LSF and NTF-1 Share a Conserved DNA Recognition Motif yet Require Different Oligomerization States to Form a Stable Protein-DNA Complex

(Received for publication, September 10, 1997, and in revised form, May 12, 1998)

Margaret K. Shirra‡§ and Ulla Hansen‡¶
From the §Department of Microbiology and Molecular Genetics, Harvard Medical School and the ¶Division of Molecular Genetics, Dana-Farber Cancer Institute, Boston, Massachusetts 02115

The mammalian transcription factor LSF (also known as CP2 and LBP-1c) binds as a homo-oligomer to directly repeated elements in viral and cellular promoters. LSF and the Drosophila transcription factor NTF-1 (also known as Elf-1 and Grainyhead) share a similar DNA binding region, which is unlike any established DNA binding motifs. However, we demonstrate that dimeric NTF-1 can bind an LSF half-site, whereas LSF cannot. To characterize further the DNA binding and oligomerization characteristics of LSF, truncation mutants were used to demonstrate that between 234 and 320 amino acids of LSF are required for high affinity DNA binding. Mixing of a truncation mutant with full-length LSF in a DNA binding assay established that the form of LSF that binds DNA is larger than a dimer. Unexpectedly, one C-terminal deletion derivative, partially defective in oligomerization properties, could occupy odd numbers of adjacent, tandem LSF half-sites, unlike full-length LSF. The numbers of DNA-protein complexes formed on multiple half-sites with this mutant indicated that LSF binds DNA as a tetramer, although cross-linking experiments confirmed a previous report concluding that LSF is primarily dimeric in solution. The DNA binding and oligomerization properties of LSF support models depicting novel mechanisms to prevent continual, adjacent binding by a protein that recognizes directly repeated DNA sequences.

One target for modulating transcription in eukaryotic cells is the specific DNA binding activity of transcription factors. DNA binding activity of a protein can be regulated by post-translational modification, such as phosphorylation, or by formation of heteromeric complexes. Mapping DNA binding regions and identifying oligomerization states can establish the basis for defining biological regulatory pathways. Proteins displaying previously unestablished strategies for formation of oligomeric protein-DNA complexes may also elucidate novel regulatory mechanisms.

The known structures of complexes between specific DNA sequences and the DNA binding motifs of several transcription factors have provided general models for how proteins specifically interact with DNA. Many DNA-binding proteins employ an α-helix for specifically contacting bases in DNA (for review, see Ref. 1). An alternate mode of DNA recognition, exemplified by three prokaryotic repressors, is accomplished by a pair of anti-parallel β-sheets (2–4). Finally, NF-xB specifically contacts DNA by a series of peptide loops (5, 6). Therefore, DNA-binding proteins employ many strategies to specifically recognize their binding sites.

Oligomerization of sequence-specific transcription factors often contributes to the stability of protein-DNA interactions and is therefore required to obtain protein-DNA complexes. Dimerization is critical for many transcription factors, often mediated by α-helical structures (7–15) or β-strands (5, 6, 16–18). Tri- merization has only been observed for the heat shock transcription factor (19). Finally, some transcription factors are tetrameric, with a variety of types of interactions forming the interface (2, 3, 20, 21). Alteration of oligomerization surfaces, or of DNA recognition surfaces, can lead to regulation of the activity of a transcription factor in the cell.

Despite the variety of known DNA recognition strategies, the mammalian transcription factor LSF presents novel DNA binding characteristics. LSF DNA binding activity is cell growth-regulated and can be modulated by phosphorylation (22). LSF, also known as CP2 (23), UBP-1 (24), and LBP-1c (25), binds to a pair of directly repeated sequences (25–27) whose intervening spacing is restricted such that LSF binds DNA on a single face of the DNA helix (26). LSF recognizes many viral and cellular promoter sequences as a homo-oligomeric protein: simian virus 40 late promoter (26), murine α-globin promoter (23), human immunodeficiency virus-1 long terminal repeat (24, 25, 28, 29), rat γ-fibrinogen promoter (27), major histocompatibility complex class II Ea and Dra promoters (30), human c-FOS (1), and mouse thymidylate synthase (2). The primary amino acid sequence of LSF is not similar to any established DNA binding or oligomerization motifs, although LSF shows a high degree of similarity to the Drosophila transcription factor NTF-1 (31), also known as Elf-1 (32) or Grainyhead (33), which binds DNA as a dimer (33, 34). Several groups (25, 33, 35, 36), including our own, have presented data that LSF also binds DNA as a dimer. However, during further investigations of the DNA binding and oligomerization properties of LSF and NTF-1, we have shown that the oligomerization states of LSF and NTF-1 differ. Our new data are also inconsistent with the stable DNA-binding moiety of LSF being a dimer. Instead, these data indicate that LSF binds DNA as a tetramer, even though in solution it is primarily a dimer. Because there are no known

* This work was supported in part by grants from the Public Health Service, National Cancer Institute Grant CA38038, and from the Sandoz/DFCI Drug Discovery Program (to U. H.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† Supported in part by Training Grant T32 CA09361 from the National Institutes of Health. Present address: Dept. of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260.

‡ Supported in part by Training Grant T32 CA09361 from the National Institutes of Health. Present address: Dept. of Biology, Boston University, 5 Cummington St., Boston, MA 02215. Tel.: 617-353-8730; Fax: 617-353-8734; E-mail: uhansen@bio.bu.edu.

1 R. Misra, H.-C. Huang, M. E. Greenberg, and U. Hansen, unpublished data.
2 C. M. H. Powell, J. L. Volker, L. F. Johnson, and U. Hansen, submitted for publication.
structures for tetramers binding to direct repeats and because LSF does not resemble known DNA binding motifs, LSF may represent a new class of DNA binding transcription factors.

