2082–2095 *Nucleic Acids Research, 2018, Vol. 46, No. 4* *Published online 4 January 2018*
doi: 10.1093/nar/gkx1299

Structural basis of gene regulation by the Grainyhead/CP2 transcription factor family

Qianqian Ming[1,2], Yvette Roske[1], Anja Schuetz[1,3], Katharina Walentin[4], Ibraim Ibraimi[4], Kai M. Schmidt-Ott[4,5] and Udo Heinemann[1,2,3,*]

[1]Macromolecular Structure and Interaction, Max Delbrück Center for Molecular Medicine, Robert-Rössle-Str. 10, 13125 Berlin, Germany, [2]Chemistry and Biochemistry Institute, Freie Universität Berlin, Takustr. 6, 14195 Berlin, Germany, [3]Helmholtz Protein Sample Production Facility, Max Delbrück Center for Molecular Medicine, Robert-Rössle-Str. 10, 13125 Berlin, Germany, [4]Molecular and Translational Kidney Research, Max Delbrück Center for Molecular Medicine, Robert-Rössle-Str. 10, 13125 Berlin, Germany and [5]Department of Nephrology, Charité Medical University, Charitéplatz 1, 10117 Berlin, Germany

Received September 26, 2017; Revised December 14, 2017; Editorial Decision December 16, 2017; Accepted December 20, 2017

ABSTRACT

Grainyhead (Grh)/CP2 transcription factors are highly conserved in multicellular organisms as key regulators of epithelial differentiation, organ development and skin barrier formation. In addition, they have been implicated as being tumor suppressors in a variety of human cancers. Despite their physiological importance, little is known about their structure and DNA binding mode. Here, we report the first structural study of mammalian Grh/CP2 factors. Crystal structures of the DNA-binding domains of grainyhead-like (Grhl) 1 and Grhl2 reveal a closely similar conformation with immunoglobulin-like core. Both share a common fold with the tumor suppressor p53, but differ in important structural features. The Grhl1 DNA-binding domain binds duplex DNA containing the consensus recognition element in a dimeric arrangement, supporting parsimonious target-sequence selection through two conserved arginine residues. We elucidate the molecular basis of a cancer-related mutation in Grhl1 involving one of these arginines, which completely abrogates DNA binding in biochemical assays and transcriptional activation of a reporter gene in a human cell line. Thus, our studies establish the structural basis of DNA target-site recognition by Grh transcription factors and reveal how tumor-associated mutations inactivate Grhl proteins. They may serve as points of departure for the structure-based development of Grh/CP2 inhibitors for therapeutic applications.

INTRODUCTION

The Grh/CP2 family of transcription factors (TF) comprises two distinct divisions, CP2 (CCAAT box-binding protein 2) and Grh (grainyhead). Members of this TF family are widely found in diverse taxa, ranging from fungi to animals. The first member of the Grh/CP2 family was identified in *Drosophila* when mutant embryos had slack and fragile cuticles, as well as 'grainy' and discontinuous head skeletons (1,2). In humans, six homologs are known with LSF, LBP-1a and LBP-9 belonging to the CP2 subfamily (3), and grainyhead-like (Grhl) 1–3 constituting the Grh subfamily (4). Proteins of the CP2 subfamily are generally expressed ubiquitously (5), while the expression pattern of the Grh subfamily is rather tissue- and developmental stage-specific (4).

In animals, Grhl proteins are predominantly expressed in epithelial tissues and are essential regulators of epithelial development and extracellular barrier repair after tissue damage (4,6,7). *GRHL2* or *GRHL3* null mutations in mice lead to embryonic lethality with defects in dorsal/ventral closure (8,9), while mice lacking *GRHL1* exhibit delayed hair coat growth, defective hair anchoring and palmoplantar keratoderma. Many studies have been devoted to identifying Grhl target genes. Grhl1 was described to specifically regulate expression of the desmosomal cadherin desmoglein-1 (Dsg1), and phenotypes of *GRHL1*-null mice are similar to that of a *DSG1* mutation (10). In developing epithelia, Grhl2 regulates genes encoding components of the apical junctional complex of epithelial cells, like E-cadherin (Cdh1) and claudin-4 (Cldn4) (11,12). In the placenta, Grhl2 transactivates the serine protease inhibitor Kunitz type 1 (Spint1), controlling trophoblast integrity and labyrinth formation (13). Grhl3 can regulate the production of transglutaminase (TGase) 1, the enzyme re-

*To whom correspondence should be addressed. Tel: +49 30 9406 3420; Fax: +49 30 9406 2548; Email: heinemann@mdc-berlin.de

quired for covalent crosslinking of cuticular structural components (14–16).

In recent years, Grhl1–3 have also been implicated in several different types of cancer (6,17). Grhl1, for example, acts as a tumor suppressor in squamous cell carcinoma (SCC) of the skin (18) and neuroblastoma (19). Grhl2 and Grhl3, respectively, directly regulate diverse genes relevant to cancer (20–23). In breast cancer and colorectal cancer, Grhl2 is involved in controling the epithelial–mesenchymal transition (EMT) during tumor progression (24–27). Additionally, Grhl2 is reported as a regulator of human telomerase reverse transcriptase (hTERT), the catalytic subunit of the telomerase which plays a critical role in human carcinogenesis through the maintenance of telomeres (28,29).

Phylogenetic analysis indicates that the Grh/CP2 family originated prior to the metazoan-fungal divergence and the diversification of the two subfamilies probably occurred in an ancient animal lineage (30). Sequence analyses suggest that all Grh/CP2 family members share a common domain architecture (Figure 1A) characterized by an intrinsically unstructured N-terminal transactivation domain (TAD), a conserved DNA-binding domain (DBD), also termed CP2 binding domain, and a C-terminal dimerization domain (DD) (30,31). The Grh and CP2 subfamilies share only about 20% sequence conservation between their DBDs. However, the DBD is more conserved within the Grh family, with a sequence identity of 81% between Grhl2 and Grhl3, and of 63% comparing Grhl1 to Grhl2 or Grhl3. Sequence-specific binding of a Grhl DBD to its recognition element is a crucial event in transcription initiation of a target gene. Mammalian CP2 binds as a tetramer (5,32) to a DNA core sequence 5′-CNRG-$N_{5/6}$-CNRG-3′ (N = any nucleotide, R = purine) (33), while Grhl factors specifically recognize the consensus sequence 5′-AACCGGTT-3′ (14,15,34) and bind to it as dimers (35).

To date, no experimental structure has been reported for Grh or CP2 factors or any of their domains. Here, we present crystal structures of Grhl1 DBD and Grhl2 DBD. The structures are closely similar, and their fold resembles the DBD of the tumor suppressor p53 DNA-binding domain. Moreover, we determined the crystal structure of Grhl1 DBD bound to its consensus DNA response element. The structure analyses combined with mutagenesis studies, identify the molecular basis of DNA binding and provide a direct explanation of how a single mutation in the DNA binding site is linked to cancer.

MATERIALS AND METHODS

DNA and expression constructs

Gel-filtration-purified DNA oligonucleotides used for crystallization and biochemical assays were obtained from Biotez (Berlin, Germany). The DNA oligonucleotides were dissolved in DNA buffer containing 20 mM HEPES–NaOH, 100 mM NaCl, at pH 7.5. To prepare duplex DNA, well-mixed solutions of single strands were heated to 95°C and kept for 5 min. Strands were then annealed by slowly cooling down to RT in a thermomixer. For electrophoretic mobility shift assays (EMSA), DNA oligonucleotides were 5′ labeled with cyanine-5 fluorescent dye. All DNA oligonucleotides were stored at –20°C.

The Grhl1-DBD construct (aa 248–485) and the N-terminal truncation construct (ΔN_Grhl1, aa 248–618) of Grhl1 were PCR-amplified from human *GRHL1* cDNA (DNASU, Tempe, USA). Human Grhl2 DBD (aa 217–492) was cloned from *GRHL2* cDNA (Source BioScience, Berlin, Germany). Both Grhl1 DBD and Grhl2 DBD were cloned into the pQlinkH vector (36) between the BamHI and NotI restriction sites for expression as N-terminally His_7-tagged recombinant proteins. The ΔN_Grhl1 construct was subcloned into modified pET28a, yielding a product protein with a non-cleavable C-terminal His tag.

Protein expression and purification

For overexpression, each plasmid was transformed into competent *Escherichia coli* Rosetta (DE3) cells. In large-scale culture, cells were grown to an OD_{600} of 1.0 in terrific broth (TB) medium and subsequently induced at 17°C with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) overnight. All steps of protein purification were carried out at 4°C. The harvested cells were resuspended using lysis buffer (1× PBS pH 7.5, 500 mM NaCl, 5% (v/v) glycerol, 0.5 mM 1,4-dithiothreitol (DTT), 1 mg ml^{-1} lysozyme and adding 1 protease inhibitor cocktail tablet (Roche) per 50 ml buffer). The cell lysate was centrifuged to remove the debris. The supernatant was applied to a Ni^{2+}-nitrilotriacetic acid (NTA) column, followed by His-tag cleavage using tobacco etch virus (TEV) protease which was subsequently removed using a Ni^{2+}-NTA column. Grhl1 DBD was further purified by cation-exchange chromatography prior to Superdex 75 (GE Healthcare) size-exclusion chromatography (SEC) in buffer C (20 mM HEPES–NaOH, 125 mM NaCl, 2 mM DTT, pH 7.2). Grhl2 DBD was applied to Superdex 75 directly after the second Ni^{2+}-NTA step. For ΔN_Grhl1, the expression and purification protocol was essentially the same as described for Grhl1 DBD, except that the TEV-cleavage step was omitted. Finally, purified proteins were concentrated and stored at –80°C.

Selenomethionine (SeMet)-labeled Grhl1 DBD was expressed in *E. coli* strain Rosetta (DE3) (Novagen). Initially, 50 ml of Grhl1-DBD overnight culture with LB medium were incubated at 37 °C. The cells were spun down, then washed once with M9 medium. Subsequently, the pellet was resuspended using 2 l M9 medium and cultured at 37°C until the OD_{600} reached 0.8. The culture was cooled to 17°C, and lysine, threonine, phenylalanine at 100 mg l^{-1}, leucine, isoleucine, valine, and SeMet at 50 mg l^{-1} were added to down-regulate methionine synthesis (37). After incubation for 15 min, cell growth was induced as described for the native protein.

