




A Krüppel-like factor 1 (KLF1) Mutation Associated with Severe Congenital Dyserythropoietic Anemia Alters Its DNA-Binding Specificity

Klaudia Kulczynska,^a James J. Bieker,^b  Miroslawa Siatecka^a

^aDepartment of Genetics, Institute of Experimental Biology, University of Adam Mickiewicz, Poznan, Poland

^bDepartment of Cell, Developmental, and Regenerative Biology, Icahn School of Medicine at Mount Sinai, New York, New York, USA

ABSTRACT Krüppel-like factor 1 (KLF1/EKLF) is a transcription factor that globally activates genes involved in erythroid cell development. Various mutations are identified in the human KLF1 gene. The E325K mutation causes congenital dyserythropoietic anemia (CDA) type IV, characterized by severe anemia and non-erythroid-cell-related symptoms. The CDA mutation is in the second zinc finger of KLF1 at a position functionally involved in its interactions with DNA. The molecular parameters of how CDA-KLF1 exerts its biological effects have not been addressed. Here, using an *in vitro* selection strategy, we determined the preferred DNA-binding site for CDA-KLF1. Binding to the deduced consensus sequence is supported by *in vitro* gel shifts and by *in vivo* functional reporter gene studies. Two significant changes compared to wild-type (WT) binding are observed: G is selected as the middle nucleotide, and the 3' portion of the consensus sequence is more degenerate. As a consequence, CDA-KLF1 did not bind the WT consensus sequence. However, activation of ectopic sites is promoted. Continuous activation of WT target genes occurs if they fortuitously contain the novel CDA site nearby. Our findings provide a molecular understanding of how a single mutation in the KLF1 zinc finger exerts effects on erythroid physiology in CDA type IV.

KEYWORDS consensus binding site, Krüppel-like factor 1, KLF1, transcription factor, DNA-protein interaction, gene expression, zinc finger, *in vitro* selection, CASTing, erythropoiesis, dyserythropoietic anemia, aberrant transcription, transcription factors

Krüppel-like factor 1 (KLF1/EKLF) is an erythroid cell-specific transcription factor essential for red blood cell development (1, 2). It has a typical modular structure for transcription factors, with its transactivation domain at the N terminus and three zinc fingers (ZnFs; C2H2) generating its DNA-binding domain at the C terminus. KLF1 belongs to the KLF family of transcription factors that binds the G-rich strand of so-called CACCC-box motifs located in regulatory regions of numerous erythroid genes. KLF1 is involved in activation of globins, globin chaperones, cytoskeleton and membrane proteins, ion and water channels, iron metabolism and heme synthesis enzymes, and cell cycle regulators (3, 4). Mouse knockout studies have shown that KLF1 ablation is embryonic lethal at embryonic day 14.5 (E14.5) (5, 6). KLF1 exerts its function by interaction with chromatin-modifying and -remodeling factors, such as P/CAF, CBP/p300, and the SWI/SNF complex (7–9; reviewed in reference 10), that help to coordinate opening of chromatin structure, for example, at the β -like globin locus (11). Furthermore, KLF1 functions are regulated by various posttranslational modifications such as phosphorylation (12), sumoylation (13), acetylation (7, 14), and ubiquitination (15).

The importance of KLF1 for human erythropoiesis is supported by the hematologic diseases and disorders that arise due to mutations occurring along the whole *KLF1*

Citation Kulczynska K, Bieker JJ, Siatecka M. 2020. A Krüppel-like factor 1 (KLF1) mutation associated with severe congenital dyserythropoietic anemia alters its DNA-binding specificity. *Mol Cell Biol* 40:e00444-19. <https://doi.org/10.1128/MCB.00444-19>.

Copyright © 2020 American Society for Microbiology. All Rights Reserved.

Address correspondence to Miroslawa Siatecka, msiatecka@amu.edu.pl.

Received 17 September 2019

Returned for modification 16 October 2019

Accepted 4 December 2019

Accepted manuscript posted online 9 December 2019

Published 12 February 2020

gene. These mutations lead to a range of phenotypic pathologies from benign to severe (16, 17). Some mutations in KLF1 lead to haploinsufficiency that affects expression of certain genes (*Lu/BCAM*, *Bcl11A*, *HBD*, and *HBA2*) that are highly sensitive to KLF1 expression levels. These effects are observed in hereditary persistence of fetal hemoglobin (HPFH), the rare In(Lu) blood type, or borderline hemoglobin A2 (HbA2) (17–21).

The mutations causing the most severe phenotypes affect functionally important amino acids. Two examples of such monoallelic mutations have been described. They replace the amino acids in the zinc finger DNA-binding domain at positions that are involved in direct interactions with regulatory elements of KLF1's target genes. These mutations alter the properties of the protein such that it acquires novel dominant characteristics (reviewed in references 17 and 22). They adjust its ability to recognize and bind novel target DNA sequences and lead to serious transcriptional consequences. In both described cases, KLF1 mutations are responsible for development of severe anemia. One, found in the mouse and called Nan-KLF1, leads to hemolytic neonatal anemia with hereditary spherocytosis (23–25). The second is found in humans and leads to congenital dyserythropoietic anemia (CDA) type IV (26, 27).

The mouse mutation (E339D), although conservative, on the one hand limits recognition of the normal set of wild-type KLF1 (WT-KLF1) targets but, on the other hand, leads to acquisition of a novel ability to recognize neomorphic genes not normally activated by WT-KLF1 (25, 28, 29).

The human substitution is not conservative (E325K) although it is on the orthologous amino acid residue as the mutation of the Nan mice. This charge change on the protein-DNA interface suggests that the consequences for the organism would be much more serious than for the Nan mutation.

To date, only seven patients suffering from CDA type IV have been described (26, 27, 30–38). In general, CDAs are a heterogeneous group of rare hereditary diseases. The CDA type IV is an autosomal dominant inherited blood disorder (27) characterized by ineffective erythropoiesis and by distinct morphological anomalies in the erythroid compartment (blood and bone marrow), such as multinucleated erythroblasts, (39–41), euchromatin areas connecting the nuclear membrane, atypical cytoplasmic inclusions, and intercellular bridges (27, 32). In addition, nonerythroid phenotypes such as growth retardation and disturbance in organ development, particularly urogenital anomalies, are manifested in some of these patients (27, 30).

In the long run, we are interested in illuminating the molecular mechanism underlying CDA type IV disease and the involvement of mutated CDA-KLF1 in the phenotype. As an essential step toward this ultimate goal, we determined the consensus binding site for this mutant. For this purpose, we used cyclic amplification and selection of targets (CASTing) (42–45), which we combined with next-generation sequencing (NGS; CASTing-seq) (46). The identified binding motif was verified by gel retardation assays, where we followed complex formation with the newly determined sites. Next, reporter gene assays allowed us to verify transcriptional functionality of these sites.

Our findings reveal that CDA-KLF1 recognizes and binds sites that are mutually exclusive compared to those of the WT-KLF1 transcription factor. These altered properties have functional consequences that may help explain some of the phenotypic changes seen in the CDA patient's erythroid cells.

RESULTS

Identification of the CDA-KLF1 consensus DNA-binding sequence. To empirically determine the DNA-binding consensus site to which CDA-KLF1 binds, first we introduced the CDA mutation (E339K) to WT-KLF1, which is an equivalent of E325 position in human. This mutation is found in the second zinc finger (ZnF2) in the Y position directly involved in the interactions with DNA (47, 48) (Fig. 1A). The human and mouse KLF1 orthologues share a very high homology, with 91% identity within the zinc finger domain (Fig. 1B). Since most of the KLF1 analyses were performed on mice, we wanted to change only one variable and follow its effect by direct comparison with the WT- and Nan-KLF1 variants. The amino acid located in the Y position of ZnF2 interacts with the

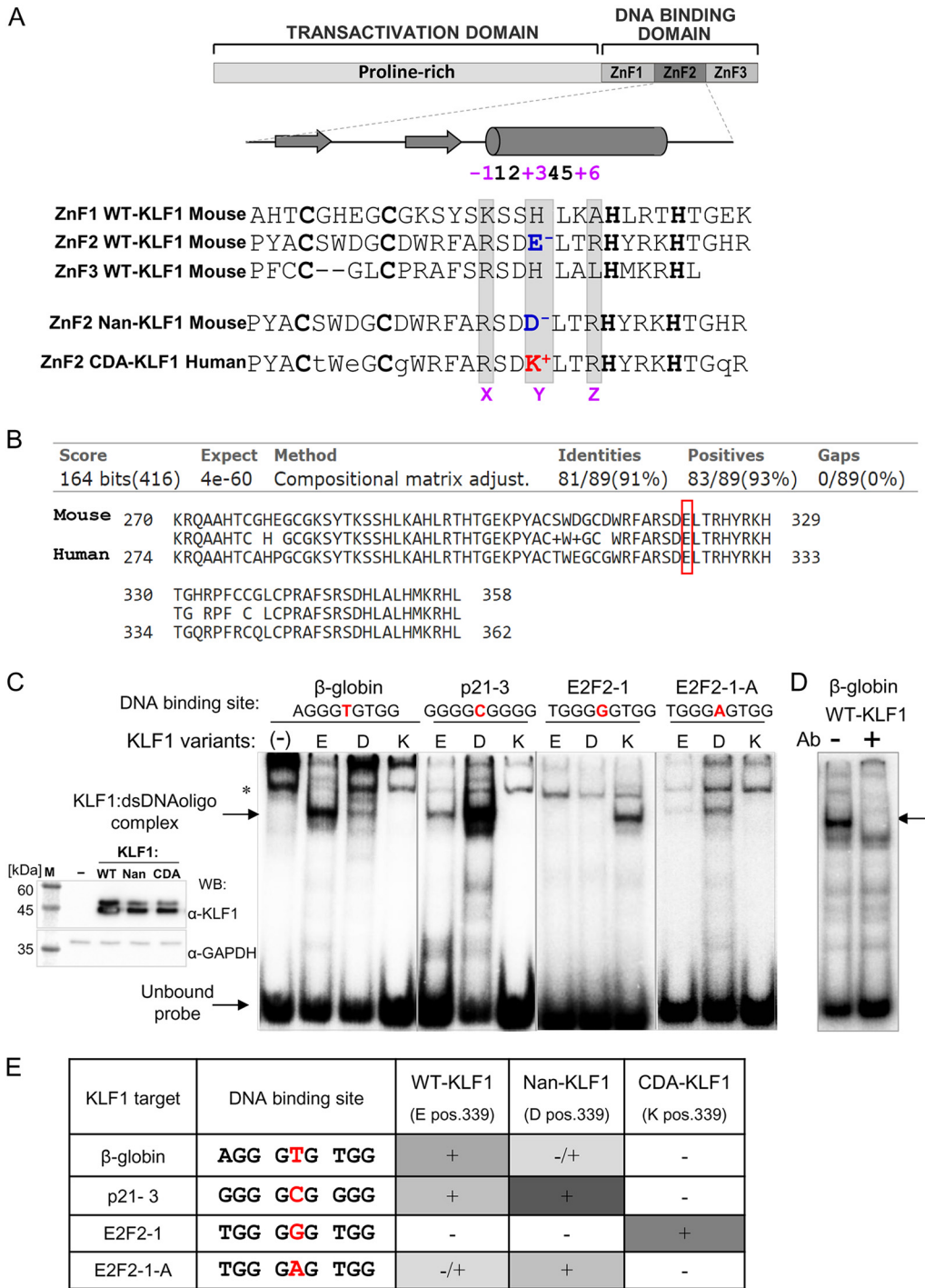


FIG 1 DNA recognition and binding specificity of CDA-KLF1 versus that of WT-KLF1 and Nan-KLF1. (A) A schematic representation of KLF1 consisting of two domains. In the secondary structure of ZnF, arrows indicate β -sheets, and the cylinder indicates the α -helix. The alignment of three ZnFs for mouse WT-KLF1, together with mouse ZnF2 of Nan-KLF1 and human ZnF2 of CDA-KLF1, is done based on Cys and His amino acids (bold) involved in zinc coordination. The amino acids denoted X, Y, and Z in positions -1, +3, and +6 of the α -helix involved in DNA interactions (47) are shaded. The crucial amino acids in position +3 of ZnF2 are in blue (negative charge) and red (positive charge). (B) Amino acid comparison of the murine and human KLF1 proteins (blastp). The sequence alignment comprises the DNA-binding domain; the top line represents the mouse KLF1 sequence, and the bottom line is the human KLF1 sequence. The box indicates conservative glutamic acid position 339 in mouse and 325 in human. (C) Complex formation (shown by arrow) between KLF1 variants (WT, E; Nan, D; CDA, K) and dsDNA oligonucleotides with 4 different nucleotides in 5th position (in red). The asterisk indicates a nonspecific band. A Western blot (inset) shows the expression level of KLF1 variants (2 μ l) used in EMSAs, with the loading control visualized by antibodies against glyceraldehyde-3-phosphate dehydrogenase (GAPDH). KLF1 migrates as a double band (79). (D) Identification of complex formed between WT-KLF1 and the dsDNA β -globin binding site by

(Continued on next page)

nucleotide in the middle (5th) position of the 9-nucleotide (nt)-long binding site (48, 49). We started our research by analyzing this relationship. We performed a gel retardation assay and followed complex formation between KLF1 variants which differ in the Y position and binding sites with the altered middle (5th) position. We examined the binding ability of WT-KLF1 with glutamic acid (E) at position 339 and mutants: CDA-KLF1 with lysine (K) and Nan-KLF1 with aspartic acid (D) (Fig. 1C). As binding sites, we used natural KLF1 targets: β -globin with T in the middle (5th) position, p21 with C, and E2F2 site 1 with G (although WT-KLF1 binds only the E2F2-2 and E2F2-3 sites of E2F2 [50]). For an A residue in the middle (5th) position, we prepared an artificial E2F2-1A site. To be sure where the KLF1–double-stranded DNA (dsDNA) oligonucleotide complex migrates, we added antibodies against the KLF1 DNA-binding domain epitope, which prevents complex formation (Fig. 1D).

