Multiple GTF2I-like Repeats of General Transcription Factor 3 Exhibit DNA Binding Properties

EVIDENCE FOR A COMMON ORIGIN AS A SEQUENCE-SPECIFIC DNA INTERACTION MODULE

Detlef Vullhorst‡ and Andres Buonanno

From the Section on Molecular Neurobiology, NICHD, National Institutes of Health, Bethesda, Maryland 20892

A hallmark of general transcription factor 3 (GTF3) is the presence of multiple GTF2I-like repeats that were suggested to mediate protein-protein interactions. However, we have recently demonstrated that repeat 4 is necessary and sufficient for binding of GTF3 to the bicoid-like motif of the Troponin I slow enhancer. Given the sequence similarity between different GTF2I-like repeats we hypothesized that DNA binding might be a common property of this domain type. We subjected five repeats of GTF3 to random oligonucleotide selection (SELEX) to assess their DNA binding potentials. We delineated the consensus sequence G$_{1}$G$_{4}$GATTA$_{3}$B$_{4}$G$_{4}$ for repeat 4 and showed that binding sites for GTF3 in enhancers for Troponin I and homeobox c8 (HOXc8) are in very good agreement with this motif. SELEX selections for repeats 5 and 2 enriched for oligonucleotides that were also bound by R4, suggesting that they share common sequence preferences, whereas repeat 3 exhibited relaxed sequence requirements for DNA binding. No binding was observed for repeat 1. We also show that GTF2I-like repeats 4 and 6 of transcription factor II-I (TFII-I) exhibit modest DNA binding properties. Lastly, we identified several amino acids of GTF3 repeat 4 required for high affinity protein-DNA interaction. Based on the ability of many repeats to bind DNA in vitro, we suggest that GTF2I-like domains evolved by duplication and diversification of a prototypic DNA-binding ancestor.

Transcription of eukaryotic genes generally requires the assembly of multiple nuclear factors bound directly or indirectly to enhancer elements to regulate the recruitment of the basal transcription machinery to the core promoter. General transcription factor 3 (GTF3); also known as GTF2ird1, MusTRD1, BEN) belongs to a family of nuclear proteins including TFII-I as its prototype which has been implicated in organizing transcription factor complexes by virtue of its multiple GTF2I-like repeats that structurally resemble helix-loop-helix (HLH) domains (1). These motifs exhibit extensive sequence similarities to each other and to the corresponding domains in TFII-I (see Fig. 1). As potential HLH domains they are structurally unusual in that their loop regions are long (~40 residues) compared with conventional HLH domains (~30 residues) (2, 3). GTF3 has been implicated to interact directly with several proteins, such as retinoblastoma protein (Rb), HDAC3, PI-ASβ, and MEF-2 (4–6). However, it is not known which, if any, of the GTF2I-like repeats are involved in these interactions. The second repeat of TFII-I is preceded by a stretch of basic residues and therefore constitutes a classical basic HLH domain that binds to the upstream-derived E-box of the AdML (adenovirus major late) promoter (7). No such consensus basic HLH motif has been functionally identified for GTF3.

We and others isolated GTF3/MusTRD1 as a nuclear factor that binds the upstream enhancer of the rodent (SURE) and human (USE) troponin I (TnI) slow gene, a regulatory element that controls slow fiber-specific transcription in transgenic mice (8–10). The SURE consists of two functional units that cooperate to establish slow fiber-restricted transcription. The downstream half, including binding sites for MEF-2 and basic helix-loop-helix transcription factors such as MyoD, is required for muscle-specific activity but lacks slow fiber specificity. The upstream half is necessary to restrict the pan-muscle activity of the downstream half to slow twitch muscle fibers (10). GTF3 was isolated in yeast one-hybrid screens as a nuclear factor that binds to a bicoid-like sequence motif (BML; B1b in the human USE) within the upstream halves of the SURE and the USE. In in vivo transfected rat muscles, GTF3 represses transcriptional activity of the SURE (10). GTF3 is expressed in muscle primarily during early development and is subjected to extensive alternative splicing, generating proteins with 6 (α- and β-isomers) or 5 (γ-isomers) GTF2I-like repeats (10, 11). Together, these data suggest that GTF3 is involved in the down-regulation of the TnI slow gene in developing fast muscles. Like GTF3, BEN was isolated in a yeast one-hybrid screen as a factor that binds to the EFG site of the Hoxc8 enhancer (12).