To prepare the protein–DNA complex for crystallization, purified Grhl1 DBD was mixed with double-stranded DNA at a molar ratio of 1:0.6 by adding protein slowly into DNA solution on ice. Excess unbound DNA was removed by SEC with Superdex 75 in buffer D (20 mM HEPES, 100 mM NaCl, 2 mM DTT). The complex was finally concentrated to 18 mg ml^{-1} and stored at –80°C.

Protein crystallization

All crystallization screens were performed using the sitting-drop vapor-diffusion method by mixing 200 nl protein solu-

tion and 200 nl reservoir buffer at 4°C, with a reservoir volume of 75 μl in 96-well plates. A solution of Grhl1-DBD at 18 mg ml^{-1} was supplemented with 2 mM KI and crystallized in a reservoir solution containing 2.1 M $(NH_4)_2SO_4$ and 0.1 M NaOAc, pH 5.5. Under the same conditions, crystals of SeMet-incorporated Grhl1 DBD and mutant Grhl1-DBD R427Q were obtained as well. The above crystals were frozen directly in liquid nitrogen. Grhl2 DBD was crystallized at 15 mg ml^{-1} in a solution containing 21% PEG 3350, 0.2 M sodium formate. Before flash-freezing, the crystal was transferred into a cryoprotectant consisting of reservoir solution supplemented with 30% (v/v) glycerol. Crystals of the Grhl1-DBD:DNA complex were obtained in the presence of 10% (v/v) isopropanol as precipitant with and 0.1 M citric acid, pH 4.0. They were soaked in 20% ethylene glycol for cryoprotection prior to freezing.

Structure analysis and refinement

X-ray diffraction data from single crystals were collected on beamline 14.1 at BESSY II (Berlin, Germany) (38). The wavelengths chosen for data collection were 0.9764 Å for Grhl1-DBD SeMet, 1.5000 Å for Grhl1 DBD and 0.9184 Å for all other datasets, including Grhl1-DBD R427Q, Grhl2 DBD and the Grhl1-DBD:DNA complex. Initial indexing and determination of an optimal data collection strategy were performed using iMOSFLM (39). Data collection and refinement statistics are reported in Table 1. The recorded diffraction data were integrated and scaled with XDSapp (40).

The crystal structure of Grhl1 DBD was determined in two steps. First, an initial atomic model was built with phases calculated from the SeMet derivative by the SAD method using PHENIX AutoSol (41). This model was then used as a search template for molecular replacement to phase the 2.34-Å resolution dataset of native Grhl1 DBD using the program MOLREP (42). The model was further extended using PHENIX Autobuild (43) and completed manually within Coot (44). After several cycles of refinement within PHENIX, the final model was obtained yielding the statistics shown in Table 1. The Ramachandran map produced by MolProbity (45) shows that 94% of the residues are in favored regions with no outliers.

The crystal structures of Grhl2 DBD, Grhl1-DBD R427Q and Grhl1-DBD:DNA were determined by molecular replacement using the program PHASER (46) and the Grhl1-DBD crystal structure as a search model. The Grhl2-DBD structure was refined with autoBuster (47). In all four crystal structures determined, several terminal and loop residues could not be modeled, presumably due to disorder. The structure of Grhl1 DBD comprises residues 249–481 except for three segments without electron density (aa 386–387, aa 290–297, aa 437–454). In the structure of Grhl2 DBD, residues 247–485 are present in chain A and residues 247–484 in chain B. In chain A, aa 258–264, aa 284–287, aa 381–385 and aa 431–463 are missing, and in chain B aa 261–264, aa 284–287, aa 379–387 and aa 433–459 are missing. In the structure of the Grhl1-DBD:DNA complex, residues 250–478 of chain A (aa 264–266, aa 288–297, aa 401–402, aa 404–406 and aa 439–453 missing) and residues 250–477 of chain B (aa 266–268, aa 289–297 and aa 441–454 missing) are present. The structure of the Grhl1-R427Q mutant comprises residues 249–480, and aa 291–297, aa 386–387 and aa 437–454 are not included.

The data and refinement statistics are presented in Table 1. No protein model has residues in the disallowed region of the Ramachandran map. The DNA conformation was analyzed with 3DNA (48). Atomic coordinates were aligned based on matching Cα positions with the online server PDBeFOLD (49). Figures were generated with PyMOL and ESPript.

Electrophoretic mobility shift assay

The binding reaction mixture (final volume of 10 μl) contained binding buffer (20 mM HEPES–NaOH pH 7.4, 100 mM NaCl, 2 mM DTT and 5% (w/v) glycerol), 1 μM of fluorescently labeled Cy5-consensus (cons) DNA, 100 μM of individual non-labeled DNA (Mut_12mer DNA as non-specific competitor and Cons_12mer DNA as specific competitor) and 4 μM of purified Grhl1 protein: DBD, DBD R427Q, DBD R427A/R430A, DBD R427A, DBD R430A, ΔN_Grhl1. The samples were incubated at 4°C for 1 h. Subsequently, protein–DNA complexes were separated from unbound DNA by electrophoresis on an 8% polyacrylamide gel in ice-cold TAE buffer (Tris acetate EDTA) for 40 min at 100 V. The fluorescence signal was recorded with a Typhoon™ FLA 9500 biomolecular imager (GE Healthcare) at an excitation wavelength of 635 nM.

Size-exclusion chromatography coupled to right-angle light scattering (SEC-RALS)

Each sample, (Grhl1 DBD, Grhl1-DBD:DNA complex, ∼3 mg ml^{-1}, 200 μl) was injected on a pre-equilibrated Superdex 75 (10/300 GL, GE Healthcare) column at a flow rate of 0.5 ml min^{-1} in buffer D. The column was coupled in-line with 270 RALS and VE3580 RI detectors. This technique allows calculation of the absolute molecular mass and oligomeric state of a protein or protein complex at any point in the chromatogram. Data were analyzed with OmniSEC (Malvern) using 4 mg ml^{-1} BSA as a standard.

Isothermal titration calorimetry (ITC)

ITC experiments were performed on a MicroCal VP-ITC microcalorimeter. Before any experiment, all protein and DNA samples were dialyzed overnight against ITC buffer (20 mM HEPES NaOH pH 7.4, 100 mM NaCl, 0.5 mM Tris [2-carboxyethylphosphine] (TCEP)) and degassed. The experiment was performed with an initial half injection (2 μl) followed by a series of injections (7 μl) by adding protein solution (100–300 μM) into the DNA sample (5–10 μM) contained in the cell compartment. The binding curves were fit to the model of one set of binding sites except for Grhl1-DBD R427Q or Grhl1-DBD R427A/R430A, which were fit to the model of sequential binding sites within Origin 7.0.

Thermal shift assay

The environmentally sensitive fluorescent dye Sypro Orange® (Invitrogen, Carlsbad, CA, USA) was used to

Table 1. X-ray data collection and structure refinement statistics

	Grhl1 DBD_SeMet	Grhl1 DBD	Grhl1 DBD_R427Q	Grhl2 DBD	Grhl1 DBD:DNA
Data collection					
Space group	$P6_122$	$P6_122$	$P6_122$	$P2_1$	$P4_1$
Cell dimensions					
a, *b*, *c* (Å)	103.41, 103.41, 116.21	106.75, 106.75, 115.63	104.41, 104.41, 116.26	54.19, 48.20, 101.30	59.03, 59.03, 198.85
α, β, γ (°)	90, 90, 120	90, 90, 120	90, 90, 120	90, 102.92, 90	90, 90, 90
Resolution (Å)	47.24–2.99 (3.17–2.99)[a]	49.02–2.34 (2.42–2.34)	47.62–2.35 (2.49–2.35)	43.37–2.50 (2.59–2.50)	44.09–2.92 (3.09–2.92)
R_{merge}	0.118 (1.397)	0.060 (1.920)	0.119 (3.394)	0.105 (0.910)	0.132 (0.975)
$< I / \sigma(I) >$	17.45 (1.57)	35.62 (1.88)	20.65 (1.23)	11.01 (1.70)	10.63 (1.84)
Completeness (%)	98.7 (97.7)	99. 9 (99.9)	99.7 (98.8)	99.03 (97.6)	99.2 (98.7)
Multiplicity	8.2 (8.2)	24.9 (24.3)	15.7 (15.8)	4.4 (4.3)	4.1 (4.1)
Refinement					
Resolution (Å)		49.0–2.34	47.62–2.35	43.37–2.50	44.09–2.92
No. reflections		31189	16178	17837	14601
R_{work}/R_{free}		0.192 / 0.229	0.222 / 0.251	0.196 / 0.226	0.235 / 0.279
No. atoms					
Protein		1680	1613	3194	3590
DNA					485
Water		39	10	45	20
Mean *B* factors (Å^2)					
Protein		95.1	79.9	89.8	71.4
DNA					46.0
Water		82.1	79.9	54.1	54.4
R.m.s. deviations					
Bond lengths (Å)		0.008	0.011	0.010	0.002
Bond angles (°)		1.05	1.12	1.10	0.53

[a]Values in parenthesis for highest resolution shell.

monitor protein unfolding with respect to temperature. The assay was conducted in quadruplicates using the iQ5 Multicolor real-time PCR detection system (Bio-Rad, Munich, Germany). All experiments were carried out in 25 mM phosphate pH 7.2, 225 mM NaCl, 10% glycerol, and 5 mM DTT. Sample size was 50 μl and the protein concentration 25 μM. The sample was heated from 25°C to 80°C with a rate of 0.5°C min^{-1}. Changes in fluorescence intensity were monitored, and the wavelengths for excitation and emission were 490 and 530 nm, respectively. Melting temperatures (Tm, the temperature midpoint for the protein unfolding transition) were determined from the peak value of the first-derivative curve.

Reporter gene assay

Mouse inner medullary collecting duct cells (mIMCD-3) were purchased from ATCC (CRL-2123). Subsequently, a knockout of *GRHL2* was induced by CRISPR-Cas9 technology. Briefly, two oligonucleotides designed to yield sgRNAs targeting exon 2 of the *GRHL2* gene (5′- CACCG TCGCC TTGGT GGCCG CAGTC-3′; 5′-AAACG ACTGC GGCCA CCAAG GCGAC-3′) were annealed and cloned into BbsI-digested pX330-Cas9-T2A-mCherry plasmid (Addgene, Cambridge, MA, USA). mIMCD-3 cells were transiently transfected using Lipofectamine® 3000 transfection reagent (Thermo Fisher Scientific, Waltham, USA). After 48 h, mCherry-positive cells were selected by fluorescence-activated cell sorting and replated for culturing. Single-clonally derived colonies were picked and analyzed for mutations at the target region of *GRHL2* exon 2. Three clones (designated 21b, 24b and 26c) were found to carry frame shift-inducing exon 2 mutations resulting in homozygous inactivation of *GRHL2*.