The results were very interesting (Fig. 1C and E) as they showed that the CDA-KLF1 mutant was able to bind only the one motif containing G in the middle (5th) position. At the same time, two other variants, WT-KLF1 and Nan-KLF1, were unable to bind such a motif, demonstrating the mutually exclusive preferences for binding of the tested proteins. After obtaining these results, we decided to define the binding site for CDA-KLF1 more precisely.

We performed an *in vitro* CASTing-seq assay, which was combined with high-throughput sequencing to characterize the high-resolution DNA-binding specificity of the KLF1 mutant (Fig. 2A). In CASTing, we used purified recombinant proteins comprising the DNA-binding domain (DBD) of KLF1, which consists of three zinc fingers of the C2H2 type (ZnF-KLF1). The transactivation domain contains numerous prolines that make the structure unstable and prone to degradation, rendering it impossible to use the full-length protein in this assay. The DBD domain had His and Flag tags fused at the N terminus, enabling its purification by His tag affinity followed by removal of the His tag. The resultant recombinant protein was bound via its Flag tag to magnetic beads coated with anti-Flag M2 monoclonal antibody and used in a CASTing assay. In parallel to CDA-ZnF-KLF1, we used wild-type DBD (WT-ZnF-KLF1) as a positive control, for which the consensus binding site is already known (4). The magnetic beads alone without any protein attached served as a negative control. At first, we used the library that contained 12 random nucleotides flanked by constant sequences needed for amplification. The magnetic properties of the beads allowed for a very rapid separation of the beads coated by protein with bound DNA oligonucleotides from a suspension containing unbound oligonucleotides. The remaining DNA oligonucleotides that interacted with the particular ZnF domains were PCR amplified, purified, and used for the next round of the assay. We performed five such cycles.

The DNA oligonucleotides enriched after the fifth cycle were sequenced by NGS Ion Torrent. The results showed that CDA-ZnF-KLF1 recognizes a very G-rich sequence, containing guanine in the middle (5th) position of the 9-nt binding motif (data not shown). This nucleotide is directly involved in interactions with the amino acid in position 339 of the second zinc finger of KLF1 (25, 51). In the wild-type scenario, WT-KLF1, with glutamic acid (E) in position 339, recognizes a 5'-NGG-G(**C/T**)G-(T/G)GG-3' consensus site, that is, a T or C in the middle position (in boldface). Thus, oligonucleotides containing G in the 5th position were not expected to form a complex with WT-KLF1 (25). Our results from the CASTing-seq assay suggested that the CDA mutation (E339K) alters the binding specificity of KLF1. This was in agreement with our initial observation that CDA-KLF1 recognizes G in the middle position of the binding site (Fig. 1C). However, because of the long stretches of G repeats found in the CASTing-selected CDA-ZnF-KLF1 binding motifs and due to Ion Torrent system limita-

FIG 1 Legend (Continued)

addition of antibodies (Ab) against KLF1 that prevents complex formation. (E) Summary of the interactions between KLF1 variants and the dsDNA oligonucleotides based on EMSA (C). All possible nucleotides in the 5th position (in red) were tested. The shade of gray represents the intensity of the generated complex.

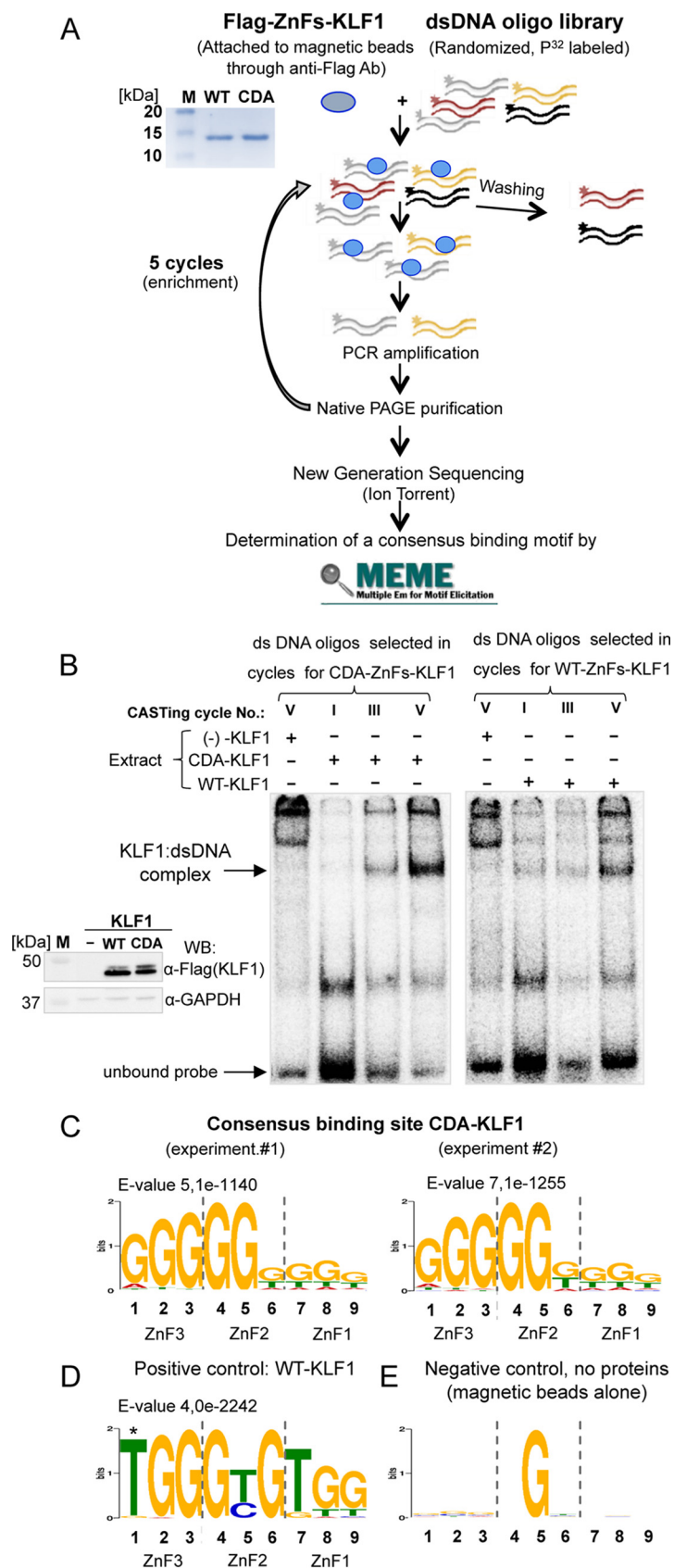


FIG 2 Determination of a consensus binding sites for the CDA variant of KLF1 using the CASTing-seq technique and MEME motif-based sequence analysis tools. (A) The scheme of the CASTing-seq technique is

(Continued on next page)

tions related to sequencing of homopolymer repeats, it was difficult to unambiguously determine the CDA-KLF1 consensus binding motif.

In order to overcome this problem, we repeated the CASTing assay, but this time we applied a more accurate library. In this approach, flanking sequences surround 9-nt-long DNA oligonucleotides (5'-NNN-NGN-NNN-3'), which are precisely the length of the binding site for three zinc fingers. In addition, we forced a G in the middle position (underlined), leaving the rest of the residues random, and the entire CASTing-seq experiment was performed once again. The gradual enrichment of selected DNA oligonucleotides from the library in cycles I, III, and V was monitored by electrophoretic mobility shift assay (EMSA). The EMSA complexes were formed with full-length CDA-KLF1 or WT-KLF1 expressed in extracts of transfected COS-7 cells (Fig. 2B), which therefore verified that the CASTing oligonucleotides selected by the zinc finger domain interact with the full-length proteins.

The DNA oligonucleotides obtained after the final round were sequenced by Ion Torrent. The resultant ~30,000 sequences were analyzed on a Galaxy platform, and then the MEME suite was applied. The final results are summarized in Fig. 2C to E. The new consensus binding site for CDA-KLF1 was determined to be 5'-NGG-**GG**(T/G)-(T/G)(T/G)(T/G)-3' (middle G in boldface). The data for CDA-ZnF-KLF1 were obtained in two biological repeats (Fig. 2C). The reliability of the results is based on the positive-control data (WT-ZnF-KLF1) (Fig. 2D), which are in agreement with published consensus site 5'-NGG-G(**C/T**)G-(T/G)GG-3' (middle position in boldface) (4, 25, 28). The thymine in the first position marked with a star in Fig. 2D comes from flanking sequence of the library. By way of explanation, the second library of the random oligonucleotides with G fixed in the middle (5th) position of the motif was designed especially for CDA-ZnF-KLF1. Thus, in order to be bound by WT-ZnF-KLF1, a shift has to occur to have C or T in the middle position. In such a situation, the binding motif had to expand to the flanking sequence. Indeed, the consensus sequence motif for the WT zinc finger domain contains C or T in the 5th position, as expected, and a flanking T in the 1st position. Also, T is preferably selected over G in the 7th position. The negative control (magnetic beads alone) did not select any sequence motif; it remained random (Fig. 2E).

In conclusion, we have newly identified the CDA-KLF1 binding motif. It is G rich, it has G in the middle (5th) position crucial for interactions with ZnF2 (position 339 of KLF1), and it shows rather high degeneracy at the 3' end of the motif. The degeneracy refers to positions 6, 7, 8, and 9 on the G-rich strand that are mostly involved in interactions with ZnF1 (Fig. 2C).

Next, we validated the *in vitro* selection results by EMSAs, functional reporter assays in K562 cells, and *in vivo* analyses using a cell line-based KLF1 variant overexpression assay. Several important issues were analyzed in this set of experiments, including the following: (i) confirmation of the specificity of KLF1 variants based on the 5th (middle) position of the binding site, (ii) testing of individual oligonucleotide sequences with 3' degeneracy and their effect on interactions with ZnF domains, (iii) testing of the specificity of the interactions between KLF1 variants and particular oligonucleotides, and (iv) *in vivo* analysis of the expression levels of selected KLF1 targets upon induction of CDA- or WT-KLF1.

FIG 2 Legend (Continued)

shown, from protein/library incubation to cycling and determination of the consensus binding motif. The inset shows Coomassie-stained purified recombinant proteins that were the starting material and comprise the three zinc finger domains of WT-KLF1 and CDA-KLF1. Lane M, molecular mass marker. (B) Enrichment of oligonucleotide pools selected in CASTing cycles I, III, and V for ZnF-KLF1 variants. Gel shift assay analyses show increased efficiency in complex formation (shown by arrow) between radiolabeled dsDNA oligonucleotides from cycle numbers I, III, and V of the CASTing experiment and full-length CDA-KLF1 (left panel) and WT-KLF1 (right panel) in extracts from transfected COS-7 cells. The inset shows a Western blot of extracts used for the gel shift analysis. (C) A new consensus DNA-binding motif for CDA-ZnF-KLF1 obtained in two biological repeats of CASTing-seq experiment and MEME motif-based sequence analysis. WB, Western blotting; GAPDH, glyceraldehyde-3-phosphate dehydrogenase. (D) The DNA-binding motif obtained for WT-ZnF-KLF1 in the CASTing-seq experiment that serves as a positive control. The asterisk marks a thymidine that comes from the flanking sequence of the random core of the oligonucleotide library. (E) Lack of selection of any DNA-binding motif for the magnetic beads alone that serves as a negative control in the CASTing-seq experiment.

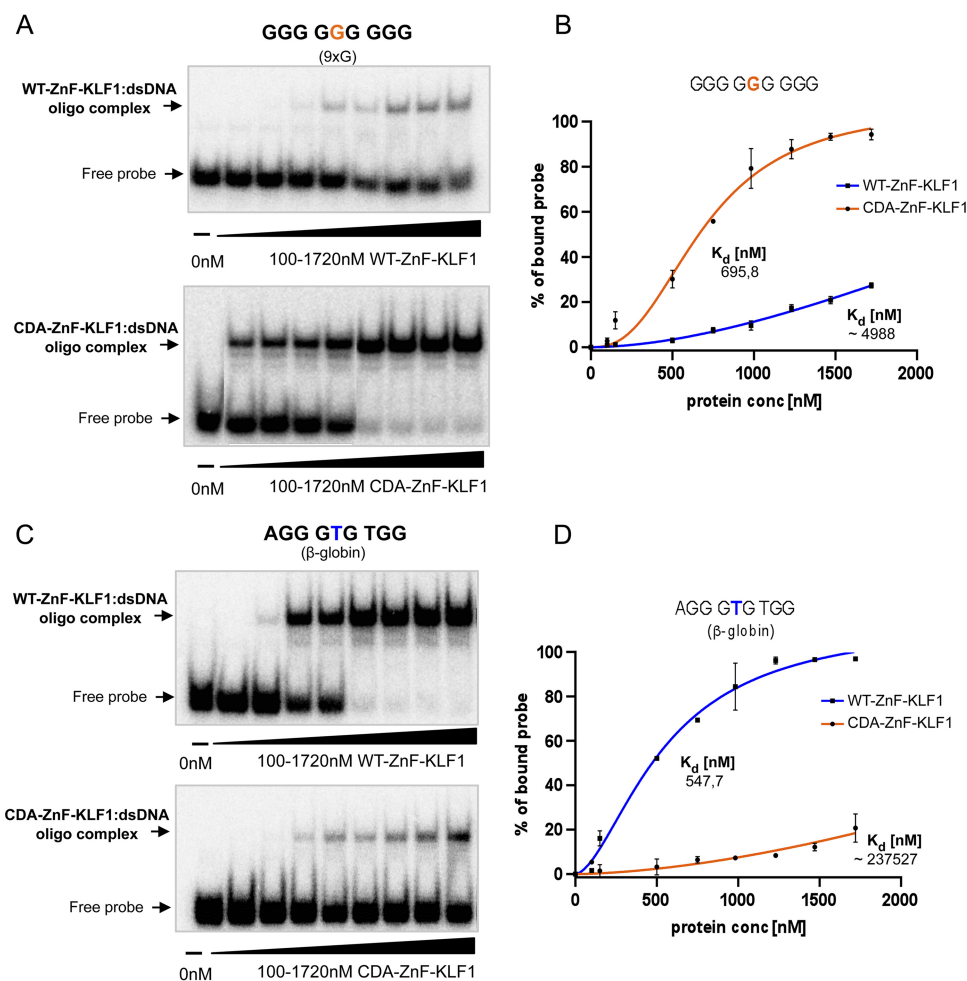


FIG 3 Comparison of the binding affinities of the zinc finger (ZnF) domain of WT-KLF1 and CDA-KLF1 toward a newly identified binding site with G in the middle (5th) position (9×G) and β -globin binding site. (A) Complex formation between an increased amount (0 to 1,720 nM) of purified recombinant ZnF domain (WT or CDA) and the radiolabeled dsDNA 5'-GGG-GGG-GGG-3' binding site newly identified for CDA-KLF1. (B) K_d binding curves for the 9×G dsDNA oligonucleotide and KLF1 variants as indicated, obtained using the Hill slope in GraphPad Prism. Average K_d values were calculated based on at least two EMSA gels. (C) Complex formation between an increased amount (0 to 1,720 nM) of purified recombinant ZnF domain (WT or CDA) and the radiolabeled dsDNA 5'-AGG-GTG-TGG-3' binding site for β -globin. (D) K_d binding curves for β -globin dsDNA oligonucleotide and KLF1 variants, as indicated, obtained using the Hill slope in GraphPad Prism. Average K_d values were calculated based on at least two EMSA gels.

