⌘ *Author's Choice*

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 284, NO. 43, pp. 29310–29319, October 23, 2009

Translational Isoforms of FOG1 Regulate GATA1-interacting Complexes*

Received for publication, July 13, 2009, and in revised form, July 28, 2009 Published, JBC Papers in Press, August 4, 2009, DOI 10.1074/jbc.M109.043497

Jonathan W. Snow[‡§¶1] **and Stuart H. Orkin**[‡§¶‖2]

From the [‡]*Division of Hematology/Oncology, Children's Hospital, and* [§]*Dana Farber Cancer Institute,* [¶]*Harvard Medical School, and* [‖]*Howard Hughes Medical Institute, Boston, Massachusetts 02115*

Erythropoietic and megakaryocytic programs are directed by the transcription factor GATA1. Friend of GATA1 (FOG1), a protein interaction partner of GATA1, is critical for GATA1 function in multiple contexts. Previous work has shown that FOG1 recruits two multi-protein complexes, the nucleosome remodeling domain (NuRD) complex and a C-terminal binding protein (CTBP)-containing complex, into association with GATA1 to mediate activation and repression of target genes. To elucidate mechanisms that might differentially regulate the association of FOG1, as well as GATA1, with these two complexes, we characterized a previously unrecognized translational isoform of FOG1. We found that an N-terminally truncated version of FOG1 is produced from an internal ATG and that this isoform, designated FOG1S, lacks the nucleosome remodeling domain-binding domain, altering the complexes with which it interacts. Both isoforms interact with the C-terminal binding protein complex, which we show also contains lysine-specific demethylase 1 (LSD1). FOG1S is preferentially excluded from the nucleus by unknown mechanisms. These data reveal two novel mechanisms for the regulation of GATA1 interaction with FOG1-dependent protein complexes through the production of two translational isoforms with differential interaction profiles and independent nuclear localization controls.

Erythropoietic and megakaryocytic programs are specified from multipotential progenitors by the transcription factor GATA1, a zinc finger transcription factor first identified by its ability to bind globin gene regulatory regions (1, 2). The importance of this factor in the development of these lineages is underscored by the presence of functionally relevant GATA sites in promoters and enhancers of virtually all erythropoietic and megakaryocytic specifically expressed genes (3, 4). In fact, GATA1 is essential for erythropoiesis (5, 6) and megakaryopoiesis (7, 8), as revealed by mouse knock-out studies. Beyond its essential role in normal development, inherited mutations in GATA1 in rare individuals cause hematological disorders, including anemia and thrombocytopenia. Among those with trisomy 21/Down syndrome, somatic mutation of GATA1 leads to expression of a truncated form of GATA1 that is associated with acute megakaroblastic leukemia (9).

FOG1,[3] or Friend of GATA1, interacts physically with GATA1 and is critical to its function in multiple contexts (10, 11). In addition to GATA1, FOG1 has been shown to interact with the CTBP-containing (12–14) and NuRD repression complexes (15, 16).[4] Although the interaction of FOG1 with these complexes may account in part for GATA-mediated gene repression, no clear mechanism has been put forward for how FOG1 contributes to GATA1-dependent gene activation.

One model to explain how FOG1 may act as either a coactivator or a corepressor in the same cellular context posits that transcriptional activation by GATA1 is mediated by GATA1 in association with FOG1 but without the described repressive complexes. Accordingly, GATA1 might interact with a form of FOG1 that may not recruit repression complexes. Consequently, complexes containing activators would prevail at target genes.

Here we characterize an alternate, translational isoform of FOG1, designated FOG1S, which is truncated at its N terminus. The translation of FOG1S from an internal ATG is regulated *in cis* by both the 5′-UTR and the Kozak sequences surrounding the canonical start codon. The N-terminally shortened isoform lacks the NuRD-binding domain and, as a result, fails to bind the NuRD complex. Both isoforms interact with CTBP-containing complexes, which we also show contain the lysine-specific histone demethylase LSD1. FOG1S is preferentially excluded from the nucleus in an erythroid cell-specific manner. These data reveal several novel mechanisms for the regulation of GATA1 interaction with FOG1-dependent protein complexes through the production of two translational isoforms.

EXPERIMENTAL PROCEDURES

Cell Lines and Plasmids—Mouse erythroleukemia (MEL) and 293T cells were grown in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum. MEL cells expressing BirAV5his and a vector containing the FLAG-biotin tag have been published previously (17). BirA-expressing MEL cells were electroporated with plasmid constructs containing FLAG-biotin-tagged wild type or mutant forms of FOG1. MEL cells expressing tagged molecules were confirmed by Western blotting with anti-streptavidin-horseradish peroxidase of the

* This work was supported, in whole or in part, by National Institutes of Health Grants HL32259 and HL32262 (to S. H. O.).
⌘ *Author's Choice*—Final version full access.
[1] Supported in part by a Leukemia and Lymphoma Society Fellowship.
[2] Investigator of the Howard Hughes Medical Institute. To whom correspondence should be addressed: Division of Hematology, Children's Hospital, 300 Longwood Ave., Boston, MA 02115. E-mail: stuart_orkin@dfci.harvard.edu.

[3] The abbreviations used are: FOG1, Friend of GATA 1; UTR, untranslated region; HA, hemagglutinin; LSD1, lysine-specific demethylase 1; TACC3, transforming, acidic coiled-coil containing protein 3; MEL, murine erythroleukemia; CTBP, C-terminal binding protein; NuRD, nucleosome remodeling domain; ORF, open reading frame; WT, wild type.
[4] J. W. Snow and S. H. Orkin, unpublished observations.

total lysates or nuclear extracts. Plasmids for expression of untagged wild type or mutant FOG1 cDNA in 293T (pEF1α-V5his series) were purchased from Invitrogen. Wild type FOG1 cDNA was cloned into the pEF1α vector using EcoR1 from MT2-FOG1 (10) to generate pEF1α-FOG1 (cDNA). To generate constructs containing wild type or N-terminal truncations of FOG1 with 5′-UTR replacement, forward primers containing a 5′ BamHI site spanning the start codon of interest were used in conjunction with a common reverse primer to generate a BamHI/PflMI fragment, which was cloned into pEF1α-FOG1, replacing the wild type FOG1 N terminus and nascent 5′-UTR. To generate constructs lacking the endogenous 3′-UTR, an internal forward primer and a reverse primer spanning the stop codon with a 3′ engineered XbaI site were used to generate an Sbf1/XbaI fragment that was cloned into pEF1α-FOG1, replacing the wild type FOG1 C terminus and nascent 3′-UTR. To generate constructs containing N-terminally FLAG-tagged FOG1L and FOG1S, forward primers containing a 5′ BamHI site spanning the start codon of interest were used in conjunction with a common reverse primer to generate a BamHI/PflMI fragment that was cloned into the pEF1α-FOG1 vector. A BamHI/XbaI fragment was then cloned into the FL-Bio vector to allow for generation of in-frame tagged FOG1. Mutant versions of FOG1 were generated from FOG1 cDNA or fragments thereof using the QuikChange II XL mutagenesis kit (Stratagene, La Jolla, CA) per the manufacturer's instructions. In cases where fragments were used, they were cloned back into pEF1α-FOG1 (cDNA) or pEF1α-FOG1 (WT). HA-FOG1L has been previously described (14). All of the constructs were verified by DNA sequencing.

Total Lysate and Nuclear Extract Preparation and Western Blot Analyses—Total lysates and nuclear extracts were prepared as described (18). For Western blot analysis, aliquots of total lysates or nuclear extracts (10–30 μg) were fractionated on a SDS-polyacrylamide gel and electroblotted onto polyvinylidene difluoride membrane. Antibody incubation and chemiluminescence detection were performed according to the manufacturer's instructions (Amersham Biosciences). The antibodies used were purchased from Santa Cruz Biotechnology (Santa Cruz, CA) unless otherwise specified and include those directed to FOG1 M-20 (sc-9361), FOG1 A-20 (sc-9362), MTA2 C-20 (sc-9447), V5 (Invitrogen), GATA1 N-6 (sc-265), CTBP1 (BD Biosciences, Franklin Lakes, New Jersey), CTBP2 (BD Biosciences, Franklin Lakes, NJ), LSD1 (Abcam, Cambridge, MA), and HA Y-11 (sc-805), FLAG (anti-ECS; Bethyl Labs, Montgomery, TX). An additional antibody directed to FOG1 was produced previously in our lab (10).