Mapping of the DNA binding domain using in vitro-translated proteins revealed that GTF2I-R4 is necessary and sufficient for interaction of GTF3 with the BLM (13). The absence of a nearby basic region in R4 suggested an intrinsic ability of the GTF2I-like repeat motif to interact with DNA in a sequence-specific manner that is distinct from the basic HLH/E-box interaction observed for TFII-I. Based on the sequence similarities between GTF2I-like repeats and the observed DNA binding capability of R4, we speculated that DNA-protein interactions might be a common property of this domain. In agreement with this notion it was recently reported that GTF3/MusTRD1 interacts via R4 and, albeit weakly, via R2 with the B1b element of the USE of the human TnI slow gene (6). In this study we...
investigated whether individual repeats can interact with DNA in a sequence-specific fashion. We purified GTF2I-like repeats as glutathione S-transferase (GST) fusion proteins and used oligonucleotide selection by exponential enrichment (SELEX) to isolate binding-competent sequences from a random oligonucleotide library. We delineated a consensus binding site for R4 and showed that all repeats except for R1 display some degree of DNA binding activity in vitro. The binding is sequence-specific for R4, R5, and R2 but appears to be largely sequence-independent for R3. These findings are important for in silico-based prediction of GTF3 binding sites in other genes and thus may contribute to our understanding of GTF3 functions in different tissues and possibly of the pathogenesis underlying Williams syndrome in which both GTF3 and TFII-I are hemizygotically deleted (14). Moreover, we also demonstrated that some repeats of TFII-I exhibit DNA binding properties. Based on the ability of most repeats to interact with DNA, we speculate that this property was inherent to the prototypic GTF2I-like motif.

MATERIALS AND METHODS

Oligonucleotides—Oligonucleotides were made by Bioserve (Laurel, MD) or IDT DNA technologies (Coralville, IA). For a list of oligonucleotides used in this work, see Table I.

Subcloning of GTF2I-like Repeats for Prokaryotic and Eukaryotic Expression—Sequences corresponding to the 75-amino acid GTF2I-like repeats plus 10 flanking residues on either side (encompassing the complete sequences suggested to represent HLH domains) were amplified by PCR from mouse cDNAs encoding the β isoform of GTF3/BEN and the Δ isoform of TFII-I (Image clones 555547 and 3157775, respectively (15)). PCR products were then subcloned into pGEM-T vector (Promega, Madison, WI). Inserts were excised with BglII and EcoRI restriction enzymes using flanking sites introduced by PCR and placed between the BamHI and EcoRI sites of pGEX-2T (Amersham Biosciences) for in-frame expression of GTF2I-like repeats domain as GST fusion proteins in Escherichia coli. Point mutations resulting in single amino acid substitutions of GTF3 repeat 4 were introduced into the corresponding pGEX-2T (GTF3-R4 expression plasmid using the Gene-Editor mutagenesis kit (Promega)).

For reporter analysis in eukaryotic cells, mouse GTF3-R4 was expressed as an amino-terminal VP16 fusion protein in vector pCMV-Sport2 (Invitrogen). The sequence encoding the repeat exactly matched the sequence in the corresponding GST fusion protein (see above). It was amplified using SalI (5′)- and ApaI (3′)-linked oligonucleotides and inserted between the corresponding sites of pCMV-Sport2. Next, the VP16 transactivation domain from herpes simplex virus (amino acids 412–490) was PCR amplified and inserted as an ApaI-HindIII fragment into the sequence in the corresponding GST fusion protein (see above). It was expressed as an amino-terminal VP16 fusion protein in vector pCMV-Sport2. Next, the VP16 transactivation domain from herpes simplex virus (amino acids 412–490) was PCR amplified and inserted as an ApaI-HindIII fragment in-frame with the upstream sequence (GTF3-R4-VP16). Because preliminary experiments revealed that this protein exhibited poor nuclear residence, a fragment of GTF3 encompassing the carboxyl-terminal 65 amino acids including the nuclear localization signal was inserted as a PCR product into the ApaI site of GTF3-R4-VP16.

**GTF3-R4 single amino acid point mutations**

| K700V | AGACCTGGTATTAACGTGGGAAATACGTTGGA |
| K708A | GCTCTCGCGGATCGACATCGCCAAC |
| K717E | GGTGCCCTACGAGAGAATCAAAA |
| K720L | GCTACCAGGATTAACAGCAGG |
| E729Q | GGTAATCATTCAAGGCCTACCCC |
| P733E | ATTGAAGGCCTACCCGAAGGGATCCCATTCCGC |
| N747A | CTTTCTTTGAAATCGGATTATAGGAATGTTTTGTCT |
| S745P | CCCTGCACCTTTGGCCCCCAGAACCTGGAAAGG |
| K709A | GCTCTGGGCATCGCATACCCAGTGCAG |
| S753E | CCAGAACCTGGAAAGGATTCTCGAGGTGGCTGACAAGATCAAG |
| K700V | AGACCTGGTATTAACGTGGGAAATACGTTGGA |
| K708A | GCTCTCGCGGATCGACATCGCCAAC |
| K717E | GGTGCCCTACGAGAGAATCAAAA |
| K720L | GCTACCAGGATTAACAGCAGG |
| E729Q | GGTAATCATTCAAGGCCTACCCC |
| P733E | ATTGAAGGCCTACCCGAAGGGATCCCATTCCGC |
| N747A | CTTTCTTTGAAATCGGATTATAGGAATGTTTTGTCT |
| S745P | CCCTGCACCTTTGGCCCCCAGAACCTGGAAAGG |
| K709A | GCTCTGGGCATCGCATACCCAGTGCAG |
| S753E | CCAGAACCTGGAAAGGATTCTCGAGGTGGCTGACAAGATCAAG |