For reporter assays, wildtype and *GRHL2* knockout mIMCD-3 cells were transiently transfected using Lipofectamine® 3000 transfection reagent. Reporter assays were performed according to manufacturer instructions using the firefly luciferase reporter vector pGL3-Basic and the Renilla luciferase plasmid pRL-SV40 (Promega, Madison, WI, USA) for normalization. The *CLDN4* reporter construct was generated by inserting bases –611 to +174 relative to the mouse *CLDN4* TSS into pGL3-Basic vector as described previously (11). Full-length human *GRHL1* and *GRHL2* expression vectors (EX-I2097-M02, EX-W2222-M02) were purchased from GeneCopoeia (Rockville, MD, USA). Mutagenized constructs (Grhl1 R427Q, Grhl1 R427A, Grhl2 R423Q and Grhl2 R423A) were generated using the QuikChange II XL Site-Directed Mutagenesis Kit (Agilent Technologies, Santa Clara, USA) and primers listed in Supplementary Table S1. For generating the empty control vector the human *GRHL1* sequence was excised from the wildtype EX-I2097-M02 plasmid. Subsequently, the backbone was blunt-ended and religated. Cells were harvested 72 h after transfection, and reporter activity was assayed via the Dual-Luciferase® Reporter Assay System (Promega) and a BioTek Synergy™ HT Multi-Detection Microplate Reader.

RESULTS

Overall architecture of Grhl1 DBD and Grhl2 DBD

Prior to this study, no structural data were available for any member of the Grh/CP2 transcription factor family. An initial prediction of disorder and secondary structure with PrDOS (50) and psipred (51) suggested that the DNA-binding and dimerization domains of the Grhl proteins are more structured than their transactivation domains. Based on this observation, we produced two soluble constructs of Grhl1: ΔN_Grhl1 (aa 248–618) and Grhl1 DBD (aa 217–492) (Supplementary Figure S1A). Both constructs retained efficient DNA binding as detected by electrophoretic mobility-shift assays (EMSA, see below). Due to problems with degradation of ΔN_Grhl1, and failure to crystallize that protein construct, we focused on Grhl1 DBD to study the structural basis for DNA binding and target-site recognition. The Grhl1-DBD crystal structure was determined at 2.3 Å resolution by single-wavelength anomalous diffraction (SAD) phasing using selenomethionine-labeled protein (Table 1). The asymmetric unit of the crystal contained one molecule of the Grhl1 DBD (Figure 1B).

Grhl1 DBD contains an immunoglobulin- (Ig-) like core with two long, twisted antiparallel β sheets of five (β1, β3, β8, β5 and β11) and six (β2, β4, β6, β7, β9 and β10) strands (Figure 1C), respectively. This core is decorated by three α helices and a series of surface loops. A C-terminal extension wrapping halfway around of the core (Supplementary Movie S1) was found to be important for the solubility and also for the stability of Grhl1 DBD during purification.

To assess the degree of structural similarity between the Grhl factors, the crystal structure of Grhl2 DBD (aa 247–484) was determined at 2.5 Å resolution. The asymmetric unit of the Grhl2-DBD crystal contains two protein chains with closely similar conformation. The DBD of Grhl1 and Grhl2 could be superimposed with a mean root-mean-square deviation (RMSD) of 0.5 Å between 160 matching Cα positions, demonstrating that the two structures are nearly identical (Supplementary Figure S1B).

Surprisingly, a search of the Protein Data Bank (PDB) using PDBeFold identified the tumor suppressor protein p53 as the closest structural homolog of Grhl1 (Q score = 0.26, 2GEQ, chain A). Although the two proteins' DNA-binding domains share only 10% sequence identity, their matching Cα positions are superimposed with an RMSD of 2.2 Å, indicating a significant structural similarity. The structural homology is most pronounced in the Ig-like core of the DBD, whereas surface loops and decorating α helices appear more variable (Figure 1E). The DNA-binding loop-strand-helix (L1, S10, and H2) motif of p53 is in a different orientation compared to the corresponding region in Grhl1 DBD, underlining different recognition element preferences for DNA binding by the two transcription factors.

DNA binding and target-site recognition by Grhl1 DBD

To reveal the structural basis of DNA target-site recognition and binding by Grhl1, we determined the crystal structure of Grhl1 DBD bound to a symmetric 12-base-pair DNA duplex centered about the Grhl consensus binding sequence (AACCGGTT) at 2.9 Å resolution. The crystal contained two Grhl1-DBD molecules and one DNA duplex in each asymmetric unit, where the protein chains and the strands of the DNA double helix (or two 6-bp half helices) are related by non-crystallographic dyad symmetry (Figure 2A). The two Grhl1-DBD chains are arranged nearly symmetrically to opposite faces of the standard B-form DNA and contact each other only via the L10 loop by forming hydrogen-bond interactions between the central Lys386 and the carbonyl oxygens of residue Ser383 and Thr380 in the opposite chain (Supplementary Figure S2A). The minor groove is widened at the central CG base pairs up to a cross-groove P-P distance of 15.7 Å to accommodate the L10 residues, whereas the major groove width remains fairly constant over the entire duplex with cross-groove P-P distances between 17.0 and 18.4 Å. The variation in DNA groove widths is linked to the slightly unusual O4'-*endo* sugar pucker of the central cytosines and χ torsion angles around -99° of the central guanosines. In this arrangement, Lys386 may contribute to the overall stabilization of the DBD–DNA complex. The C-termini of the two Grhl1-DBD chains are oriented towards the same face of the DNA, from where the dimerization domains would extend to interact with each other (Figure 2A).

When superimposed, the two Grhl1-DBD chains present in the asymmetric unit have essentially identical conformation, suggesting that their structure is not strongly influenced by different crystal environments. Similarly, comparing the structures of free with DNA-bound Grhl1 DBD reveals high structural similarity with a mean RMSD of 0.54 Å for 158 matching Cα atoms (Supplementary Figures S3A and B). These results establish that the Grhl1 DBD adopts a pre-formed conformation, permitting target DNA binding without major structural rearrangement. Upon DNA binding, Grhl1 DBD marginally changes the orientation of its DNA-contacting helix α3, which tilts slightly to dock optimally into the major groove, and the loop L10 undergoes a more obvious conformational change. Loop L10 in the Grhl1 DBD:DNA complex is less flexible than in the absence of DNA where it lacks well-defined electron density and displays elevated atomic displacement factors.

The total size of the protein–DNA interface is ∼1400 Å^2 with the Grhl1 DBD presenting a positively charged surface to the DNA (Figure 2B). This surface is decorated mainly with residues from the α3 helix and the L10 loop, which interact with the DNA major and minor groove, respectively (Supplementary Movie S2). Direct hydrogen-bonded or Coulomb interactions with the DNA backbone, supporting unspecific Grhl1-DBD binding to DNA, are mediated by Thr380, Gln385, Lys386, Lys389, Cys421, Lys428 and Arg430 (Figure 2C). Due to the limited resolution of the crystal structure, no water-mediated contacts could be identified. Sequence-specific contacts between the Grhl1 DBD and DNA bases are formed primarily by three conserved amino acids (Supplementary Figure S3D), Arg427 and Gly387, as well as Arg430. The Arg427 guanidinium group plays a pivotal role in anchoring the recognition helix α3 to the DNA major groove via hydrogen bonding to the O6 carbonyl and N7 imine of guanine G8, the most strongly conserved nucleotide in the Grhl consensus binding motif. The same G8 is further interacting with the carbonyl oxygen of Gly387 from the opposite protein chain, forming a