**Experimental Procedures**

**Plasmid Construction**—Plasmids are derivatives of the expression clone pET-LSF (35), which contains the entire LSF coding sequence in a modified pET-11c (Novagen) vector. Except as described below, constructs coding N-terminal and C-terminal deletion mutants resulted from exonuclease digestion of pET-LSF beginning at appropriate restriction sites. Oligonucleotides were ligated to the 5′ deletions, as necessary, to maintain the proper reading frame for translation. pET-LSF324 was described previously (35). pET-LSFΔ164 was constructed by digesting pET-LSF with NheI and AvaII and religating the compatible cohesive ends of the larger fragment, containing both vector and LSF sequences. pET-LSFΔ566, pET-LSF3397, and pET-LSFΔ396–3963 were obtained by polymerase chain reaction amplification of the relevant LSF sequences and ligation into a pET-LSF vector that had been digested with appropriate enzymes (gifts of Q. Zhu). Sequences were verified by chain-termination sequencing (Sequenase; U. S. Biochemical Corp.). Fig. 2D shows the additional encoded amino acid sequences in some of the resulting LSF derivatives, at either the N- or C-terminals. Complete plasmid DNA sequences and maps are available upon request.

**Oligonucleotides**—The following oligonucleotides were used in the electrophoretic mobility shift assays (EMSA).3 LSF-280 (26), 5′-GATCGACTGTTCTTCCGCTTA 3′ and 3′-GATCCACTATGGCGTC-CTGGACCTAA 5′; 5′-CTAGCCATATGGCGTCCT-CTGGACCTAA 5′ and 3′-GATCCACTATGGCGTCCT-CTGGACCTAA 5′; and 5′-GATCCACTAGTCCTGTT-CTCTTCGTTAGCTG 3′ and 3′-GATCCACTATGGCGTCCT-CTGGACCTAA 5′. The following oligonucleotides represent base changes from the consensus oligonucleotide; UBX (31), 5′-GATCCAAACATCTGGGTTTGGAGCGGTGTTA 3′ and 3′-TTCGGTTAGCCACATCTGGGTTTGGAGCGGTGTTA 5′; 5′-GATCCTAGTGCCTTCCTGGTCT-CTCTTGTTAGCTG 3′ and 3′-GATCCACTATGGCGTCCT-CTGGACCTAA 5′; 5′-GATCCTAGTGCCTTCCTGGTCT-CTCTTGTTAGCTG 3′ and 3′-GATCCACTATGGCGTCCT-CTGGACCTAA 5′. The following oligonucleotides are modified from the HIV-1 promoter—6 to 15, see Ref. 25.) Each pair of oligonucleotides was annealed and labeled with polynucleotide kinase and [γ-32P]ATP for use in EMSA, as described below.

**In Vitro Transcription-Translation Reactions**—Proteins were synthesized in reticulocyte lysate with the coupled transcription-translation system from Promega according to manufacturer’s instructions, using [35S]methionine. This system produces LSF at approximately 0.2 to 0.5 ng/μl. LSF derivatives are produced at similar levels.

**Electrophoretic Mobility Shift Assays**—2 μl of protein from the intron in vitro transcription-reaction reactions was added to a buffer that contained 10 mM Tris-HCl (pH 8.0), 10% glycerol, 2% polyvinyl alcohol, 0.1 mM EDTA, 1 mM KCl, 1 mM diithiothreitol, and 5 μg/ml poly[dI-C][dI-C] in the final reaction volume of 20 μl. The protein was incubated without the labeled DNA for 15 min, when reactions were performed at room temperature, or 30 min, when reactions were performed at 4 °C, to permit nonspecific DNA-binding proteins to be absorbed to the poly[dI-C][dI-C]. 15 fmoles of labeled DNA was added, and the mixture was incubated for an additional 15 min (at room temperature) or 30 min (at 4 °C) prior to electrophoresis through a 5% polyacrylamide gel containing 44.5 mM Tris base, 44.5 mM boric acid, and 1 mM EDTA. Dried gels were analyzed with a Molecular Dynamics model 400E PhosphorimaggeR ImageQuant software (Molecular Dynamics). The following oligonucleotides were added to the poly[dI-C][dI-C]. 15 fmoles of labeled DNA was added, and the mixture was incubated for an additional 15 min (at room temperature) or 30 min (at 4 °C) prior to electrophoresis through a 5% polyacrylamide gel containing 44.5 mM Tris base, 44.5 mM boric acid, and 1 mM EDTA. Dried gels were analyzed with a Molecular Dynamics model 400E PhosphorimaggeR ImageQuant software (Molecular Dynamics). The following oligonucleotides were added to the poly[dI-C][dI-C]. 15 fmoles of labeled DNA was added, and the mixture was incubated for an additional 15 min (at room temperature) or 30 min (at 4 °C) prior to electrophoresis through a 5% polyacrylamide gel containing 44.5 mM Tris base, 44.5 mM boric acid, and 1 mM EDTA. Dried gels were analyzed with a Molecular Dynamics model 400E PhosphorimaggeR ImageQuant software (Molecular Dynamics).

**GST-LSF Oligomerization Assays**—Assays were performed as described (35). Briefly, glutathione- Sepharose beads containing approximately 1 μg of glutathione S-transferase (GST)-LSF were incubated with 2 μl of in vitro translated proteins for 1 h at room temperature. The beads were washed four times in a buffer containing 500 mM NaCl. An equal volume of 2× SDS sample buffer was added, and samples were incubated at 100 °C to release the bound proteins. Bound proteins were then separated by SDS-polyacrylamide gel electrophoresis. Radiolabeled proteins were visualized and quantitated using the PhosphoImager as described above.

**Results**

**Dimeric NTF-1 Stably Interacts with an LSF DNA-binding Half-site**—LSF (CP2; LBP-1c) and its closely related mammalian family member LBP-1a/b (25) are similar to only one other protein in the database, the Drosophila transcription factor NTF-1 (Elf-1; product of the Grainyhead gene). Fig. 2D illustrates the regions of similarity, with the shaded box reflecting the 25% identity in sequence across 427 amino acids (residues 65–502 of LSF and residues 631–1058 of NTF-1) and the solid boxes reflecting short regions of 66–79% identity. The Drosophila protein is dimeric both in solution, as shown by cross-linking analysis, and when bound to DNA, as shown by EMSA of DNA-protein complexes formed from mixtures of differently sized NTF-1 derivatives (33, 34). To determine whether the primary amino acid similarity between LSF and NTF-1 reflected identical modes of binding DNA, we tested whether NTF-1 would bind an LSF site and, vice versa, whether LSF would bind an NTF-1 site. The high affinity LSF sites tested for binding an LSF site and, vice versa, whether LSF would bind an NTF-1 site. The high affinity LSF sites tested for binding an LSF site and, vice versa, whether LSF would bind an NTF-1 site. The high affinity LSF sites tested for binding an LSF site and, vice versa, whether LSF would bind an NTF-1 site. The high affinity LSF sites tested for binding an LSF site and, vice versa, whether LSF would bind an NTF-1 site.

