-hydroxy-3-methylglutaryl-coenzyme A reductase

**(transcription factors/DNA-binding proteins/regulation of cholesterol metabolism)**

GREGORIO GIL*†, JEFFREY R. SMITH*, JOSEPH L. GOLDSTEIN*, CLIVE A. SLAUGHTER‡, KIM ORTH‡, MICHAEL S. BROWN*, AND TIMOTHY F. OSBORNE*

*Departments of Molecular Genetics and Internal Medicine and ‡Howard Hughes Medical Institute, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX 75235

*Contributed by Michael S. Brown, August 31, 1988*

**ABSTRACT** **DNA-binding proteins of the nuclear factor 1 (NF1) family recognize sequences containing TGG. Two of these proteins, termed reductase promoter factor (RPF) proteins A and B, bind to the promoter for hamster 3-hydroxy-3-methylglutaryl-coenzyme A reductase, a negatively regulated enzyme in cholesterol biosynthesis. In the current study, we determined the sequences of peptides derived from hamster RPF proteins A and B and used this information to isolate a cDNA, designated pNF1/Red1, that encodes RPF protein B. The peptide sequence of RPF protein A, the other reductase-related protein, suggests that it is the hamster equivalent of NF1/L, which was previously cloned from rat liver. We also isolated a hamster cDNA for an additional member of the NF1 family, designated NF1/X. Thus, the hamster genome contains at least three genes for NF1-like proteins. It is likely to contain a fourth gene, corresponding to NF1/CTF, which was previously cloned from the human. The $NH_2$-terminal sequences of all four NF1-like proteins (NF1/Red1, NF1/L, NF1/X, and NF1/CTF), which are virtually identical, contain the DNA-binding domain that recognizes TGG. Functional diversity may arise from differences in the COOH-terminal sequences. We hypothesize that the COOH-terminal domain interacts with adjacent DNA-binding proteins, thereby stabilizing the binding of a particular NF1-like protein to a particular promoter. This protein–protein interaction confers specificity to a class of proteins whose DNA-recognition sequence is widespread in the genome. Sterols may repress transcription of the reductase gene by disrupting this protein–protein interaction.**

Among the factors that regulate transcription of eukaryotic genes is a family of proteins known as nuclear factor 1 (NF1), CCAAT-box-binding transcription factor (CTF), or CCAAT-box-binding protein (1–4). These proteins are related in that they all bind to double-stranded sequences containing TGG and its complement, CCA, making contact with the adjacent guanosines (4). NF1-like proteins bind to the promoters of many genes, including the gene for 3-hydroxy-3-methylglutaryl-coenzyme A reductase (HMG-CoA reductase), the rate-determining enzyme in cholesterol synthesis (5).

NF1 binds to a TGG-containing sequence in adenovirus and thereby stimulates its replication (6, 7). The same protein (designated NF1/CTF) binds to the CCAAT box, making contact with TGG on the opposite strand and activating transcription of the human $\alpha$-globin gene (1). NF1/CTF binds with highest affinity to sequences that contain the inverted repeat $TGGN_7CCA$. It also binds to "half-sites" that contain only one copy of TGG. Purified NF1/CTF consists of a family of proteins in the molecular weight range of 55,000–62,000. Santoro *et al.* (3) isolated three cDNAs for NF1/CTF proteins from human HeLa cells. All of the mRNAs were derived from a single gene by alternative splicing and differed by the presence or absence of an insertion in the middle of the protein and the presence of various COOH termini. All three proteins, when produced in *Escherichia coli*, bound to the NF1/CTF recognition sequence, and at least two of them stimulated replication of adenovirus *in vitro*.

Another member of the NF1/CTF family is the TGGCA-binding protein, which was described in chicken liver (8) and believed to have a molecular weight in the 30,000 range. The cDNA for this protein, designated pNF1/L, was isolated from rat liver (9). It encodes a protein of 505 amino acids, which was proteolyzed during its isolation. The first 175 amino acids of the rat TGGCA-binding protein show 98% identity with the $NH_2$ terminus of human NF1/CTF, and the remaining portion shows an identity of $\approx$50% (3, 9).

The 5′ flanking region of the reductase gene contains the information necessary for transcription of the gene as well as for feedback repression by sterols. This region contains eight sequences that bind nuclear proteins as revealed by DNase I protection assays (10). We have purified (5) a protein doublet of 33 and 35 kDa, designated reductase promoter factor (RPF) that produces six of the eight footprints. The two proteins were active as monomers, and both recognized all six footprinted sequences. The only sequence shared by all six footprinted regions is the trinucleotide TGG. Methylation-interference analysis showed that both of the adjacent guanosines in this sequence made contact with RPF (5).

The HMG-CoA reductase sequence that binds RPF with highest affinity (designated footprint 2B) contains the inverted repeat $TGGN_7CCA$ (5). The other five RPF-binding sites contain only half-sites. We speculated that RPF belonged to the NF1 family, and we demonstrated by gel retardation assays that RPF bound with high affinity to an oligonucleotide corresponding to the NF1-binding site in the adenovirus origin of replication (5). The question remains as to whether NF1/CTF and RPF are products of the same gene with differences in behavior attributable to alternative splicing, or whether multiple genes for this protein family exist.