**Table I**

Oligonucleotides used in this study

| Subcloning and expression of GTF2I-like repeatsa | Subcloning and expression of GTF2I-like repeatsa |
|---------------------------------------------|---------------------------------------------|
| GTF3R1.FWD | CTGGAGGACAGTGCCTCGGAGCTGTC |
| GTF3R1.REV | CGGCTTCGTTGTCTGCTGCCACCCTG |
| GTF3R2.FWD | ATCAAGAAAGATGGAACATCAAT |
| GTF3R2.REV | ATTCGCCCTCTGAGAATGCTGTC |
| GTF3R3.FWD | GAGAGGCCTGTGACTGAGTCAAG |
| GTF3R3.REV | AGGTTCCTTGGACCCGTCAGTAG |
| GTF3R4.FWD | CTGGTCGAGACTTCGCCGAAAAGG |
| GTF3R4.REV | GTCATTGGAGACTTCGCCGAAAAGG |
| GTF3R5.FWD | CTCAACAGCTCGGGAAAGAGGAG |
| GTF3R5.REV | ATTTGTTGGAGCTCCCTGAAAGAG |
| TFI1-R4.FWD | GAAGATTGGAATGTCAGAATCACC |
| TFI1-R4.REV | CGTGCACCAACATTGAGAACAG |
| TFI1-R6.FWD | CTGTGACCAACATTGAGAACAG |
| TFI1-R6.REV | CTGTGACCAACATTGAGAACAG |

**a** Oligonucleotides included linker sequences that are not shown in the table (coding, BglII, noncoding, EcoRI).

**b** Double-stranded; sequences correspond to upper strands.
DNA Binding of GTF2I-like Repeats

RESULTS

Delineation of the Consensus Binding Sequence for GTF3-R4—GTF2I-like-repeats 1–5 of mouse GTF3 β were expressed in E. coli and purified as GST fusion proteins. Most GTF3 isoforms contain an additional sixth repeat whose core repeat is identical in amino acid sequence to R5 and whose flanking sequences are highly similar (13; see also Fig. 1A). We included only R5 in our experiments because the properties of repeats 5 and 6 were expected to be very similar. To confirm their ability to interact with double-stranded DNA oligonucleotides, fusion proteins were incubated with a probe corresponding to the SURE enhancer encompassing the sequences between nucleotide positions −844 and −808 and then subjected to electrophoretic mobility shift assays (Fig. 1, B and C). In agreement with our previous data using in vitro-translated proteins (13), this probe gave rise to a robust shift with R4. Additional bands of similar mobility were obtained with R5, and to a lesser extent with R3, suggesting that these domains also exhibit some DNA binding activity. In contrast, R1 did not interact with the 844/808 probe. As expected, no complexes were obtained with unfused GST protein. R2 was not included in this assay because of low solubility of the corresponding GST fusion protein (see “Materials and Methods”). The interaction between R4 and the 844/808 probe was confirmed in GST pull-down assays, demonstrating that these DNA-protein interactions were stable in free solution (data not shown). The apparent ability of several GTF3 repeats to interact with the 844/808 probe supports our hypothesis that DNA binding might be a common property of GTF2I-like domains.

We then used the random oligonucleotide selection technique SELEX to determine the consensus DNA binding sequence for GTF3-R4. Starting from a library of random 14-bp oligonucleotide sequences flanked by invariant primer binding sites, we enriched for oligonucleotide pools capable of interacting with this repeat using a GST pull-down approach (17). Five cycles of selection and subsequent amplification were performed. Fig. 2A shows an oligonucleotide mobility shift assay with R4 protein and radioactively labeled probes derived from the unselected SELEX library and from oligonucleotide pools obtained after each round of selection and amplification. The enrichment of oligonucleotides competent in binding to this repeat is evident from the increase in signal intensity of the corresponding DNA-protein complexes. The signal intensities of the shifts in lanes 5 and 6 were similar and only slightly weaker than those of a complex formed between R4 and the SURE 844/808 probe (lane 7), suggesting that the selection was mostly completed after four rounds of SELEX. Moreover, competition experiments demonstrated that the selected sequences could be competed with SURE-derived unlabeled oligonucleotides that included the BLM (844/808, lane 9; 842/815, lane 10), but not with an oligonucleotide that lacked the BLM (827/780, lane 11). We then cloned the oligonucleotide pool from SELEX cycle 4 and analyzed 30 random clones for the presence of shared sequences (Fig. 2B). The alignment revealed a 9-bp motif, including the conserved core sequence G_7GATTAA_2, a stringency of >90% in all six positions (Fig. 2C). The second guanosine is the only completely invariant position in all clones analyzed. The 6-bp core is flanked on the upstream side by a single and on the downstream side by two more relaxed positions preferentially occupied by guanosines. The core consensus sequence is in perfect agreement with the sequence of the BLM (CGGATTAC), which serves as the binding site for GTF3 in the Tn1 enhancer SURE. The absence of guanosines from the flanking sequences apparently does not significantly affect DNA-protein interaction, giving the ability of the BLM to bind to R4 (see Fig. 1C).
To corroborate the results from the SELEX analysis of R4, we tested the interactions between wild-type and mutated versions of oligonucleotides derived from the TnI SURE (844/823) and from the EGF site of the HOXc8 enhancer that was demonstrated previously to bind GTF3/BEN (12). As shown in Fig. 3A, the sequence of the HOXc8 enhancer used to isolate GTF3/BEN by functional cloning includes a motif that conforms to the consensus motif for GTF3 R4. The complexes formed between individual repeats and a HOXc8 oligonucleotide probe that contains this sequence were very similar to those seen with the BLM of the TnI SURE (Fig. 3A, lanes 1–4). Moreover, this oligonucleotide competed for binding of R4 to the TnI SURE probe 844/823 that harbors the R4 binding motif (lane 8). In contrast, both HOXc8M and Mut837, in which the invariant G of the core consensus motif was mutated to T, failed to bind R4 (lanes 10 and 12). However, unlike Mut837, HOXc8M modestly competed for binding to the 844/823 wild-type probe (lanes 7 and 9), possibly because of the presence of the GGATAG sequence right downstream from the mutated consensus that closely resembles the GAGATTAG motif.