Affinity Capture and Immunoprecipitation—Transient co-transfection of 293T cells with plasmids expressing FOG1 cDNA, FOG1 mutants, or empty vector was performed with FuGENE 6 according to the manufacturer's protocol (Roche Applied Science). Two days after transfection, total lysates or nuclear extracts were prepared and incubated with anti-FOG1 antibody and with protein G-agarose beads (Roche Applied Science) or with anti-M2 FLAG directly conjugated to agarose beads (Sigma) overnight. On day 2, unbound material was washed away, and bound material was eluted by boiling in Laemmli buffer and subjected to Western blot analyses.

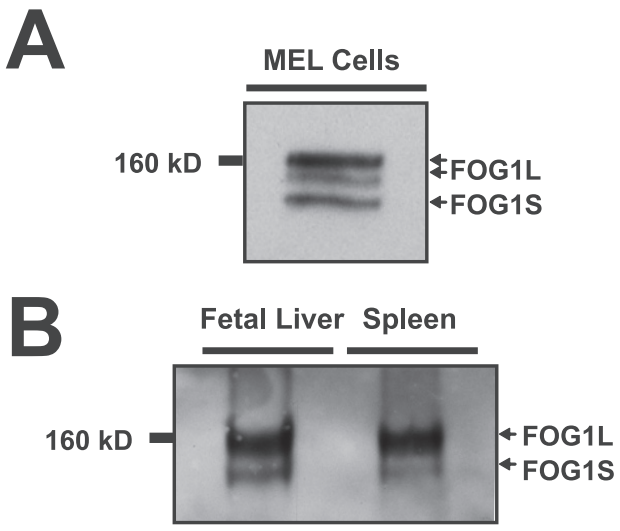

FIGURE 1. **FOG1 exists as two isoforms in an erythroid cell line and in primary erythroid cells.** *A*, two bands are recognized by α-FOG1 antibody in MEL cell nuclear extracts: the canonical FOG1 and a faster migrating form, delineated as FOG1L and FOG1S, respectively. *B*, both bands are found in nuclear extracts of embryonic day 14.5 fetal livers and in the spleen of phenylhydrazine-treated mice.

Coimmunoprecipitation of MEL cells stably transfected with various FLAG-biotin-tagged cDNAs using anti-streptavidin-agarose or anti-M2 FLAG-agarose was performed as described (19). In brief, nuclear extracts from MEL cells expressing BirA and biotin-tagged FOG1L were incubated with anti-streptavidin-agarose or anti-M2 FLAG-agarose in a buffer containing 20 mM Tris-HCl, 139 mM KCl, 12 mM NaCl, and 20% glycerol, and 0.5% Nonidet P-40. Binding was performed at 4 °C for 1 h to overnight on a rocking platform, followed by six washes in binding solution. Bound material was eluted by boiling for 5 min in Laemmli buffer. Preparation of samples for mass spectrometry, liquid chromatography-tandem mass spectrometry, and peptide sequence analysis was performed as described previously (20).

Sephacryl S400 Gel Filtration Chromatography—Crude nuclear extracts were prepared from uninduced MEL cells and prepared as previously described (17). 15 mg of total protein was injected into a 5-ml loop of a DuoFlow (Bio-Rad) fast protein liquid chromatography apparatus and separated on a HiPrep Sephacryl S400 26/60 column (Amersham Biosciences) in cold BC100 buffer at 0.5 ml/min with a collection of 1-ml fractions. Molecular mass standards were catalase (240 kDa), ferritin (438 kDa), and thyroglobulin (670 kDa).

RESULTS

FOG1 Exists as Full-length and N-terminally Truncated Isoforms—Polyclonal antibody to FOG1 (10) recognizes multiple distinct isoforms in Western blots of MEL cell nuclear extracts (Fig. 1*A*). The primary band, previously referred to as FOG1 (hereafter referred to as FOG1L), migrates at an apparent molecular mass of 160 kDa, whereas another prominent band exhibits an apparent molecular mass of 120 kDa. The FOG1L band appears here as a doublet, representing differentially phosphorylated species of the same FOG1 peptide.[3] To determine whether the FOG1S and FOG1L isoforms are pres-

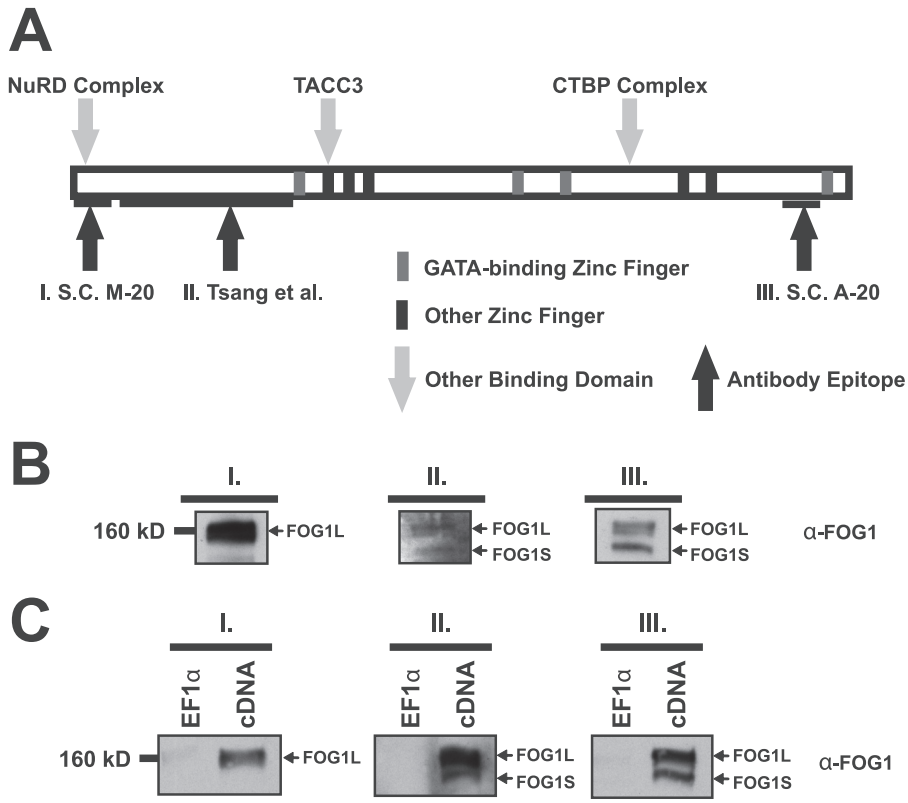

FIGURE 2. **FOG1S lacks an N-terminal domain, and both isoforms are produced from a single transcript.** *A*, schematic showing location of epitopes recognized by various α-FOG1 antibodies relative to other structural features of FOG1. *B*, isoforms recognized by these antibodies in MEL cell nuclear extracts. *C*, antibodies are used to illuminate FOG1 exogenously expressed in 293T cells transfected with FOG1 cDNA or vector alone.

ent in primary cells, we collected fetal liver cells from day 14.5 mouse embryos and probed nuclear extracts with a C-terminally directed monoclonal antibody, A-20. Greater than 95% of fetal liver cells at this stage are committed erythroid precursors and express CD71, Ter119, or both (data not shown). We found that both isoforms were detected in fetal liver nuclear extracts (Fig. 1*B*). As in MEL cells, FOG1L predominated. To determine whether both isoforms are present also in adult erythroid progenitors, we harvested spleen cells and prepared nuclear extracts 4 days after phenylhydrazine-induced hemolytic stress. At this time point ~70% of cells are committed to the erythroid lineage and express CD71, Ter119, or both (data not shown). Again, we detected both isoforms by Western blot using the A-20 antibody (Fig. 1*B*). In addition to erythropoietic tissue, both isoforms appear to be present in Western blot analysis of *in vitro* stimulated primary T-cells (21).

To further characterize the isoforms, we employed antibodies recognizing epitopes located at different sites along the full-length FOG1 protein (represented in Fig. 2*A* as *I*, *II*, and *III*). Western blotting of MEL cell nuclear extracts reveals that the M-20 antibody, which is directed to an epitope within the N-terminal 50 amino acids (Fig. 2*A*, *I*), specifically recognizes FOG1L (Fig. 2*B*). In contrast, a polyclonal antibody directed to an epitope between amino acids 19–248 (10) (Fig. 2*A*, *II*) and the A-20 antibody, directed to an epitope within amino acids 900–950 (Fig. 2*A*, *III*), recognize FOG1L and FOG1S (Fig. 2*B*). These data indicate that the FOG1S isoform lacks the N-terminal portion of FOG1.