In the current study, we obtained partial amino acid sequences for the two RPF proteins and used this information to isolate cDNAs§ for two related but genetically distinct NF1-like proteins. The evidence suggests that the hamster

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: HMG-CoA, 3-hydroxy-3-methylglutaryl-coenzyme A; NF1, nuclear factor 1; PCR, polymerase chain reaction; RPF, reductase promoter factor; CTF, CCAAT-box-binding transcription factor.

†Present address: Department of Biochemistry, University of Massachusetts Medical School, Worcester, MA 01605.

§The sequences reported in this paper are being deposited in the EMBL/GenBank data base (IntelliGenetics, Mountain View, CA, and Eur. Mol. Biol. Lab., Heidelberg) (accession nos. J04122 and J04123).

genome contains at least three genes of the NF1/CTF family and that at least two of these proteins bind to the HMG-CoA reductase promoter.

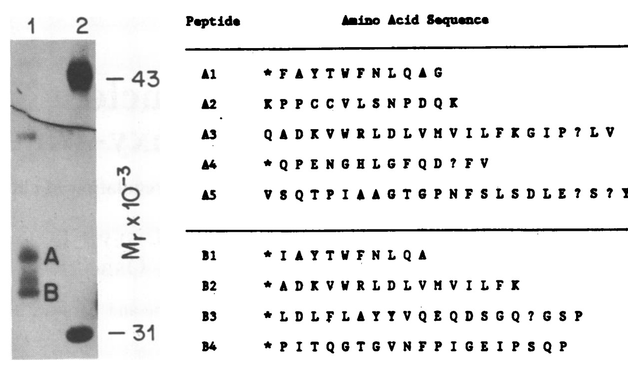
## MATERIALS AND METHODS

**General Methods.** Standard molecular biology techniques were used (11). cDNA clones were sequenced by the dideoxy chain-termination method (12) with either the M13 universal sequencing primer or specific oligonucleotides after subcloning into bacteriophage M13 vectors. Sequencing reactions were performed with the Klenow fragment of *E. coli* DNA polymerase I (12) or a modified bacteriophage T7 DNA polymerase (13). Total cellular RNA was isolated by guanidinium thiocyanate extraction followed by centrifugation through a cesium chloride cushion (14). Blot hybridization of RNA was performed as described (15) with single-stranded $^{32}$P-labeled probes (16).

**Amino Acid Sequence of RPF Peptides.** Approximately 300 pmol of purified RPF (5) was subjected to $NaDodSO_4$/polyacrylamide gel electrophoresis and transferred to nitrocellulose sheets for solid-phase tryptic digestion (17). Peptides were separated by reverse-phase HPLC on a Brownlee RP300 (2.1 × 1000 mm) $C_8$ column in 0.1% trifluoroacetic acid with a gradient of 0–50% (vol/vol) acetonitrile for 120 min at 50 $\mu$l/min. Peaks were collected manually on 1-cm Whatman GF/C discs. Cysteines were reduced and alkylated (18). Peptides were sequenced on an Applied Biosystems (Foster City, CA) model 470A sequencer (Fig. 1).

**cDNA Library.** A double-stranded cDNA library from Syrian hamster liver was synthesized from 10 $\mu$g of poly(A)$^+$ RNA (19) by using a kit from Bethesda Research Laboratories. After second-strand synthesis, the cDNAs were methylated with *Eco*RI methylase and *S*-adenosylmethionine and ligated to *Eco*RI linkers [8 base pairs (bp)]. After *Eco*RI cleavage, excess linkers were removed on a Sephadex G-100 column. An aliquot (150 ng) of the cDNA was ligated with 2 $\mu$g of *Eco*RI-cleaved λgt10 DNA. One-fourth of the mixture was packaged *in vitro* by using a λ DNA packaging extract from Stratagene. Phage were plated on host strain *E. coli* C600. Approximately 95% of the plaques were clear, and $3.5 \times 10^5$ plaques were screened. Replica filters were hybridized in 25% (vol/vol) formamide (11) containing $^{32}$P-end-labeled oligonucleotide probe at $5 \times 10^6$ cpm/ml (see below). Filters were washed in 4× SSC (1× SSC = 150 mM NaCl/15 mM sodium citrate, pH 7) plus 0.2% $NaDodSO_4$. After several rounds of plaque purification, 7 positive clones remained. Plate-lysate DNA from each clone was subcloned into either a plasmid vector (pGem3) or a M13 vector (M13mp18 and -19). Four of the 7 clones were sequenced in the region that hybridized to the probe, and 3 clones were authentic.