**Fig. 1.** Binding of GTF2I-like repeats to an oligonucleotide probe derived from the TnI enhancer SURE. A, alignment of the 95-amino acid sequences of all six GTF2I-like repeats of GTF3 and repeats 4 and 6 of TFII-I. Alignments were generated with ClustalX (23). The colors used to indicate conserved positions (defined as >60% conservation) were as follows: orange, Gly; yellow, Pro; blue, small and hydrophobic (Ala, Val, Ile, Leu, Met, Phe, Trp); green, hydroxyl and amine (Ser, Thr, Asn, Gln); red, charged (Asp, Glu, Arg, Lys); cyan, His, Tyr. The highly conserved 75-amino acid GTF2I-like repeats that represent the cores of the putative HLH domains described previously (12, 21) are flanked on either side by 10 poorly conserved residues. Filled triangles indicate residues subjected to mutation analysis as shown in Fig. 7. The bars below the sequences indicate the alignment quality for each position. B, SDS-polyacrylamide gel of purified GST fusion proteins for repeats 1, 3, 4, and 5 (10 μg/lane). The calculated molecular mass is ~37 kDa for all proteins. R2 protein preparations were mostly insoluble and were therefore not included. C, gel retardation assay of GTF3 repeats 1, 3, 4, and 5 using the oligonucleotide probe 844/808 derived from the TnI enhancer SURE (lanes 2–5). Unfused GST was included as a control (lane 1).
We then sought to confirm these results in an in vivo binding assay. COS-7 cells that express very low amounts of endogenous GTF3 (6) were cotransfected with an expression plasmid for a GTF3-R4-VP16 fusion protein and luciferase reporter plasmids harboring trimerized versions of wild-type and single-point mutated SURE 844/823 sequences linked to the basal
As shown in Fig. 4, no specific enrichment of binding-competent oligonucleotides was observed for R1. This result suggests that repeat domain 1 does not exhibit detectable DNA binding properties despite its sequence similarities to R4. Likewise, R2 apparently did not interact robustly with the R2-selected SELEX oligonucleotide pools. However, incubation of these pools with R4 revealed a clear increase in the fraction of binding-competent sequences. A likely explanation for these seemingly contradictory observations is that R2 exhibits very weak but specific DNA binding properties. The affinity was apparently strong enough to allow for enrichment of specific sequences by SELEX (note that even very small amounts of selected oligonucleotides will subsequently be greatly amplified by PCR). However, the binding was probably too weak to be directly detectable in pull-down assays. Based on the observed interaction with R4, we speculate that the selected sequences conform, or are very similar to, the consensus motif delineated for R4. In contrast to R2, repeat 3 exhibited significant DNA binding activity even with the unselected SELEX oligonucleotide pools. However, the avidity of R5 interactions to the selected oligonucleotides appears to resemble R4 and R2 because it also exhibits significant DNA binding properties. The affinity of these pools for R5 was revealed by a clear increase in the fraction of binding-competent sequences. A likely explanation for these seemingly contradictory observations is that R2 exhibits very weak but specific DNA binding properties. The affinity was apparently strong enough to allow for enrichment of specific sequences by SELEX (note that even very small amounts of selected oligonucleotides will subsequently be greatly amplified by PCR). However, the binding was probably too weak to be directly detectable in pull-down assays. Based on the observed interaction with R4, we speculate that the selected sequences conform, or are very similar to, the consensus motif delineated for R4. In contrast to R2, repeat 3 exhibited significant DNA binding activity even with the unselected SELEX oligonucleotide pool.
We chose SELEX pools from the last cycles of both R3 and R5 selections to determine sequences from individual oligonucleotide clones (Fig. 5). As apparent from the tabulation of sequences obtained for R3, we could not identify a complex conserved motif for binding of R3 to DNA. This is in agreement with results from the pull-down assays that suggested a promiscuous interaction of R3 with a large fraction of the unselected oligonucleotide library and identifies R3 as a DNA binding domain with little or no selectivity for distinct recognition motifs. In stark contrast, sequences selected with R5 exhibited a stereotypical configuration of two GAGATTA motifs fused tail-to-tail to yield a palindromic arrangement of two motifs each resembling the core consensus sequence recognized by R4. In rare cases, the orientation of these motifs was inverted (such as in clones 5 and 22). This result explains why the corresponding SELEX pools were also recognized by R4. It is conceivable that the SELEX procedure selected this configuration to augment the weak affinity of R5 for the core GAGATTAG motif (see Fig. 4). However, we also observed toward later cycles of SELEX that pools increasingly exhibited a tendency to form multimers (despite using only the 43-bp band as an input for the next cycle), suggesting that selection-independent enrichment of these oligonucleotides resulting from self-priming might have contributed to this effect. In support of this idea we also found that mutating one half-site of the palindrome did not decrease the binding of GTF2i-R5 in gel retardation assays (data not shown). The significance of this particular arrangement of sequence motifs is therefore unclear. It can nevertheless be concluded with good confidence that the presence of a GTF2I-R4 core consensus sequence is obligatory for R5-DNA interaction, demonstrating that this motif is recognized by both R4 and R5 (and, by extrapolation, R6).