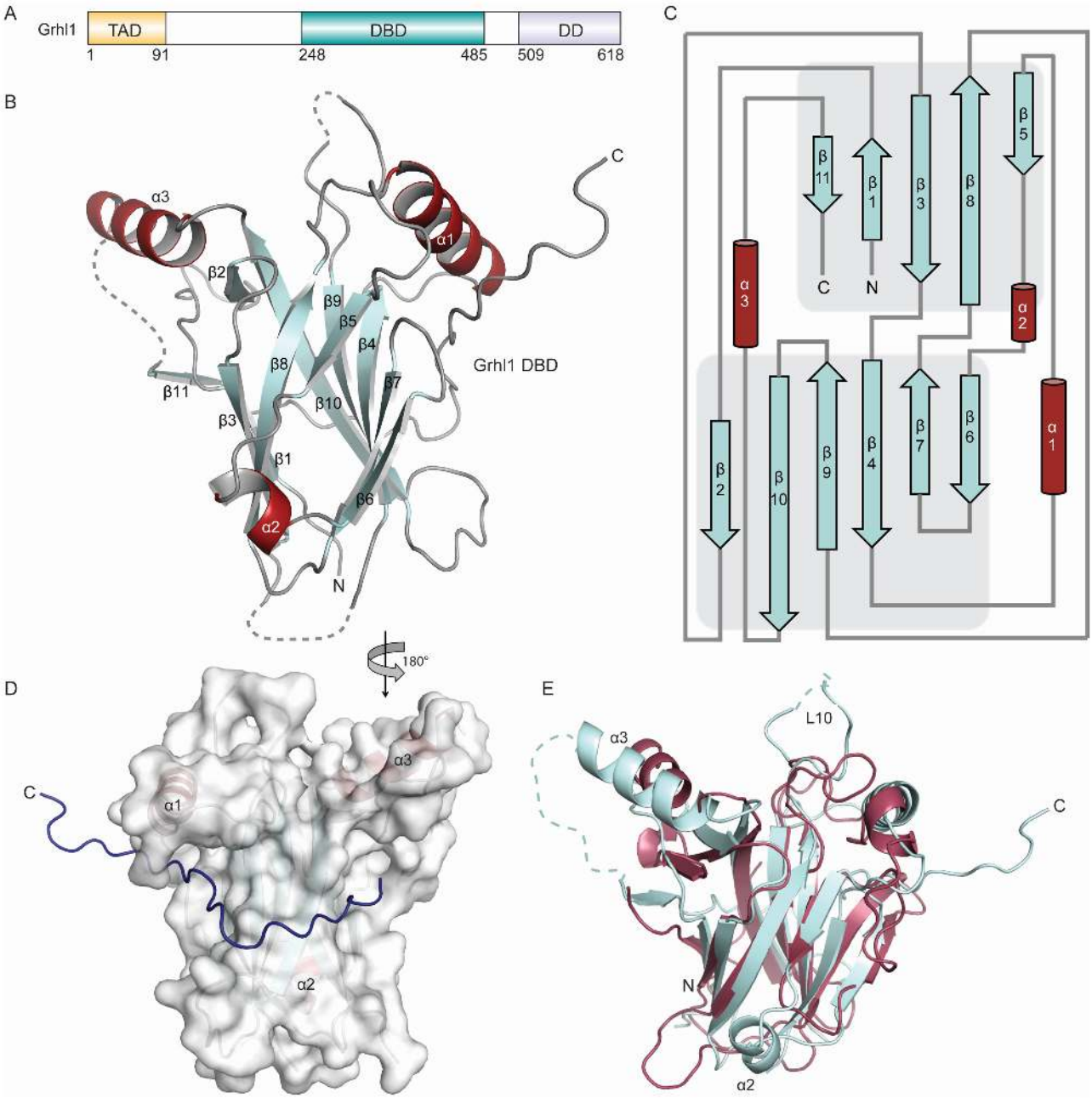

Figure 1. Conserved structure of the Grhl1 DBD shows similar fold to p53. (**A**) Schematic representation of the Grhl1 domain organization. The transactivation domain (TAD; orange), DNA-binding domain (DBD; turquoise) and dimerization domain (DD; light blue) are indicated. Segments predicted as disordered are shown in white. (B) Overall structure of the DBD in cartoon representation, with helices in red and β strands in pale cyan. Dashed lines indicate peptide segments without electron density (aa386–387, aa290–297, aa437–454), presumably due to disorder. (**C**) Topology diagram for the DBD structure: α helices and β strands are labeled as in (A). (**D**) Grhl1 DBD surface (gray) with the C-terminal extension highlighted as blue ribbon. (**E**) Least-squares superimposition of Grhl1 DBD and p53 core domain (red, 2GEQ, chain A) with matching Cα atoms deviating by 2.3 Å. The DNA-contacting helix α3 shows a different orientation, and loop L10 of Grhl1 DBD has a different structure than the corresponding elements in p53. The other two helices of Grhl1 DBD, α1 and α2, are either re-oriented or missing in the p53 structure.

hydrogen bond with the N2 amino group in the DNA minor groove. In each dimeric ensemble, the Grhl1 Arg430 additionally forms a weak hydrogen bond to adenine A3, likely providing additional selectivity for targeting. The DNA-binding interface is overall conserved among members of the Grh/CP2 family (Supplementary Figure S3C).

The crystal structures of DNA-bound Grhl1 DBD and p53 DBD reveal generally similar DNA-binding modes (Supplementary Figure S4A). Superimposing the two DBD:DNA complexes via alignment of one DBD chain from each protein shows clear similarity in the DNA-binding segments, especially the recognition helices α3 of Grhl1 and H2 of p53 Supplementary Figure S4*B*). Inter-

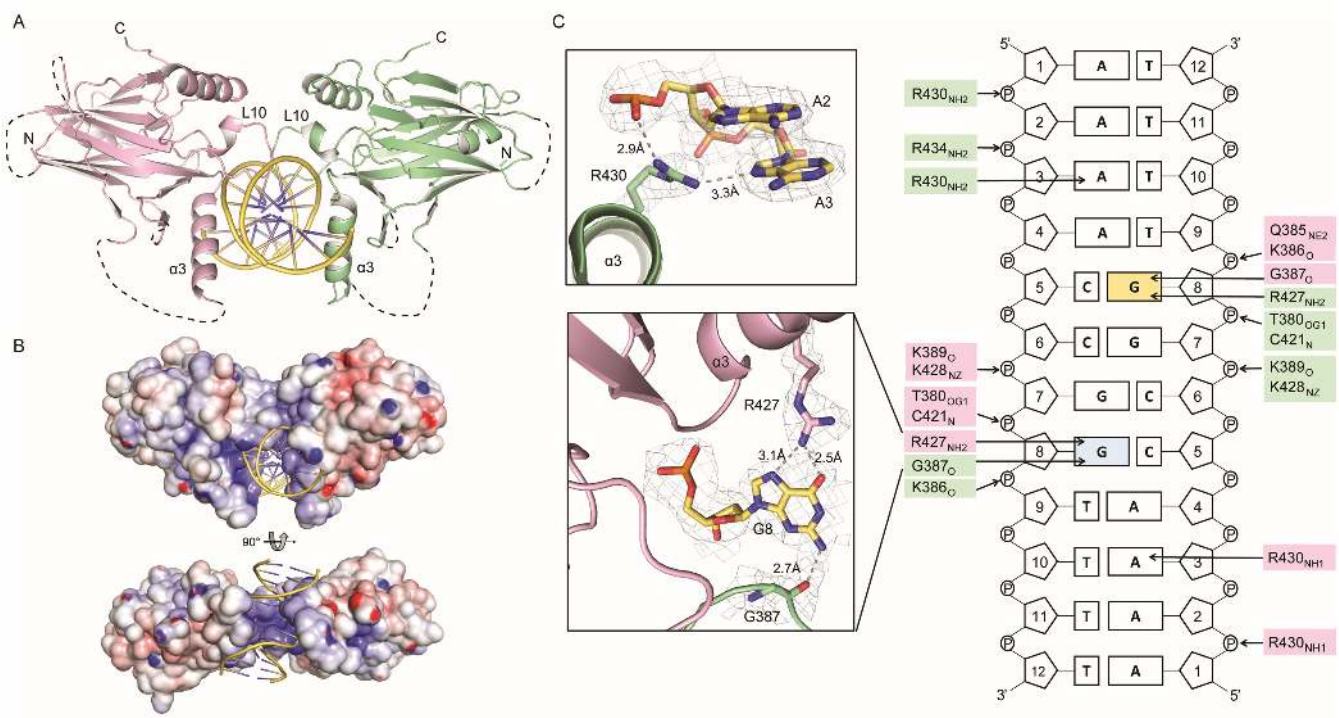

Figure 2. Overall structure of the Grhl1-DBD:DNA complex reveals a parsimonious binding mode. (**A**) Cartoon view of Grhl1 DBD bound symmetrically to a dyad-symmetric 12-mer DNA duplex. Individual monomers (pink and green) form a dimeric interface parallel to the DNA axis, with loop 10 (L10) and helix 3 (α3) interacting with the minor and major groove, respectively. (**B**) Grhl1-DBD DNA-interacting surface presented in two orientation (0° and 90°), colored by electrostatic potential, with positive potential (+10 kT) in blue and negative potential (–10 kT) in red. (**C**) Right panel, scheme showing key hydrogen-bonding and/or Coulomb interactions (arrows, 3.5 Å cut-off) between Grhl1 DBD and the DNA double helix. Grhl1 residues from different chains are shown on pink or green background. Left panels, close-up views of specific interactions with DNA by Arg430 (top) and Arg427 (bottom). $2F_o - F_c$ electron density maps are shown individually, contoured at 0.8σ (gray).

estingly, the central Grhl1 Arg427 superimposes very well with p53 Arg280, both playing essential roles in anchoring their DBD to the DNA major groove through specific interaction with a highly conserved guanine (G8 for Grhl1). However, distinct architectural differences are also apparent in the DNA binding interface. The two recognition helices (α3) in Grhl1 are more deeply buried in the major groove, resulting in a larger interface of 566 Å^2 compared to that of p53 (helix H2, PDB code 2ATA) with 277 Å^2. The different interfaces can be explained by comparing the binding angles between the two helices with respect to the DNA. The binding angle for Grhl1 is 68° versus 130° for p53 (Supplementary Figure S4A), calculated by the Python module 'anglebetweenhelices' implemented in PyMOL. Unlike in p53, where loop L3 is involved in zinc coordination and located at the edge of the DNA backbone, the equivalent L10 loops of Grhl1 fit into the DNA minor groove.

Grhl1-DBD:DNA interaction studies

In the crystal, Grhl1 DBD binds its DNA recognition site in a dimeric arrangement. To assess whether Grhl1 DBD dimerization persists in solution, we analyzed the oligomerization of two different Grhl1 fragments using size-exclusion chromatography and right-angle light scattering (SEC-RALS) in the presence or absence of DNA. In these experiments, Grhl1 DBD was monomeric, but bound target DNA as an apparent dimer (Supplementary Figures S5A and B), in agreement with the crystal structure. However, construct ΔN_Grhl1, containing the dimerization domain, formed dimers efficiently without DNA and bound target DNA as a constitutive dimer (Supplementary Figures S5C and D), which suggests that residues 486–618 contribute to dimerization.

In order to assess the contribution from both specific and non-specific Grhl1-DBD:DNA contacts observed in the crystal structure, we measured target-DNA binding by wild-type and mutant Grhl1-DBD proteins in a series of isothermal titration calorimetry (ITC) experiments (Figure 3 and Supplementary Table S2). Grhl1 DBD bound the 12-bp target DNA with a dissociation constant (K_d) of 90 nM. When single DNA-contacting protein side chains were replaced by alanine, the mutant Grhl1-DBD variants showed at least a two-fold reduction in DNA binding, with the exception of C421A, whose DNA binding remained unchanged. A significantly stronger, 12-fold drop in affinity was observed for the Grhl1-DBD R427A mutant, which eliminated two specific hydrogen bonds to G8, the most conserved nucleotide in the target DNA. Strikingly, an even stronger reduction in affinity was measured for the R427Q mutant, which is highlighted as cancer-derived in the Catalog of Somatic Mutations in Cancer (COSMIC (52)) and the Cancer Genome Atlas (53). This observation points to a central role of Arg427 in the physiological function of Grhl1 as a transcription factor.