**Polymerase Chain Reaction (PCR).** A 0.5-$\mu$g aliquot of the hamster liver cDNA library was used as a template for PCR (20, 21). Oligonucleotide primers were based on the $NH_2$- and COOH-terminal sequences of peptide A3, which was identical to peptide B2 (Fig. 1). The two primers included all degenerate codons as shown in Fig. 2. Seventy picomoles of each primer were used. One primer was end-labeled with [$\gamma$-$^{32}$P]ATP and T4 kinase. The amplification reaction mixture was boiled, after which the PCR was carried out sequentially for 1.5 min at 94°C, 2.5 min at 40°C, and 5 min at 50°C with *Taq* I polymerase. After 35 cycles, the products were loaded onto a 7% polyacrylamide gel. A major product of 59 bp was localized by autoradiography, cut from the gel, eluted (11), and subjected to Maxam–Gilbert sequencing (22). Translation of the nucleotides between the primers gave the expected amino acid sequence, and a 31-mer oligonucleotide corresponding to the sequence between the primers was synthesized and used to screen the cDNA library (Fig. 2).




| Peptide | Amino Acid Sequence |
|---|---|
| A1 | * F A Y T W F N L Q A G |
| A2 | K P P C C V L S N P D Q K |
| A3 | Q A D K V W R L D L V M V I L F K G I P ? L V |
| A4 | * Q P E N G H L G F Q D ? F V |
| A5 | V S Q T P I A A G T G P N F S L S D L E ? S ? Y |
| B1 | * I A Y T W F N L Q A |
| B2 | * A D K V W R L D L V M V I L F K |
| B3 | * L D L F L A Y Y V Q E Q D S G Q ? G S P |
| B4 | * P I T Q G T G V N F P I G E I P S Q P |

FIG. 1. Sequence of tryptic peptides from RPF. (*Left*) RPF was separated into two proteins (A and B) by electrophoresis on a 35-cm 7% polyacrylamide gel containing $NaDodSO_4$ (lane 1). This gel gave wide separation of two marker proteins, ovalbumin (43,000) and carbonic anhydrase (31,000) (lane 2). Proteins A and B were transferred to nitrocellulose, stained, cut from the filters, and separately digested with trypsin. The peptides were separated by HPLC and sequenced. (*Right*) Peptides A1–A5 and B1–B4 were derived from proteins A and B, respectively. A single sequencer run was performed on each peptide. An asterisk denotes that more than one amino acid was obtained at the indicated position. Amino acid signals ranged from 0.5 to 63 pmol.

## RESULTS

RPF was isolated as a protein doublet from hamster liver nuclei as described (5). The two major components, A and B, were separated on 35-cm $NaDodSO_4$/polyacrylamide gels (Fig. 1) and digested with trypsin. The peptide profiles from the two proteins were different after fractionation by HPLC. Amino acid sequences from several peptides in each protein were determined (Fig. 1). One peptide was identical in both proteins (peptides A3 and B2). Other peptides were similar, yet sufficiently different to suggest derivation from different genes.

To derive an unambiguous probe that would hybridize to cDNAs from both genes, we used the PCR. We prepared a cDNA library from Syrian hamster liver and incubated it with degenerate oligonucleotides predicted to hybridize to the upper and lower strands of the DNA encoding the $NH_2$- and COOH-terminal ends of peptide A3/B2 (Fig. 2). The products of the PCR were subjected to electrophoresis, and the predicted 59-bp fragment was isolated and sequenced. The DNA sequence of the codons corresponded to the sequence of the tryptic peptide in the region between the two primers (Fig. 2). This sequence was used for cDNA cloning.

A λgt10 cDNA library prepared from the liver yielded three positive clones that were subsequently subcloned as plasmids. One corresponded to the lower molecular weight protein of RPF (protein B); it is designated pNF1/Red1. Two other cDNAs had coding regions that were identical to each other, designated pNF1/X. It encoded a protein with peptides that were similar but not identical to the peptides in proteins A and B. pNF1/Red1 hybridized to a 8.7-kilobase mRNA in CHO cells (Fig. 3, lane 1) and Syrian hamster liver (lane 2). pNF1/X hybridized to a smaller 6.1-kilobase mRNA in the same tissues (lanes 3 and 4).

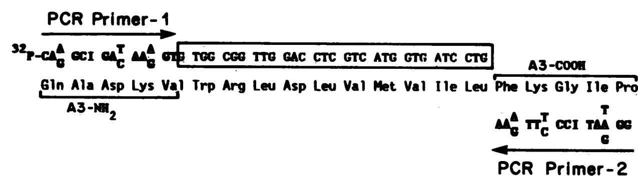



FIG. 2. Generation of a cDNA probe from the amino acid sequence of peptide A3. The boxed nucleotide sequence was generated by PCR and used as a hybridization probe for cDNA cloning.