Analysis of the essential G in the middle (5th) position of the binding site for CDA-KLF1. We started validating the *in vitro* consensus binding site for the KLF1 CDA mutant by performing a quantitative EMSA and a dissociation constant (K_d) measurement. The complexes were generated with an increased amount of purified, recombinant zinc finger domain for WT- or CDA-KLF1 (Fig. 3). We monitored the formation of complexes for one of the newly determined 5'-GGG-GGG-GGG-3' sites and for the β -globin binding site 5'-AGG-GTG-TGG-3' as a positive control for WT-KLF1. We measured the dissociation constant for all the possible combinations of complexes (Fig. 3).

Under applied conditions, the binding affinity of WT-ZnF-KLF1 to the β -globin site has a K_d of 547.7 ± 76.4 nM, and it is approximately 10-fold lower than that for the binding site with G in the middle position (estimated K_d of ~4,988 nM). The situation was reversed for CDA-ZnF-KLF1, which binds 5'-GGG-GGG-GGG-3' with an affinity (K_d) of 695.8 ± 71.1 nM, while the dissociation constant for the β -globin binding site increased dramatically to a K_d of approximately ~237,527 nM (Fig. 3). These results clearly indicate preferences in the specificity of KLF1 variants based on the middle (5th) position of the binding site.

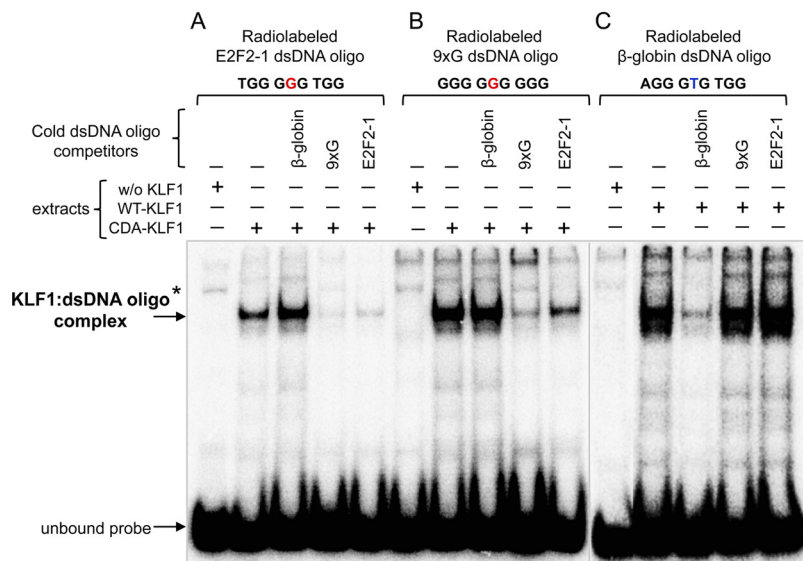


FIG 4 CDA-KLF1 binding specificity of the newly determined consensus binding site. Gel shift assay monitoring complex formation (indicated by the arrow) between full-length CDA-KLF1 (A and B) with the radiolabeled dsDNA oligonucleotides E2F2-1 (5'-TGG-GGG-TGG-3') and 9×G (5'-GGG-GGG-GGG-3'), respectively, and WT-KLF1 (C) with the radiolabeled dsDNA β-globin (5'-AGG-GTG-TGG-3') binding site. Complexes formed by both KLF1 variants were subjected to a binding competition with a 100-fold molar excess of the indicated unlabeled (cold) dsDNA oligonucleotides. The asterisk indicates a nonspecific band. The protein extracts used in these EMSAs are the same as visualized on the Western blot shown in Fig. 1C.

Next, we performed a cold competition assay which additionally supported the specificity of CDA-KLF1 binding to a new site with G in the middle position (Fig. 4). For this purpose, we used full-length WT- or CDA-KLF1 protein from transfected COS-7 extracts and performed EMSAs. Complexes between CDA-KLF1 and a radiolabeled dsDNA oligonucleotide with G in the middle (5th) position were effectively antagonized by cold competition with binding sites that contained middle G, marked as E2F2-1 and 9×G in (Fig. 4A and B). In the case of WT-KLF1 complexes with a radiolabeled β-globin binding site, they were replaced only with cold β-globin sites. Two other binding sites with G in the middle (5th) position, i.e., E2F2-1 and 9×G, did not affect the WT-KLF1 complexes (Fig. 4C).

Both experiments are self-consistent and reliably indicate that CDA-KLF1 and WT-KLF1 have mutually exclusive binding sites.

Validation of the CASTing-seq results focusing on 3' end degeneracy of the consensus binding site. Several individual CASTing-based sequences were tested in parallel for their specificity in complex formation with CDA (E339K mutation) and with WT (E339) zinc fingers (Fig. 5). We focused on testing of the 3' degeneracy of the consensus binding site and analyzed oligonucleotides differing in the three nucleotides (positions 7, 8, and 9 on the G-rich strand) that are recognized and bound by zinc finger 1. Based on EMSA results, we distinguished two major groups of binding sites depending on their ability for complex formation. First, the largest group consists of sites that are bound by CDA-ZnF-KLF1 and not bound by WT-ZnF-KLF1 (Fig. 5A); a second group consists of sites that are not bound by either CDA- or WT-ZnF-KLF1 (Fig. 5B). As a positive control for WT ZnF-KLF1, we used the β-globin binding site, which has T in the middle (5th) position, which at the same time is not a proper site for CDA-ZnF-KLF1 (Fig. 5C). The results were subjected to densitometric analysis to estimate the percentage of involvement of the tested oligonucleotides in interactions with the CDA or the WT ZnF recombinant domain.

We find that oligonucleotides having at least one G or T in the 3' triple nucleotide are efficiently bound by CDA-ZnF-KLF1, whereas complexes generated with oligonu-

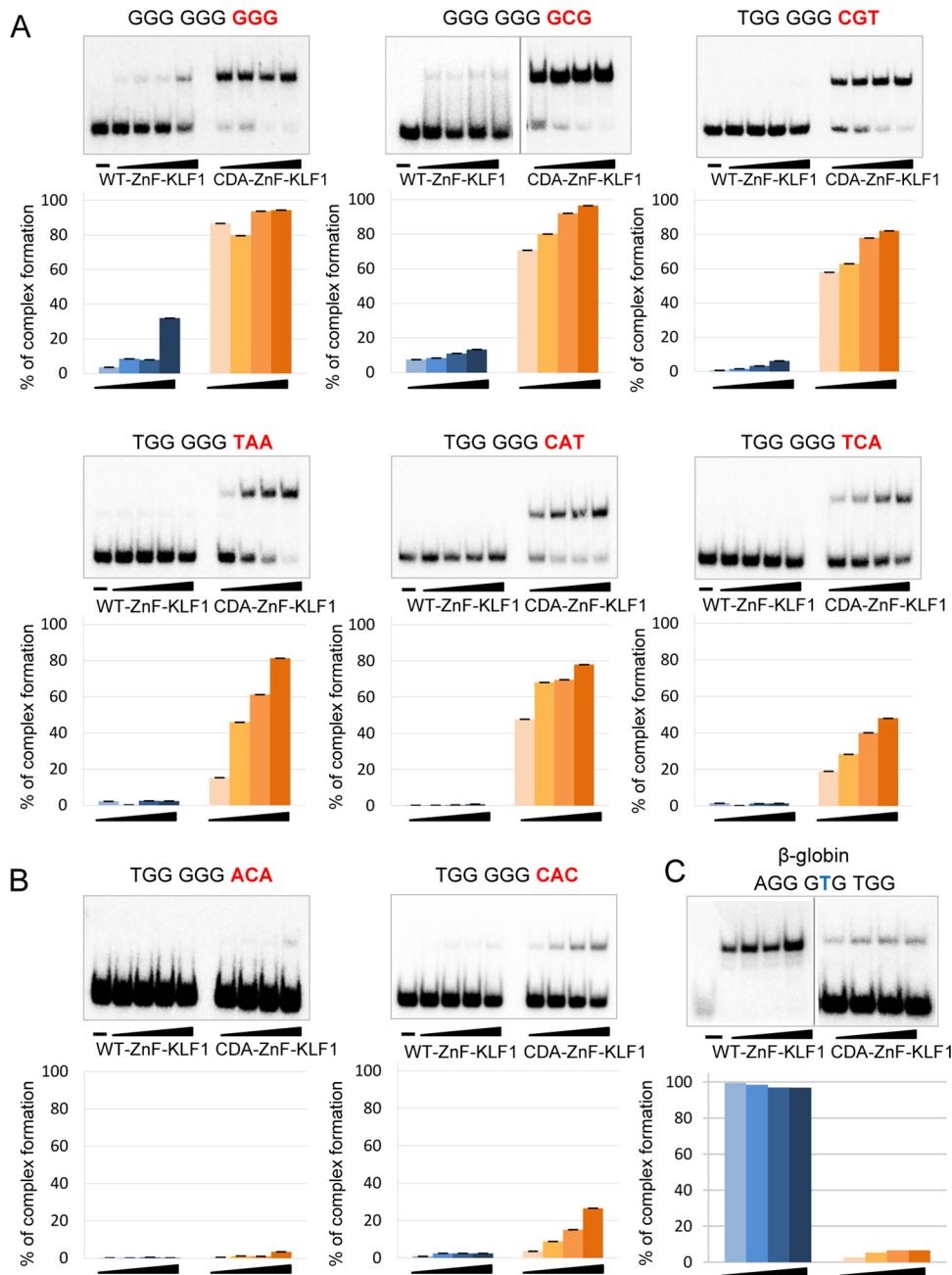


FIG 5 Comparison of the efficiencies of complex formation between individual sequences derived from the consensus binding site for CDA-KLF1 and the ZnF domains of WT-KLF1 and CDA-KLF1. Gel shift assays (upper panels) were performed with increasing amounts (0.12, 0.15, 0.18, and 0.21 μ g) of purified recombinant ZnF domain (WT or CDA) and radiolabeled dsDNA oligonucleotides comprising binding sites for CDA-KLF1. Analyzed DNA sequences are denoted above each individual gel. Densitometric analysis (lower panels) of the gel shift results show the percentage of dsDNA oligonucleotide that is involved in complex formation with WT-ZnF-KLF1 and CDA-ZnF-KLF1. The graphs are made based on a single representative EMSA gel, but gel shift analyses were performed at least twice for each tested binding site. (A) The DNA-binding sites that generate efficient complexes with CDA-ZnF-KLF1 compared to those of WT-ZnF-KLF1. (B) The DNA-binding sites that generate very weak complexes with CDA-ZnF-KLF1 comparable to those of WT-ZnF-KLF1. (C) The DNA-binding site for β -globin that generates the complex with the WT-ZnF-KLF1 domain served as a positive control.

cleotides containing only C or A in positions 7 to 9 are very weak, with complex formation estimated to be below 20% (Fig. 5B).

Analysis of the specificity of complex formation between full-length KLF1 variants (WT, CDA, and Nan) and radiolabeled oligonucleotides. As we could use

only the DNA-binding domain of KLF1 for the CASTing experiment, subsequently we tested whether full-length CDA-KLF1 is able to interact with the determined consensus binding sites. For this purpose, we chose the same individual binding sites as previously tested for the 3' degeneracy (Fig. 5) and performed EMSAs using extracts from transfected COS-7 cells expressing equivalent amounts of full-length constructs (Fig. 6).

Here, in parallel, we tested the specificity of recognition of these oligonucleotides by different KLF1 variants that vary in the amino acid at position 339: WT-KLF1 with E339, Nan-KLF1 with D339, and CDA-KLF1 with K339.

The oligonucleotides that have G in the middle (5th) position and the 3' triple nucleotide consisting of at least one T or G were bound by CDA-KLF1 (Fig. 6A), similar to results for the zinc finger domain alone (compare Fig. 6A and 5A). However, the oligonucleotides that have G in the middle (5th) position of the motif but do not have any G or T in the last triple nucleotide were not bound by CDA-KLF1 (compare Fig. 6A and 5B). Neither the WT-KLF1 nor Nan-KLF1 full-length protein is able to interact with binding sites containing G in the middle position on the G-rich strand in any case, which is consistent with our previous observations (25). As a positive control for these analyses, we again used the β -globin binding site, which was recognized and bound by WT-KLF1 and bound slightly by Nan-KLF1 (Fig. 6B).

We conclude that the novel binding properties determined by *in vitro* selection with the CDA-KLF1 DBD alone are retained when the full-length CDA-KLF1 protein is used.

Next, we focused on the degenerate 6th position of the novel CDA-KLF1 consensus binding site (Fig. 2C). We applied gel retardation testing and monitored complex formation between KLF1 variants and radiolabeled oligonucleotides that differ in this 6th position. We analyzed all four combinations of nucleotides, and the data revealed that only G or (more weakly) T in position 6 creates a suitable binding motif for CDA-KLF1 (Fig. 6C). As expected, WT-KLF1 or Nan-KLF1 did not bind to any of these as they retain G in the 5th position, which prevents the interactions.