FOG1S Is Produced through Alternate Translation from an Internal ATG—Generation of multiple isoforms of a protein can be achieved through transcriptional, translational, or post-translational mechanisms. Previous studies from our lab using Northern blot analysis (10) detected a single transcript from the FOG1 locus. To ascertain whether both FOG1 isoforms could be generated from a single transcript, we transiently expressed full-length cDNA in 293T cells under control of the pEF1α promoter. As shown in Fig. 2*C*, protein species corresponding to FOG1L and FOG1S, as seen in MEL cells, were also evident upon expression of a unique cDNA in heterologous cells. These data provide initial evidence on behalf of a translational mechanism for the generation of the two FOG1 isoforms.

In principle, other potential mechanisms might account for isoform generation. For example, alternate spliced transcripts of similar size might differ in exon utilization *in vivo*. To exclude this possibility we performed 5′-rapid amplification of cDNA ends with RNA from MEL cells. We found no evidence for transcripts of different coding exon composition that could lead to an N-terminal truncation (data not shown). Interestingly, this contrasts with findings reported for FOG2 (22). In addition, a previous *in silico* study found no evidence for alternate first exon usage for FOG1 using human genome sequence data (23). Furthermore, it is possible that FOG1S might be generated by proteolytic cleavage of full-length FOG1. However, upon treatment of MEL cells with cycloheximide to interrupt translation initiation and production of new FOG1 protein, we failed to observe an increased ratio of FOG1S, arguing against a proteolytic mechanism (data not shown). Finally, although FOG1 is heavily modified by post-translational modifications, these modifications do not play a role in the production of the two isoforms described here.[5] Taking these data *in toto*, the generation of FOG1S appears to be mediated through a translational mechanism.

The canonical translation initiation sites for mouse and human FOG1 mRNA species are conserved (Fig. 3*A*) and conform to a weak Kozak sequence (24). Both species have multiple alternate in-frame AUG codons of varying degrees of Kozak strength downstream of the predominant start codon (Fig. 3*A*). The presence of multiple downstream AUGs may provide alternate sites of translation initiation. Full-length FOG1 migrates at an apparent molecular mass of 160 kDa, despite a predicted

[5] J. W. Snow and S. J. Orkin, manuscript in preparation.

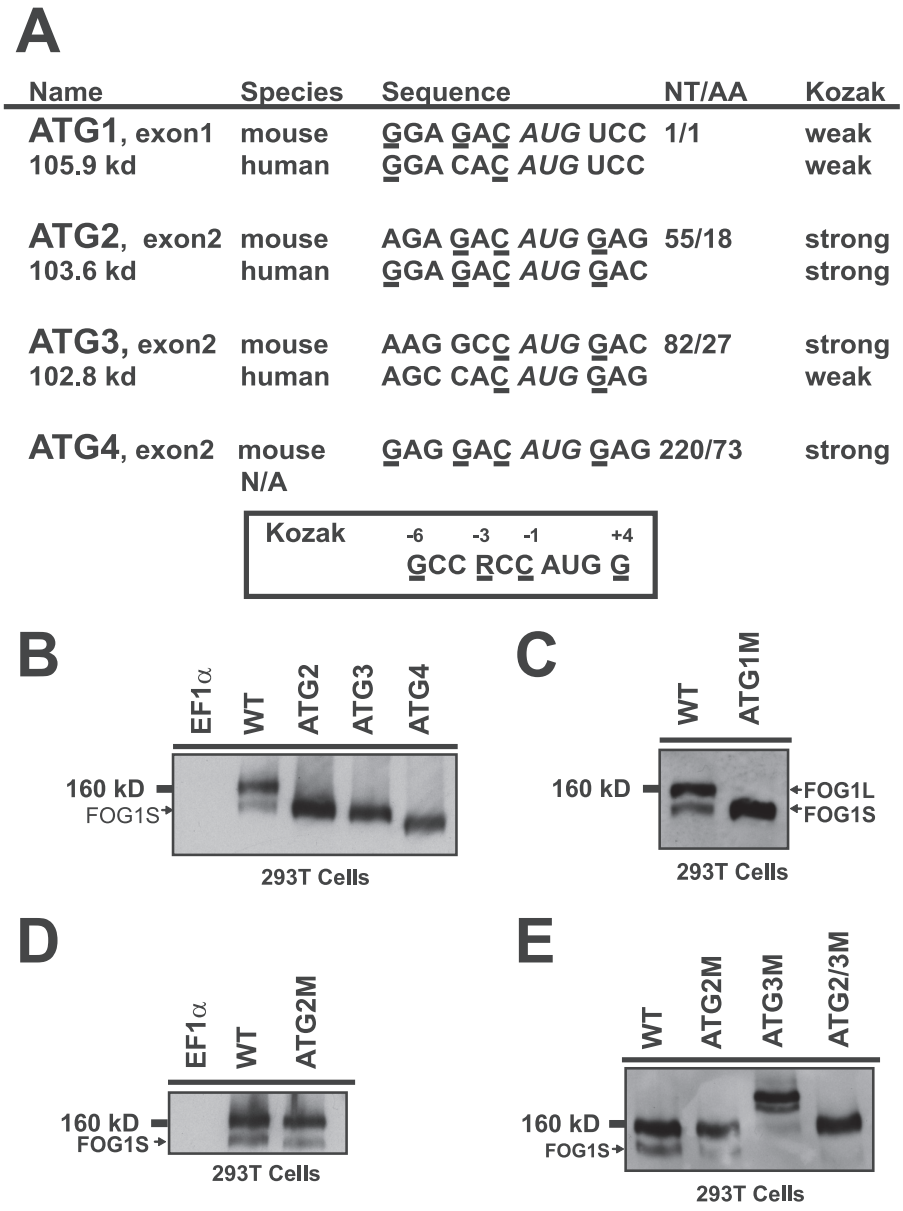

FIGURE 3. **FOG1S is translated from the second in-frame ATG.** *A*, table showing the first four in-frame ATGs from the mouse FOG1 ORF, along with Exon location, predicted molecular mass of the product, Kozak sequence and strength, and mouse and human conservation. *B*, vectors expressing the wild type FOG1 ORF (WT) or truncations of FOG1 starting from the second (ATG2), third (ATG3), or fourth (ATG4) internal ATG were transfected into 293T cells and Western blotted with α-FOG1 (A-20) antibody. *C*, vectors expressing wild type FOG1 (WT) or FOG1 mutated at the first ATG to ACG (ATG1M) were transfected into 293T cells and Western blotted with FOG1 (A-20) antibody. *D*, vectors expressing wild type FOG1 ORF (WT) or FOG1 mutated at the second ATG (ATG2M) were transfected into 293T cells and Western blotted with α-FOG1 (A-20) antibody. *E*, vectors expressing wild type FOG1 ORF (WT) or FOG1 mutated at the second ATG (ATG2M), third ATG (ATG3M), or both the second and third ATG (ATG2/3M) were transfected into 293T cells and Western blotted with α-FOG1 (A-20) antibody.

molecular mass of 105 kDa. In the absence of a linear relationship between amino acid sequence and apparent molecular mass, it is difficult to predict which downstream AUG might account for FOG1S translation. We first attempted to map the second start site through the use of Edman degradation N-terminal sequencing of FOG1S, affinity-purified following expression of a FOG1 construct containing the native 5′-UTR and start codon with a C-terminal V5 tag. However, because no useable sequence was recovered, it is likely that the N-terminal methionine is chemically blocked. Therefore, we generated N-terminal truncations corresponding to potential alternate start sites (Fig. 3*B*) and expressed each exogenously in 293T cells. Blotting of whole cell lysates with the A-20 antibody revealed that the product most closely corresponding to FOG1S was expressed from the second ATG (ATG2) and lacked only 17 amino acids from the N terminus. It is clear from the size of the protein produced from the fourth ATG (ATG4) that FOG1S must be generated from one of the previous two start codons (ATG2 or ATG3) (Fig. 3*B*). In support of this, only the first three ATGs are conserved in mouse and human (Fig. 3*A*). In addition, expression of a construct in which the canonical initiation codon was mutated to ACG, and therefore translation was presumably initiated at the next ATG (M2), produced a protein indistinguishable in size from FOG1S (Fig. 3*C*). Expression of FOG1 from a construct in which the second start codon was mutated produced FOG1L and a molecule migrating somewhat faster FOG1S, presumably starting at the next downstream ATG (Fig. 3*D*). Finally, we examined expression of FOG1 from constructs in which the second start codon, the third codon, or both were mutated to ACG (Fig. 3*E*). Again, the construct containing mutation of the second ATG produced FOG1L and a band running slightly faster than FOG1S. Unfortunately, mutation of the third ATG alone was not interpretable because mutation of this amino acid results in a shift in the apparent molecular mass. Mutation of both the second and third ATG together results in production of a faint product running visibly faster than FOG1S. This protein product is consistent in size with initiation from the fourth ATG. These data from the double mutant provide strong evidence that FOG1S is derived from the second or third ATG, and the molecular mass shift of FOG1S in the absence of the second ATG is consistent with novel initiation from this position.