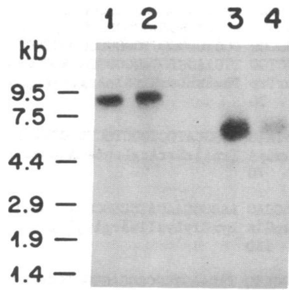



FIG. 3. Blot hybridization of RNA. Total RNA (20 $\mu$g) from CHO cells (lanes 1 and 3) and hamster liver (lanes 2 and 4) was denatured with glyoxal and fractionated on a 1.5% agarose gel, transferred to a Nylon membrane, and hybridized with probes corresponding to nucleotides 39–763 of pNF1/Red1 (lanes 1 and 2) or nucleotides 98–993 of pNF1/X (lanes 3 and 4). The filters were exposed to x-ray film for 16 hr at room temperature.

The cDNA sequence of pNF1/Red1 contained an open reading frame encoding 561 amino acids that predicts a protein of $M_r$ 62,723 (Fig. 4). The putative initiator methionine is preceded by a terminator codon located 12 bp upstream. The predicted 5′ untranslated region is unusually long. Primer extension and S1 nuclease mapping indicated that the cloned cDNA was colinear with the mRNA and that the 5′ untranslated region is about 820 nucleotides (data not shown). The amino acid sequence included all four of the peptides that were found in protein B of RPF.

The sequence of pNF1/X contained an open reading frame encoding 441 amino acids with a predicted molecular weight of 48,842 (Fig. 5). Preceding the putative initiator methionine was an extremely G+C-rich 5′ untranslated region (86% G+C). The first upstream terminator codon occurred 168 nucleotides upstream. S1 nuclease mapping showed that the mRNA was colinear with the cDNA and predicts a 5′ untranslated region of at least 337 nucleotides. The predicted amino acid sequence encoded by pNF1/X included sequences that corresponded to peptides A3/B2 (residues 123–138 in Fig. 5) and A2 (residues 99–111). There was no identical match with any other peptides obtained from RPF proteins A and B.

## DISCUSSION

The sequences of all NF1-like proteins studied to date are compared in Fig. 6. The hamster genome contains at least three NF1-like genes. RPF proteins A and B are derived from