These results clearly demonstrate that the newly determined CDA-KLF1 consensus binding site, 5'-NGG-GG(T/G)-(T/G)(T/G)(T/G)-3', crucially prefers G in the middle (5th) position (underlined) that directly interacts with the mutated E339K residue in the CDA-KLF1 protein. In addition, this single amino acid change yields a protein with an unexpectedly high level of binding degeneracy for the sequence that interacts with zinc finger 1 of CDA-KLF1 (positions 7, 8, and 9 on the G-rich strand), as well as in position 6. The last three positions at the 3' end of the binding site may tolerate C or A if there is at least one G or T in this triplet (Fig. 6A). This differs from the WT-KLF1 preference, which exhibits a strong discriminatory ability toward potential binding sites and does not allow for 3' end degeneracy (Fig. 6E).

Functional analysis of individual CDA-KLF1 binding sites by luciferase reporter gene assays in K562 cells. To test if the newly identified consensus site for CDA-KLF1 is functional, we used a luciferase reporter gene system that provides a robust and specific readout of KLF1 activity (52). We constructed a set of luciferase reporters where, upstream of the luciferase coding sequence with its minimal TK (thymidine kinase) promoter, we placed six copies of repeated individual 9-nt-long binding sites selected from CASTing (Fig. 7A, top). We used them for cotransfection of K562 cells together with particular KLF1 variants (WT, Nan, or CDA). In our K562 cells, the endogenous KLF1 is not expressed (53); thus, we can easily investigate the individual activity of the KLF1 variants by measuring the activation of luciferase and monitoring the relative light units (RLUs).

Our positive control, which is a reporter containing six copies (6 \times) of the β -globin binding site 5'-AGG-GTG-TGG-3', is solely activated by WT-KLF1 but not by Nan- or CDA-KLF1, as expected (Fig. 7A). The negative control, which consists of the same six copies of the β -globin site but with a thalassemia point mutation 5'-AGG-ATG-TGG-3' (underlined) (54), is not activated by any KLF1 construct (Fig. 7A). These data demonstrate that our assay system is robust and selective.

For the functional analyses, we used the same binding site sequences as for the *in vitro* EMSAs described above (Fig. 5 and 6). To make the results more convincing, we

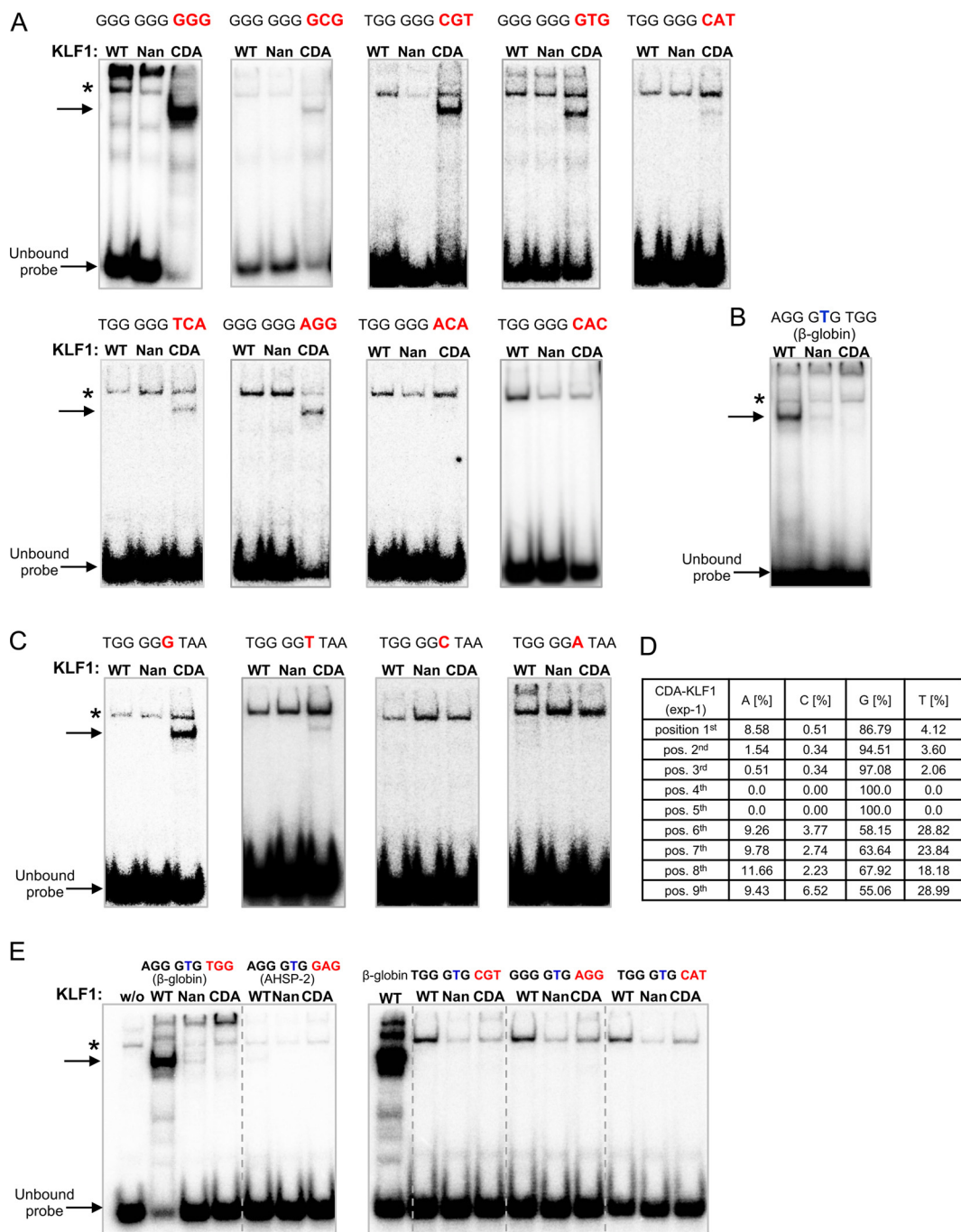


FIG 6 Analysis of the new consensus DNA-binding site for CDA-KLF1 and investigation of the binding specificity by KLF1 variants (CDA versus WT and Nan) with an emphasis on the analysis of the degenerate 3' end of CDA-KLF1 consensus binding site. (A) Analysis of variations in positions 7, 8, and 9 within the CDA-KLF1 binding site that interact with zinc finger 1 of KLF1. Gel shift assays compare complex formation (indicated by arrows) between full-length WT-KLF1, Nan-KLF1, and CDA-KLF1 from extracts of transfected COS-7 cells and radiolabeled dsDNA oligonucleotide binding sites, as indicated. The asterisk indicates a nonspecific band. (B) Complex formation with the β-globin binding site served as a positive control for WT-KLF1. (C) Comparable analysis of the degenerate position 6 of the consensus binding site for CDA-KLF1. Four binding sites that differ in position 6 were subjected to complex formation with KLF1 variants, indicated by the arrow. The asterisk indicates a nonspecific band. The protein extracts used in these EMSAs are the same as visualized on the Western blot shown in Fig. 1C. (D) Table generated by the MEME suite with the percent probability of occurrence of a possible nucleotide in each position of the motif for CDA-KLF1. (E) Analysis of variations in positions 7, 8, and 9 within the WT-KLF1 binding site that interact with zinc finger 1 of KLF1. Shown are gel shift assays comparing complex formation (indicated by arrow) between full-length WT-KLF1, Nan-KLF1, and CDA-KLF1 from extracts of transfected COS-7 cells and radiolabeled dsDNA oligonucleotide binding sites, as indicated. The AHSP-2 site is from Keys et al. (80). The asterisk indicates a nonspecific band.

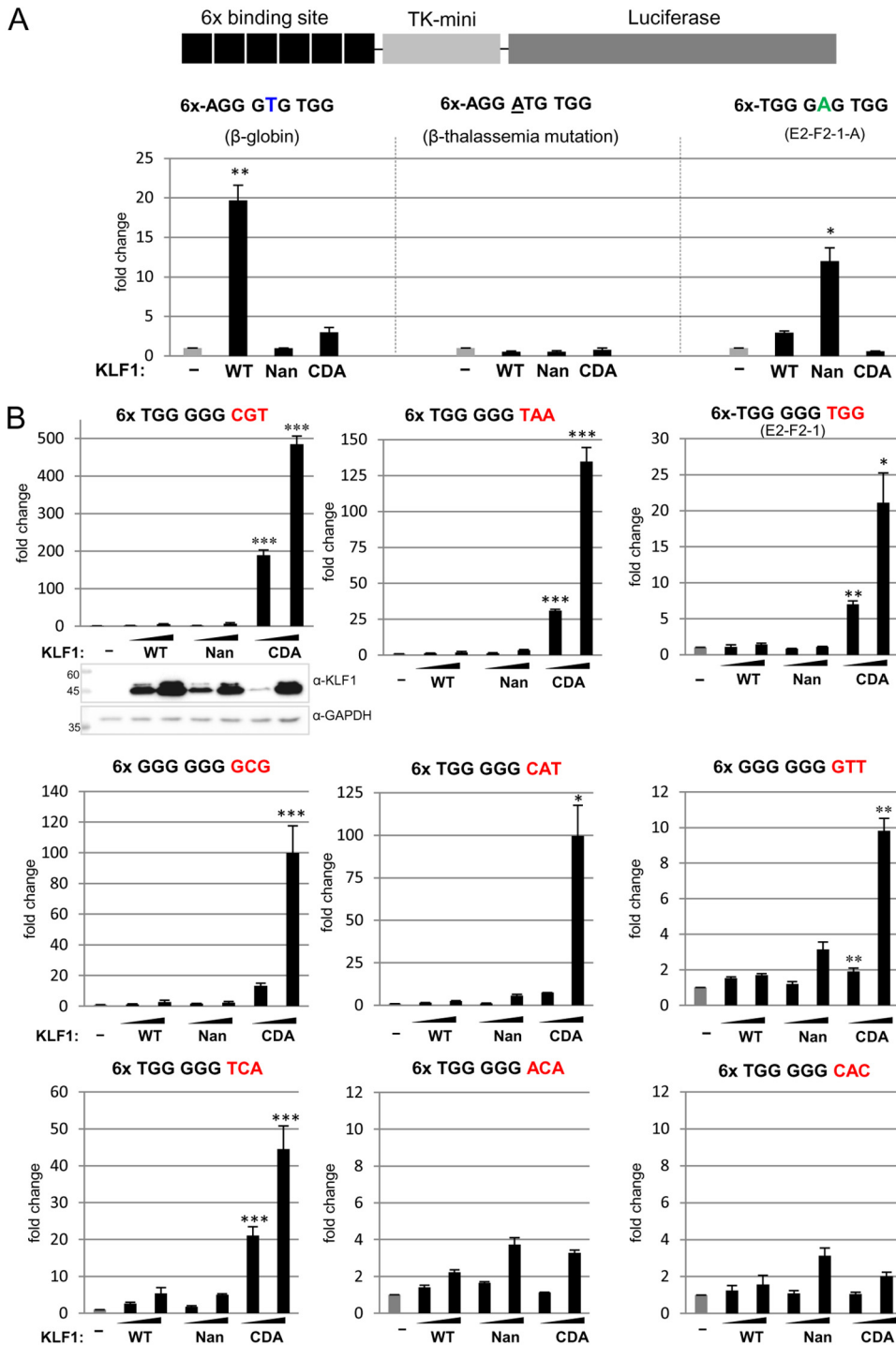


FIG 7 Functional analyses of the specificity of KLF1 variants in recognition and binding of CASTing-selected sites for CDA-KLF1 by activation of a luciferase reporter. K562 erythroleukemia cells that do not express endogenous KLF1 (53) were cotransfected with plasmids expressing the luciferase reporter gene under the control of the 63-bp fragment of the thymidine kinase promoter (TKmini) and six copies (6×) of the selected binding sites is shown at the top. Data represent a comparison of the fold change in transcriptional abilities of full-length WT-KLF1, Nan-KLF1, and CDA-KLF1 in binding and activation of luciferase reporter genes: binding site from the β-globin promoter (1) with a blue T in the middle (5th) position that serves as a positive control for WT-KLF1; binding site the from β-globin promoter (1) with an underlined A for the thalassemia mutation (54) that serves as a negative control for WT-KLF1; binding site from E2F2 with a green A in the middle (5th) position that serves as a positive control for Nan-KLF1. (B) Comparison of the fold change in transcriptional abilities of full-length WT-KLF1, Nan-KLF1, and CDA-KLF1 in binding and activation a set of luciferase reporter genes that were constructed based on CASTing-selected binding sites as shown at the top of each panel. A *Renilla* reporter construct was included as

(Continued on next page)

transfected K562 cells with increasing amounts of expression plasmids carrying KLF1 variants. We found that the luciferase reporter genes were activated only from binding sites that are recognized and bound by CDA-KLF1, as proven by a gel retardation assay (compare Fig. 6A and 7B). Such sites need to have G in the middle (5th) position of the binding motif, and the 3' end of the sequence (positions 7, 8, and 9) has to contain at least one G or T. Neither WT-KLF1 nor Nan-KLF1 is able to activate any of the studied reporter genes because of the G in the middle position.

These results evidently show that only CDA-KLF1 is able to activate the reporter genes in a pattern totally consistent with the previous analyses and thus are entirely confirmatory with the *in vitro* EMSA results.

Quantitative analysis of the endogenous levels of KLF1 target genes, expressed in stable K562 cell lines with induced expression of WT- or CDA-KLF1.