Isoform Generation Is Regulated by Multiple Factors—Examination of published mRNA sequences of mouse (NM_009569) and human (NM_153813) FOG1 reveals the presence of a

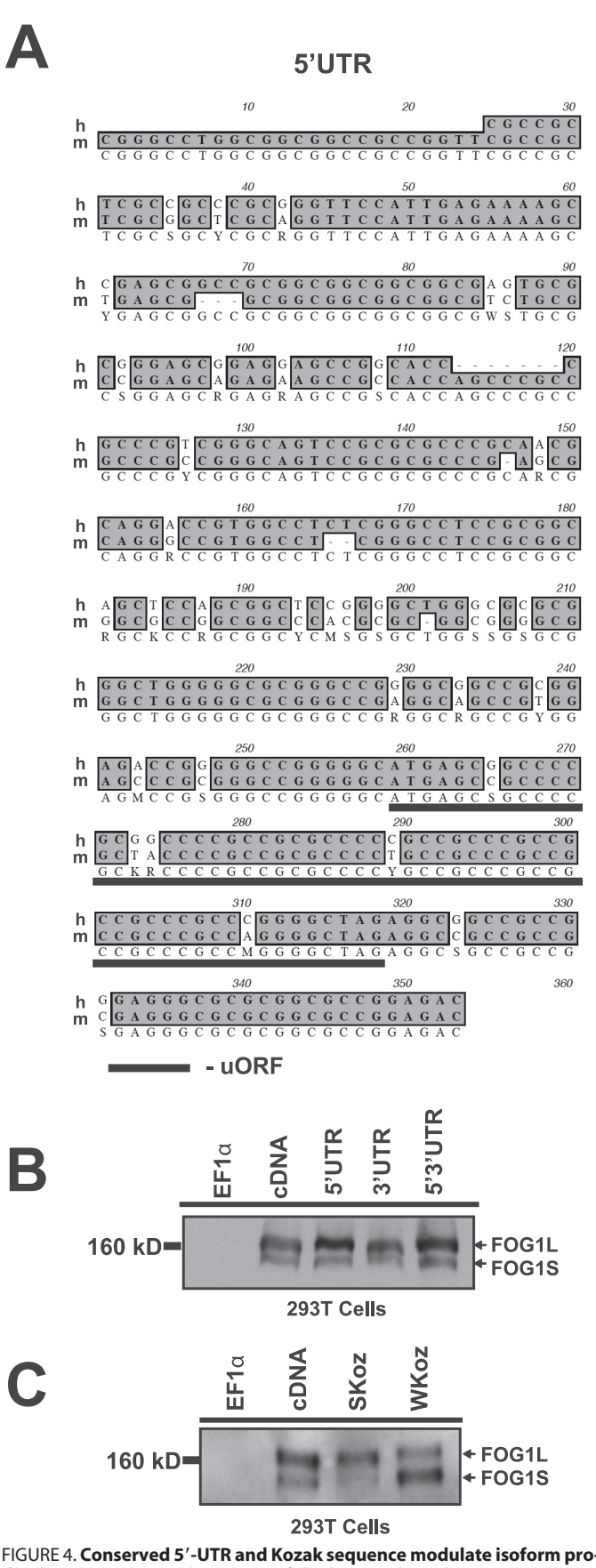

FIGURE 4. **Conserved 5′-UTR and Kozak sequence modulate isoform production.** *A*, sequence alignments of the human (*h*) and mouse (*m*) 5′-UTR from the FOG1 cDNA. The conserved upstream ORF sequence in the 5′-UTR is *underlined*. *B*, vectors expressing wild type FOG1 cDNA or FOG1 constructs in which the 5′-UTR, 3′-UTR, or both (*5′3′-UTR*) were replaced by the generic UTR found in the pEF1α vector were transfected into 293T cells and Western blotted with α-FOG1 (A-20) antibody. *C*, vectors expressing wild type FOG1 cDNA or FOG1 cDNA with the Kozak sequence mutated from GGAGACATGTCC (WT) to a stronger (GGAGACATGgCC) (SKoz) or a weaker version (tGAtACATGTCC) (WKoz) were transfected into 293T cells and Western blotted with α-FOG1 (A-20) antibody.

highly conserved 5′-UTR (Fig. 4*A*). The 5′-UTRs are relatively long (347 bp for the mouse and 323 bp for the human), have a high GC content (84.7% for the mouse and 93.7% for the human), and are predicted, using the mFOLD program (25), to form complex secondary structures in cells (data not shown). In addition, investigation of the 5′-UTR of mouse FOG1 mRNA revealed an upstream ORF that is conserved in human with a high degree of homology (Fig. 4*A*). Upstream ORFs contribute to the regulation of general translation, as well as the production of alternative translational isoforms (26). Additionally, the 3′-UTR of the murine FOG1 mRNA, although not particularly long, shares a short region of conservation in the human mRNA (data not shown). This sequence is predicted to form a secondary structure *in vivo* using the mFOLD program. Because 3′-UTRs often regulate translation through recruitment of RNA-binding proteins, we considered that the 3′-UTR might contribute to the generation of FOG1S as well.

To test potential roles of the 5′-UTR and 3′-UTR in FOG1 mRNA translation, we generated constructs containing combinations of the endogenous 5′- and 3′-UTR and the "generic" 5′-UTR and 3′-UTR from within the pEF1α expression vector. Expression of these constructs in 293T and subsequent Western blotting with antibody recognizing FOG1 (A-20) revealed that absence of the native 5′-UTR results in slight alteration of the ratio of FOG1S and FOG1L, such that increased amounts of FOG1L are generated (Fig. 4*B*). We also found that replacement of the normal 3′-UTR with a generic version did not appear to alter the ratio of FOG1L/FOG1S (Fig. 4*B*), and as expected, the construct containing generic versions of both the 5′-UTR and 3′-UTR behaved in a similar fashion to the construct with only the 5′-UTR replaced. These results indicate that the 5′-UTR contributes to the generation of FOG1S and is likely important for modulating its expression in cells. Because the canonical start codon does not conform to a strong Kozak consensus, we reasoned that this feature might be relevant to expression of FOG1S. To examine this possibility, we generated constructs in which we replaced the endogenous Kozak sequence (GGAGACATGTCC) with either a stronger (GGAGACATGgCC) or a weaker version (tGAtACATGTCC) (27) in the context of the native cDNA structure. Expression of constructs containing WT or mutant cDNA (Skoz and WKoz) in 293T cells revealed that alteration of the Kozak sequence had the predicted effect, such that the stronger Kozak resulted in reduced amount of FOG1S, whereas the weakened Kozak led to relatively more FOG1S (Fig. 4*C*). These data support a model consistent with the prevailing mechanistic understanding of translation and the generation of alternate translational isoforms in which FOG1S is pro-