A Large Portion of LSF Is Required for Stable, Specific Interaction with DNA—LSF and NTF-1 are not structurally similar to any known DNA binding motifs. The region of NTF-1 required to bind DNA has previously been mapped to a minimal region from residues 778 to 837, with optimal DNA binding encompassing residues 632 to 948 or greater (33, 34). However, the differences in DNA binding site recognition between LSF and NTF-1, as well as the unusual requirements for DNA recognition by LSF, a pair of stringently spaced direct repeats, prompted us to map separately the regions of LSF necessary for DNA-protein complex formation. To determine the N- and C-terminal boundaries of the DNA binding region, a series of truncation mutants of LSF were translated in vitro in rabbit reticulocyte lysates. The proteins, which were all expressed at equivalent levels (data not shown), were incubated with the LSF-280 site, and protein-DNA complexes were analyzed by
mements is that whereas the minimal DNA binding region of NTF-1 can form both monomeric and dimeric interactions with DNA (33), no LSF truncation derivatives generated faster migrating protein-DNA species. We infer that the interactions between monomeric LSF and DNA are not sufficient to produce a stable protein-DNA complex. Therefore, oligomerization must be critical for LSF-DNA complex formation.

**LSF Binds DNA as a Homotetramer**—From several previous studies, it was concluded that LSF binds DNA as a dimer (25, 33, 35, 36) (these previous reports are analyzed in more detail under “Discussion”). However, the different DNA-binding site specificities of LSF and NTF-1, together with definitive results that NTF-1 binds DNA as a dimer (33, 34), indicated that a more critical examination of the oligomerization state of LSF on DNA was warranted. In particular, because NTF-1 binds a single LSF half-site as a dimer, it seemed doubtful that the similar DNA binding region of LSF would bind two half-sites, the normal LSF recognition sequence, also as a dimer. We first examined the oligomerization state of LSF on DNA by mixing two DNA binding competent LSF derivatives of different sizes and analyzing the electrophoretic mobility of the heteromeric protein-DNA complexes (38). Unfortunately, GST-LSF fusion proteins are heterogeneous in size when isolated from bacteria (data not shown), and mixtures of LSF and GST-LSF form too many DNA-protein complexes to determine accurately the multimeric state of the protein on the DNA (36). Therefore, this experiment necessitated use of LSF deletion mutants. Although little of LSF is dispensable for its DNA binding activity (Fig. 2), this experiment was deemed feasible due to slightly differing mobilities of the complexes of LSFΔ64 and LSF with DNA. If LSF did bind DNA as a dimer, mixing LSF and LSFΔ64, which each form a single, distinct complex with the LSF-280 DNA (Fig. 3A, lanes 1 and 2), should generate only one heterodimeric complex with intermediate mobility in addition to the two homodimers. Instead, at least two heteromers were revealed (lane 3). The limits of resolution on the polyacrylamide gel of the closely migrating heteromeric complexes prevented an unequivocal determination of the oligomerization state. Nonetheless, this result proved that the LSF oligomer that binds DNA is larger than a dimer. We note that homooligomers of LSF and LSFΔ64 were not readily detected in the mixing experiment, which is expected if the DNA binding moiety contains more than two subunits but not if it is dimeric. If LSF bound DNA as a trimer or tetramer and both LSF derivatives bound DNA with the same affinity, each of the homooligomers would be represented by one-eighth or one-sixteenth of the total number of LSF-DNA complexes, respectively. In contrast, for a dimer, one-fourth of the protein-DNA complexes would be represented by each homodimer. This is certainly not the case for LSF.

Additional data to delineate the oligomerization state of LSF on DNA derived from unexpected DNA binding properties of one of the DNA binding competent LSF deletion mutants employed in the mapping studies, LSF448ΔA (Fig. 2). In studies to determine whether LSF bound in a head-to-tail fashion to the directly repeated half-site recognition sequences, we compared the binding of LSF to DNAs containing two, three, or four adjacent, tandem copies of the LSF half-site. Incubation of LSF with LSF-280 DNA (two half-sites, with a similar overall binding affinity as the two-half-site LSF consensus DNA, see Fig. 1) and 3× DNA (containing three half-sites) produced complexes of similar mobilities (Fig. 3B, lanes 1 and 2; the complex to the longer 3× site migrating slightly slower). Incubation with 4× DNA (containing four tandem half-sites) produced two complexes (lanes 3, asterisks). The faster migrating of the two LSF-4× DNA complexes essentially comigrated with the
LSF-3DNA complex. Similar complexes were obtained upon binding of LSFΔ24 or LSFΔ64 to this set of DNAs (data not shown). These data indicate that LSF binds a complete recognition site but cannot efficiently add by contiguous half-site units in a head-to-tail fashion along the DNA.