![Consensus for GTF3-R5](image)

FIG. 5. SELEX for other GTF2I-like repeat domains. A, tabulation of SELEX-derived DNA sequences for GTF3-R3/-R5 and TFII-I-R4/-R6 domains. Invariant flanking sequences were removed. Sequences for GTF3-R5 and TFII-I-R4 were aligned to highlight shared motifs. For TFII-I-R4, 12 clones that did not align were omitted. In TFII-I-R6, clones that contained GTF3-R5-like motifs are in bold. Note that some clones contain sequences slightly shorter or longer than the 14 bp present in the random portion of the SELEX-1 oligonucleotide; these aberrations likely represent amplification or oligonucleotide synthesis artifacts. B, compilation of the consensus sequence for GTF3-R5. Note that the motif contains an inverted arrangement of two motifs closely resembling R4 consensus binding sites.
they are the most similar to the four cycles of SELEX. Repeats 4 and 6 were chosen because we subjected two repeats of TFII-I, a close homolog of GTF3, to whether DNA binding is limited to GTF2I-like repeats of GTF3, GTF3-R4 (GAGATTAG). SELEX for TFII-I-R6, in contrast, analyzed) that conform perfectly to the core consensus sequence of Fig. 5, TFII-I-R4 selected 27 sequences (of a total of 39 analyzed) from GTF3 and TFII-I. Highly conserved in all GTF2I-like repeats analyzed in this study—

**SELEX for Repeats 4 and 6 of TFII-I**—To investigate whether DNA binding is limited to GTF2I-like repeats of GTF3, we subjected two repeats of TFII-I, a close homolog of GTF3, to four cycles of SELEX. Repeats 4 and 6 were chosen because they are the most similar to the bona fide DNA binding domain GTF3-R4 (68 and 74% similarity, respectively). As shown in Fig. 5A, TFII-I-R4 selected 27 sequences (of a total of 39 analyzed) that conform perfectly to the core consensus sequence of GTF3-R4 (G₃GATTAG). SELEX for TFII-I-R6, in contrast, mostly did not yield sequences with identifiable consensus motifs. Only 2 of 31 sequences contained the G₃GATTAG core motif. Interestingly, they were arranged as direct or inverted repeats (clones 2 and 13, respectively), similar to sequences obtained for GTF3-R5. We conclude that sequence-specific DNA-protein interaction is not confined to GTF2I-like repeats present in GTF3.

**Relative Binding Affinities of GTF2I-like Repeats**—To approximate the relative binding affinities of repeats 3–5 of GTF3, as well as repeat 4 and 6 of TFII-I, we used gel retardation assays with the SURE 844/823 probe and variable amounts of protein and asked at which concentrations specific complexes of comparable intensities are obtained. As shown in Fig. 6, 1.35–2.7 μM R3 protein (50–100 ng/μl), 0.68–1.35 μM R5 protein, and 88 nM (2.5 ng/μl) GTF3-R4 protein yielded complexes of similar intensities, indicating that R4 binds with 20–40-fold and 10–20-fold higher avidity than GTF3-R3 and -R5, respectively. GTF3-R5 protein yielded complexes of similar intensities, indicating that R4 binds with 20–40-fold and 10–20-fold higher avidity than GTF3-R3 and -R5, respectively. DNA binding activity of TFII-I-R4 was at least 80-fold weaker than GTF3-R4, whereas the affinity of repeat 6 was between those determined for GTF3-R3 and GTF3-R5 (±20–fold; lanes 10–15). The results for TFII-I-R4 demonstrate that repeats can have distinct sequence preferences despite weak DNA affinity. In contrast, GTF3-R3 and TFII-I-R6 exhibit moderate DNA binding affinity but no or very modest sequence preferences. Although these findings clearly demonstrate the ability of GTF2I-like repeats from GTF3 and TFII-I to interact with DNA targets containing the consensus motif G₃GATTAG, they also demonstrate that GTF3-R4 binds to this motif with significantly greater affinity than any other repeat tested.