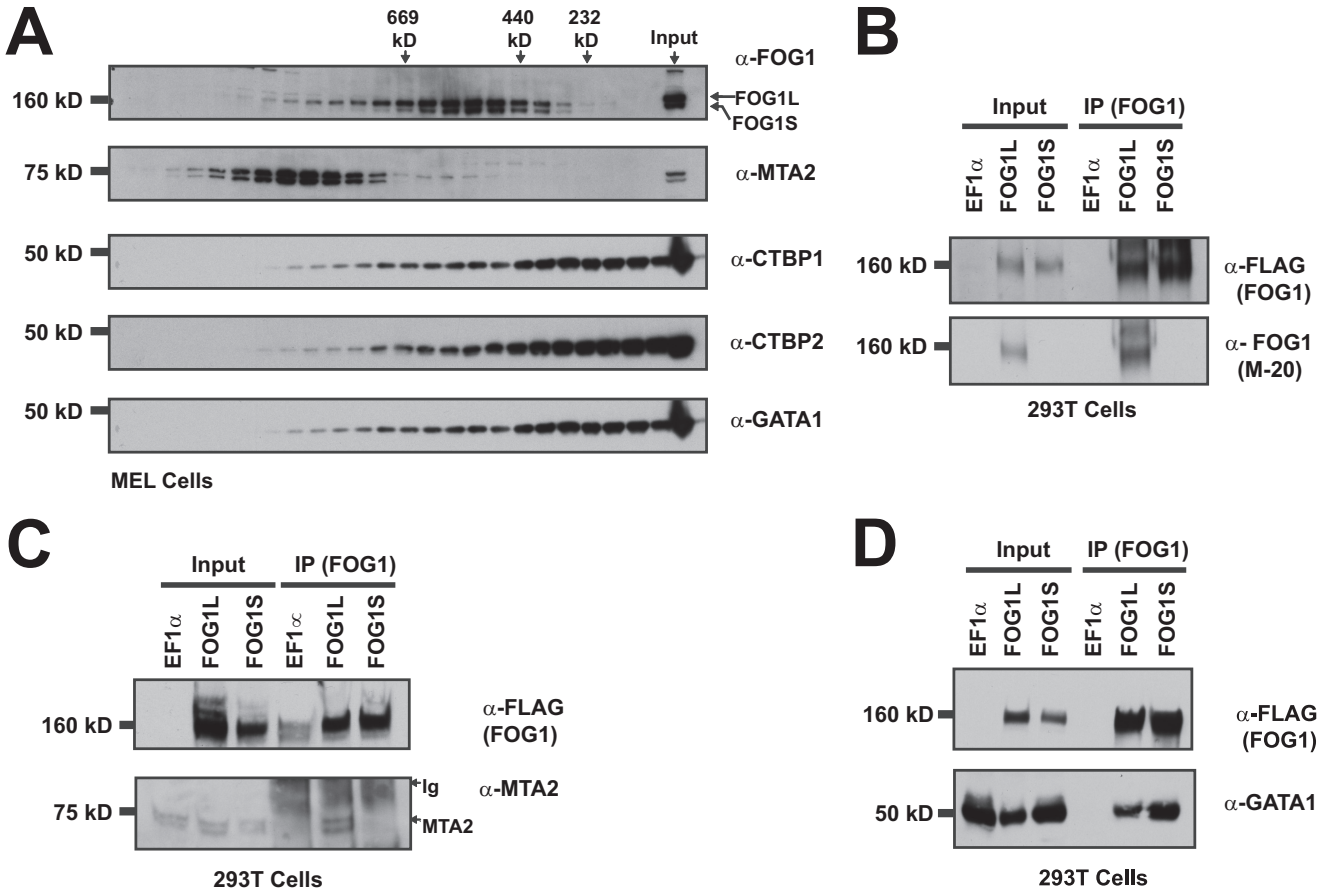

FIGURE 5. **FOG1L and FOG1S form distinct complexes with known FOG1 interaction partners.** *A*, native complexes were fractionated from MEL cell nuclear extracts using a Sephacryl S400 26/60 column (every fourth fraction of 1-ml fractions is shown on the blot). Larger to smaller molecular mass fractions are shown from *left* to *right* (fractions 48, 68, and 84 correspond to sizes of ~670, 438, and 240 kDa, respectively). *B*, constructs containing either FOG1L or FOG1S with an N-terminal FLAG-Bio tag were transfected into 293T cells. After immunoprecipitation (*IP*) with the α-FOG1 antibody recognizing the C terminus (A-20), input and immunoprecipitates were run for Western blot with the FOG1 antibody specific for the N terminus (M-20), followed by blotting with anti-FLAG (FOG1L and FOG1S). *C*, constructs containing either the tagged FOG1L or FOG1S were transfected into 293T cells. After immunoprecipitation with α-FOG1 antibody, input and immunoprecipitates were run for Western blot with antibodies directed to FLAG (FOG1L and FOG1S) and MTA2. *D*, constructs containing either the tagged FOG1L or FOG1S were cotransfected into 293T cells with a construct expressing GATA1. After immunoprecipitation with α-FOG1 antibody, input and immunoprecipitates were run for Western blot with antibodies to FLAG (FOG1L and FOG1S) and GATA1.

duced from the second ATG via the ribosome scanning mechanism.

FOG1 Isoforms Interact with Distinct Complexes—Full-length FOG1 interacts with members of the NuRD complex by coimmunoprecipitation in the MEL cells (15, 16).[3] Consistent with this, we find that FOG1L, but not FOG1S, associates with MTA2, a core component of the NuRD complex, as revealed by gel filtration (Fig. 5*A*). FOG1L and FOG1S exhibit overlapping, but distinct, size distributions. These data imply that the two isoforms of FOG1 are found in different multiprotein complexes. Interestingly, CTBP1 and 2 proteins, which are known to interact with FOG1(12–14), do not overlap with MTA2, perhaps implying temporally or spatially exclusive interaction of FOG1 with either one or the other of these two complexes. Additionally, FOG1L and FOG1S, as well as GATA1, are present in size fractions containing CTBPs. Thus, both isoforms appear to form complexes with GATA1 and CTBP family members.

To explore further the nature of these interactions, we performed coimmunoprecipitation experiments. First, FLAG-tagged versions of the FOG1 isoforms, as well as GATA1, were coexpressed in 293T cells. Upon immunoprecipitation with anti-FOG1 antibody, we recovered similar amounts of FOG1L and FOG1S protein, as revealed by FLAG antibody. These tagged versions of the isoforms only differ in sequence by 18 amino acids. For these technical reasons, we were not able to achieve adequate separation of tagged FOG1 isoforms. To demonstrate loss of the N-terminal NuRD-binding domain in the tagged FOG1S, we then blotted with the FOG1 antibody specific for the N terminus (M-20) (Fig. 5*B*). Consistent with localization of the NuRD-binding domain to the N-terminal domain of FOG1 (16, 22), we found that FOG1S failed to coimmunoprecipitate metastasis-associated 1 family, member 2, MTA2 (Fig. 5*C*) or another component of the NuRD Complex, p66 (data not shown). To visualize endogenous MTA2, we needed to overload lanes to achieve adequate MTA2 signal, leading to background in this instance. In contrast to the loss of MTA2 binding, we observed that similar amounts of GATA1 were coimmunoprecipitated by FOG1L and FOG1S (Fig. 5*C*). These results demonstrate that FOG1S does not interact with the NuRD complex, whereas it maintains the capacity to interact with GATA1.

FOG1 Interacts with LSD1 through the CTBP Complex—We previously generated a MEL cell clone stably expressing *Esche-*