Unexpectedly, LSF448Δ bound in a distinct and unusual manner to the three- and four-half-site DNAs. Although LSF448Δ bound the two-half-site DNA as a single complex (compare lanes 1 and 7), it was unlike LSF in its capacity to generate two additional, detectable complexes of slower mobility on the 3DNA (lane 8, see bullets). This indicates that LSF448Δ, which is partially defective in LSF-LSF oligomerization (see Fig. 5), can bind to DNA in a tandem fashion. On 4DNA, LSF448Δ formed five distinguishable complexes (Fig. 3B, lane 9, bullets), three of which were efficiently produced. Even the remaining two complexes were discrete, however, and could

**FIG. 2.** A large segment of LSF is required to bind DNA. The indicated N-terminal and C-terminal deletion mutants of LSF, translated in vitro, were tested for binding to LSF-280 DNA by EMSA. Representative experiments are shown here. Results with each mutant were reproducibly obtained in 2–10 independent experiments. A, DNA binding of N-terminal deletion mutants of LSF at room temperature. Only derivatives LSFΔ24 and LSFΔ64 retained the ability to bind DNA (lanes 2 and 3, respectively). One mutant of LSF lacking regions of both N and C termini was also analyzed (lane 10). Rabbit reticulocyte lysate that was programmed with vector DNA alone (unprog.) contained a low level of endogenous LSF (lane 11). FP designates the migration of free DNA; NS designates the nonspecific protein-DNA complex formed by proteins in the rabbit reticulocyte lysate; and the brackets specify the region of migration of specific complexes between DNA and LSF or LSF derivatives. The nomenclature of the LSF derivatives, in this and all subsequent figures, is as in D, except that the “LSF” prefix is omitted. B, DNA binding of C-terminal deletion mutants of LSF at room temperature. Only LSF448Δ retained DNA binding activity at room temperature. The bands are marked as in A. C, DNA binding of C-terminal deletion mutants of LSF at 4 °C. In contrast to B, both the incubation with DNA and the electrophoresis were performed at 4 °C. Under these conditions, LSF383Δ bound DNA. D, summary of protein-DNA and protein-protein interactions by LSF derivatives. LSF is diagrammed at the top. The regions that are similar to NTF-1 (stippled regions) or highly similar to NTF-1 (solid boxes) are indicated. LSF derivatives are diagrammed below. Derivatives are named using a Δ to indicate the position of the deletion and numbers to specify the first (N-terminal deletions) or last (C-terminal deletions) amino acid of LSF present in the derivative. Lettering before or after the solid bars designates the one-letter amino acid identifications for residues whose coding sequence was added during the cloning process. In addition, the in vitro translated LSF and all C-terminal deletions contain an eight-amino acid leader, MASRGGSG, before the first methionine of native LSF. Relative activity is indicated by “+” or “–” signs with “–” symbolizing undetectable activity. The asterisk indicates the ability to bind DNA at 4 °C but not at room temperature. Oligomerization, as detected by chemical cross-linking analysis, is indicated in the column labeled XLK. Oligomerization, as detected by the GST-LSF oligomerization assay, is indicated in the column labeled GST. The bullet (●) indicates the ability to be retained on a GST-LSF resin at 100 mM but not at 500 mM NaCl.
LSF Binds to DNA as a Tetramer

**FIG. 3.** LSF stably binds as a tetramer to its recognition site, consisting of two directly repeated half-sites. All EMSAs were performed using in vitro translated proteins. Only the relevant portions of the gels are shown. Only 5–10% of the DNA was bound by protein in all reactions. A, the form of LSF binding LSF-280 DNA is larger than a dimer. 2 μl of LSF (lane 1), 2 μl of LSFΔ4 (lane 2), or 1 μl of LSF plus 1 μl of LSFΔ4 (lane 3) were incubated with LSF-280 DNA and analyzed by nondenaturing electrophoresis at room temperature. The arrows indicate the positions of migration of protein-DNA complexes formed by LSF or LSFΔ4 homo-oligomers. B, a derivative of LSF partially defective in oligomerization binds as sequential monomers to directly repeated LSF half-sites. 2 μl of LSF (lanes 1–3), unprogrammed reticulocyte lysate (lanes 4–6), or LSF448Δ (lanes 7–9) were incubated with the indicated DNAs at room temperature. The positions of migration of the two LSF-DNA complexes (tetramers or octamers; asterisks) and of the five distinct LSF448Δ-DNA complexes (tetramers through octamers; bullets) are indicated as diagrammed. Reflective of the lower affinity binding of LSF448Δ to DNA, as compared with that of LSF (Fig. 2B), lanes 4 and 7 represent a 25-fold darker exposure than lane 1, lanes 5 and 8 represent a 27-fold darker exposure than lane 2, and lanes 6 and 9 represent a 8-fold darker exposure than lane 3.

therefore not be due to dissociation of slower complexes during electrophoresis, as this would generate smears but not specific bands. The largest and smallest LSF448Δ complexes migrated just slightly faster than the two similar complexes containing LSF (lane 3, asterisks), due to the smaller size of LSF448Δ. All of the indicated complexes between LSF448Δ and 3× or 4× DNA were specifically apparent only in the presence of the mutant LSF protein, as shown by control reactions containing unprogrammed reticulocyte lysate, exposed at the same level of sensitivity of the autoradiogram. The faint bands in this region of the gel (lanes 4–6) that comigrated with LSF-DNA complexes are attributable to the endogenous LSF in the lysate. Additionally, the protein-DNA complexes specific to LSF448Δ experiments could not be attributed to contaminating protein species, because the in vitro translated LSF448Δ was homogeneous, migrating as a single, distinct protein upon SDS-polyacrylamide gel electrophoresis (data not shown). Therefore, the three new LSF448Δ-4× DNA complexes with intermediate mobilities can only be interpreted to represent the addition of one, two, or three monomeric units onto the normal LSF448Δ oligomer associated with two half-sites. Only the binding of a tetramer to a normal LSF-binding site (or a multiple of tetramers) would be consistent with the electrophoretic mobilities of these complexes (see schematic, Fig. 3B). By analogy, the LSF-4× DNA species are tetrameric and octameric complexes, as well. Significantly, like LSF, LSF448Δ did not form any protein-DNA complexes that migrated significantly faster than the complex of LSF with LSF-280 (Fig. 3B, lane 9; data not shown), confirming that the minimal protein-DNA interaction required for a stable complex, even for LSF448Δ, is a tetramer.