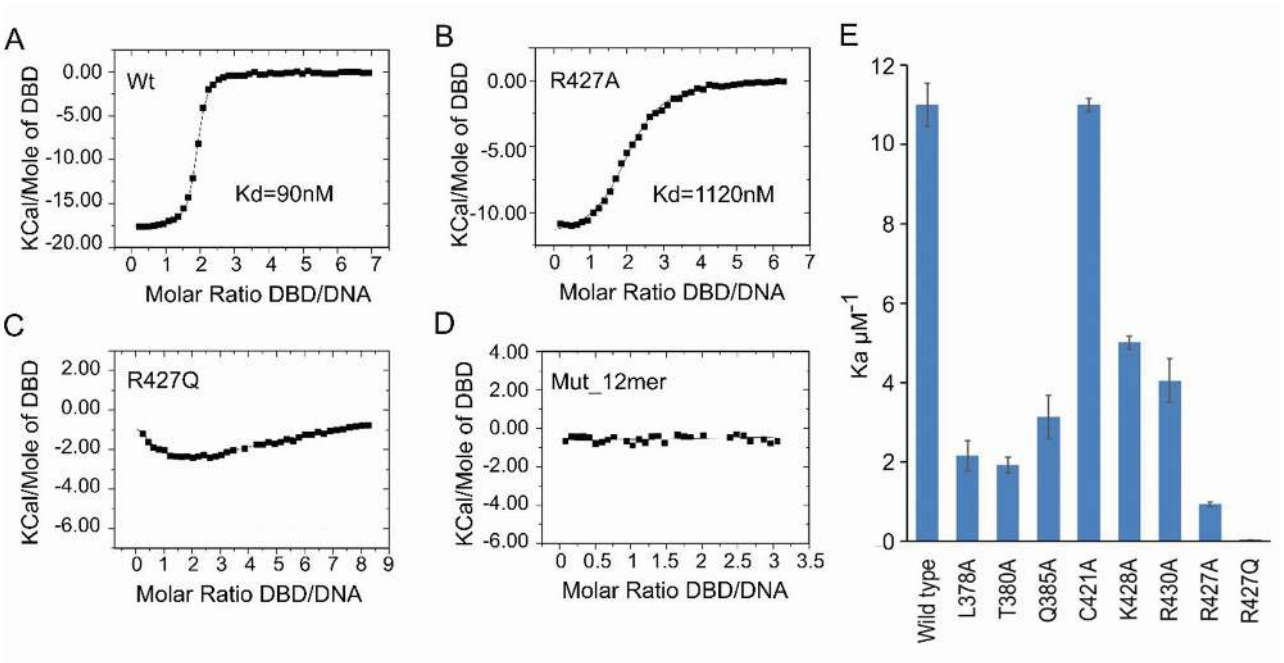

Figure 3. Effects of mutating DNA-contacting residues on DNA-binding affinity of Grhl1 DBD. ITC binding curves for injection of (**A**) Grhl1 DBD, (**B**) Grhl1-DBD R427A and (**C**) Grhl1-DBD R427Q into a micro calorimeter cell containing 12-bp consensus DNA fragment (Cons, see Methods section). Curves were fit using a one-site binding model for Grhl1 DBD and Grhl1-DBD R437A, and two sequential binding sites for Grhl1-DBD R427Q. (**D**) ITC binding curve for titrating Grhl1 DBD to mutant DNA (Mut_12mer). A fit to the binding isotherm is not possible. (**E**) Histogram showing association constants for the binding of different Grhl1-DBD proteins to consensus DNA.

Further, we analyzed the impact on DNA binding of Lys386, which is located within the DNA minor groove without directly interacting with DNA. Lys368 mediates contacts between the L10 loops of two Grhl1-DBD chains, thereby acting as a clamp stabilizing the dimeric arrangement of DNA-bound Grhl1. Interestingly, substitution of this residue by alanine resulted in a dramatic 17-fold decrease in DNA binding affinity, as determined by ITC (Supplementary Figure S2B), but the 2:1 binding stoichiometry remained unchanged. Whether protein-protein contacts mediated by Lys386 or Coulomb interactions of the lysine side chains in the DNA minor groove contribute more to Grhl1-DNA binding is unclear at present.

In the defined consensus response element (RE) for Grhl proteins (AACCGGTT), the first C and the second G nucleotide (underlined, C5 and G8 in the co-crystallized dodecamer) are most highly conserved (14). To verify the importance of this conservation, Grhl1-DBD binding experiments were performed with a mutant RE in which C5 and G8 were symmetrically replaced with T and A, respectively (Mut_12mer). In ITC measurements, no binding of Mut_12mer to Grhl1 DBD was detected (Figure 3D). This result was confirmed in an electrophoretic mobility shift assay (EMSA) in which a fluorescent-labeled RE bound to Grhl1-DBD could be outcompeted by excess unlabeled RE DNA, but not by Mut_12mer (Figure 4A and Supplementary Table S3).

It is known that in p53 the central dinucleotide of each half-recognition element is essential for DNA binding (54–57). To probe whether Grhl1 DNA target-site recognition is under a similar restraint, we tested Grhl1-DBD binding to different 12-mer DNA duplexes with variable central dinucleotides at positions 6 and 7. Mutation of these nucleotides increased K_d values by 3–6-fold (Supplementary Table S4), following a trend of CC (CG) < AA < AT (TA). These results suggest that Grhl1 DNA binding strength is also sensitive to the variation of the central dinucleotide.

An EMSA experiment as described for Mut_12mer also confirmed that a Grhl1 variant comprising the DBD and DD but lacking the N-terminal TAD (ΔN_Grhl1) shows similar DNA binding as Grhl1 DBD alone. In further EMSA experiments, Grhl1-DBD mutants displayed reduced RE binding (Figure 4*A*) in accordance with their relative affinities established by ITC. In summary, the *in vitro* DNA-binding experiments are in full agreement with the Grhl1-DBD:DNA crystal structure by corroborating the central role of both Grhl1 Arg427 and the G8 nucleotide for sequence-specific binding.

Grhl1-R427Q mutation impedes Arg430-DNA contacts

The crystal structure of the Grhl1-DBD:DNA complex did not rationalize the significantly stronger reduction in DNA affinity by the cancer-related R427Q mutation as compared to R427A. To reveal the structural basis for this effect, the crystal structure of Grhl1-DBD R427Q was determined at a nominal resolution of 2.35 Å. The effective resolution of this structure may be lower due to the inclusion of weak high-resolution diffraction data. Whereas the overall structure remained unchanged compared to wild-type Grhl1 DBD, two residues in the DNA binding interface were clearly affected. In the R427Q mutant, the neighboring

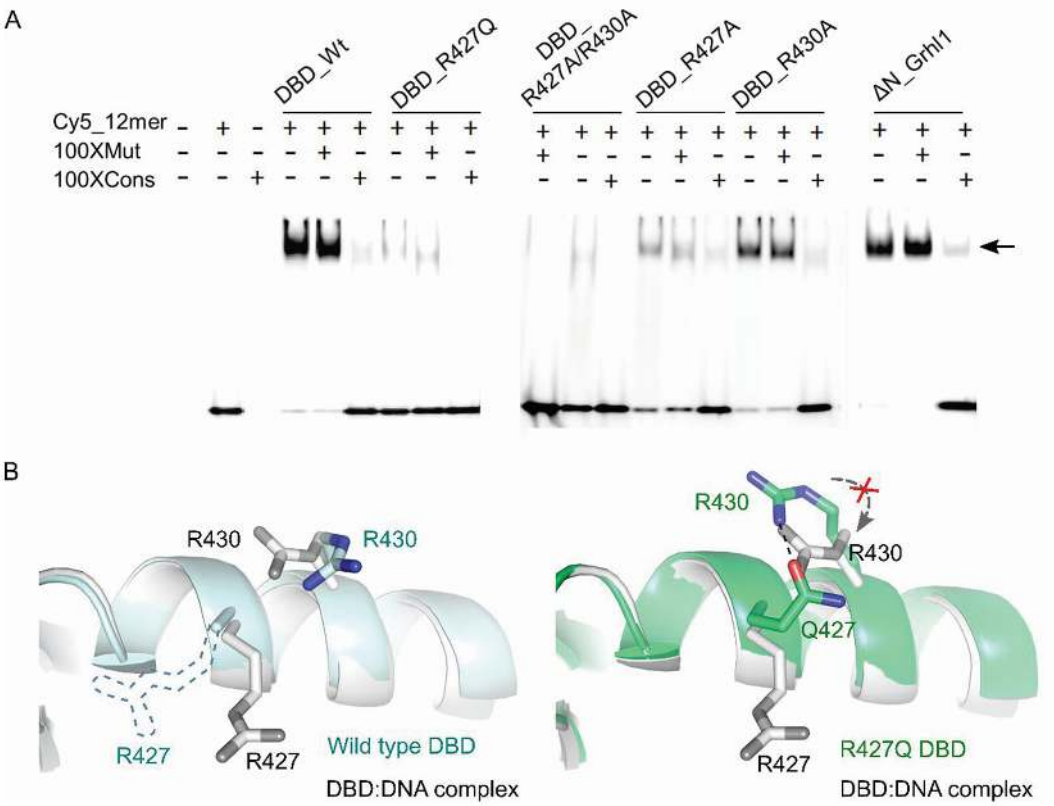

Figure 4. Electrophoretic mobility shift assays (EMSA) with different Grhl1 variants and structural consequence of the Grhl1 R427Q mutation. (**A**) EMSA with fluorophore (cyanine-5) labeled double-stranded Grhl consensus 12-mer DNA (Cy5_12mer) in the (+) presence or (−) absence of Grhl1 variants and 100-fold excess of unlabeled competitor DNA. Two types of competitor DNAs are applied, unlabeled Grhl consensus fragment (100xCons, Cons_12mer in 100-fold excess over Cy5_12mer) and unlabeled mutated fragment (100xMut, Mut_12mer in 100-fold excess over Cy5_12mer). The position of bands representing DNA-protein complexes is marked with an arrow. Grhl1-DBD R427Q and R427A/R430A mutants display comparably weak DNA binding. DNA oligonucleotides used in the experiments are listed in (Supplementary Table S3). (**B**) Left, close-up view of the Arg427 and Arg430 side-chain orientation in Grhl1 DBD (cyan) and the Grhl1-DBD:DNA complex (gray). The dashed contour indicates the floppy nature of Arg427 in the absence of DNA. Right, comparison of the same segment from the Grhl1-DBD:DNA complex (gray) as shown in the left panel with the Grhl1-DBD R427Q mutant (green). A hydrogen bond formed between Gln427 and Arg430, marked as dashed line (black), restricts the conformation of the Arg430 side chain and prevents a readjustment to a conformation facilitating DNA binding (crossed-out curved arrow) as observed in the Grhl1-DBD:DNA complex.

residue Arg430 underwent a conformational rearrangement whereby the sidechain rotated by ∼180° and formed a hydrogen bond to the side chain of Gln427 introduced by the mutation (Figure 4B). As a consequence, the R427Q mutation perturbs DNA binding by Grhl1 DBD in two different ways. Firstly, the Gln427 side chain cannot contact the important guanine base in the RE in the same way as Arg427 in wild-type Grhl1 DBD. Secondly, the Arg430 is pulled away from the DNA, eliminating two further protein–DNA hydrogen bonds (Figures 2C and 4B). The Grhl1 R427Q mutation, therefore, has a stronger effect in destabilizing DNA binding than R427A, as it removes all protein side chain-DNA base contacts observed in the crystal structure.