Next, we studied the expression levels of endogenous genes that might be affected by the presence of mutant KLF1. For this purpose, we prepared two stable K562 cell lines with doxycycline-inducible KLF1 variants, either WT-KLF1 or CDA-KLF1. After induction, we isolated total RNA and prepared reverse-transcribed cDNA for quantitative analysis. We focused on monitoring expression of known WT-KLF1 targets, notably, *Lu/BCAM* (for basal cell adhesion molecule, Lutheran blood group), *AQP1* (for aquaporin 1), *EPB4.2* (for erythrocyte protein band 4.2), *Cdkn1B* (for cyclin-dependent kinase inhibitor 1B; p27), and *E2F2* (for E2F transcription factor). Our rationale was as follows. First, it is known that *AQP1* levels are dramatically lower in samples originating from patients suffering from CDA type IV than in those of healthy controls (16, 36, 55). Second, expression of *EPB4.2* is reduced in CDA induced pluripotent stem cell (iPS)-derived erythroblasts (38); also, analysis of the transcriptome of the CDA patient revealed its decreased level, as it contributes to the fragility of CDA patients' red cell membrane abnormalities (56). Third, expression of the *BCAM* gene is lower in individuals with the In(Lu) phenotype that is caused by a variety of KLF1 mutations that lead to its haploinsufficiency (18). Fourth, expression of *E2F2* is decreased in the presence of the Nan-KLF1 mutation (25). Finally, the *Cdkn1B* gene (p27 gene) is directly regulated by KLF1 during late stages of differentiation, where it is critical for enucleation of erythroid precursors (57), and this process is disturbed in the presence of CDA mutation (38, 56). Furthermore, we expanded our analysis for *TFR2* (for transferrin receptor 2), *SLC4A1* (for anion transport protein; BAND3), and *IL17RB* (for interleukin-17 receptor B) based on a recently published article (56).

Analyses of RNA expression in the induced cells (Fig. 8A) show that *AQP1* and *EPB4.2* are not transactivated by CDA-KLF1 as efficiently as by WT-KLF1, resembling the data from CDA patients and in CDA iPS-derived erythroblasts (Fig. 8A). On the other hand, *E2F2*, *BCAM*, *Cdkn1B*, *TFR2*, *SLC4A1*, and *IL17RB* remain activated and to an even greater level by the CDA variant than by the WT. The expression of *IL17RB* is 8-fold higher in the presence of CDA-KLF1 than of WT-KLF1 (Fig. 8A). As a potential explanation, inspection of the upstream promoters and introns of these genes reveals that the regulatory regions of *BCAM*, *Cdkn1B*, *E2F2*, *TFR2*, *SLC4A1*, and *IL17RB* contain binding sites for both WT-KLF1 and CDA-KLF1 and thus are activated and transcribed (Fig. 8C). *AQP1* and *EPB4.2* contain sites only for WT-KLF1 that are not recognized by CDA-KLF1 (Fig. 8B) and thus are unsuitable as activation targets. Interestingly, the *BCAM* result suggests a resolution of paradoxical patient observations, namely, that *BCAM* expression is not affected in CDA patients while at the same time haploinsufficient levels of KLF1 are known to adversely affect *BCAM* expression (58, 59). Our data suggest that the presence of variant DNA-binding sites recognized by the CDA-KLF1 protein compensates for any drop in expression of the WT-KLF1 protein.

FIG 7 Legend (Continued)

a normalization control for transfection efficiency. An average of three biological replicates for each reporter gene (arithmetic mean \pm standard deviation) is shown (*, $P \leq 0.05$; **, $P \leq 0.01$; ***, $P \leq 0.001$). The Western blot under the first panel shows the expression levels of KLF1 variants used in the luciferase assay; the loading control was visualized by antibodies against glyceraldehyde-3-phosphate dehydrogenase (GAPDH).

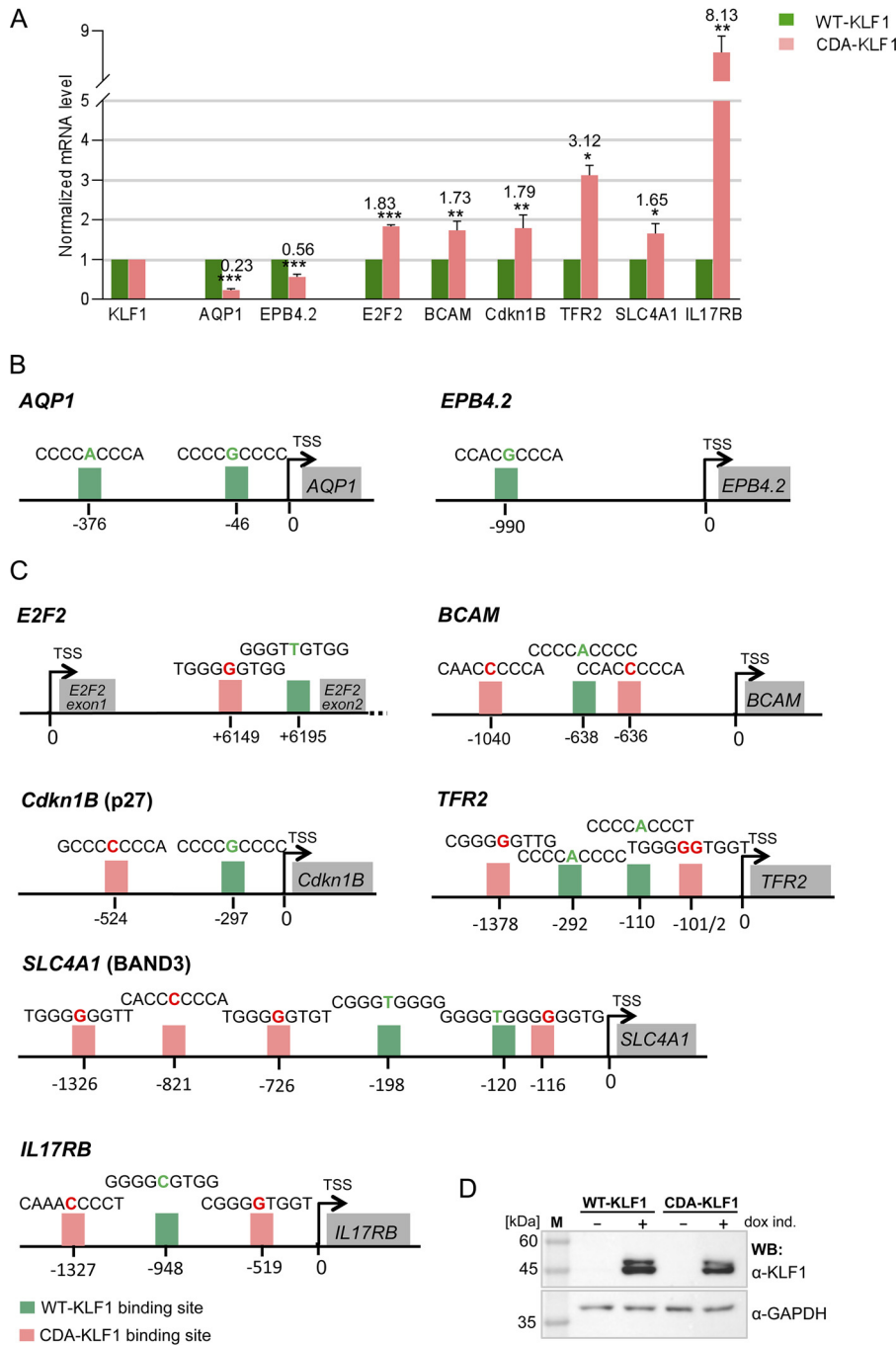


FIG 8 Quantitative analysis of the *in vivo* expression levels of KLF1 target genes, carried out in stable K562 cell lines with induced expression of WT-KLF1 or CDA-KLF1. (A) RT-qPCR analyses of KLF1 target mRNA levels in stably transfected K562 cells with inducible expression of WT-KLF1 and CDA-KLF1. Data are normalized to the level of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and to the level of expression of WT-KLF1 or CDA-KLF1. Data represent an average of triple repeats. (B) The scheme of human KLF1 target gene regulatory regions that contain only the WT-KLF1 consensus binding sites, marked in green, based on bioinformatics analysis. (C) The scheme of human KLF1 target gene regulatory regions that contain CDA-KLF1 consensus binding sites, marked in red, and WT-KLF1 binding sites, marked in green, based on bioinformatics analysis. The location of binding sites based on distance from the transcription start site (TSS). The individual DNA sequences of binding motifs are written above: *AQP1*, aquaporin 1; *EPB4.2*, erythrocyte protein band 4.2; *E2F2*, E2F transcription factor 2; *BCAM*, basal cell adhesion molecule (Lutheran blood group); *Cdkn1B*, cyclin-dependent kinase inhibitor 1B (p27); *TFR2*, transferrin receptor 2; *SLC4A1*, anion transport protein (BAND3); *IL17RB*, interleukin-17 receptor B. Data represent an average of at least three biological replicates of experiments (arithmetic mean \pm standard deviation) is shown (*, $P \leq 0.05$; **, $P \leq 0.01$; ***, $P \leq 0.001$). (D) Western blot with the expression levels of KLF1 variants from K562 stable cell lines induced 18 h with 1.5 $\mu\text{g/ml}$ doxycycline. The loading control was visualized by antibodies against glyceraldehyde-3-phosphate dehydrogenase (GAPDH).

Collectively, these results show that the CDA-KLF1 consensus binding sites, as determined in the CASTing-seq experiment, are uniquely recognized, bound, and transcriptionally functional endogenously in the cell.

DISCUSSION

Mutations of transcription factors that affect amino acids involved in direct interactions with DNA change their specificity. This leads to recognition and activation of an alternate set of genes and, as a result, may cause serious consequences for organisms. We have investigated such a mutation in KLF1, the essential transcription factor of erythropoiesis (reviewed in reference 22).

Using the *in vitro* CASTing method, we identified a new set of sequences bound by CDA-KLF1, and based on them we defined the consensus binding site as 5'-NGG-GG(T/G)-(T/G)(T/G)(T/G)-3'. It differs from the consensus binding sites for WT-KLF1, 5'-NGG-G(C/T)G-(T/G)GG-3', and for Nan-KLF1, 5'-NGG-G(C/A)N-(T/G)GG-3', as well. The CDA-KLF1 binding site is mutually exclusive, in contrast to the binding sites for WT-KLF1 and Nan-KLF1. The CDA-KLF1 binding site contains G in the middle (5th) position (boldface), whereas WT-KLF1 has C/T and Nan-KLF1 has C/A. Our results are in agreement with the structural studies conducted by Carl Pabo and colleagues, who analyzed the relationship between functionally important amino acids of zinc fingers and the DNA sequence to which they bound (51, 60, 61). They found that if lysine is located in position +3 of the α -helix, within the $\beta\beta\alpha$ structure of the middle zinc finger, it has a high affinity for guanine in the middle (5th) position of the 9-nt binding site (51, 60, 61). Other research investigating YY1 and TFIIIA also described similar results, finding that lysine in zinc finger 2 of YY1 and in finger 1 of TFIIIA bound guanine (62–65). In addition, a detailed bioinformatics analysis of biochemical preferences between amino acids and bases in the DNA α -helix proves that lysine preferentially forms a strong bond with guanine (66), observations that entirely support our data.

Another interesting feature of our newly determined consensus binding site for CDA-KLF1 is high degeneracy of the 3' end of the motif 5'-NGG-GG(T/G)-(T/G)(T/G)(T/G)-3'. On the G-rich strand, the last four positions (underlined) of the motif could contain T or G nucleotides. Both bases, guanine and thymine, can have alternate molecular structures based on different locations of a particular hydrogen atom. They can generate tautomeric keto or enol forms. Both guanine and thymine can switch easily from one tautomer to another. The change affects the three-dimensional shape of the molecule that may disturb the DNA-protein interface. The degeneracy of the CDA-KLF1 consensus site generates a much wider combination of recognized sites than WT-KLF1 or Nan-KLF1 that likely leads to a broader range of binding affinities for less specific targets. Occurrence of high degeneracy may be a solution for a too strong binding between the transcription factor and its target gene, which may translate to the speed of the transcriptional reaction (16).

The results showing that the CDA mutation located in ZnF2 affects binding of DNA sequence by ZnF1 were unexpected. We believe that the main reason for this effect is the high affinity of interactions between lysine (E339K) and G in the middle (5th) position of the 9-nt binding site. These interactions, along with the remaining interactions of ZnF2 and ZnF3 with G-rich positions 1 to 4 of the binding site, may be sufficient to effectively bind CDA-KLF1 to DNA. Thus, the E-to-K mutation does not so much affect the specificity of ZnF1 interaction with DNA; rather, it affects the binding strength/affinity of all three ZnFs, allowing for greater freedom and higher degeneracy of the DNA sequence in interactions with ZnF1. The strong interactions in the complex would make the duration of binding too long, leading to low efficient function (67) as a transcription factor. Interactions between a transcription factor and DNA that are too strong and too long can stall/block the transcriptional machinery, and because of this, the overall transcription efficiency may be reduced. The efficiency of biomolecular function is controlled by the association and dissociation rates *in vivo* (67).

While the manuscript was being prepared, another group published data on the consensus DNA-binding site of the E339K KLF1 mutant using a different approach (68).

Ilsey et al. performed chromatin immunoprecipitation sequencing (ChIP-seq) in a transgenic cell line under specific cell-induced conditions. In our case, the *in vitro* assay enabled us to obtain the potential binding capacity for CDA-KLF1 regardless of the set of genes activated in a given cell state such as proliferation, differentiation, or stress conditions. Their consensus binding motif is partly consistent with ours: 5'-NGG-G(**A/G**)G-(T/G)GG-3'. Their motif contains A as the most common nucleotide in the middle (5th) position (in boldface; G is the second most common), which differs from our results. Based on our data, A in the 5th position of the motif excludes binding by CDA-KLF1 (Fig. 1C). Similarly, their affinity binding data do not support having A in the 5th position, as the K_d value for a binding site with A is almost 2-fold higher than the K_d value for the wild-type protein (68); it is known that the WT-KLF1 prefers only C or T in the middle (5th) position (4, 25). The motif reported by Ilsey et al. does not show 3' end degeneracy; in this context, reliance on induction of WT- or CDA-KLF1 raises the issue of overexpression relative to normal endogenous levels, which may affect the binding equilibrium between KLF1 and its target DNA. However, their ChIP-seq data demonstrate a 3-fold difference in binding peaks obtained for CDA-KLF1 compared to those of the WT-KLF1 (68) that could be explained by degeneracy of the 3' end of the binding motif. Nonetheless, their modeling studies based on the KLF4 zinc finger/DNA structure performed on a sequence with a central GGG triplet are entirely consistent with our observations.