**LSF Oligomerization State in Solution**—By sedimentation analyses, LSF (CP2) formed dimers in solution (36). However, because it was unclear in this report whether any LSF was recovered as higher order forms (tetramers or higher) at the bottom of these gradients, we performed in vitro protein-protein cross-linking experiments to assess the higher order oligomerization states of LSF in solution. In vitro translated proteins were incubated with the chemical cross-linker BS3, and the reaction products were analyzed by SDS-polyacrylamide gel electrophoresis (Fig. 4). Multimers of LSF were assigned based on the apparent molecular weights, as compared with protein standards (see legend to Fig. 4), with the expectation that cross-linked species would migrate slower than their actual molecular mass, due to their branched structures. Tetramers of LSF were clearly apparent by this analysis (Fig. 4A, lane 5; Fig. 4B, lane 4; labeled Tet), although primarily dimers were detected at lower cross-linker concentrations (Fig. 4A, lane 2) or shorter time points (Fig. 4B, lane 2). To confirm that these cross-linked products were not solely the result of random collisions, a reaction was performed in a 5-fold higher volume. The same pattern of cross-linked products was obtained, suggesting that the formation of dimers, trimers, and tetramers was not concentration-dependent (data not shown). That the cross-linked products from these crude extracts are homomultimers of LSF was supported by several additional observations. First, cross-linked species of histidine-tagged LSF, purified from bacteria, comigrated with the species obtained with in vitro translated LSF. Second, the cross-linked products of N- and C-terminal deletion mutants of LSF migrated as predicted from their reduced molecular weights. The specificity of the cross-linking reactions was established by the inability of more extensive N- and C-terminal deletion mutants of LSF to form any cross-linked products under identical conditions (see summary in Fig. 2D).

**Mapping Regions of LSF Involved in Oligomerization**—Due to the higher oligomerization state of LSF on DNA, as compared with NTF-1, we sought to determine whether a new region of oligomerization could be localized for LSF. The boundaries of the region(s) involved in oligomeric protein-protein interactions were determined by two approaches. First, deletion mutants of LSF were incubated with chemical cross-linkers to form homo-oligomeric products (data not shown). These experiments were performed for 30 min at room temperature, in order to maximize the detection of oligomers. As summarized in Fig. 2D, this assay mapped the region of oligomerization in solution to between amino acids 210 and 403. For all mutants that produced cross-linked homo-oligomers, a similar set of multimeric species was obtained; thus we were unable to separate tetramerization and dimerization regions in these experiments. Second, N- and C-terminal deletion mutants of LSF were
LSF Binds to DNA as a Tetramer

FIG. 4. LSF primarily forms dimers in solution. In vitro cross-linking of in vitro translated LSF was performed using BS³, and products were analyzed by electrophoresis through a 7% SDS-polyacrylamide gel. The positions of the molecular mass markers are noted to the left of the autoradiograms. The positions of presumed monomeric (M), dimeric (Di), trimeric (Tri), and tetrameric (Tet) complexes, with apparent molecular masses from the autoradiogram of 68, 174, 232, and 268 kDa, respectively, are indicated with arrows to the right of the autoradiograms. A, cross-linking with increasing concentrations of BS³. LSF was incubated with 0 mM (lane 1), 0.1 mM (lane 2), 0.5 mM (lane 3), 1 mM (lane 4), or 5 mM (lane 5) BS³ for 1 min at room temperature. B, time course of cross-linking reaction. LSF was incubated with 0 mM (lane 1) or 5 mM (lanes 2–4) BS³ for 1 (lane 2), 5 (lane 3), or 30 min (lane 4) at 4 °C.

tested for their ability to associate with high concentrations of GST-LSF in a heteromeric complex. Immobilized GST-LSF was incubated with in vitro translated proteins, followed by extensive washing of the resin with a high salt buffer. The amounts of bound, labeled LSF derivatives were subsequently quantitated, following SDS-polyacrylamide gel electrophoresis. Results from a representative experiment are shown in Fig. 5. LSF derivatives missing 24 or 64 amino acids from the N terminus reproducibly bound GST-LSF to approximately the same degree as LSF did, with 8–11% of input protein being stable to high salt washes in this experiment. Despite similar oligomerization potentials to that of wild type LSF, these two mutants were defective in their ability to bind DNA (Fig. 2A). The reduction in DNA binding activity in these mutants is probably due to unmasking of an inhibitory region of LSF upon removal of the extreme N-terminal sequences (33). In summary, these data establish two regions in DNA binding studies of deletion mutants of NTF-1, where it was concluded that the homologous region of NTF-1 inhibited DNA binding activity in the absence of more N-terminal sequences (33). In summary, these data establish two regions in

FIG. 5. Mapping of LSF oligomerization regions by oligomerization with GST-LSF. The oligomeric interactions between GST-LSF, bound to beads, and in vitro translated, radiolabeled LSF derivatives were quantitated. The percent of input radiolabeled protein that was retained on the beads following extensive washing at high salt was determined and is shown for this representative experiment. In other experiments, the binding by LSF was equivalent to or higher than those of LSFΔ24 and LSFΔ448. The LSF derivatives, along with a summary of GST-LSF oligomerization data from multiple experiments, are presented in Fig. 2D; a subset of the derivatives were tested in this particular experiment.

The GST-LSF binding data are in general agreement with the mapping results from chemical cross-linking experiments (Fig. 2D), with the cross-linking experiments (due to the protocol used) providing a qualitative view of oligomerization, in establishing the absence or presence of protein-protein interactions, and the GST-LSF experiments providing relative efficiencies of oligomerization. Although GST contains an intrinsic oligomerization domain that can affect the ability of fusion proteins to bind DNA (39), we think it is unlikely that the GST moiety on the beads alters the oligomerization assays, because the LSF test proteins that are translated in vitro are not fusion proteins. The only apparent contradiction between the two types of mapping experiments was in the ability of LSFΔ266 to
LSF Binds to DNA as a Tetramer

oligomerize (Fig. 2D). The region between amino acids 210 and 266 includes 10 lysine residues that are potential targets for the cross-linker BS3. Thus, the apparent discrepancy between the two types of data can readily be explained by postulating that the absence of these lysines, rather than the lack of protein-protein contacts, dramatically reduced the cross-linking potential of LSFΔ266.