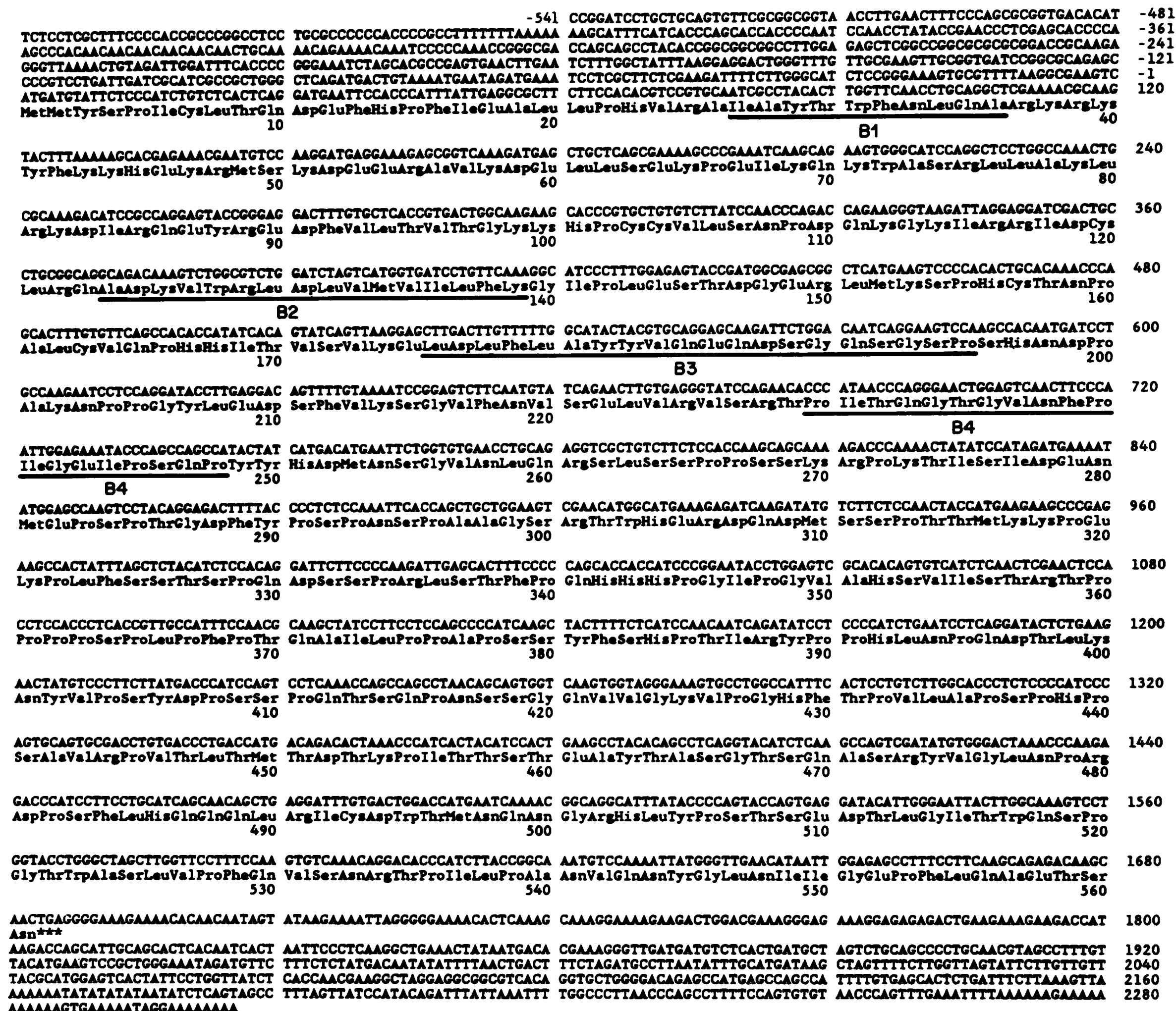


FIG. 4. Nucleotide sequence of cDNA for pNF1/Red1 and predicted amino acid sequence of the protein. Nucleotides are numbered on the right; nucleotide 1 is the adenosine of ATG that encodes the putative initiator methionine; negative numbers refer to the 5′ untranslated region. Amino acids are numbered underneath the sequence; residue 1 is the putative initiator methionine. Amino acid sequences corresponding to peptides B1–B4 are underlined. The sequence in the coding and 5′ untranslated regions and >95% of the 3′ untranslated region was determined on both strands. Although the library was constructed by oligo[d(T)] priming, the lack of a poly(A) tail and consensus poly(A) addition signal suggests that this cDNA was generated by internal priming.

```
             -337 CCTTTCA CAGCGCGGCGCGGCCGGCGGCGGCGGC GGCGGTGGTGGCGGGCGAGGGCGACCGGCC GAGCGGCGGCGGCATGGAGTAGACGCGCGG  -241
CGGCGGCGGCGGCGGACGCCGAGAGGCAGC GGCGAGCGCGGCGGCGGCCCGGAGCCCGCG GGGCCGAGCCTGTGAGCGCGGAGCGGCGGT GGGCCGAGCGGGGCCGCGGGCCGGGCGGGC  -121
GCCGCGCGGCGGAGGCCGGAGCAGCGGAGC CGGAGCCCGAGTGCCAGCCAGCGCGACGCC CGCCTGCCGGGCCTGTCCTCGCCGCGGCCG GCCTCCGCGCTCGCGCCCGGGCGCCCAGCG  -1
ATGTACTCCCCGTACTGCCTCACCCAGGAT GAGTTCCACCCGTTCATCGAGGCGCTGCTG CCTCACGTCCGAGCCTTCTCCTACACCTGG TTCAACCTGCAGGCGCGGAAGCGCAAGTAC  120
MetTyrSerProTyrCysLeuThrGlnAsp GluPheHisProPheIleGluAlaLeuLeu ProHisValArgAlaPheSerTyrThrTrp PheAsnLeuGlnAlaArgLysArgLysTyr
                            10                             20                             30                             40

TTCAAGAAGCACGAGAAGCGGATGTCAAAG GATGAGGAGCGCGCGGTGAAGGACGAGCTG CTGGGCGAGAAGCCTGAGATCAAGCAGAAG TGGGCATCCCGGCTGCTGGCCAAGCTGCGC  240
PheLysLysHisGluLysArgMetSerLys AspGluGluArgAlaValLysAspGluLeu LeuGlyGluLysProGluIleLysGlnLys TrpAlaSerArgLeuLeuAlaLysLeuArg
                            50                             60                             70                             80

AAAGACATCCGGCCTGAGTTCCGAGAGGAC TTCGTGCTGACCATCACCGGCAAGAAGCCC CCCTGCTGCGTGCTCTCCAACCCCGACCAG AAGGGCAAGATCCGGCGGATTGACTGCCTG  360
LysAspIleArgProGluPheArgGluAsp PheValLeuThrIleThrGlyLysLysPro ProCysCysValLeuSerAsnProAspGln LysGlyLysIleArgArgIleAspCysLeu
                            90                            100                            110                            120

CGCCAGGCTGACAAGGTGTGGCGGCTGGAC CTGGTCATGGTGATTTTGTTTAAGGGGATC CCCTTGGAAAGTACTGACGGGGAGCGGCTC TACAAGTCGCCCCAGTGCTCGAACCCCGGC  480
ArgGlnAlaAspLysValTrpArgLeuAsp LeuValMetValIleLeuPheLysGlyIle ProLeuGluSerThrAspGlyGluArgLeu TyrLysSerProGlnCysSerAsnProGly
                           130                            140                            150                            160

CTGTGCGTCCAGCCACATCACATTGGAGTC ACAATCAAAGAACTGGACCTTTATCTGGCT TACTTTGTCCACACTCCGGAATCCGGACAA TCAGATAGTTCAAACCAGCAAGGAGATGCG  600
LeuCysValGlnProHisHisIleGlyVal ThrIleLysGluLeuAspLeuTyrLeuAla TyrPheValHisThrProGluSerGlyGln SerAspSerSerAsnGlnGlnGlyAspAla
                           170                            180                            190                            200

GACATCAAACCACTGCCCAATGGACACTTA AGTTTCCAGGACTGCTTTGTGACTTCCGGG GTCTGGAATGTGACAGAGCTGGTGAGAGTA TCACAGACTCCTGTTGCAACTGCATCAGGG  720
AspIleLysProLeuProAsnGlyHisLeu SerPheGlnAspCysPheValThrSerGly ValTrpAsnValThrGluLeuValArgVal SerGlnThrProValAlaThrAlaSerGly
                           210                            220                            230                            240

CCCAACTTCTCGCTGGCAGACCTGGAAAGC CCCAGTTACTACAACATAAACCAAGTGACC CTGGGACGGCGGTCCATCACCTCCCCTCCT TCCACCAGCACCACCAAGCGCCCCAAGTCC  840
ProAsnPheSerLeuAlaAspLeuGluSer ProSerTyrTyrAsnIleAsnGlnValThr LeuGlyArgArgSerIleThrSerProPro SerThrSerThrThrLysArgProLysSer
                           250                            260                            270                            280

ATCGATGACAGCGAGATGGAGAGCCCAGTT GATGACGTGTTCTATCCTGGGACAGGCCGC TCCCCAGCAGCTGGCAGCAGCCAGTCGAGT GGATGGCCCAATGACGTGGATGCAGGCCCT  960
IleAspAspSerGluMetGluSerProVal AspAspValPheTyrProGlyThrGlyArg SerProAlaAlaGlySerSerGlnSerSer GlyTrpProAsnAspValAspAlaGlyPro
                           290                            300                            310                            320

GCTTCTCTAAAGAAGTCAGGAAAACTGGAC TTCTGCAGCGCCCTCTCCTCTCAGGGCAGC TCCCCACGCATGGCTTTCACCCACCACCCG CTGCCTGTGCTTGCTGGAGTCAGACCAGGG  1080
AlaSerLeuLysLysSerGlyLysLeuAsp PheCysSerAlaLeuSerSerGlnGlySer SerProArgMetAlaPheThrHisHisPro LeuProValLeuAlaGlyValArgProGly
                           330                            340                            350                            360

AGCCCCCGGGCCACAGCATCCGCCCTGCAC TTCCCCTCCACGTCCATCATTCAGCAGTCC AGCCCGTACTTTACACACCCGACCATCCGC TACCACCACCACCACGGGCAGGACTCCCTG  1200
SerProArgAlaThrAlaSerAlaLeuHis PheProSerThrSerIleIleGlnGlnSer SerProTyrPheThrHisProThrIleArg TyrHisHisHisHisGlyGlnAspSerLeu
                           370                            380                            390                            400

AAGGAGTTTGTGCAGTTTGTGTGCTCTGAC GGCTCGGGTCAGGCCACTGGACAGCATTCA CAACGACAGGCACCTCCTCTGCCAGCCGGT TTGTCGGCATCGGACCCCGGGACGGCAACT  1320
LysGluPheValGlnPheValCysSerAsp GlySerGlyGlnAlaThrGlyGlnHisSer GlnArgGlnAlaProProLeuProAlaGly LeuSerAlaSerAspProGlyThrAlaThr
                           410                            420                            430                            440
TTCTGAACATCCCACAGCAGTCTCAGTCCT GGTTCCTCTGATTAAGATCAACAAAGAAAC AACAAAATATAAAAAAAAAAAAGGACTGTT AGTAAAAATATATCAAACTGTTTCTTTGTA  1440
Phe***
AAATTTTATGTATTGAGGATATAAATAAAT ATCTCCTTCAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAA
```