We tested the biochemical specificity between CDA-KLF1 in comparison with the binding specificities of WT- and Nan-KLF1 in complex formation with our defined DNA motif. We observed that the presence of G in the middle (5th) position of the binding motif entirely precludes complex formation with WT- or Nan-KLF1. In the presence of CDA-KLF1, we detected variation in the intensities of the generated complexes, suggesting that there is a range of affinities for the individual nucleotide sequences.

Both characteristics of the newly determined CDA-KLF1 consensus site (guanine in the middle position of the motif and the 3' end degeneracy) contribute to activation of neomorphic genes by CDA-KLF1 that are never activated by WT-KLF1. This could be a potential explanation for the dominant phenotype due to the E325K mutation in the KLF1 heterozygotes (CDA^{+/+}), which disturbs wild-type gene expression by binding and activating new targets. In the case of Nan-KLF1, the 6th position of the consensus binding site, 5'-NGG-GC/AN-(T/G)GG-3', contains N (underlined) that accepts any nucleotide. Such a motif extends the range of recognizable binding sites relative to those of WT-KLF1, which accepts only G in this position. As a consequence, Nan-KLF1 activates neomorphic targets that are typically never expressed in the erythroid cells (28, 29). The Nan-KLF1 mutant activates *Hamp* and *Irf7* genes (29). This ectopic gene expression contributes to the occurrence of chronic, lifelong anemia in the heterozygote Nan^{+/+} mice (29). We can predict that a similar situation occurs in the case of a CDA-KLF1 mutant. For example, patients suffering from CDA type IV present clinical manifestations outside the erythroid system. Short stature and problems with the urinary tract and gonad development have been observed (27, 30). Neomorphic gene activation by CDA-KLF1 may be one explanation for these characteristics.

Since we did not have access to CDA type IV patients as this disease is very rare and as only seven cases have been described (26, 27, 30–38), we prepared stable K562 cell lines with doxycycline-inducible expression constructs of WT-KLF1 or CDA-KLF1. We investigated the level of transcription for known KLF1 target genes, which have both types of binding sites for the WT- and CDA-KLF1 variants in their regulatory regions (promoters and/or introns). Such coexistence of both sites in close proximity may ameliorate the final outcome of gene activation for the CDA patients as they are heterozygous for the mutation. CDA patients do not exhibit the In(Lu) phenotype (26, 27, 37, 69), in which activation of the *Lu/BCAM* gene, for the antigen of the Lutheran blood group, is sensitive to haploinsufficient levels of KLF1. Our analyses suggest that in CDA patients (that are genetically CDA^{+/+}), both WT- and CDA-KLF1 variants are able to bind and activate the *Lu/BCAM* gene via independent promoter binding sites for both WT-KLF1 and CDA-KLF1. This clarifies previously published observations that were

not entirely explained (16, 58). This scenario may not be unique as we found that the *E2F2* and *Cdkn1B* cell cycle regulator genes can be activated by both transcription factors.

It will be of interest to analyze the transcriptome of CDA type IV patients to verify that they, indeed, have ectopic expression of neomorphic, CDA-KLF1-activated genes. A recent study (56) shows that the genomic regions surrounding a number of the top ectopically expressed genes in a CDA patient (37) contain sites that, based on the present study, bind and are activated by CDA-KLF1 (5'-TGG-GGG-TGG-3') (Fig. 1 and 7). We tested some of them, for example, that study and ours found a higher level of expression from CDA-KLF1 for *IL17RB*, a gene that is critical for expression of interleukin-8 (IL-8) (56). Since the genes activated by CDA-KLF1 were not confirmed by ChIP assay, we may not exclude the indirect effect of the presence of a KLF1 mutant.

Activation of nonerythroid genes may help to explain the anomalies observed in CDA patients. It is also necessary to remember that zinc fingers of KLF1 are involved in interactions with proteins (9, 70–75), such that the CDA mutation could also disturb the interface between proteins and contribute to the pathological phenotype.

MATERIALS AND METHODS

Plasmids and molecular cloning. For mammalian expression of Flag-tagged KLF1 variants, the pSG5-Flag-WT-KLF1 (70) plasmid was used, and the mutations in amino acid position 339 of KLF1 were introduced by site-directed mutagenesis (Agilent Technologies).

For bacterial expression of the DBDs of KLF1 variants, the pMCSG7 vector was used. The Flag-tagged ZnF domains were subcloned from pSG5-Flag-WT-ZnF-KLF1 vector using the BamHI restriction enzyme. The mutations in amino acid position 339 of KLF1 were introduced by site-directed mutagenesis (Agilent Technologies); primers are available in Table S1 in supplemental material. For luciferase reporter gene construction, the pGL3 Basic vector (Promega) was used as a backbone. The 63-nucleotide-long herpes simplex virus (HSV) thymidine kinase minimal promoter (TKmini), synthesized by Sigma-Aldrich, was added upstream of the luciferase gene. To such a construct, pGL3-TKmini, six repeats of CASTing-selected binding sites synthesized by Sigma-Aldrich were introduced upstream of the TKmini promoter. Primers used to insert six repeats of indicated in Fig. 7 DNA-binding sites are available in Table S1.

Cell lines, transfection, nuclear extracts isolation, and Western blot analysis. K562 cells were cultured in RPMI medium 1640 supplemented with 10% fetal bovine serum (Corning) and 0.5% penicillin-streptomycin (BioShop). COS-7 cell lines were cultured in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum and 0.5% penicillin-streptomycin. All cells were cultured in a humidified incubator at 37°C in an atmosphere of 5% CO₂ and 95% air. To prepare protein extracts for EMSA, COS-7 cells growing on 10-cm-diameter dishes were transfected with 10 μg of plasmid pSG5-WT-KLF1, -Nan-KLF1, or -CDA-KLF1 using Lipofectamine LTX according to the manufacturer's protocol. After 48 h of incubation, crude extracts were prepared as described previously (76). For Western blot analysis of Flag-tagged KLF1 variants, polyvinylidene difluoride (PVDF) membranes were probed with anti-Flag tag monoclonal M2-horseradish peroxidase (HRP) antibody (Sigma) or anti-KLF1 7B2 antibody as described previously (13).

Recombinant protein expression and purification. KLF1 Flag-tagged zinc finger domains, CDA-ZnF-KLF1 and WT-ZnF-KLF1, were cloned into pMCSG7 to generate an in-frame fusion to a 6×His tag and tobacco etch virus (TEV) cleavage sites. The CDA mutation (E to K) was introduced in the mouse ZnF domain at amino acid position 339 by site-directed mutagenesis, according to the manufacturer's protocol (Agilent Technologies). Primers for mutagenesis are listed in Table S1.

To purified DBD of KLF1 variants, the pMCSG7 plasmid containing Flag-CDA-ZnF-KLF1 and Flag-WT-ZnF-KLF1 was used. BL21-CodonPlus(DE3)-RIL *Escherichia coli* cells were transformed with these plasmids and induced with 0.3 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG). Bacteria were cultured in the presence of 100 μM ZnCl₂ to assist in protein folding for 17 h at 18°C before collection. Cell pellets were resuspended in cold lysis buffer [50 mM Tris-HCl (pH 8.0), 1.5 M NaCl, 10% glycerol, and 0.5 mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP)] with addition of sarcosyl to a final concentration of 2%. Cell lysates were sonicated for 4 min (2 s on and 20 s off) and centrifuged at 30,500 × *g* for 30 min at 4°C. Supernatants were diluted 10 times in lysis S-2 buffer and loaded on a column with Zn resin equilibrated with lysis buffer (77). The column was washed with lysis buffer with increasing salt concentrations (150 mM to 1.5 M). Then the salt concentration was decreased to 150 mM, allowing for efficient, on-column TEV protease cleavage for 2 h at room temperature. For every 1 mg of protein, 120 μg of TEV protease was used. Recombinant zinc finger domains were eluted from the column with elution buffer (50 mM Tris-HCl, 150 mM NaCl, 10% glycerol, 1 M NH₄Cl). Fractions containing recombinant DBD were combined and purified on Superdex HiLoad 200 column equilibrated with buffer (50 mM HEPES [pH 7.9], 100 mM KCl, 10% glycerol, 10 μM ZnCl₂, 0.5 mM TCEP) on an automated fast protein liquid chromatography (FPLC) system (AKTApurifier). Chromatography was carried out at 4°C.

CASTing for the CDA-KLF1 consensus DNA-binding site. CASTing was performed as previously described (42, 45) with some modifications. In the first CASTing experiment, we used a library with 12 random nucleotides, 5'-CAAGCTACTGCAGATGCNNNNNNNNNNCGTAGGATCCATCTAGAGT-3', flanked by constant sequences needed for amplification. In a second attempt, we used a library consisting of 8 random

nucleotides with a fixed G in the middle position: 5'-GCTCAAGCTTACTGCAGATATNNNGNNNNTTATAGGATC CATCTAGAGTCCGA-3' (Sigma-Aldrich). Fifteen micrograms of the library was made double stranded with 5 μg of 3' primer (CAST1_R-*ACTCTAGATGGATCCTACG* or CAST2_R-*TCGGACTCTAGATGGATCCTA*) in a reaction mixture that contained 200 mM (each) the four deoxynucleoside triphosphates and 20 U of DreamTaq DNA polymerase in 1 \times DreamTaq buffer (Thermo Scientific) for one cycle of 98°C for 3 min, 94°C for 1 min, 47°C for 2 min, and 72°C for 30 min. In the CASTing procedure, 12 μl of anti-Flag M2 magnetic beads (Sigma-Aldrich) was washed twice in 40 μl of ESB buffer (20 mM HEPES [pH 7.9], 40 mM KCl, 6 mM MgCl₂, 1 mM dithiothreitol, 1 mM phenylmethylsulfonyl fluoride, 0.1% NP-40, 10% glycerol), and then 800 ng of purified ZnF-KLF1 (WT or CDA) protein in 12 μl of binding buffer (10 mM Tris-HCl [pH 7.5], 150 mM NaCl, 20 mM KCl, 10 mM MgCl₂, and 5 mM ZnCl₂) was added and incubated for 1 h at 4°C. Immunoprecipitates were washed three times in EBC buffer (50 mM Tris-HCl [pH 8.0], 120 mM NaCl, 0.5% NP-40, 5 mM NaF, 1 mM Na₃VO₄, 10 $\mu\text{g}/\text{ml}$ phenylmethylsulfonyl fluoride, 10 $\mu\text{g}/\text{ml}$ leupeptin) and twice in ESB buffer. Next, beads with attached ZnF domains were suspended in 25 μl of ESB buffer with 150 ng of sonicated salmon sperm DNA (sssDNA; Fermentas), and 150 ng of the double-stranded library was added. The mixture was incubated for 30 min at room temperature on a low-speed shaking platform. The bead-protein-dsDNA oligonucleotide complexes were washed three times in ESB buffer, and the resulting complexes served as a template for amplification in 50 μl of PCR mixture containing 200 ng of CAST_F and CAST_R primers, 0.5 mCi of [α -³²P]dCTP, 20 mM dCTP, 50 mM each dATP, dTTP, and dGTP, 10 μg of bovine serum albumin, 1 \times Dream Taq buffer, and 1 μl of Dream Taq polymerase (Thermo Scientific). Selected dsDNA oligonucleotides were amplified for 15 cycles of 1 min at 94°C, 1 min at 62°C, 1 min at 72°C, and 7 min for final extension. Amplified dsDNA oligonucleotides were purified on a Sephadex G-50 spin column (GE Health Care) and resolved on 8% nondenaturing polyacrylamide gels. Radiolabeled oligonucleotides were excised from the gels, eluted overnight, and used in subsequent CASTing cycles. After five rounds of selection, adaptors and barcodes for NGS sequencing were added by fusion PCR (primers in Table S1). The quantity and quality of the pool of selected dsDNA oligonucleotides were determined using a Qubit fluorometric assay and Agilent BioAnalyzer High-Sensitivity DNA kit (Agilent Technologies). Emulsion PCR and ion sphere particle (ISP) enrichment were done using an Ion PGM Hi-Q kit for 200 bp (Life Technologies) according to the manufacturer's instructions. Next-generation sequencing was carried out on an Ion Torrent Personal Genome Machine sequencer (Life Technologies) using an Ion 314 chip (Life Technologies) and Ion PGM Hi-Q sequencing kit (Life Technologies) according to the manufacturer's instructions. The obtained ~30,000 sequences were analyzed by the MEME software to perform *de novo* motif discovery (78).

Electrophoretic mobility shift assay EMSA. The gel retardation assay was performed as previously described (25) with some modifications. Briefly, double-stranded synthetic oligonucleotides containing specific KLF1 binding sites found by CASTing were radiolabeled by [α -³²P]CTP using Klenow polymerase and used as a probes or competitors for gel retardation analysis. Binding reactions were performed in the presence of ~10 μg (5 $\mu\text{g}/\mu\text{l}$) of the whole-cell extracts from COS-7 cells transfected with a WT-, CDA-, or Nan-KLF1 variant or purified recombinant ZnF domains of WT- or CDA-KLF1. The mix was incubated for 20 min in binding buffer (25 mM HEPES [pH 7.5], 32 mM KCl, 50 mM NaCl, 2 μM ZnCl₂, 0.7 mM β -mercaptoethanol, 8% glycerol) on ice (whole extract) or at room temperature (recombinant protein). Nonspecific interactions were quenched with 100 ng/ μl (whole extract) or 40 ng/ μl (recombinant protein) of sssDNA (Fermentas) and resolved in a nondenaturing 8% polyacrylamide gel in 0.5 \times Tris-borate-EDTA (TBE) buffer at 4°C. Anti-KLF1 antibody 4B9 was used to identify the KLF1 band position (7).

For quantitative measurement of the dissociation constant, the following increasing concentrations of the recombinant zinc finger domain of WT- or CDA-ZnF-KLF1 protein were used: 0, 100, 150, 500, 750, 980, 1,230, 1,470, or 1,720 nM. We kept the DNA probe constant. EMSA gels were scanned, and band intensities were quantified by densitometry. The ratio resulting from the upper band (bound probe) to the lower band (free probe) was calculated. K_d represents the protein concentration at which 50% of DNA is bound. Curves were fit using the Hill slope in GraphPad. Averaged K_d and its standard deviation were calculated based on at least two EMSA gel shifts. Oligonucleotide sequences used in EMSAs are listed in Table S1 in supplemental material.