To further corroborate this interpretation of the effect of the R427Q mutation, a Grhl1-DBD R427A/R430A double mutant was generated. As expected, both in ITC and EMSA, Grhl1 R427A/R430A and Grhl1 R427Q showed a comparable, significant reduction in DNA-binding affinity (Figures 3E and 4A), supporting the view that a single cancer-associated mutation, Grhl1 R427Q, is sufficient to completely abrogate target DNA binding by this transcription factor.

Arg427 determines Grhl1 recognition of *CLDN4* promoter

In order to verify that the determinants for target DNA binding by Grhl1 identified by the structural and biochemical study are relevant for transcriptional activation in cells, reporter gene assays were performed with both Grhl1 and Grhl2. As previously reported, inactivation of Grhl2 significantly reduces *CLDN4* promoter activity in mIMCD-3 cells (11,58). We used CRISPR/Cas9 technology to generate mIMCD-3 cells with frameshift-inducing mutations within exon 2 of the *GRHL2* gene. Three clones carrying homozygous null mutations of *GRHL2* revealed profoundly reduced *CLDN4* reporter activity when compared to wild-type mIMCD-3 cells (Figure 5A). Human *GRHL1* or *GRHL2* with or without point mutations were then re-expressed, as indicated, in *GRHL2* knockout mIMCD-3 cells. While wildtype *GRHL1* overexpression resulted in markedly increased *CLDN4* promoter activity, *GRHL1 R427Q* or *GRHL1 R427A* overexpression resulted in *CLDN4* promoter activity similar to *GRHL2* knockout cells. Mutations of the analogous arginine residue in Grhl2 (Arg423) had a similar effect. *CLDN4* promoter activity increased in response to overexpression of wildtype

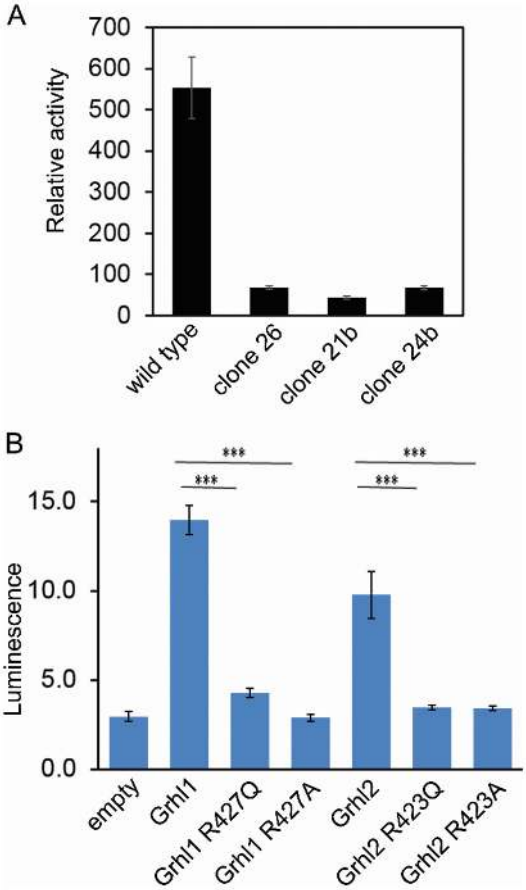

Figure 5. Point mutations affecting Arg427 of Grhl1 or the corresponding Arg423 of Grhl2 result in a complete inactivation of Grhl-dependent *CLDN4* promoter activity in mIMCD-3 cells. (**A**) *GRHL2* knockout in mIMCD-3 cells leads to reduced *CLDN4* promoter activity. *CLDN4* reporter activity was analyzed in wild-type mIMCD-3 cells and in three subclones of mIMCD-3 cells carrying *GRHL2* null mutations (clones 26c, 21b, and 24b). (**B**) Luciferase assays were carried out in mIMCD3 cells carrying a CRISPR/Cas9-induced *GRHL2* null mutation (clone 24b) transiently transfected with the indicated overexpression constructs to assay the ability of human wild-type Grhl1, Grhl2 and point-mutated versions to activate *CLDN4* promoter activity. Point mutations were induced to achieve the indicated changes of the critical arginine residue to glutamine (Grhl1 R427Q; Grhl2 R423Q)) and alanine (Grhl1 R427A; Grhl2 R423A), respectively. Overexpression constructs were as follows: pEZ-M02-empty, empty control overexpression plasmid; pEZ-M02 carrying wildtype Grhl1 and Grhl2 and mutated versions as indicated. Bars represent mean ± standard deviation; $n = 3$ biological replicates. ***$P < 0.001$ point-mutated versus wild-type constructs (ANOVA with post-hoc Tukey test).

GRHL2, but not *GRHL2 R423Q* or *GRHL2 R423A* (Figure 5B). These results demonstrate that Grhl1- and Grhl2-dependent transcriptional activation of the *CLDN4* promoter is entirely dependent on an intact Arg427 (Grhl1) or Arg423 (Grhl2).

Grhl1-DBD R427Q shows significantly lower binding affinity to target DNA compared to Grhl1-DBD R427A, while the luciferase assay shows that the two mutant proteins have very similar effects on transactivation of the *CLDN4* promoter. There are two main reasons for this apparent discrepancy. First, the *in vitro* assays (ITC and EMSA, see Figures 3 and 4) measure DNA binding activity of the Grhl1-DBD instead of the full-length protein. We cannot exclude the possibility that the presence of the other two functional domains, especially the dimerization domain, may have an influence on the DNA binding affinity of Grhl1. Second, transcriptional activation in the cellular context is not exclusively determined by DNA target site affinity, but also depends on protein-protein interactions predominantly mediated by the transcription factor's transactivation domain. It seems possible that the Grhl1 R427Q mutation affects these interactions in a slightly different way than the R427A mutation. In addition, the Grhl1 R427A/R427Q (and the corresponding Grhl2 R423A/R423Q) mutants all show very weak transactivation activity, very close to the 'empty' control, such that small differences between them cannot be significant.

DISCUSSION

Here, we report the first structures of DNA-binding domains of Grh/CP2 transcription factors and elucidate the molecular basis for their specific DNA binding activity. Despite low sequence conservation, the DBDs of Grhl1/2 and p53 show a conserved fold. This result is in agreement with previous bioinformatics predictions (59,60), thus establishing a common evolutionary origin of the two transcription factor families. The Grhl1-DBD:DNA complex structure reveals a strictly parsimonious mode of DNA sequence recognition that relies on a small number of protein side chain-DNA base contacts, whose significance for transcriptional gene activation was demonstrated by a thorough study of structure-function relationship. We shall discuss the relevance of these findings for the entire Grh/CP2 family.

Structural basis for DNA targeting

Crystallographic analysis shows that Grhl1 and Grhl2 share a highly conserved three-dimensional structure characterized by an IgG-like core, a fold that has been identified in six other transcription-factor families (61). The conserved structures of Grhl1 and Grhl2, and their conserved DNA RE sequence (31,62) provide the basis for their biological cooperativity (63). Grhl1 DBD binds the target-site DNA via extensive electrostatic interactions, while sequence recognition is mediated by two evolutionarily conserved arginines, Arg427 and Arg430, hydrogen-bonded to the most conserved guanine in the core motif and an adenine in the flanking sequence, respectively. In addition, the main-chain carbonyl of Gly387 forms a hydrogen bond in the minor groove to the conserved G8 base. Although this is a protein backbone-DNA contact, it is amino-acid specific due to the Gly387 torsion angles of $\varphi = 93^\circ$ and $\psi = 19^\circ$ (average values for the two protein chains) which are favored for glycine but disfavored for all other amino acid residues. The flexible nature of Gly387 thus helps loop L10 to properly bind in the DNA minor groove. The minor groove hydrogen bond from Gly387 to DNA requires the presence of the guanine 2-amino group and helps discriminating against AT base pairs at the center of the binding sequence.

We also find that Lys386 at the L10 loop of Grhl1 mediates an important DBD-DBD contact that stabilizes DNA

binding through hydrophobic and polar interactions in the minor groove. A comprehensive review of the origin of specificity in protein–DNA recognition (64) discusses various aspects of protein-loop binding to the minor groove. Interestingly, it shows several examples of arginine residues contributing to DNA shape recognition through interactions with the narrow minor groove of AT-rich DNA. In the structure presented here, we observe a variation of that theme in the recognition of the widened minor groove of GC-rich DNA by a tandem lysine loop.

Transcription factors such as HSTF (heat shock transcription factors) and TFIID-1 (TATA box-binding transcription initiation factor) show conformational changes during DNA binding for stabilizing the interaction (65–67). In contrast and similarly to p53, Grhl1-DBD monomers bind to B-form DNA without undergoing major conformational changes. However, full-length p53 forms tetramers through C-terminal oligomerization (68–70), while our data suggest that Grhl1 binds DNA in a dimeric arrangement, which proceeds without a major change of the DNA structure.

Comparison of DNA binding by Grhl DBD and p53

Grhl1 DBD and p53 DBD share a common protein fold, and their DNA-binding modes are similar. In addition, their DNA target sequences also present common features. The half-site binding sequence of p53 is defined as RRRCWWGYYY (R = A, G; W = A, T; Y = C, T), which is related to the CP2 DNA-binding site (CNRG-$N_{5/6}$-CNRG) (32,54,56). P53 and Grhl proteins cooperate during epidermal development and function; for instance, Grhl3, p53 and p63 (a p53 family homolog) regulate *CLDN1* expression, and Grhl3 cooperates with p63 in transcriptional *PTEN* gene regulation (71). Based on their consensus target sequences, we may speculate that Grhl proteins also show binding to p53 recognition elements, and *vice versa*. ITC data reveal that Grhl1 shows a preference for recognition elements with a central CG or CC dinucleotide (see Supplementary Figure S4), while p53 prefers AT or AA which shows a higher propensity for DNA kinking, an essential feature induced by p53 binding (72). The ITC data provide evidence for decreasing affinity of Grhl1 when the central dinucleotide is modified from CC (CG) to AA to AT (TA). Although the bases of the central dinucleotide are not in direct contact with Grhl1, they may, thus, contribute to DNA target-site recognition by the transcription factor.