In summary, the results of the oligomerization experiments indicated that a core oligomerization region maps between amino acids 266 and 403 of LSF (Fig. 6). Additional contributions to oligomerization activity derive from sequences between amino acids 403 and 502, at the C terminus, and amino acids 64 and 210, at the N terminus. The C-terminal region of LSF (amino acids 280–502) has also been shown to interact with itself in a yeast two-hybrid system (33).

DISCUSSION

Investigations into the DNA binding region and oligomerization state of LSF were motivated by previous biochemical and biological findings as follows: 1) the LSF protein sequence cannot readily be categorized into any of the known DNA binding structures; 2) the consensus LSF DNA-binding site, consisting of strictly spaced direct repeats, suggests an unusual mechanism of DNA recognition; 3) LSF DNA binding activity is regulated both during cell growth stimulation (22) and during the cell cycle; and 4) as a heteromeric complex with other protein(s), LSF can generate new DNA-binding site specificities (40). Given that LSF is strikingly similar over a large portion of the protein to the Drosophila NTF-1 and that the DNA binding region and oligomerization state of NTF-1 were previously characterized, we initially focused on a comparison of LSF with NTF-1. Ordinarily, DNA-binding proteins that share a similar DNA recognition motif bind similar DNA sequences and bind with the same oligomerization state. However, our studies established that NTF-1 not only binds an LSF site but also an LSF half-site and that LSF is unable to stably interact with an NTF-1-binding site. Consistent with the requirement for a larger DNA-binding site for LSF, mixing experiments established that LSF, unlike dimeric NTF-1, is larger than a dimer on its site. Finally, serendipitous results with a C-terminal LSF truncation mutant, LSFΔ448, indicated that LSF bound its DNA site as a tetramer. However, LSF is predominantly dimeric in solution, as revealed by chemical cross-linking analysis. These and other data led to models of novel DNA-protein interactions that would prevent tandem binding by LSF subunits to adjacent, directly repeated half-sites. These data also are critical for an understanding of the biological regulation of LSF activities.

The DNA Binding Region of LSF Consists of Greater Than 230 Amino Acids—By analysis of truncation mutants, the N-terminal boundary of the region of LSF that binds to DNA mapped between amino acids 64 and 144, which is in a similar region to the boundary of the DNA binding region for NTF-1. However, the C-terminal boundary was more extended beyond that for the similar protein region in NTF-1. Optimal binding was only achieved with the entire C terminus. At room temperature, the core C-terminal boundary mapped between amino acids 403 and 448 of LSF, although at 4 °C, a smaller LSF derivative, LSFΔ383, bound DNA as well.

In general, the decrease in DNA binding activity correlated with decrease in oligomerization potential (Fig. 6). The only exception was LSFΔ403, which did not bind DNA at either temperature but was apparently unchanged in its ability to oligomerize in solution (Fig. 2D). Therefore, we hypothesize that LSFΔ403 generates an altered, nonproductive structure that interferes with the DNA recognition motif of LSF; for example, it may form new protein-protein contacts in an individual subunit.

By using GST-LSF fusion proteins, the C-terminal boundary of the DNA binding region of LSF has previously been mapped between amino acids 239 and 276 (33, 36), although the DNA binding activity of GST-LSFΔ276 (in our nomenclature) was severely reduced compared with binding by the full-length protein. Mapping of LSF DNA binding activity in the context of GST fusion proteins most likely corresponds to mapping the true boundaries of the DNA interaction region of LSF, because GST can dimerize (for example, see Refs. 39 and 41), thereby stabilizing the protein-DNA interaction. In support of this interpretation, we have observed binding by GST-LSF to DNA sites with which histidine-tagged LSF will not detectably interact (data not shown). The core DNA interaction region of LSF would therefore correspond, as expected, to that of NTF-1, whose C-terminal boundary maps to an amino acid analogous to LSF residue 273.

LSF Binds DNA as a Tetramer, Not as a Dimer—By analyzing the DNA-protein complexes from mixtures of LSF derivatives of different sizes, and of one C-terminal deletion mutant to DNAs containing three and four LSF half-sites, we concluded that LSF binds DNA as a tetramer. As this conflicts with the conclusions of previous reports, we have carefully reanalyzed those data as follows. 1) Our own previous interpretation that LSF bound to DNA as a dimer was based on an epitope-counting method, using an antibody to supershift complexes examined by EMSA (35). Only two supershifted complexes were resolved when LSF and LSFΔ24 mixtures were incubated with an antipeptide antibody that recognized LSF but not LSFΔ24. We cautiously noted at the time that although the simplest interpretation of these data was that LSF was dimeric, they would also be consistent with a tetrameric interaction, if the antibody required two epitopes to supershift the protein-DNA complexes, instead of only one epitope. 2) Yoon et al. (25) analyzed mixtures of LSF (LB1-1) proteins of different sizes by EMSA and determined that complexes of intermediate mobility appeared, again offering the simplest interpretation that LSF was dimeric. However, the complexes were not sufficiently resolved to determine the number of heteromers and therefore the exact oligomerization state. 3) Zhong et al. (36)
demonstrated that a dimer of the bacterially produced LSF (CP2), isolated by glycerol gradient sedimentation, could bind DNA. However, dimers formed in solution could clearly form tetramers upon interacting with DNA, again consistent with our current interpretation. Multimeric complexes of GST-LSF and LSF were also examined by EMSA, with insufficient resolution of the numbers of complexes to be conclusive, especially given that multiple complexes were formed between GST-LSF alone and DNA. 4) Uv et al. (33) mapped the dimerization domain of NTF-1 (Grainyhead) to its C terminus and suggested that, by homology, this region would be involved in dimerization of LSF (CP2). By using the yeast two-hybrid system, either LSF or LSF2380 (in our nomenclature), could interact with itself (33). However, this assay only measures the ability of a protein to oligomerize and does not directly address the oligomerization state. Therefore, although the simplest interpretation of all these experiments was that LSF bound DNA as a dimer, none of them were definitive enough to contradict our current interpretation that LSF binds a pair of direct repeats as a tetramer.