Reporter gene analysis dual-luciferase assays. K562 cells seeded in 24-well plates were transfected with suitable reporter genes and increasing amounts of plasmid carrying WT-, Nan-, or CDA-KLF1 using Lipofectamine LTX. After 36 h of incubation, cells were lysed and assayed for luciferase activities with a dual-luciferase system (Promega). Plasmid pRLTK (*Renilla*) (Promega) was included as a normalization control for transfection efficiency. Luminescence was quantified with a luminometer (GloMax 96 microplate; Promega). The results are the average of at least three experiments performed in triplicates.

Generation of stable K562 cell lines. Plasmid pSG5 containing mouse Flag-WT/CDA-KLF1 served as a source of full-length KLF1 variant genes which were amplified by PCR and introduced into the pSAM vector (kind gift from Rita Perlingeiro, University of Minnesota). For lentiviral production, HEK-293T cells were cotransfected with gene-of-interest (GOI)-containing vectors with a doxycycline-dependent promoter and components of 2nd-generation packaging vectors: a pPAX2 packaging vector and pMD2G envelope vector using linear polyethylenimine (PEI 25000; Sigma-Aldrich). At 48 h posttransfection, the lentiviruses were collected, filtered, and added to the K562 cells. After 72 h the expression of KLF1 variants was induced by 1 $\mu\text{g}/\text{ml}$ doxycycline. To select KLF1 variant-containing cells, sorting for fluorescence-positive cells using a FACSAria III cell sorter (BD Biosciences, La Jolla, CA) was performed. The primers used to insert KLF1 variants into pSAM vectors are available in Table S1.

RNA isolation and real-time PCR analyses. Total RNA was extracted from stable K562 cell lines with induced expression of WT- or CDA-KLF1 (1.5 $\mu\text{g}/\text{ml}$ doxycycline for 18 h) using 3-Zone reagent (Novazym) according to the manufacturer's instructions. Extracted RNA was treated with DNase I (ThermoFisher

Scientific) and subjected to cDNA synthesis using a Moloney murine leukemia virus (M-MLV) reverse transcriptase kit from Verto according to the manufacturer's instructions (Novazym). Gene expression was quantified using real-time quantitative PCR (RT-qPCR) with SsoAdvanced Universal SYBR green Supermix on a CFX Connect real-time PCR detection system (BioRAD). The primers used for RT-qPCR are in Table S1.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 0.5 MB.

ACKNOWLEDGMENTS

We are grateful to Barbara Imiolczyk (Inst. of Bioorganic Chemistry, PAS) for help with protein purification, to Alexey Bryzgalov and Izabela Makalowska from Adam Mickiewicz University (AMU) for help with bioinformatics analyses, to Anna Witucka (AMU), Jacek Stepniewski, Alicja Jozkowicz, and Jozef Dulak (Jagiellonian University, Krakow) for help with generation of stable cell lines, and to the Molecular Biology Techniques Laboratory (AMU) for NGS.

This work was supported by the Polish National Science Center 2013/09/B/NZ1/01879 to M.S. and by the National Institutes of Health DK046865 to J.J.B.

We declare that we have no conflicts of interest.

REFERENCES

- Miller IJ, Bieker JJ. 1993. A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Kruppel family of nuclear proteins. *Mol Cell Biol* 13:2776–2786. <https://doi.org/10.1128/mcb.13.5.2776>.
- Jenkins NA, Gilbert DJ, Copeland NG, Gruzglin E, Bieker JJ. 1998. Erythroid Kruppel-like transcription factor (Eklf) maps to a region of mouse chromosome 8 syntenic with human chromosome 19. *Mamm Genome* 9:174–176. <https://doi.org/10.1007/s003359900716>.
- Pilon AM, Ajay SS, Kumar SA, Steiner LA, Cherukuri PF, Wincovitch S, Anderson SM, Comparative Sequencing Program N, Mullikin JC, Gallagher PG, Hardison RC, Margulies EH, Bodine DM. 2011. Genome-wide ChIP-Seq reveals a dramatic shift in the binding of the transcription factor erythroid Kruppel-like factor during erythrocyte differentiation. *Blood* 118:e139. <https://doi.org/10.1182/blood-2011-05-355107>.
- Tallack MR, Whittington T, Shan Yuen W, Wainwright EN, Keys JR, Gardiner BB, Nourbakhsh E, Cloonan N, Grimmond SM, Bailey TL, Perkins AC. 2010. A global role for KLF1 in erythropoiesis revealed by ChIP-seq in primary erythroid cells. *Genome Res* 20:1052–1063. <https://doi.org/10.1101/gr.106575.110>.
- Perkins AC, Sharpe AH, Orkin SH. 1995. Lethal beta-thalassaemia in mice lacking the erythroid CACCC-transcription factor EKLf. *Nature* 375:318–322. <https://doi.org/10.1038/375318a0>.
- Nuez B, Michalovich D, Bygrave A, Ploemacher R, Grosveld F. 1995. Defective haematopoiesis in fetal liver resulting from inactivation of the EKLf gene. *Nature* 375:316–318. <https://doi.org/10.1038/375316a0>.
- Zhang W, Kadam S, Emerson BM, Bieker JJ. 2001. Site-specific acetylation by p300 or CREB binding protein regulates erythroid Kruppel-like factor transcriptional activity via its interaction with the SWI-SNF complex. *Mol Cell Biol* 21:2413–2422. <https://doi.org/10.1128/MCB.21.7.2413-2422.2001>.
- Armstrong JA, Bieker JJ, Emerson BM. 1998. A SWI/SNF-related chromatin remodeling complex, E-RC1, is required for tissue-specific transcriptional regulation by EKLf in vitro. *Cell* 95:93–104. [https://doi.org/10.1016/S0092-8674\(00\)81785-7](https://doi.org/10.1016/S0092-8674(00)81785-7).
- Kadam S, McAlpine GS, Phelan ML, Kingston RE, Jones KA, Emerson BM. 2000. Functional selectivity of recombinant mammalian SWI/SNF subunits. *Genes Dev* 14:2441–2451. <https://doi.org/10.1101/gad.828000>.
- Yien YY, Bieker JJ. 2013. EKLf/KLF1, a tissue-restricted integrator of transcriptional control, chromatin remodeling, and lineage determination. *Mol Cell Biol* 33:4–13. <https://doi.org/10.1128/MCB.01058-12>.
- Drissen R, Palstra RJ, Gillemans N, Splinter E, Grosveld F, Philipsen S, de Laat W. 2004. The active spatial organization of the beta-globin locus requires the transcription factor EKLf. *Genes Dev* 18:2485–2490. <https://doi.org/10.1101/gad.317004>.
- Ouyang L, Chen X, Bieker JJ. 1998. Regulation of erythroid Kruppel-like factor (EKLf) transcriptional activity by phosphorylation of a protein kinase casein kinase II site within its interaction domain. *J Biol Chem* 273:23019–23025. <https://doi.org/10.1074/jbc.273.36.23019>.
- Siatecka M, Xue L, Bieker JJ. 2007. Sumoylation of EKLf promotes transcriptional repression and is involved in inhibition of megakaryopoiesis. *Mol Cell Biol* 27:8547–8560. <https://doi.org/10.1128/MCB.00589-07>.
- Zhang W, Bieker JJ. 1998. Acetylation and modulation of erythroid Kruppel-like factor (EKLf) activity by interaction with histone acetyltransferases. *Proc Natl Acad Sci U S A* 95:9855–9860. <https://doi.org/10.1073/pnas.95.17.9855>.
- Quadrini KJ, Bieker JJ. 2006. EKLf/KLF1 is ubiquitinated in vivo and its stability is regulated by activation domain sequences through the 26S proteasome. *FEBS Lett* 580:2285–2293. <https://doi.org/10.1016/j.febslet.2006.03.039>.
- Singleton BK, Lau W, Fairweather VS, Burton NM, Wilson MC, Parsons SF, Richardson BM, Trakarnsanga K, Brady RL, Anstee DJ, Frayne J. 2011. Mutations in the second zinc finger of human EKLf reduce promoter affinity but give rise to benign and disease phenotypes. *Blood* 118:3137–3145. <https://doi.org/10.1182/blood-2011-04-349985>.
- Perkins A, Xu X, Higgs DR, Patrinos GP, Arnaud L, Bieker JJ, Philipsen S, Workgroup KL, F. C. 2016. Kruppeling erythropoiesis: an unexpected broad spectrum of human red blood cell disorders due to KLF1 variants. *Blood* 127:1856–1862. <https://doi.org/10.1182/blood-2016-01-694331>.
- Singleton BK, Burton NM, Green C, Brady RL, Anstee DJ. 2008. Mutations in EKLf/KLF1 form the molecular basis of the rare blood group In(Lu) phenotype. *Blood* 112:2081–2088. <https://doi.org/10.1182/blood-2008-03-145672>.
- Borg J, Patrinos GP, Felice AE, Philipsen S. 2011. Erythroid phenotypes associated with KLF1 mutations. *Haematologica* 96:635–638. <https://doi.org/10.3324/haematol.2011.043265>.
- Liu D, Zhang X, Yu L, Cai R, Ma X, Zheng C, Zhou Y, Liu Q, Wei X, Lin L, Yan T, Huang J, Mohandas N, An X, Xu X. 2014. KLF1 mutations are relatively more common in a thalassemia endemic region and ameliorate the severity of beta-thalassemia. *Blood* 124:803–811. <https://doi.org/10.1182/blood-2014-03-561779>.
- Perseu L, Satta S, Moi P, Demartis FR, Manunza L, Sollaino MC, Barella S, Cao A, Galanello R. 2011. KLF1 gene mutations cause borderline HbA(2). *Blood* 118:4454–4458. <https://doi.org/10.1182/blood-2011-04-345736>.
- Siatecka M, Bieker JJ. 2011. The multifunctional role of EKLf/KLF1 during erythropoiesis. *Blood* 118:2044–2054. <https://doi.org/10.1182/blood-2011-03-331371>.
- Heruth DP, Hawkins T, Logsdon DP, Gibson MI, Sokolovsky IV, Nsumu NN, Major SL, Fegley B, Woods GM, Lewing KB, Neville KA, Cornetta K, Peterson KR, White RA. 2010. Mutation in erythroid specific transcription factor KLF1 causes hereditary spherocytosis in the Nan hemolytic anemia mouse model. *Genomics* 96:303–307. <https://doi.org/10.1016/j.ygeno.2010.07.009>.