**Mutation Analysis of Amino Acids Involved in DNA Binding of GTF3-R4**—Numerous amino acid residues are invariant or highly conserved in all GTF2I-like repeats analyzed in this study and are probably critical to establish the overall structure of the repeat (see Fig. 1). However, additional sequence determinants must contribute to confer DNA binding properties because some repeats of GTF3 interact poorly (R2, R3) or not at all (R1) with DNA. To identify such residues that impart DNA binding properties to GTF2I-like repeats such as GTF3-R4, we swapped poorly conserved residues in nine positions of R4 for amino acids from the corresponding position of GTF3-R1 (Fig. 7A). Emphasis was placed on basic residues located in the 5' end of the loop that could conceivably serve to stabilize DNA-protein interactions by binding to the phosphate backbone of DNA (K700V, K709A, K717E, K720L) and on a number of nonbasic residues in the 3' end that were absent from any of the other repeats in GTF3 (E728Q, P733E, S745P, N747A, S753E). As

---

**Fig. 6. Relative DNA binding affinities of GTF2I-like repeats from GTF3 and TFII-I.** Gel retardation assays used a radiolabeled SURE 844/823 oligonucleotide as a probe and GST fusion proteins for GTF3 repeat 4 (lanes 1–3), repeat 3 (lanes 4–6), and repeat 5 (lanes 7–9), as well as TFII-I repeats 4 (lanes 10–12) and 6 (lanes 13–15). Protein concentrations used in the binding reaction are indicated.

**Fig. 7. Mutation analysis of selected GTF3-R4 residues.** A, top, sequence alignment of GTF3-R4 and -R1. Filled arrowheads indicate positions that were subject to amino acid substitutions of residues in R4 against the corresponding residue in R1. Targeted residues were selected based on their uniqueness, i.e. their absence in most or all of the other repeats in GTF3, and with particular emphasis on basic residues. Bottom, representative gel retardation assay showing the effect of the nine single amino acid substitutions on the ability of recombinant R4 protein to shift the oligonucleotide probe 844/823 efficiently. SDS-PAGE loading controls using equivalent amounts of the respective GST fusion proteins are shown below (3 μg/lane). B, quantification of binding efficiencies. Values were obtained from three independent experiments. Binding data are relative to wild-type R4 (defined as 1). *, p < 0.01; **, p < 0.001 (paired t test).
shown in Fig. 7, gel retardation assays using corresponding point-mutated GST fusion proteins revealed distinct effects of these mutations on binding of the SURE probe 844/823. K717E dramatically decreased the affinity of R4 for the 844/823 probe (86%), thus revealing a critical role for this residue in DNA-protein interaction. Surprisingly, substitutions of the three other targeted lysines had no or modest effects on DNA binding. The importance of nonbasic residues for DNA-protein interaction was evident from the detrimental effects of S745P and N747A mutations that reduced DNA binding to 9 and 10% of wild-type levels, respectively. In contrast, the substitution of a proline at position 733 surprisingly did not affect the avidity of DNA binding despite the important role of prolines for secondary structure formation.

DISCUSSION

Our work represents the first systematic and unbiased analysis of individual GTF2I-like repeats with regard to DNA binding. A key aspect of our approach was the delineation of consensus sequences utilizing the SELEX procedure that allows for the selective enrichment of binding-competent oligonucleotides from a library of random sequences. The core 6-bp consensus sequence for repeat 4, _G₄GATTAC_, is in perfect agreement with _bona fide_ binding sites for GTF3 in enhancers for Tn1 and HOXc8 and therefore is likely to be a major recognition site for GTF3 in gene promoters and enhancers. Knowledge of this motif will be useful for the identification of more such sites in regulatory regions of other genes and thereby will enhance our understanding of the target genes for GTF3. We also demonstrated that mutating the second G within this core motif completely abolished binding of GTF3-R4 to Tn1 and HOXc8 enhancers, thereby confirming the stringency of the consensus for this position and establishing a straightforward assay to test the specificity of GTF3-R4 interactions with DNA. GTF2I-like repeats 2 and 5, and by extrapolation repeat 6, bind to core sequences that are very similar to the _G₄GATTAC_ motif identified for repeat 4. This suggests that amino acids conserved between these domains determine their DNA specificity, whereas others might regulate their affinity for the _G₄GATTAC_ motif. In contrast to the above repeats, we found no complex consensus sequence for GTF3-R3. It is therefore conceivable that this repeat lacks residues that impart sequence selectivity to protein-DNA interaction. Interestingly, part of the preferred core sequence (ATTA) matches the consensus for the homeodomain of homeobox transcription factors (19). It is therefore tempting to speculate that GTF2I-like repeats and homeodomains might share certain structural features determining their DNA-protein interactions. The homeodomain consists of three tightly packed helices that are separated by a loop and a turn from each other, respectively. Intriguingly, it has recently been proposed that GTF2I-like repeats contain a third helix located between the flanking 2 helices (2). A detailed understanding of the structural determinants of sequence specificity and binding affinity will have to await the resolution of the crystal structure of the complex between GTF3-R4 and its cognate DNA motif. It is also noteworthy that GTF3 was recently isolated as a factor capable of interacting directly with the DICE element of the V1 promoter present in many IgH genes (20). Its sequence bears no similarity to the _G₄GATTAC_ consensus we delineated for GTF2I-like repeats. Although the DNA binding domain was not mapped, it is possible that GTF3 contains an as yet unidentified additional DNA binding motif located outside of the GTF2I-like repeat.