FIG. 5. Nucleotide sequence of cDNA for pNF1/X and predicted amino acid sequence of the protein. Nucleotides and amino acids are numbered as described in Fig. 4. The entire sequence was determined on both strands. A poly(A) tail and consensus poly(A) addition signal (AATAAA) are located at the 3′ end.

two of these genes. Protein B is encoded by NF1/Red1, one of the cDNAs that was isolated in the current study. Protein A appears to be encoded by pNF1/L, which was cloned from the rat (9). The peptides from protein A show only 3 mismatches out of 81 amino acids when compared with rat NF1/L (Fig. 6). These 3 mismatches, which involve amino acids at the extreme COOH terminus of two of the five peptides, may reflect peptide sequencing error or species divergence. The peptides of protein A show a lower level of identity with the other NF1-like proteins (at least 14 mismatches out of 81 residues).

The third hamster gene is the one encoding pNF1/X. This cDNA was obtained by screening a hamster cDNA library with an oligonucleotide corresponding to a protein sequence shared by all known members of the NF1 family. The sequence of pNF1/X did not correspond to either of the hamster RPF proteins and, therefore, pNF1/X must be encoded by a third hamster NF1-like gene. A fourth member of the NF1 family is NF1/CTF, which has been isolated so far only from human cells (3). This gene undergoes alternative splicing to generate three proteins, none of which matches the sequence of the hamster genes or rat NF1/L. It seems likely that the hamster genome contains a fourth gene for the NF1 family which would encode NF1/CTF.