In the Grhl1-DBD:DNA complex, the two recognition helices (α3) penetrate more deeply into the major groove and form a larger interface with DNA compared to p53. Consequently, thymine methyl groups at the center of the binding site may sterically interfere with binding of Grhl1, providing a second explanation for the observed weaker binding to sites with central AA, AT or TA steps and a second mechanism to discriminate against p53 target sites. Besides the central dinucleotide, other factors like the oligomerization behavior and flanking sequences also play roles for the target specificity of Grhl1 and p53.

Structural insight into Grh/CP2-linked disease

We identified a common mode of DNA binding in Grhl1–3, which may be shared by the CP2 branch of Grh/CP2 transcription factors. Accordingly, the Grhl1-DBD:DNA complex structure can reveal structure-function relationships of the whole Grh/CP2 family. For example, we identified a key role of Grhl1 Arg427 (Arg423 in Grhl2) in trans-activating gene expression, unveiling the mechanistic basis linking the corresponding R427Q mutation to oncogenesis. The dramatic effect of the Grhl1 R427Q mutation on DNA binding and transcription is directly linked to the parsimonious binding mode of Grhl1 where redundant and potentially compensating protein–DNA contacts are absent. By superimposing the structures of DNA-bound Grhl1 and p53, we show that Arg427 in Grhl1 is matched with Arg280 in p53, which both play central roles in DNA binding. It is known that mutations of Arg280 to Ala or Lys in p53 are directly linked to loss of DNA binding and transactivation and to human breast cancer (59,73).

Several disease-related mutations in Grhl transcription factors, mainly in the DBD, were reported in recent studies (74–76) (Figure 6A). Missense mutations in Grhl2 (Y398H and I482K) and Grhl3 (R298H, R391C and T454M) are linked to abnormalities in ectodermal structures, as found in the Ectodermal Dysplasia Syndrome and Van der Woude Syndrome, respectively. Mapping these five mutations to the corresponding amino acids in Grhl1 (Ile474, Tyr402, Arg319, Arg412 and Thr468) reveals that they are not involved in DNA binding, but rather in stabilizing the protein structure (Figure 6B).

Arg412 in Grhl1 forms a salt bridge with Asp465 and is further hydrogen-bonded to Thr468 at the C-terminal loop and Asp398 in strand β9 (Figure 6C). A mutation of this arginine to cysteine (as found in one of the disease-related Grh3 mutations, R391C) would disturb these extensive interactions, resulting in the destabilization of the C-terminal loop. Another disease-related mutation, T454M, was identified in Grhl3 at the site corresponding to Grhl1 Thr468. Ile474 forms a hydrophobic interface with Phe473 and Tyr356 (Figure 6D); a substitution of isoleucine by a charged lysine (as found in the Grh2 I482K mutation) will disrupt the hydrophobic contacts and result in a charge repulsion with the adjacent His325. We note that all these mutations interfere with the anchoring of the C-terminal extension, whose deletion leads to insolubility of the Grhl DBD according to our observation, highlighting its fundamental role in stabilizing the protein (see Figure 1D). Most oncogenic mutations in p53 do not directly interfere with DNA binding, but are deleterious because they further destabilize an intrinsically unstable protein (77). Thermal shift assays reveal melting temperatures (T_m values) of 51.0°C for Grhl1 DBD and 54.5°C for Grhl2 DBD (Supplementary Figure S7). The thermal stability these two DNA-binding domains is thus intermediate between p53 DBD (T_m = 45.6°C) and its homolog p63 (T_m = 61.5°C) (78) suggesting that structure destabilizing mutations may contribute to Grhl1- or Grhl2-linked disease in a similar way.

The other two mutations, R319H (R298H of Grhl3) and Y402H (Y398H of Grhl2), would not be predicted to destabilize the protein because they are surface-exposed. We

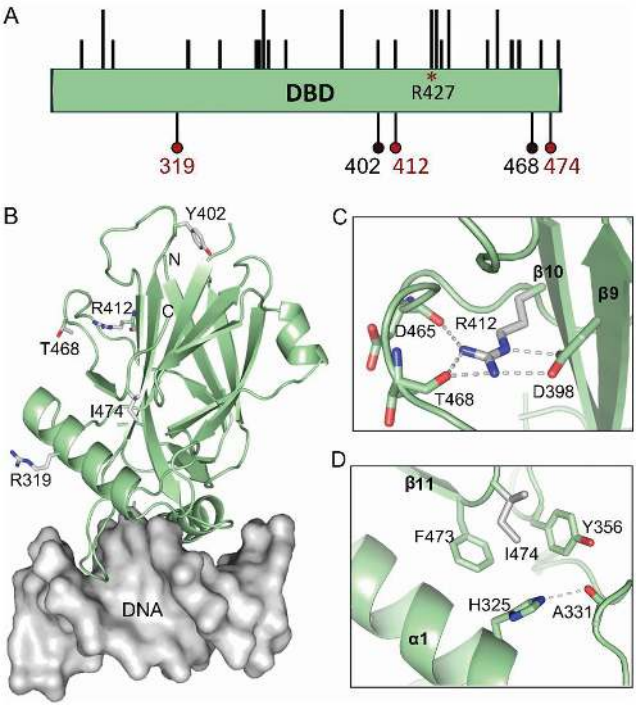

Figure 6. Insight into the structural effects of disease-associated mutations. (**A**) Summary of mutations of Grhl1 DBD. The upper bars in black indicate mutations of Grhl1 DBD that are predicted as pathogenic summarized from the COSMIC database. The length of the bars indicates the mutation frequency according to COSMIC. Residues of Grhl1 corresponding to mutations of Grhl2 and Grhl3 found in Ectodermal Dysplasia Syndrome and Van der Woude Syndrome, respectively, are shown at the bottom. Numbers in red indicate sequence conservation among Grhl1–3. (**B**) Mapping the Grhl2 and Grhl3 mutations in ectodermal Dysplasia syndrome and Van der Woude syndrome related mutations on the Grhl1-DBD:DNA structure. Two of these mutated residues, Arg412 and Ile474 are involved in intricate interaction networks. (**C**) Arg412 forms hydrogen bonds with the backbone of Thr468 and the Asp465 side chain. (**D**) Ile474 is involved in hydrophobic contacts with Phe473 and Tyr356. The R412C and I474K mutations are likely to destabilize the C-terminal loop and reduce overall stability.

would envisage that they may affect the interaction with other proteins. Based on the structural studies presented here, more disease-related mutations in Grh/CP2 proteins can be defined and predicted. Studies on the transcription factor LSF (TFCP2) from the CP2 branch of Grh/CP2 transcription factors demonstrate that the double mutant Q234L/K236E cannot bind to DNA and acts as a dominant negative inhibitor (79,80). These two residues are conserved in Grhl proteins and localize close to the DNA-binding interface. This observation is consistent with a comparable DNA binding mode of CP2 proteins.

A recent search identified small molecules inhibiting the DNA binding activity of LSF, which is an oncogene in hepatocellular carcinoma (81–83). Further progress has been hampered by the lack of structural information on LSF. The structural and functional studies presented here not only reveal the basis of promoter recognition and transcriptional activation by Grhl proteins, but may also provide guidance to efforts that aim at developing structure-based inhibitors of Grh/CP2 transcription factors for therapeutic applications.

DATA AVAILABILITY

The COSMIC database is the Catalogue of Somatic Mutations in Cancer (http://cancer.sanger.ac.uk/cosmic). The Pfam database is a large collection of protein families, each represented by multiple sequence alignments and hidden Markov models (HMMs) (http://pfam.xfam.org). ESPript, Easy Sequencing in PostScript, is a program which renders sequence similarities and secondary structure information from aligned sequences for analysis and publication purpose (http://espript.ibcp.fr/ESPript/ESPript). Coordinates and structure factors have been deposited in the Protein Data bank under accession number 5MPH (Grhl1 DBD), PDB 5MPI (Grhl1-DBD R427Q mutant), PDB 5MPF (Grhl1-DBD:DNA complex) and PDB 5MR7 (Grhl2 DBD).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

ACKNOWLEDGEMENTS

Expert technical assistance with protein purification of Grh2 protein by T. Dornblut and J. Tischer is gratefully acknowledged. We especially thank Prof. Oliver Daumke for reading the manuscript and giving valuable suggestions. The Protein Sample Production Facility at the Max Delbrück Center is funded by the Helmholtz Association of German Research Centers.

FUNDING

German Research Foundation (DFG) [SFB 740] (in part). Funding for open access charge: Max Delbrück Center for Molecular Medicine Berlin.
Conflict of interest statement. None declared.

REFERENCES

1. Bray,S.J. and Kafatos,F.C. (1991) Developmental function of Elf-1: an essential transcription factor during embryogenesis in Drosophila. *Genes Dev.*, **5**, 1672–1683.
2. Nüsslein-Volhard,C., Wieschaus,E. and Kluding,H. (1984) Mutations affecting the pattern of the larval cuticle in Drosophila melanogaster. *Wilhelm Roux's Arch. Dev. Biol.*, **193**, 267–282.
3. Venkatesan,K., McManus,H.R., Mello,C.C., Smith,T.F. and Hansen,U. (2003) Functional conservation between members of an ancient duplicated transcription factor family, LSF/Grainyhead. *Nucleic Acids Res.*, **31**, 4304–4316.
4. Auden,A., Caddy,J., Wilanowski,T., Ting,S.B., Cunningham,J.M. and Jane,S.M. (2006) Spatial and temporal expression of the Grainyhead-like transcription factor family during murine development. *Gene Expr. Patterns*, **6**, 964–970.
5. Jane,S.M, Nienhuis,A.W. and Cunningham,J.M. (1995) Hemoglobin switching in man and chicken is mediated by a heteromeric complex between the ubiquitous transcription factor CP2 and a developmentally specific protein. *EMBO J.*, **14**, 97–105.
6. Mlacki,M., Kikulska,A., Krzywinska,E., Pawlak,M. and Wilanowski,T. (2015) Recent discoveries concerning the involvement of transcription factors from the Grainyhead-like family in cancer. *Exp. Biol. Med.*, **240**, 1396–1401.