24. White RA, Sokolovsky IV, Britt MI, Nsumu NN, Logsdon DP, McNulty SG, Wilmes LA, Brewer BP, Wirtz E, Joyce HR, Fegley B, Smith A, Heruth DP. 2009. Hematologic characterization and chromosomal localization of the novel dominantly inherited mouse hemolytic anemia, neonatal anemia (Nan). *Blood Cells Mol Dis* 43:141–148. <https://doi.org/10.1016/j.bcmd.2009.03.009>.
25. Siatecka M, Sahr KE, Andersen SG, Mezei M, Bieker JJ, Peters LL. 2010. Severe anemia in the Nan mutant mouse caused by sequence-selective disruption of erythroid Kruppel-like factor. *Proc Natl Acad Sci U S A* 107:15151–15156. <https://doi.org/10.1073/pnas.1004996107>.
26. Singleton BK, Fairweather VS, Lau W, Parsons SF, Burton NM, Frayne J, Brady RL, Anstee DJ. 2009. A novel EKLK mutation in a patient with dyserythropoietic anemia: the first association of EKLK with disease in man. *Blood* 114:162–162. <https://doi.org/10.1182/blood.V114.22.162.162>.
27. Arnaud L, Saison C, Helias V, Lucien N, Steschenko D, Giarratana M-C, Prehu C, Foliguet B, Montout L, de Brevern AG, Francina A, Ripocce P, Fenneteau O, Da Costa L, Peyrard T, Coghlan G, Illum N, Birgens H, Tamary H, Iolascon A, Delaunay J, Tchernia G, Cartron J-P. 2010. A dominant mutation in the gene encoding the erythroid transcription factor KLF1 causes a congenital dyserythropoietic anemia. *Am J Hum Genet* 87:721–727. <https://doi.org/10.1016/j.ajhg.2010.10.010>.
28. Gillinder KR, Ilseley MD, Nebor D, Sachidanandam R, Lajoie M, Magor GW, Tallack MR, Bailey T, Landsberg MJ, Mackay JP, Parker MW, Miles LA, Graber JH, Peters LL, Bieker JJ, Perkins AC. 2017. Promiscuous DNA-binding of a mutant zinc finger protein corrupts the transcriptome and diminishes cell viability. *Nucleic Acids Res* 45:1130–1143. <https://doi.org/10.1093/nar/gkw1014>.
29. Planutis A, Xue L, Trainor CD, Dangeti M, Gillinder K, Siatecka M, Nebor D, Peters LL, Perkins AC, Bieker JJ. 2017. Neomorphic effects of the neonatal anemia (Nan-Eklf) mutation contribute to deficits throughout development. *Development* 144:430–440. <https://doi.org/10.1242/dev.145656>.
30. Ravindranath Y, Johnson RM, Goyette G, Buck S, Gadgeel M, Gallagher PG. 2018. KLF1 E325K-associated congenital dyserythropoietic anemia type IV: insights into the variable clinical severity. *J Pediatr Hematol Oncol* 40:e405–e409. <https://doi.org/10.1097/MPH.0000000000001056>.
31. Ortolano R, Forouhar M, Warwick A, Harper D. 2018. A case of congenital dyserythropoietic anemia type IV caused by E325K mutation in erythroid transcription factor KLF1. *J Pediatr Hematol Oncol* 40:e389–e391. <https://doi.org/10.1097/MPH.0000000000001042>.
32. de-la-Iglesia-Iñigo S, Moreno-Carralero M-I, Lemes-Castellano A, Molero-Labarta T, Méndez M, Morán-Jiménez M-J. 2017. A case of congenital dyserythropoietic anemia type IV. *Clin Case Rep* 5:248–252. <https://doi.org/10.1002/ccr3.825>.
33. Tang W, Cai SP, Eng B, Poon MC, Waye JS, Illum N, Chui DH. 1993. Expression of embryonic zeta-globin and epsilon-globin chains in a 10-year-old girl with congenital anemia. *Blood* 81:1636–1640. <https://doi.org/10.1182/blood.V81.6.1636.1636>.
34. Agre P, Asimos A, Casella JF, McMillan C. 1986. Inheritance pattern and clinical response to splenectomy as a reflection of erythrocyte spectrin deficiency in hereditary spherocytosis. *N Engl J Med* 315:1579–1583. <https://doi.org/10.1056/NEJM198612183152504>.
35. Wickramasinghe SN, Illum N, Wimberley PD. 1991. Congenital dyserythropoietic anaemia with novel intra-erythroblastic and intra-erythrocytic inclusions. *Br J Haematol* 79:322–330. <https://doi.org/10.1111/j.1365-2141.1991.tb04541.x>.
36. Parsons SF, Jones J, Anstee DJ, Judson PA, Gardner B, Wiener E, Poole J, Illum N, Wickramasinghe SN. 1994. A novel form of congenital dyserythropoietic anemia associated with deficiency of erythroid CD44 and a unique blood group phenotype [In(a-b)-, Co(a-b)-]. *Blood* 83:860–868. <https://doi.org/10.1182/blood.V83.3.860.860>.
37. Jaffray JA, Mitchell WB, Gnanapragasam MN, Seshan SV, Guo X, Westhoff CM, Bieker JJ, Manwani D. 2013. Erythroid transcription factor EKLK/KLF1 mutation causing congenital dyserythropoietic anemia type IV in a patient of Taiwanese origin: review of all reported cases and development of a clinical diagnostic paradigm. *Blood Cells Mol Dis* 51:71–75. <https://doi.org/10.1016/j.bcmd.2013.02.006>.
38. Kohara H, Utsugisawa T, Sakamoto C, Hirose L, Ogawa Y, Ogura H, Sugawara A, Liao J, Aoki T, Iwasaki T, Asai T, Doisaki S, Okuno Y, Muramatsu H, Abe T, Kurita R, Miyamoto S, Sakuma T, Shiba M, Yamamoto T, Ohga S, Yoshida K, Ogawa S, Ito E, Kojima S, Kanno H, Tani K. 2019. KLF1 mutation E325K induces cell cycle arrest in erythroid cells differentiated from congenital dyserythropoietic anemia patient-specific induced pluripotent stem cells. *Exp Hematol* 73:25–37.e8. <https://doi.org/10.1016/j.exphem.2019.03.001>.
39. Renella R, Wood WG. 2009. The congenital dyserythropoietic anemias. *Hematol Oncol Clin North Am* 23:283–306. <https://doi.org/10.1016/j.hoc.2009.01.010>.
40. Iolascon A, Russo R, Delaunay J. 2011. Congenital dyserythropoietic anemias. *Curr Opin Hematol* 18:146–151. <https://doi.org/10.1097/MOH.0b013e32834521b0>.
41. Heimpel H, Wendt F, Klemm D, Schubotho H, Heilmeyer L. 1968. Congenital dyserythropoietic anemia. *Arch Klin Med* 215:174–194. (In German.)
42. Tao Y, Kassatly RF, Cress WD, Horowitz JM. 1997. Subunit composition determines E2F DNA-binding site specificity. *Mol Cell Biol* 17:6994–7007. <https://doi.org/10.1128/mcb.17.12.6994>.
43. Corbi N, Perez M, Maione R, Passananti C. 1997. Synthesis of a new zinc finger peptide; comparison of its “code” deduced and “CASTing” derived binding sites. *FEBS Lett* 417:71–74. [https://doi.org/10.1016/S0014-5793\(97\)01257-X](https://doi.org/10.1016/S0014-5793(97)01257-X).
44. Thiesen HJ, Bach C. 1990. Target detection assay (TDA): a versatile procedure to determine DNA binding sites as demonstrated on SP1 protein. *Nucleic Acids Res* 18:3203–3209. <https://doi.org/10.1093/nar/18.11.3203>.
45. Shields JM, Yang VW. 1998. Identification of the DNA sequence that interacts with the gut-enriched Kruppel-like factor. *Nucleic Acids Res* 26:796–802. <https://doi.org/10.1093/nar/26.3.796>.
46. Gu G, Wang T, Yang Y, Xu X, Wang J. 2013. An improved SELEX-Seq strategy for characterizing DNA-binding specificity of transcription factor: NF-kappaB as an example. *PLoS One* 8:e76109. <https://doi.org/10.1371/journal.pone.0076109>.
47. Kleit RE. 1991. Recognition of DNA by Cys2, His2 zinc fingers. *Science* 253:1367, 1393. <https://doi.org/10.1126/science.1896847>.
48. Feng WC, Southwood CM, Bieker JJ. 1994. Analyses of beta-thalassemia mutant DNA interactions with erythroid Kruppel-like factor (EKLK), an erythroid cell-specific transcription factor. *J Biol Chem* 269:1493–1500.
49. Pabo CO, Sauer RT. 1992. Transcription factors: structural families and principles of DNA recognition. *Annu Rev Biochem* 61:1053–1095. <https://doi.org/10.1146/annurev.bi.61.070192.005201>.
50. Tallack MR, Keys JR, Humbert PO, Perkins AC. 2009. EKLK/KLF1 controls cell cycle entry via direct regulation of E2f2. *J Biol Chem* 284:20966–20974. <https://doi.org/10.1074/jbc.M109.006346>.
51. Wolfe SA, Nekudova L, Pabo CO. 2000. DNA recognition by Cys2His2 zinc finger proteins. *Annu Rev Biophys Biomol Struct* 29:183–212. <https://doi.org/10.1146/annurev.biophys.29.1.183>.
52. Siatecka M, Lohmann F, Bao S, Bieker JJ. 2010. EKLK directly activates the p21WAF1/CIP1 gene by proximal promoter and novel intronic regulatory regions during erythroid differentiation. *Mol Cell Biol* 30:2811–2822. <https://doi.org/10.1128/MCB.01016-09>.
53. Donze D, Townes TM, Bieker JJ. 1995. Role of erythroid Kruppel-like factor in human gamma- to beta-globin gene switching. *J Biol Chem* 270:1955–1959. <https://doi.org/10.1074/jbc.270.4.1955>.
54. Orkin SH, Antonarakis SE, Kazazian HH, Jr. 1984. Base substitution at position -88 in a beta-thalassemic globin gene. Further evidence for the role of distal promoter element ACACCC. *J Biol Chem* 259:8679–8681.
55. Agre P, Smith BL, Baumgarten R, Preston GM, Pressman E, Wilson P, Illum N, Anstee DJ, Lande MB, Zeidel ML. 1994. Human red cell Aquaporin CHIP. II. Expression during normal fetal development and in a novel form of congenital dyserythropoietic anemia. *J Clin Invest* 94:1050–1058. <https://doi.org/10.1172/JCI117419>.
56. Varricchio L, Planutis A, Manwani D, Jaffray J, Mitchell WB, Migliaccio AR, Bieker JJ. 2019. Genetic disarray follows mutant KLF1-E325K expression in a congenital dyserythropoietic anemia patient. *Haematologica* 104:2372. <https://doi.org/10.3324/haematol.2018.209858>.
57. Gnanapragasam MN, McGrath KE, Catherman S, Xue L, Palis J, Bieker JJ. 2016. EKLK/KLF1-regulated cell cycle exit is essential for erythroblast nucleation. *Blood* 128:1631–1641. <https://doi.org/10.1182/blood-2016-03-706671>.
58. Helias V, Saison C, Peyrard T, Vera E, Prehu C, Cartron JP, Arnaud L. 2013. Molecular analysis of the rare In(Lu) blood type: toward decoding the phenotypic outcome of haploinsufficiency for the transcription factor KLF1. *Hum Mutat* 34:221–228. <https://doi.org/10.1002/humu.22218>.
59. Singleton BK, Frayne J, Anstee DJ. 2012. Blood group phenotypes resulting from mutations in erythroid transcription factors. *Curr Opin Hematol* 19:486–493. <https://doi.org/10.1097/MOH.0b013e328358f92e>.
60. Pavletich NP, Pabo CO. 1991. Zinc finger-DNA recognition: crystal struc-

- ture of a Zif268-DNA complex at 2.1 Å. *Science* 252:809–817. <https://doi.org/10.1126/science.2028256>.
61. Wolfe SA, Greisman HA, Ramm EI, Pabo CO. 1999. Analysis of zinc fingers optimized via phage display: evaluating the utility of a recognition code. *J Mol Biol* 285:1917–1934. <https://doi.org/10.1006/jmbi.1998.2421>.
 62. Houbaviy HB, Usheva A, Shenk T, Burley SK. 1996. Cocystal structure of YY1 bound to the adeno-associated virus P5 initiator. *Proc Natl Acad Sci U S A* 93:13577–13582. <https://doi.org/10.1073/pnas.93.24.13577>.
 63. Wuttke DS, Foster MP, Case DA, Gottesfeld JM, Wright PE. 1997. Solution structure of the first three zinc fingers of TFIIIA bound to the cognate DNA sequence: determinants of affinity and sequence specificity. *J Mol Biol* 273:183–206. <https://doi.org/10.1006/jmbi.1997.1291>.
 64. Nolte RT, Conlin RM, Harrison SC, Brown RS. 1998. Differing roles for zinc fingers in DNA recognition: structure of a six-finger transcription factor IIIA complex. *Proc Natl Acad Sci U S A* 95:2938–2943. <https://doi.org/10.1073/pnas.95.6.2938>.
 65. Segal DJ, Dreier B, Beerli RR, Barbas CF. 1999. Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5'-GNN-3' DNA target sequences. *Proc Natl Acad Sci U S A* 96:2758–2763. <https://doi.org/10.1073/pnas.96.6.2758>.
 66. Luscombe NM, Laskowski RA, Thornton JM. 2001. Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res* 29:2860–2874. <https://doi.org/10.1093/nar/29.13.2860>.
 67. Chu X, Wang J. 2014. Specificity and affinity quantification of flexible recognition from underlying energy landscape topography. *PLoS Comput Biol* 10:e1003782. <https://doi.org/10.1371/journal.pcbi.1003782>.
 68. Ilsley MD, Huang S, Magor GW, Landsberg MJ, Gillinder KR, Perkins AC. 2019. Corrupted DNA-binding specificity and ectopic transcription underpin dominant neomorphic mutations in KLF/SP transcription factors. *BMC Genomics* 20:417. <https://doi.org/10.1186/s12864-019-5805-z>.
 69. Mitchell WB, Gnanapragasam MN, Jaffray JA, Bieker J, Manwani D. 2011. Case report of erythroid transcription factor EKLK mutation causing a rare form of congenital dyserythropoietic anemia in a patient of Taiwanese origin. *Blood* 118:2154. <https://doi.org/10.1182/blood.V118.21.2154.2154>.
 70. Soni S, Pchelintsev N, Adams PD, Bieker JJ. 2014. Transcription factor EKLK (KLF1) recruitment of the histone chaperone HIRA is essential for beta-globin gene expression. *Proc Natl Acad Sci U S A* 111:13337–13342. <https://doi.org/10.1073/pnas.1405422111>.
 71. Sengupta T, Chen K, Milot E, Bieker JJ. 2008. Acetylation of EKLK is essential for epigenetic modification and transcriptional activation of the beta-globin locus. *Mol Cell Biol* 28:6160–6170. <https://doi.org/10.1128/MCB.00919-08>.
 72. Brown RC, Pattison S, van Ree J, Coghill E, Perkins A, Jane SM, Cunningham JM. 2002. Distinct domains of erythroid Kruppel-like factor modulate chromatin remodeling and transactivation at the endogenous beta-globin gene promoter. *Mol Cell Biol* 22:161–170. <https://doi.org/10.1128/mcb.22.1.161-170.2002>.
 73. Chen X, Bieker JJ. 2004. Stage-specific repression by the EKLK transcriptional activator. *Mol Cell Biol* 24:10416–10424. <https://doi.org/10.1128/MCB.24.23.10416-10424.2004>.
 74. Quadrini KJ, Bieker JJ. 2002. Kruppel-like zinc fingers bind to nuclear import proteins and are required for efficient nuclear localization of erythroid Kruppel-like factor. *J Biol Chem* 277:32243–32252. <https://doi.org/10.1074/jbc.M205677200>.
 75. Sengupta T, Cohet N, Morle F, Bieker JJ. 2009. Distinct modes of gene regulation by a cell-specific transcriptional activator. *Proc Natl Acad Sci U S A* 106:4213–4218. <https://doi.org/10.1073/pnas.0808347106>.
 76. Southwood CM, Downs KM, Bieker JJ. 1996. Erythroid Kruppel-like factor exhibits an early and sequentially localized pattern of expression during mammalian erythroid ontogeny. *Dev Dyn* 206:248–259. [https://doi.org/10.1002/\(SICI\)1097-0177\(199607\)206:3<248::AID-AJA3>3.0.CO;2-I](https://doi.org/10.1002/(SICI)1097-0177(199607)206:3<248::AID-AJA3>3.0.CO;2-I).
 77. Vorackova I, Suchanova S, Ulbrich P, Diehl WE, Ruml T. 2011. Purification of proteins containing zinc finger domains using immobilized metal ion affinity chromatography. *Protein Expr Purif* 79:88–95. <https://doi.org/10.1016/j.pep.2011.04.022>.
 78. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* 37:W202–W208. <https://doi.org/10.1093/nar/gkp335>.
 79. Quadrini KJ, Gruzglin E, Bieker JJ. 2008. Non-random subcellular distribution of variant EKLK in erythroid cells. *Exp Cell Res* 314:1595–1604. <https://doi.org/10.1016/j.yexcr.2008.01.033>.
 80. Keys JR, Tallack MR, Hodge DJ, Cridland SO, David R, Perkins AC. 2007. Genomic organisation and regulation of murine alpha haemoglobin stabilising protein by erythroid Kruppel-like factor. *Br J Haematol* 136:150–157. <https://doi.org/10.1111/j.1365-2141.2006.06381.x>.