The NF1-like proteins shown in Fig. 6 show a nearly complete identity over the first 170–190 amino acids at the $NH_2$-terminal end, and a lower level of identity in their COOH-terminal portions. All of the peptides isolated from proteins A and B of RPF are found in the $NH_2$-terminal half of the proteins. This suggests that both proteins had been cleaved proteolytically, and only the $NH_2$-terminal half was isolated. Inasmuch as these proteins were isolated by oligonucleotide affinity chromatography, the $NH_2$-terminal half must contain the DNA-binding domain. A similar conclusion was reached by Paonessa *et al.* (9).

The COOH-terminal portion of each of the four NF1-like proteins is enriched in serine, threonine, and proline. They account for 32–40% of the amino acids in this region. Jackson and Tjian (24) showed that many transcription factors, including NF1/CTF, are subjected to O-linked glycosylation. The serine/threonine-rich COOH-terminal half of the NF1-like proteins may be the site of this modification.

Why are there so many different NF1-like proteins? The $NH_2$-terminal regions, which contain the DNA-binding domains, are nearly identical in all of the proteins, implying similar DNA-binding properties. Indeed, all of these proteins recognize TGG. Purified RPF or NF1/CTF does not bind to every TGG-containing sequence, and thus nucleotides outside of this core element must influence DNA-binding specificity, raising the possibility that some of the protein sequence divergence at the $NH_2$ terminus may contribute to differential DNA binding. However, it is unlikely that this is the sole reason for the sequence divergence. A more likely hypothesis, based on the current studies and those of others (3, 9), is that the divergent COOH-terminal regions of the proteins may engage in protein–protein interactions that modify the DNA binding.

A working model to explain this diversity is shown in Fig. 7. This model shows two promoters with TGG sequences that are capable of binding any NF1-like protein. *In vivo*, these two genes bind different members of the NF1 family. Differential binding occurs because gene A has a downstream sequence (designated by the box) that binds to a hypothetical protein, Y, which in turn binds to the COOH terminus of

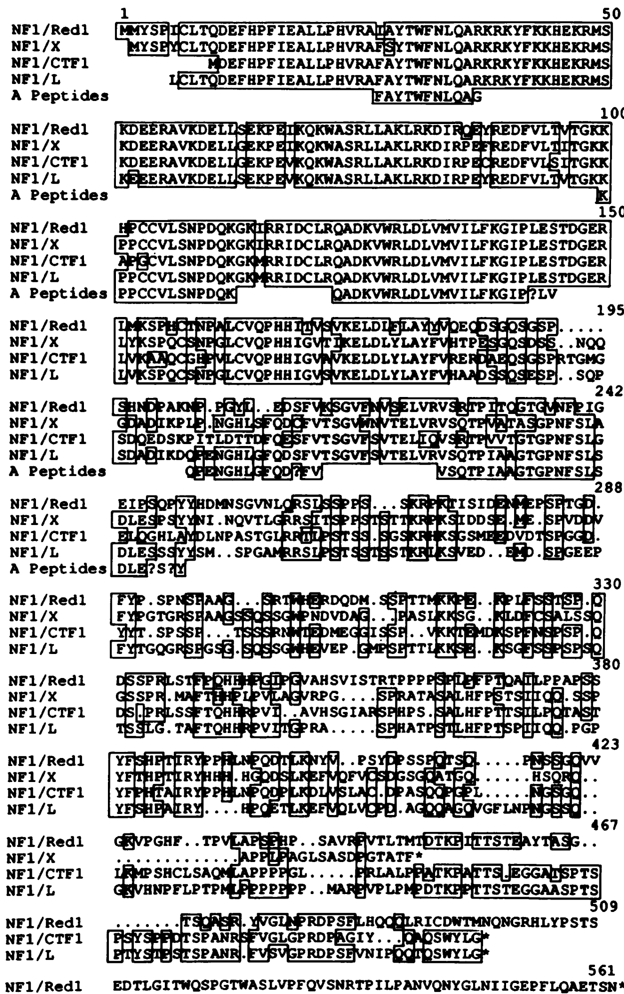


FIG. 6. Comparison of the sequences of four NF1-like proteins. Sequences were aligned with the use of the UWGCG computer program (23). Gaps introduced to maximize identity are represented by dots. Residue numbers refer to NF1/Red1. Residues are boxed when the amino acid is identical in three of the four proteins (or two of the three proteins at the COOH terminus). An asterisk denotes the end of the protein. Amino acid sequences for hamster NF1/Red1 and hamster NF1/X were deduced from the cDNA sequences in Figs. 3 and 4, respectively. Amino acid sequences for human NF1/CTF1 and rat NF1/L were deduced from cDNA sequences in refs. 3 and 9, respectively. The partial sequence of protein A of RPF was derived from the amino acid sequences of tryptic peptides A1–A5 in Fig. 1.