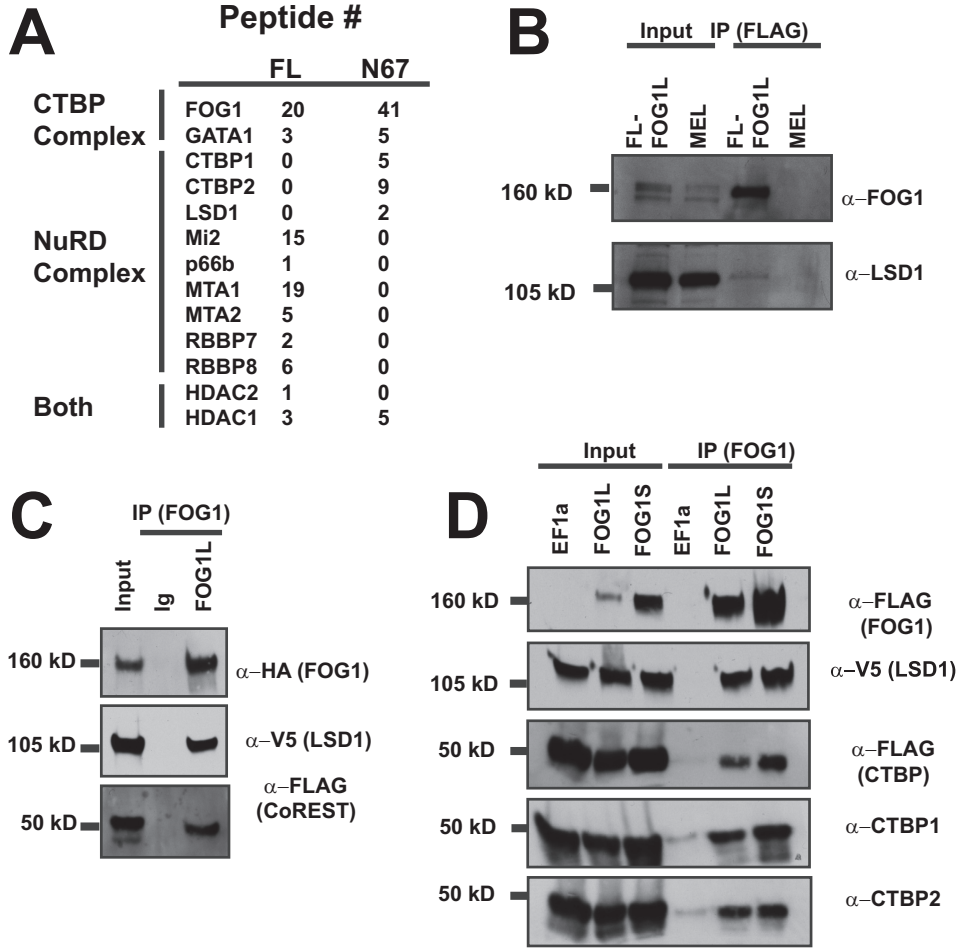

FIGURE 6. **FOG1 interacts with LSD1-containing CTBP complex.** *A*, table showing number of peptides recovered by tandem affinity purification using FLAG-Bio-tagged WT or N-truncated (N67) FOG1. *B*, coimmunoprecipitation (*IP*) of endogenous LSD1 from MEL cells expressing FLAG-Bio-Tagged FOG1 immunoprecipitated using α-FLAG-agarose. *C*, coimmunoprecipitation of LSD1 and CoREST by FOG1 using antibody directed to FOG1 or normal goat Ig, after cotransfection of 293Ts with constructs expressing HA-tagged FOG1, V5-tagged LSD1, and FLAG-tagged CoREST. *D*, constructs containing either FOG1L or FOG1S with an N-terminal FLAG-Bio tagged FOG1 were cotransfected with constructs expressing V5-tagged LSD1, FLAG-tagged CTBP1, and FLAG-tagged CTBP2. After immunoprecipitation was performed on whole cell lysates with a FOG1 antibody, input and immunoprecipitates were run for Western blot with an antibodies to FLAG (FOG1L or FOG1S, and both CTBP family members) and V5 (LSD1).

richia coli BirA biotin ligase (17). Upon coexpression of cDNA bearing the target sequence, the resulting protein will be subjected to *in vivo* biotinylation (19). We performed mass spectrometry on samples immunoprecipitated from MEL clones expressing tagged FOG1L and identified peptides from NuRD complex proteins (Fig. 5*A*) but found no evidence of CTBP family members or their interacting proteins (except HDAC1, which is common to both complexes). We used a tagged version of FOG1, N67, which lacks the first 67 amino acids (28), to interrogate FOG1 interaction partners in the absence of NuRD binding. Truncated FOG1 no longer recovers peptides from the NuRD complex and instead displays robust capture of peptides from CTBP1 and CTBP2. In addition, peptides from HDAC1 and LSD1 were recovered (Fig. 6*A*). Interaction of FOG1 with CTBP family members has been described (13, 29). However, CTBP proteins interact with molecules comprising a number of distinct complexes (30, 31). The molecular nature of the CTBP complex associated with FOG1 has not heretofore been elucidated. To confirm these mass spectrometry results, we performed immunoprecipitation on MEL cells expressing the tagged FOG1L (FOG1L) or the parental cell line (MEL) using α-FLAG antibodies. We found that FOG1 and LSD1 are coimmunoprecipitated from the tagged cells (Fig. 6*B*). Interestingly, we did not recover peptides derived from CoREST, which has previously been shown to be required for LSD1 demethylase function (32, 33). To determine whether CoREST is able to associate with FOG1, we coexpressed HA-tagged FOG1, V5-tagged LSD1, and FLAG-tagged CoREST in 293T cells (which express low levels of endogenous CTBP1 and CTBP2) and performed immunoprecipitation with α-FOG1 antibody or normal goat Ig. Again, we found that FOG1L efficiently pulls down LSD1, indicating a robust interaction with this protein (Fig. 6*C*). FOG1 is able to coimmunoprecipitate CoREST as well (Fig. 6*C*), demonstrating that FOG1 can interact with all of the necessary components for K4 histone demethylation. When we repeated the above experiment using FOG1L and FOG1S, we found that both isoforms efficiently coimmunoprecipitate LSD1 and CTBP1 and 2 (Fig. 6*D*). These findings contrast with our mass spectrometric results, in which FOG1L did not recover peptides for CTBP complex components. However, because the interaction of FOG1 with CTBP family members has been well documented (13, 29), we can only conclude that issues specific to the tandem purification coupled with mass spectrometric analysis are responsible. Mass spectrometric analysis allows for identification of proteins represented by the most abundant peptides. Our unpublished data have demonstrated a remarkably robust interaction between FOG1 and the NuRD complex, refractory to stringent wash conditions and long incubation times. The interaction with CTBP is not as robust. In support of this conclusion, when less stringent single step affinity purification protocols were used, members of the CTBP complex were recovered by FOG1L. Second, there may be some mutual exclusivity of the two interactions (see gel filtration data). Both of these factors would be expected to favor the recovery and identification of the more tightly bound complex in the context of immunoprecipitation with full-length FOG1.

Differential Nuclear Localization of FOG1L and FOG1S—To investigate potential biological differences of FOG1 isoforms,

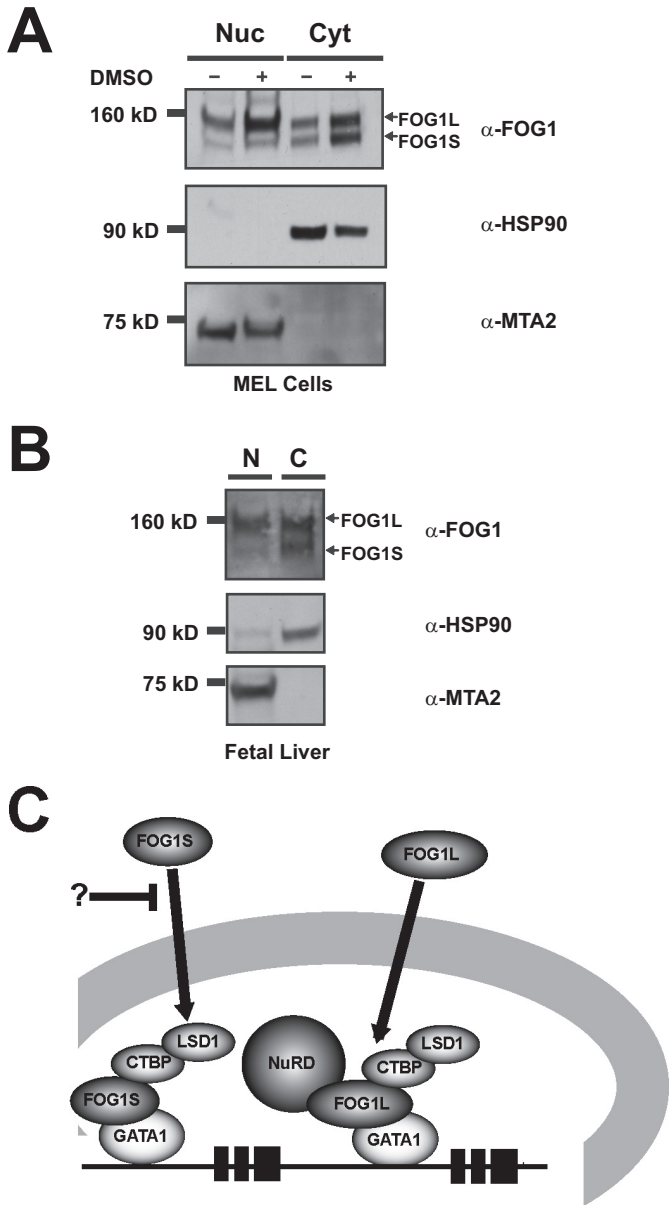

FIGURE 7. **FOG1 isoforms possess differential nuclear localization patterns.** *A* and *B*, nuclear (*Nuc*) and cytoplasmic (*Cyt*) extracts from MEL cells (*A*) and primary fetal liver cells (*B*) were prepared. Equal amounts of protein were run on SDS-PAGE gels for Western blot analysis antibody against FOG1, as well as antibodies to MTA2 and Hsp90 to demonstrate purity of cellular fractions. *C*, model of differential FOG1L and FOG1S regulation and complex formation in erythroid cells.

we examined their cellular localization in erythroid cells. Because FOG1 plays a crucial role in GATA1-dependent differentiation of erythroid cells, we first examined how differentiation might affect their relative production, reasoning that they might play differential regulatory roles during this process. Probing nuclear extracts of MEL cells that were treated with Me_2SO for 3 days with the A-20 antibody revealed that FOG1L and FOG1S appear at roughly equivalent proportions during differentiation, arguing that changes in the ratios of the two isoforms are not critical in cellular maturation (Fig. 7*A*). We observed similar results in the megakaryocyte cell line L8057 with and without 12-*O*-tetradecanoylphorbol-13-acetate-induced differentiation (data not shown).

However, a striking observation to emerge from these experiments was the differential cellular localization of FOG1L and FOG1S. In MEL cells, we observed that FOG1L predominates in the nucleus, whereas in the cytoplasmic fraction FOG1L and FOG1S are present at similar levels (Fig. 6*A*). This observation was confirmed in primary fetal erythroid cells (Fig. 6*B*). These data also suggest that selective regulation of nuclear FOG1S may be important for control of gene expression in the appropriate context.