**LSF and NTF-1 Differ in Their Oligomerization Requirements for DNA Binding Activity**—Although LSF and NTF-1 share a striking 66% similarity over 445 amino acids (per the FASTA program on the EMBL server), these two proteins differ in their oligomerization state on DNA and, in parallel, in the structure of their respective DNA-binding sites. Our results suggest that the similarity between the proteins reflects a similarity in recognition of the consensus sequence CTGG half-site. In particular, the region of highest similarity between LSF and NTF-1, amino acids 235–246 of LSF, has been predicted to form an a-helix that recognizes DNA, based on a comparison with DNA-recognition helices found in crystal structures of several transcription factors (42). In support of this prediction, double amino acid substitutions at residues conserved between LSF and NTF-1, either at positions 234 and 236 (LSF234QL/236KE) or at positions 233 and 235, abolished LSF DNA binding activity (35, 43).

Although both proteins have a dimerization interface, LSF and NTF-1 cannot form heteromers. First, NTF-1 and LSF were unable to functionally interact in the yeast two-hybrid system (33). Second, when LSF and NTF-1 were mixed and analyzed by EMSA, no intermediate bands were revealed between the position of the NTF-1 homodimer and the LSF homotetramer. This was true both for LSF endogenous to rabbit liver (42) and for LSF cotranslated with NTF-1 in vitro (data not shown). These data provide additional evidence for differences between the oligomerization requirements of the two proteins. Therefore, although the similarity between LSF and NTF-1 through the C-terminal half of the protein may also indicate a similarity in the mode of dimerization, specific interactions have not been conserved, and LSF must make additional protein-protein contacts to form a tetramer. Consistent with this point, amino acids 364–384 of LSF (just within the minimal oligomerization and DNA binding regions of LSF) diverge from sequences in NTF-1 but are highly homologous with LBP-1a, with which LSF (LBP-1c) can oligomerize (25).

The absence of any LSF complex on the DNA retaining a single LSF half-site (Fig. 1, lane 4) indicated again that the LSF dimer-DNA complex, if formed, was unstable. The ability of NTF-1 to stably bind DNA as a dimer suggests that LSF DNA recognition contacts must be inherently weaker than those of NTF-1. Presumably the nonconserved amino acids in the DNA interaction region are reflective of residues that, in LSF, diminish DNA-protein contacts important for the high affinity interaction in NTF-1. In this regard, LSF can be viewed as a site-directed mutant of NTF-1 with a lower inherent DNA binding affinity. Due to this defect, formation of a tetrameric structure with its additional contacts is required to obtain a stable LSF-DNA complex.

**The C Terminus of LSF Prevents Tandem Binding**—Based on the data we have obtained, several models can be postulated for how tetrameric LSF binds directly repeated half-sites (Fig. 7). Because of the similarities between LSF and NTF-1, binding to DNA most likely involves recognition of an LSF half-site by a protein dimer. The two dimers of LSF could be situated either symmetrically, as diagrammed in B, or asymmetrically, as diagrammed in C. A critical element in these models is that the binding by a pair of dimers on the same face of DNA to directly repeated half-sites requires that the protein has some mechanism to prevent repetitive oligomerization along the DNA. Repeated oligomerization would be a potential problem when oligomerization regions are positioned in the same orientation.
(B and C). Repetitive oligomerization is clearly restricted for LSF, because LSF does not form a stable hexamer on DNA containing three half-sites (Fig. 7B).

Three possible solutions that would prevent continual, tandem binding are presented. First, C-terminal tetramerization of LSF could limit further protein-protein interactions (Fig. 7A). The modeled tetramerization region, based on the four-helical bundle found in lac repressor (21) or p53 (20), forms a closed structure that does not permit higher order oligomerization. In fact, LSF sequences between 471 and 492 are strongly predicted to form an α-helix (44, 45). This model predicts that the protein-protein interactions between LSF448Δ dimers, which lack these potential α-helices, will be reduced, allowing at least one monomer to be less constrained and available to interact with protein subunits on adjacent half-sites (Fig. 7A).

A second method of preventing continual, tandem binding of LSF monomers or dimers would be to spatially prevent the exposed oligomerization regions from further association (Fig. 7B). This model builds upon the interactions found in the crystal structure of Arc repressor and is dependent on the protein-mediated bending of DNA (3). Tetrameric interactions in Arc are mediated by loop structures derived from one monomer of each of the antiparallel dimers that recognizes the half-site. A generalized model for LSF is diagrammed in Fig. 7B. LSF-LSF interactions would bind the DNA site to maximize the energetically favorable protein-DNA interactions. In the illustrated version of this model, the protein-protein interactions would produce a conformational change in LSF (as predicted by the pointed half-oval) but not in LSF448Δ. Alternatively, the C terminus of LSF could participate by forming additional, stabilizing protein-protein interactions, which would be missing in LSF448Δ. In either case, binding by LSF448Δ would not bend the DNA and would no longer restrict the outward oligomerization regions from forming contacts with additional LSF448Δ molecules binding to DNA at adjacent half-sites. This model is intriguing since LSF apparently bends DNA when bound to the LSF-280 site,4 although whether LSF448Δ bends DNA remains to be determined.

Finally, a third possibility is that the C terminus of LSF promotes capping, such that formation of a tetramer promotes a conformational change in the overall protein structure so that the “free” tetrameric interface is not accessible for further oligomerization. This capping function would be absent in LSF448Δ, thereby permitting further oligomerization.

No structures of tetrameric proteins binding to directly repeated DNA sequences have yet been solved. LSF therefore represents a novel DNA binding and oligomerization strategy. The requirement for multiple protein-protein and protein-DNA interactions explains why the region of LSF that is necessary to form a stable protein-DNA complex is so large.