Given the ability of multiple repeats to mediate sequence-specific protein-DNA interactions, it will be important to know which one is actually required for DNA binding _in vivo_. This question has not yet been addressed specifically, but preliminary biochemical evidence points to the importance of GTF2I-R4 as the major domain mediating binding to the _G₄GATTAC_ motif. First, we have shown previously in gel retardation assays that R4 is necessary in the context of the full-length GTF3 protein and sufficient as a minimal protein fragment to mediate binding to the BLM of the Tn1 SURE (13). Moreover, in the present work we have approximated the relative binding affinities of several of the GTF2I-like repeats and found that GTF3-R5 bound to DNA with 10–20-fold lower affinity than R4. The binding of R2 to DNA was only inferred indirectly through its ability to enrich sequences in the SELEX procedure which were robustly bound by R4, indicating that in our hands its affinity for the DNA was below detectable levels. This result differs somewhat from a recent report showing weak binding of an _in vitro_ translated fragment of the mouse MusT1D1a1 isoform including R2 but lacking carboxyl-terminal sequences to a trimerized B1 element of the human USE (6). It is conceivable that trimerization of the probe or the inclusion of neighboring sequences in this protein fragment helped to stabilize the interaction under gel retardation conditions. Lastly, we have failed to detect any transactivation in cells transfected with the 3×(844/823) reporter and VP16 fusions of GTF3-R3 and -R5 which exhibited DNA binding properties in gel retardation assays, indicating that these repeats do not bind the BLM motif _in vivo_ (data not shown). This does not preclude the possibility that R4 can oligomerize with other repeats to bind DNA _in vivo_ (see below).

Because TFII-I was cloned as the first member of this gene family, it has been proposed that the GTF2I-like repeats represent HLH structures with unusually long loop regions, similar to those described for Myc and upstream stimulatory factor (21). Interestingly, a recent bioinformatics study found that the sequence of the repeats does not conform to the consensus for HLH domains of characterized proteins (2). Rather, the second helix appears to be short and is followed by a β-strand. The authors suggest the existence of a third helix within the central loop region (residues 726–730 in GTF3-R4) and propose that dimerization of repeats could bring these helices into close proximity, possibly mediating key aspects of protein-DNA interactions. In support of this idea, we found that the substitution of a glutamate residue within this helix (E729Q) significantly weakened the affinity of R4 for its DNA target. Immediately upstream of this helix lies a short stretch of three basic residues (K₇₁⁷RIK⁷₂⁰) not conserved in all GTF2I-like repeats and therefore possibly involved in making contact with the phosphate groups of the DNA backbone (2). We identified Lys⁷₁⁷ as being critically important for DNA binding, whereas mutating Lys⁷₂⁰ had no effect. Lys⁷₁⁷, which is also present in DNA binding repeats GTF3-R5/6, is therefore likely to be crucial to stabilize protein-DNA complexes. A second critical region is located in the carboxyl-terminal half of the repeat as identified by two mutations, S745P and N747A. A better understanding of the involvement of this area in DNA-protein interaction will require knowledge of the crystal structure of GTF3-R4 bound to its target DNA sequence.

Our analysis of the DNA binding properties of individual GTF2I-like repeats of GTF3 has revealed a general ability of many of these motifs to interact in a sequence-specific fashion with DNA under defined conditions _in vitro_. SELEX experiments for TFII-I repeats 4 and 6 have expanded this concept beyond GTF3. We do not propose, however, that these repeats represent functional DNA binding domains in TFII-I. Binding of TFII-I to targets in _c-fos_ and _V₁_ promoters depends not on R4 or R6, but on sequences located in the amino-terminal half including a basic stretch upstream of R2, reminiscent of classical basic HLH factors (22). Taken together with the low affinity of TFII-I-R4 and R6 for the _G₄GATTAC_ motif, it seems that the important role of prolines for secondary structure formation.
unlikely that these repeats play a major role for DNA binding of TFII-I in situ.