FOG1 has been shown previously to interact with TACC3, a protein thought to be involved in regulation of microtubule dynamics during cell division (34). Interaction with TACC3 may sequester FOG1 in the cytoplasm (35). To determine whether preferential interaction with TACC3 by the short isoform could account in part for differential cellular localization, we coexpressed a construct containing a V5-tagged TACC3 with constructs containing either a FLAG-tagged FOG1L or FLAG-tagged FOG1S. Unexpectedly, immunoprecipitation with α-FOG1 antibody revealed that FOG1L bound TACC3 robustly enough to allow for coimmunoprecipitation, whereas FOG1S did not (data not shown). These data exclude TACC3 interaction as a mechanism for preferential nuclear exclusion of FOG1S and support the existence of novel modulators of FOG1 cellular localization.

DISCUSSION

FOG1, an interaction partner of GATA1 necessary for GATA1-mediated regulation of gene transcription in multiple contexts (10), recruits two repression complexes into association with GATA1: a CTBP-containing complex (12–14) and the NuRD repression complex (15, 16).[3] Although the interaction of FOG1 with repression complexes may account for repression of genes by GATA1, how FOG1 might contribute to GATA1-dependent gene activation has yet to be elucidated.

We demonstrate here the existence of a novel isoform of FOG1, FOG1S, in both erythroleukemia cells and primary fetal liver erythroblasts. Using antibody based mapping, we determined that FOG1S lacks the extreme N terminus. FOG1S can be generated along with FOG1L upon expression of full-length cDNA (10) in a heterologous cell line. These data imply that formation of FOG1S is not due to alternative RNA processing, a conclusion consistent with 5′-rapid amplification of cDNA ends experiments. Furthermore, cycloheximide experiments in MEL cells provide evidence that FOG1S is not a cleavage product of FOG1L, because the ratio of the isoforms does not change after translation is blocked. Together, these data indicate that FOG1S is created by alternate translation of the canonical transcript from a downstream, internal initiator codon, most likely through use of the next downstream ATG.

Translationally mediated generation of FOG1S appears to differ from the manner in which a similar variant of full-length FOG2 is produced (22). Production of FOG1S is regulated by the 5′-UTR and the Kozak sequence, probably through modulation of ribosome scanning (27). However, it is unclear what cellular cues, if any, may be involved in generating different ratios of the two translation products or whether maintaining a consistent ratio is biologically important.

As anticipated, FOG1S is unable to bind components of the NuRD complex. The N-terminal domain of FOG proteins bears a highly conserved NuRD-binding motif (36). The NuRD complex is most often associated with repression, although instances in which it potentiates activation have been described (37, 38). However, FOG1 (16) and FOG2 (22), mutated to abrogate NuRD binding, show decreased GATA-mediated repression in transient reporter assays. Additionally NuRD is present, along with FOG1 and GATA1, at repressed genes in erythroid cells (16), indicating that NuRD functions largely in repression of GATA1-dependent genes. Recent data indicate that mice expressing only an N-terminally truncated form of FOG1 display perturbed erythroid development, specifically demonstrating defects in repression of select GATA1-target genes (39). These data show that the ability of FOG to repress genes is not wholly explained by NuRD binding and that the FOG1-NuRD interaction is dispensable for FOG1-dependent activation of erythroid target genes. These, and our own, data lead to speculation that FOG1L may be required for regulation of specific genes in a NuRD-dependent fashion, whereas FOG1S may be involved in the regulation of a subset of genes through a NuRD-independent mechanism. This model would suggest that other FOG1 interaction partners play a role in this regulation or that FOG1 is able to mediate regulation itself in the absence of NuRD recruitment.

Previously, our lab demonstrated that mice in which the canonical CTBP-binding motif of FOG1 was mutated exhibited normal erythroid development (14). This result was unanticipated. Our recent data, however, indicate that this mutant version of FOG1 possesses residual CTBP binding activity, most likely through a noncanonical CTBP-binding motif.[5] Thus, reassessment of the involvement of CTBP in FOG1-dependent activities is warranted. Here, we have shown that, although FOG1L is able to interact with CTBP, it may bind the CTBP and NuRD complexes in a mutually exclusive fashion. Furthermore, utilizing a tagged N-terminal truncation of FOG1, we found that only in the absence of NuRD binding were peptides of the CTBP complex recovered, an observation that provides preliminary evidence of exclusive binding to these distinct repressive complexes. Additionally, LSD1 peptides are recovered with CTBP proteins by purification of an N-terminally truncated FOG1 (N67). We confirmed this interaction by immunoprecipitation with FOG1L, which elucidates the molecular identity of the CTBP complex that associates with FOG1 in erythroid cells. CTBP family members have been shown to interact with a number of molecules (30, 31) that may comprise distinct functional complexes. LSD1 is already known to play an important role in the biology of erythroid development as an interaction partner of Gf11b (40). LSD1 can act as a repressor or an activator of transcription in a context-dependent manner (41–43). To understand what functional consequences the interaction between FOG1 and LSD1 may have on erythroid gene expression, it will be important to characterize the molecular components that are brought into association with GATA1 through the interaction of FOG1 and CTBP at specific gene loci.

Translational regulation of protein expression affects other hematopoietic transcription factors, including T-cell acute lymphocytic leukemia 1 (SCL/TAL) and CCAAT/enhancer binding protein (C/EBP), α (44–46). In these cases, translationally generated isoforms exhibit biochemical and functional properties distinct from those of the canonical protein. Such appears to the case for FOG1 also, because FOG1L and FOG1S display different capacities to bind the NuRD complex and are regulated differentially with respect to their cellular localization. Translational regulation also allows for an additional level of control that is sensitive to the translational status of the cell. This status can serve as a finely tuned sensor for multiple cellular cues, such as cell cycle stage, nutrient deprivation, DNA damage, and oxidative stress (44, 47, 48). Erythroid differentiation is characterized by a number of cellular changes that impact stress pathways that alter translation, including oxidative stress associated with hemoglobin synthesis (49), cell cycle exit (50), and energetic stresses as mitochondria are removed through autophagy (51) during terminal differentiation. These extrinsic and intrinsic cues may result in differential production of the two isoforms. These data lead to conjecture that favoring translation of FOG1L or FOG1S might allow for regulation of erythroid gene expression that is carefully linked to the cell state.

The cellular localization of FOG1L and FOG1S also appears to be regulated, such that, in erythroid cells, FOG1L predominates in the nucleus, whereas the FOG1L and FOG1S are present equally in the cytoplasm. We show that TACC3, previously shown to be able to sequester FOG1 in the cytoplasm, is not responsible for this differential localization. We hypothesize that other molecules are involved in retaining FOG1S preferentially in the cytoplasm in erythroid cells. Further work is needed to test this speculation. Utilization of nuclear import and export pathways to regulate transcriptional regulators is common (52) and has been shown for at least one erythroid-specific factor, Kruppel-like factor 1 (erythroid) (KLF1) (53). Although we have been unable to determine the mechanism through which FOG1S is selectively retained in the cytoplasm, it may be that cellular stimuli are also involved in regulation of this process. In addition, it is also possible that maintaining equivalent amounts of FOG1L and FOG1S in the cytoplasm is biologically important, although no cytoplasmic role for either FOG1 isoform has been demonstrated.

Hence, we propose a model in which FOG1 isoforms are produced in a specific ratio, regulated by cues acting upon cellular translation machinery. Further control of the nuclear ratio of FOG1L and FOG1S may allow for adaptive adjustment of the relative amounts of biologically distinct GATA1-interacting proteins and hence the fine-tuning of gene expression during erythroid differentiation (Fig. 7*C*).

Acknowledgments—We thank A. Cantor for biotin tagging constructs, MEL cells expressing BirA, and advice on affinity purification and chromatography. We also thank Y. Shi for generously providing the FLAG-tagged CoREST vector. We thank R. Tomaino and S. Gygi for performing liquid chromatography-tandem mass spectrometry and providing advice and assistance in data collection and analysis and acknowledge Jason Zhang and David Zhang for excellent technical assistance. In addition, we thank Jianlong Wang, Janni Chu, Tyler Moran, Andrew Woo, Jennifer Trowbridge, Thomas Akie, and Stephan Chan for technical advice and helpful discussions and Jian Xu for critical comments in the preparation of this manuscript.