Biological Rationale for Tetrameric DNA Binding Transcription Factors—The requisite formation of tetramers of LSF for generation of stable protein-DNA complexes opens the door for regulation of DNA binding activity at a variety of steps. For example, phosphorylation, which is known to enhance the DNA binding activity of LSF (22), might modulate DNA recognition, oligomerization, or both. In addition, given that LSF is not stable in solution as a tetramer, but as a dimer, regulation of the interactions between LSF and other partner proteins could prevent tetrameric LSF DNA binding and/or allow recognition of new DNA-binding sites. Examples of complexes containing LSF with other partner proteins are emerging (25, 40, 46, 47), some of which define new DNA site specificities. Whether the tissue-specific complexes between LSF and partner proteins use similar oligomerization and DNA-binding interfaces to those of LSF remains an open question, which can be addressed with the methodologies and reagents we have presented here.

Acknowledgments—We thank Laura Attardi and Robert Tjian for the NTF-1 expression vector and UBX oligonucleotides; Quan Zhu for plasmid constructs pET-Δ266, pET-3266, and pET-3370 and for design of consensus and mutant consensus oligonucleotides; and Christina Powell, Elise Drouin, Thomas Ellenberger, Robert Kingston, and Robert Sauer for critically reading the manuscript.

REFERENCES

1. Nelson, H. C. M. (1995) Curr. Opin. Genet. & Dev. 5, 180–189
2. Beckwith, J. R. B., and Gilbert, N. E. (1995) J. Mol. Biol. 252, 179–191
3. Jenq, H., and Sauer, R. T. (1994) Science 263, 15036–15045
4. Schmitt, S. C., Mietz, G. C., and Steitz, T. A. (1991) Cell 65, 135–145
5. Jordan, S. R., and Pabo, C. O. (1988) Science 240, 893–899
6. Schultz, S. C., Shields, G. C., and Steitz, T. A. (1991) Science 253, 1001–1007
7. Brenner, R. G., Roderick, S. L., Takeda, Y., and Matthews, B. W. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, 8165–8169
8. Hegde, R. S., Grossman, S. R., Laimes, L. A., and Sigler, P. B. (1992) Nature 359, 505–512
9. Falcioni, S., and Tson, T. J. (1995) Nature 376, 469–468
10. Eismont, R. J., and Nelsen, H. R. (1992) Biochemistry 31, 12272–12276
11. Jeffrey, P. D., Gorina, S., and Pavletich, N. P. (1995) Science 267, 1498–1502
12. Friedman, A. M., Fischman, T. O., and Steitz, T. A. (1995) Science 268, 1720–1727
13. Volker, J. J., Rameh, L. E., Zhu, Q. DeCaprio, J., and Hansen, U. (1997) Genes Dev. 11, 1435–1446
14. Lim, C. P., Swedendam, S. L., and Sheffer, M. (1992) Mol. Cell. Biol. 12, 825–835
15. Wu, F. K., Garcia, J. A., Harrich, D., and Gaynor, R. B. (1988) EMBO J. 7, 2117–2129
16. Yoon, J.-B., Li, G., and Roeder, R. G. (1994) Mol. Cell. Biol. 14, 1776–1785
17. Huang, H.-C., Sundseth, R., and Hansen, U. (1990) EMBO J. 9, 287–298
18. Freeman, H. C., Sundseth, R., and Hansen, U. (1990) Genes Dev. 4, 287–298
19. Lim, C. F., Fang, L., Swedendam, S. L., and Sheffer, M. (1993) J. Biol. Chem. 268, 16096–16107
20. Malim, M. H., French, D., Ballard, D. W., Hauber, J. B., and Callen, B. R. (1989) J. Virol. 63, 3213–3219
21. Jones, K. A., Lucie, C. A., and Dungan, N. P. (1998) Embo J. 2, 1101–1114
22. Falcioni, S., and Tson, T. J. (1995) Biochemistry 34, 8156–8169
23. Mantovani, R. (1996) Mol. Cell. Biol. 16, 503–512
24. Deylach, B. D., Attardi, L. D., Admon, A., Freeman, M., and Tijan, R. (1989) Genes Dev. 3, 1677–1686
25. Bray, S. J., Burke, B., Brown, N. H., and Hirsch, J. (1989) Genes Dev. 3, 1130–1145
26. Uv, A. E., Thompson, C. R., and Bray, S. J. (1994) Mol. Cell. Biol. 14, 4020–4031
27. Attardi, L. D., and Tijan, R. (1993) Genes Dev. 7, 1341–1353
28. Shirra, M. K., Zhu, Q., Huang, H.-C., Pallas, D., and Hansen, U. (1994) Mol. Cell. Biol. 14, 5076–5087
29. Zifcowsky, J., Popik, W., and Sheffer, M. (1994) J. Biol. Chem. 269, 21269–21276
30. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Siedman, J. G., Smith, J. A., and Struhl, K. (1995) Current Protocols in Molecular Biology, pp. 10.21–10.29, Greene Publishing Associates/Wiley InterScience, New York
31. Hope, I. A., and Struhl, K. (1987) Genes Dev. 1, 2709–2724
32. Barnov, M., Schwartz, E., and Felsenhok, B. (1997) J. Biol. Chem. 272, 15521–15526
33. Jane, S. M., Nienhaus, A. W., and Cunningham, J. M. (1995) EMBO J. 14, 97–105
34. Walker, J., Crowley, P., Moreman, A. D., and Barrett, J. (1993) Mol. Biochem. Parasitol. 61, 255–264
35. Suzuki, M. (1993) EMBO J. 12, 3221–3226
36. Shirra, M. K. (1995) Characterization of DNA binding and Oligomerization Properties of Mammalian transcription Factor LSF, Ph.D. thesis, Harvard University
37. Chow, F. Y., and Fasman, G. D. (1978) Ann. Rev. Biochem. 47, 251–276
38. Garnier, J., Osguthorpe, D. J., and Robson, B. (1978) J. Mol. Biol. 120, 97–120
39. Sueyoshi, T., Kobayashi, R., Nishio, K., Aida, K., Moore, R., Wada, T., Handa, H., and Negishi, M. (1995) Mol. Cell. Biol. 15, 4158–4166
40. Romero, F., Gabriel, M. N., and Margolis, D. M. (1997) J. Virol. 71, 9375–9382
41. V. Bajaj and U. Hansen, unpublished data.