NF1/Red1. This binding stabilizes the protein–DNA complex and assures preferential binding of NF1/Red1 to gene A. Gene B contains a different downstream sequence that binds protein Z, which binds to the COOH terminus of NF1/CTF. Through these protein–protein interactions, a relatively nonspecific TGG sequence, which can bind many NF1-like proteins, is transformed into a highly specific sequence that will stably bind only one of them. If the downstream recognition element is TGG on the opposite strand, then two TGG-binding proteins can interact to form a stable complex. This would explain why the purified RPF or NF1/CTF binds to the rotationally symmetric site $TGGN_7CCA$ *in vitro* with a higher affinity than the TGG half-sites (1, 5). If a TGG sequence occurs in the absence of a neighboring protein-binding sequence, then it will bind NF1-like proteins only with low affinity. This requirement for multicomponent binding would prevent NF1-like proteins from forming stable complexes with TGG sequences throughout the genome.

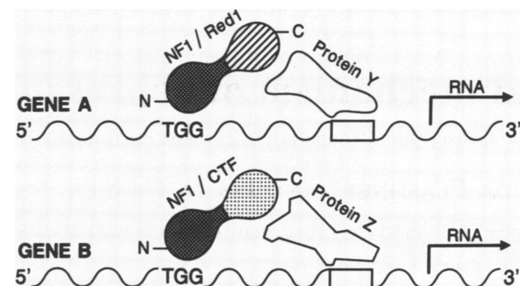



FIG. 7. Model for diversity of transcriptional regulation by NF1-like proteins.

This model would be consistent with the data of Chodosh *et al.* (2) who found that NF1-like proteins bind to DNA with highest affinity when present as heterodimers. The second component of the heterodimer would be analogous to proteins Y and Z in Fig. 7. The model also raises the possibility that the Y protein that binds to the HMG-CoA reductase gene is sterol-sensitive, thus allowing the transcriptional activity of its NF1-like protein to be negatively regulated by sterols.

We thank Drs. Ricardo Cortese and Robert Tjian for communicating results prior to publication. Daphne Norsworthy and Carolyn R. Moomaw provided excellent technical assistance. This work was supported by grants from the National Institutes of Health (HL 20948), the Perot Family Foundation, and the Lucille P. Markey Charitable Trust. T.F.O. is a Lucille P. Markey Scholar. J.R.S. is a M.D./Ph.D. student supported by a Medical Scientist Training Grant from the National Institutes of Health (GM 08014).

1. Jones, K. A., Kadonaga, J. T., Rosenfeld, P. J., Kelly, T. J. & Tjian, R. (1987) *Cell* **48**, 79–89.
2. Chodosh, L. A., Baldwin, A. S., Carthew, R. W. & Sharp, P. A. (1988) *Cell* **53**, 11–24.
3. Santoro, C., Mermod, N., Andrews, P. C. & Tjian, R. (1988) *Nature (London)* **334**, 218–224.
4. de Vries, E., van Driel, W., van den Heuvel, S. J. L. & van der Vliet, P. C. (1987) *EMBO J.* **6**, 161–168.
5. Gil, G., Osborne, T. F., Goldstein, J. L. & Brown, M. S. (1988) *J. Biol. Chem.*, in press.
6. Gronostajski, R. M., Adhya, S., Nagata, K., Guggenheimer, R. A. & Hurwitz, J. (1985) *Mol. Cell. Biol.* **5**, 964–971.
7. Rosenfeld, P. J. & Kelly, T. J. (1986) *J. Biol. Chem.* **261**, 1398–1408.
8. Rupp, R. A. W. & Sippel, A. E. (1987) *Nucleic Acids Res.* **15**, 9707–9726.
9. Paonessa, G., Gounari, F., Frank, R. & Cortese, R. (1988) *EMBO J.* **7**, 3115–3123.
10. Osborne, T. F., Gil, G., Brown, M. S., Kowal, R. C. & Goldstein, J. L. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 3614–3618.
11. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
12. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
13. Tabor, S. & Richardson, C. C. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 4767–4771.
14. Gil, G., Goldstein, J. L., Slaughter, C. A. & Brown, M. S. (1986) *J. Biol. Chem.* **261**, 3710–3716.
15. Gil, G., Brown, M. S. & Goldstein, J. L. (1986) *J. Biol. Chem.* **261**, 3717–3725.
16. Church, G. M. & Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 1991–1995.
17. Aebersold, R. H., Leavitt, J., Saavedra, R. A., Hood, L. E. & Kent, S. B. H. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 6970–6974.
18. Andrews, P. C. & Dixon, J. E. (1987) *Anal. Biochem.* **161**, 524–528.
19. Gubler, U. & Hoffman, B. J. (1983) *Gene* **25**, 263–269.
20. Lee, C. C., Wu, X., Gibbs, R. A., Cook, R. G., Muzny, D. M. & Caskey, C. T. (1988) *Science* **239**, 1288–1291.
21. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. & Erlich, H. A. (1988) *Science* **239**, 487–491.
22. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
23. Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395.
24. Jackson, S. P. & Tjian, R. (1988) *Cell* **55**, 125–133.