Our findings raise the interesting possibility that the ancestral GTF2I-like repeat served as a DNA binding motif in much the same way that the repeat 4 of GTF3 serves as the DNA binding domain for interaction with enhancers for TnI and HOXc8. By undergoing multiple domain duplication events, newly added repeats were available for other types of interactions and lost some (such as GTF3-R3 and R5), most (GTF3-R2), or all of their DNA binding abilities (GTF3-R1). Future experiments will focus on the interactions of non- or weakly DNA binding repeats with other nuclear factors to regulate transcription of target genes.

Acknowledgment—We thank Connie Zhi for expert technical assistance.

REFERENCES
1. Roy, A. L. (2001) Gene (Amst.) 274, 1–13
2. Hinsley, T. A., Cunliffe, P., Tipney, H. J., Brass, A., and Tassabehji, M. (2004) Protein Sci. 13, 2588–2599
3. Ferre-D’Amare, A. R., Prendergast, G. C., Ziff, E. B., and Burley, S. K. (1993) Nature 363, 38–45
4. Yan, X., Zhao, X., Qian, M., Guo, N., Gong, X., and Zhu, X. (2000) Biochem. J. 345, 749–757
5. Tussie-Luna, M. I., Bayarsaihan, D., Seto, E., Ruddle, F. H., and Roy, A. L. (2002) Proc. Natl. Acad. Sci. U. S. A. 99, 12807–12812
6. Polly, P., Haddadi, L. M., Issa, L. L., Subramaniam, N., Palmer, S. J., Tay, E. S., and Hardeman, E. C. (2003) J. Biol. Chem. 278, 36603–36610
7. Roy, A. L., Meisterernst, M., Pognonec, P., and Roeder, R. G. (1991) Nature 354, 245–248
8. O’Maloney, J. V., Guven, K. L., Lin, J., Joya, J. E., Robinson, C. S., Wade, R. P., Hardeman, E. C. (1998) Mol. Cell. Biol. 18, 6641–6652
9. Nakayama, M., Stauffer, J., Cheng, J., Banerjee-Basu, S., Wavrousek, E., and Buonanno, A. (1996) Mol. Cell. Biol. 16, 2408–2417
10. Calvo, S., Vullhorst, D., Veneraply, P., Cheng, J., Karavanova, I., and Buonanno, A. (2001) Mol. Cell. Biol. 21, 8490–8503
11. Tay, E. S., Guven, K. L., Subramaniam, N., Polly, P., Issa, L. L., Gunning, P. W., and Hardeman, E. C. (2003) Biochem. J. 374, 359–367
12. Bayarsaihan, D., and Ruddle, F. H. (2000) Proc. Natl. Acad. Sci. U. S. A. 97, 7342–7347
13. Vullhorst, D., and Buonanno, A. (2003) J. Biol. Chem. 278, 8370–8379
14. Tussie-Luna, M. I., Roy, A. L., and Sharp, P. A. (2001) J. Biol. Chem. 276, 5460–5469
15. Lennon, G., Auffray, C., Polymeropoulos, M., and Soares, M. B. (1996) Genomics 33, 151–152
16. Feledy, J. A., Morasso, M. I., Jang, S. I., and Sargent, T. D. (1999) Nucleic Acids Res. 27, 764–770
17. Chittenden, T., Livingston, D. M., and Kaelin, W. G., Jr. (1991) Cell 65, 1073–1082
18. Nordeen, S. K. (1988) BioTechniques 6, 454–458
19. Gehring, W. J., Qian, Y. Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A. F., Resendez-Perez, D., Affolter, M., Otting, G., and Wuthrich, K. (1994) Cell 78, 211–223
20. Tantin, D., Tussie-Luna, M. I., Roy, A. L., and Sharp, P. A. (2004) J. Biol. Chem. 279, 5460–5469
21. Roy, A. L., Du, H., Gregor, P. D., Novina, C. D., Martinez, E., and Roeder, R. G. (1997) EMBO J. 16, 7091–7104
22. Cheriyath, V., and Roy, A. L. (2001) J. Biol. Chem. 276, 8377–8383
23. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeannotte, P., and Higgins, D. G. (1997) Nucleic Acids Res. 25, 4876–4882
Multiple GTF2I-like Repeats of General Transcription Factor 3 Exhibit DNA Binding Properties: EVIDENCE FOR A COMMON ORIGIN AS A SEQUENCE-SPECIFIC DNA INTERACTION MODULE

Detlef Vullhorst and Andres Buonanno

J. Biol. Chem. 2005, 280:31722-31731.
doi: 10.1074/jbc.M500593200 originally published online June 28, 2005

Access the most updated version of this article at doi: 10.1074/jbc.M500593200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2005/07/28/M500593200.DC1

This article cites 23 references, 11 of which can be accessed free at
http://www.jbc.org/content/280/36/31722.full.html#ref-list-1