REFERENCES

1. Tsai, S. F., Martin, D. I., Zon, L. I., D'Andrea, A. D., Wong, G. G., and Orkin, S. H. (1989) *Nature* **339,** 446–451
2. Evans, T., and Felsenfeld, G. (1989) *Cell* **58,** 877–885
3. Weiss, M. J., and Orkin, S. H. (1995) *Exp. Hematol.* **23,** 99–107
4. Orkin, S. H. (1992) *Blood* **80,** 575–581
5. Pevny, L., Simon, M. C., Robertson, E., Klein, W. H., Tsai, S. F., D'Agati, V., Orkin, S. H., and Costantini, F. (1991) *Nature* **349,** 257–260
6. Fujiwara, Y., Browne, C. P., Cunniff, K., Goff, S. C., and Orkin, S. H. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93,** 12355–12358
7. Shivdasani, R. A., Fujiwara, Y., McDevitt, M. A., and Orkin, S. H. (1997) *EMBO J.* **16,** 3965–3973
8. Vyas, P., Ault, K., Jackson, C. W., Orkin, S. H., and Shivdasani, R. A. (1999) *Blood* **93,** 2867–2875
9. Vyas, P., and Crispino, J. D. (2007) *Curr. Opin. Pediatr.* **19,** 9–14
10. Tsang, A. P., Visvader, J. E., Turner, C. A., Fujiwara, Y., Yu, C., Weiss, M. J., Crossley, M., and Orkin, S. H. (1997) *Cell* **90,** 109–119
11. Tsang, A. P., Fujiwara, Y., Hom, D. B., and Orkin, S. H. (1998) *Genes Dev.* **12,** 1176–1188
12. Turner, J., and Crossley, M. (1998) *EMBO J.* **17,** 5129–5140
13. Fox, A. H., Liew, C., Holmes, M., Kowalski, K., Mackay, J., and Crossley, M. (1999) *EMBO J.* **18,** 2812–2822
14. Katz, S. G., Cantor, A. B., and Orkin, S. H. (2002) *Mol. Cell Biol.* **22,** 3121–3128
15. Rodriguez, P., Bonte, E., Krijgsveld, J., Kolodziej, K. E., Guyot, B., Heck, A. J., Vyas, P., de Boer, E., Grosveld, F., and Strouboulis, J. (2005) *EMBO J.* **24,** 2354–2366
16. Hong, W., Nakazawa, M., Chen, Y. Y., Kori, R., Vakoc, C. R., Rakowski, C., and Blobel, G. A. (2005) *EMBO J.* **24,** 2367–2378
17. Woo, A. J., Moran, T. B., Schindler, Y. L., Choe, S. K., Langer, N. B., Sullivan, M. R., Fujiwara, Y., Paw, B. H., and Cantor, A. B. (2008) *Mol. Cell Biol.* **28,** 2675–2689
18. Gobert, S., Chretien, S., Gouilleux, F., Muller, O., Pallard, C., Dusanter-Fourt, I., Groner, B., Lacombe, C., Gisselbrecht, S., and Mayeux, P. (1996) *EMBO J.* **15,** 2434–2441
19. de Boer, E., Rodriguez, P., Bonte, E., Krijgsveld, J., Katsantoni, E., Heck, A., Grosveld, F., and Strouboulis, J. (2003) *Proc. Natl. Acad. Sci. U.S.A.* **100,** 7480–7485
20. Wang, J., Rao, S., Chu, J., Shen, X., Levasseur, D. N., Theunissen, T. W., and Orkin, S. H. (2006) *Nature* **444,** 364–368
21. Zhou, M., Ouyang, W., Gong, Q., Katz, S. G., White, J. M., Orkin, S. H., and Murphy, K. M. (2001) *J. Exp. Med.* **194,** 1461–1471
22. Dale, R. M., Remo, B. F., and Svensson, E. C. (2007) *Biochem. Biophys. Res. Commun.* **357,** 683–687
23. Tan, J. S., Mohandas, N., and Conboy, J. G. (2006) *Blood* **107,** 2557–2561
24. Kozak, M. (1987) *Nucleic Acids Res.* **15,** 8125–8148
25. Zuker, M. (2003) *Nucleic Acids Res.* **31,** 3406–3415
26. Meijer, H. A., and Thomas, A. A. (2002) *Biochem. J.* **367,** 1–11
27. Kozak, M. (2005) *Gene* **361,** 13–37
28. Cantor, A. B., Katz, S. G., and Orkin, S. H. (2002) *Mol. Cell Biol.* **22,** 4268–4279
29. Holmes, M., Turner, J., Fox, A., Chisholm, O., Crossley, M., and Chong, B. (1999) *J. Biol. Chem.* **274,** 23491–23498
30. Shi, Y., Sawada, J., Sui, G., Affar el, B., Whetstine, J. R., Lan, F., Ogawa, H., Luke, M. P., Nakatani, Y., and Shi, Y. (2003) *Nature* **422,** 735–738
31. Chinnadurai, G. (2007) *Int. J. Biochem. Cell Biol.* **39,** 1593–1607
32. Shi, Y. J., Matson, C., Lan, F., Iwase, S., Baba, T., and Shi, Y. (2005) *Mol. Cell* **19,** 857–864
33. Lee, M. G., Wynder, C., Cooch, N., and Shiekhattar, R. (2005) *Nature* **437,** 432–435
34. Peset, I., and Vernos, I. (2008) *Trends Cell Biol.* **18,** 379–388
35. Garriga-Canut, M., and Orkin, S. H. (2004) *J. Biol. Chem.* **279,** 23597–23605
36. Lin, A. C., Roche, A. E., Wilk, J., and Svensson, E. C. (2004) *J. Biol. Chem.* **279,** 55017–55023
37. Denslow, S. A., and Wade, P. A. (2007) *Oncogene* **26,** 5433–5438
38. Koipally, J., Heller, E. J., Seavitt, J. R., and Georgopoulos, K. (2002) *J. Biol. Chem.* **277,** 13007–13015
39. Gregory, G. D., Wang, Y., Hong. W., Miccio, A., Bersenev, A., Yu, X., Wang, H., Choi, J., Shelat, S., Tong, W., Poncz, M., and Blobel, G. (2008) *Blood (ASH Annual Meeting Abstracts*), Blood (ASH Annual Meeting Abstracts)
40. Saleque, S., Kim, J., Rooke, H. M., and Orkin, S. H. (2007) *Mol. Cell* **27,** 562–572
41. Garcia-Bassets, I., Kwon, Y. S., Telese, F., Prefontaine, G. G., Hutt, K. R., Cheng, C. S., Ju, B. G., Ohgi, K. A., Wang, J., Escoubet-Lozach, L., Rose, D. W., Glass, C. K., Fu, X. D., and Rosenfeld, M. G. (2007) *Cell* **128,** 505–518
42. Metzger, E., Wissmann, M., Yin, N., Müller, J. M., Schneider, R., Peters, A. H., Günther, T., Buettner, R., and Schüle, R. (2005) *Nature* **437,** 436–439
43. Wang, J., Scully, K., Zhu, X., Cai, L., Zhang, J., Prefontaine, G. G., Krones, A., Ohgi, K. A., Zhu, P., Garcia-Bassets, I., Liu, F., Taylor, H., Lozach, J., Jayes, F. L., Korach, K. S., Glass, C. K., Fu, X. D., and Rosenfeld, M. G. (2007) *Nature* **446,** 882–887
44. Calkhoven, C. F., Müller, C., and Leutz, A. (2002) *Trends Mol. Med.* **8,** 577–583
45. Calkhoven, C. F., Müller, C., and Leutz, A. (2000) *Genes Dev.* **14,** 1920–1932
46. Calkhoven, C. F., Muller, C., Martin, R., Krosl, G., Pietsch, H., Hoang, T., and Leutz, A. (2003) *Genes Dev.* **17,** 959–964
47. Wek, R. C., Jiang, H. Y., and Anthony, T. G. (2006) *Biochem. Soc. Trans.* **34,** 7–11
48. Lackner, D. H., and Bähler, J. (2008) *Int. Rev. Cell Mol. Biol.* **271,** 199–251
49. Weiss, M. J., Zhou, S., Feng, L., Gell, D. A., Mackay, J. P., Shi, Y., and Gow, A. J. (2005) *Ann. N.Y. Acad. Sci.* **1054,** 103–117
50. Sankaran, V. G., Orkin, S. H., and Walkley, C. R. (2008) *Genes Dev.* **22,** 463–475
51. Sandoval, H., Thiagarajan, P., Dasgupta, S. K., Schumacher, A., Prchal, J. T., Chen, M., and Wang, J. (2008) *Nature* **454,** 232–235
52. Ziegler, E. C., and Ghosh, S. (2005) *Sci. STKE* **2005,** re6
53. Quadrini, K. J., Gruzglin, E., and Bieker, J. J. (2008) *Exp. Cell Res.* **314,** 1595–1604