7. Powell,C.M., Rudge,T.L., Zhu,Q., Johnson,L.F. and Hansen,U. (2000) Inhibition of the mammalian transcription factor LSF induces S-phase-dependent apoptosis by downregulating thymidylate synthase expression. *EMBO J.*, **19**, 4665–4675.
8. Goldie,S.J., Arhatari,B.D., Anderson,P., Auden,A., Partridge,D.D., Jane,S.M. and Dworkin,S. (2016) Mice lacking the conserved transcription factor Grainyhead-like 3 (Grhl3) display increased apposition of the frontal and parietal bones during embryonic development. *BMC Dev. Biol.*, **16**, 37–48.
9. Adhikari,N., Neupane,S., Gwon,G.J., Kim,J.Y., An,C.H., Lee,S., Sohn,W.J., Lee,Y. and Kim,J.Y. (2017) Grhl3 modulates epithelial structure formation of the circumvallate papilla during mouse development. *Histochem. Cell Biol.*, **147**, 5–16.
10. Wilanowski,T., Caddy,J., Ting,S.B., Hislop,N.R., Cerruti,L., Auden,A., Zhao,L.L., Asquith,S., Ellis,S., Sinclair,R. *et al.* (2008) Perturbed desmosomal cadherin expression in grainy head-like 1-null mice. *EMBO J.*, **27**, 886–897.
11. Werth,M., Walentin,K., Aue,A., Schonheit,J., Wuebken,A., Pode-Shakked,N., Vilianovitch,L., Erdmann,B., Dekel,B., Bader,M. *et al.* (2010) The transcription factor grainyhead-like 2 regulates the molecular composition of the epithelial apical junctional complex. *Development*, **137**, 3835–3845.
12. Varma,S., Cao,Y., Tagne,J.B., Lakshminarayanan,M., Li,J., Friedman,T.B., Morell,R.J., Warburton,D., Kotton,D.N. and Ramirez,M.I. (2012) The transcription factors Grainyhead-like 2 and NK2-homeobox 1 form a regulatory loop that coordinates lung epithelial cell morphogenesis and differentiation. *J. Biol. Chem.*, **287**, 37282–37295.
13. Walentin,K., Hinze,C., Werth,M., Haase,N., Varma,S., Morell,R., Aue,A., Potschke,E., Warburton,D., Qiu,A. *et al.* (2015) A Grhl2-dependent gene network controls trophoblast branching morphogenesis. *Development*, **142**, 1125–1136.
14. Ting,S.B., Caddy,J., Hislop,N., Wilanowski,T., Auden,A., Zhao,L.L., Ellis,S., Kaur,P., Uchida,Y., Holleran,W.M. *et al.* (2005) A homolog of Drosophila grainy head is essential for epidermal integrity in mice. *Science*, **308**, 411–413.
15. Yu,Z.Q., Bhandari,A., Mannik,J., Pham,T., Xu,X.M. and Andersen,B. (2008) Grainyhead-like factor Get1/Grhl3 regulates formation of the epidermal leading edge during eyelid closure. *Dev. Biol.*, **319**, 56–67.
16. Yu,Z.Q., Mannik,J., Soto,A., Lin,K.K. and Andersen,B. (2009) The epidermal differentiation-associated Grainyhead gene Get1/Grhl3 also regulates urothelial differentiation. *EMBO J.*, **28**, 1890–1903.
17. Frisch,S.M., Farris,J.C. and Pifer,P.M. (2017) Roles of Grainyhead-like transcription factors in cancer. *Oncogene*, **36**, 6067–6073.
18. Mlacki,M., Darido,C., Jane,S.M. and Wilanowski,T. (2014) Loss of Grainy Head-Like 1 is associated with disruption of the epidermal barrier and squamous cell carcinoma of the skin. *PloS ONE*, **9**, e89247.
19. Fabian,J., Lodrini,M., Oehme,I., Schier,M.C., Thole,T.M., Hielscher,T., Kopp-Schneider,A., Opitz,L., Capper,D., von Deimling,A. *et al.* (2014) GRHL1 acts as tumor suppressor in neuroblastoma and is negatively regulated by MYCN and HDAC3. *Cancer Res.*, **74**, 2604–2616.
20. Xiang,X.Y., Deng,Z.B., Zhuang,X.Y., Ju,S.W., Mu,J.Y., Jiang,H., Zhang,L.F., Yan,J., Miller,D. and Zhang,H.G. (2012) Grhl2 determines the epithelial phenotype of breast cancers and promotes tumor progression. *PLoS ONE*, **7**, e50781.
21. Yang,X.A., Vasudevan,P., Parekh,V., Penev,A. and Cunningham,J.M. (2013) Bridging cancer biology with the clinic: Relative expression of a GRHL2-mediated gene-set pair predicts breast cancer metastasis. *PLoS ONE*, **8**, e56195.
22. Bhandari,A., Gordon,W., Dizon,D., Hopkin,A.S., Gordon,E., Yu,Z. and Andersen,B. (2013) The Grainyhead transcription factor Grhl3/Get1 suppresses miR-21 expression and tumorigenesis in skin: modulation of the miR-21 target MSH2 by RNA-binding protein DND1. *Oncogene*, **32**, 1497–1507.
23. Darido,C., Georgy,S.R., Wilanowski,T., Dworkin,S., Auden,A., Zhao,Q., Rank,G., Srivastava,S., Finlay,M.J., Papenfuss,A.T. *et al.* (2011) Targeting of the tumor suppressor GRHL3 by a miR-21-dependent proto-oncogenic network results in PTEN loss and tumorigenesis. *Cancer Cell*, **20**, 635–648.
24. Cieply,B., Riley,P., Pifer,P.M., Widmeyer,J., Addison,J.B., Ivanov,A.V., Denvir,J. and Frisch,S.M. (2012) Suppression of the epithelial–mesenchymal transition by Grainyhead-like-2. *Cancer Res.*, **72**, 2440–2453.
25. Cieply,B., Farris,J., Denvir,J., Ford,H.L. and Frisch,S.M. (2013) epithelial–mesenchymal transition and tumor suppression are controlled by a reciprocal feedback loop between ZEB1 and Grainyhead-like-2. *Cancer Res.*, **73**, 6299–6309.
26. Pifer,P.M., Farris,J.C., Thomas,A.L., Stoilov,P., Denvir,J., Smith,D.M. and Frisch,S.M. (2016) Grainyhead-like 2 inhibits the coactivator p300, suppressing tubulogenesis and the epithelial–mesenchymal transition. *Mol. Biol. Cell*, **27**, 2479–2492.
27. Xiang,J., Fu,X., Ran,W. and Wang,Z. (2017) Grhl2 reduces invasion and migration through inhibition of TGF beta-induced EMT in gastric cancer. *Oncogenesis*, **6**, 284–290.
28. Kang,X., Chen,W., Kim,R.H., Kang,M.K. and Park,N.H. (2009) Regulation of the hTERT promoter activity by MSH2, the hnRNPs K and D, and GRHL2 in human oral squamous cell carcinoma cells. *Oncogene*, **28**, 565–574.
29. Zhang,Y., Toh,L., Lau,P. and Wang,X.Y. (2012) Human telomerase reverse transcriptase (hTERT) Is a novel target of the Wnt/beta-catenin pathway in human cancer. *J. Biol. Chem.*, **287**, 32494–32511.
30. Traylor-Knowles,N., Hansen,U., Dubuc,T.Q., Martindale,M.Q., Kaufman,L. and Finnerty,J.R. (2010) The evolutionary diversification of LSF and Grainyhead transcription factors preceded the radiation of basal animal lineages. *BMC Evol. Biol.*, **10**, 101.
31. Kim,M. and McGinnis,W. (2011) Phosphorylation of Grainy head by ERK is essential for wound-dependent regeneration but not for development of an epidermal barrier. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 650–655.
32. Attardi,L.D. and Tjian,R. (1993) Drosophila tissue-specific transcription factor Ntf-1 contains a novel isoleucine-rich activation motif. *Genes Dev.*, **7**, 1341–1353.
33. Murata,T., Nitta,M. and Yasuda,K. (1998) Transcription factor CP2 is essential for lens-specific expression of the chicken alpha A-crystallin gene. *Genes to Cells*, **3**, 443–457.
34. Nevil,M., Bondra,E.R., Schulz,K.N., Kaplan,T. and Harrison,M.M. (2017) Stable binding of the conserved transcription factor Grainy Head to its target genes throughout Drosophila melanogaster development. *Genetics*, **205**, 605–620.
35. Shirra,M.K. and Hansen,U. (1998) LSF and NTF-1 share a conserved DNA recognition motif yet require different oligomerization states to form a stable protein–DNA complex. *J. Biol. Chem.*, **273**, 19260–19268.
36. Scheich,C., Kummel,D., Soumailakakis,D., Heinemann,U. and Bussow,K. (2007) Vectors for co-expression of an unrestricted number of proteins. *Nucleic Acids Res.*, **35**, e43.
37. Vanduyne,G.D., Standaert,R.F., Karplus,P.A., Schreiber,S.L. and Clardy,J. (1993) Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J. Mol. Biol.*, **229**, 105–124.
38. Heinemann,U., Bussow,K., Mueller,U. and Umbach,P. (2003) Facilities and methods for the high-throughput crystal structural analysis of human proteins. *Acc. Chem. Res.*, **36**, 157–163.
39. Leslie,A.G.W. (1992) Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography*, **26**.
40. Krug,M., Weiss,M.S., Heinemann,U. and Mueller,U. (2012) XDSAPP: a graphical user interface for the convenient processing of diffraction data using XDS. *J. Appl. Crystallogr.*, **45**, 568–572.
41. Adams,P.D., Afonine,P.V., Bunkoczi,G., Chen,V.B., Davis,. W., Echols,N., Headd,J.J., Hung,L.W., Kapral,G.J., Grosse-Kunstleve,R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D*, **66**, 213–221.
42. Collaborative Computational Project, Number 4 (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. D*, **50**, 760–763.
43. Adams,P.D., Grosse-Kunstleve,R.W., Hung,L.W., Ioerger,T.R., McCoy,A.J., Moriarty,N.W., Read,R.J., Sacchettini,J.C., Sauter,N.K. and Terwilliger,T.C. (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D*, **58**, 1948–1954.

44. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D*, **60**, 2126–2132.
45. Davis,I.W., Leaver-Fay,A., Chen,V.B., Block,J.N., Kapral,G.J., Wang,X., Murray,L.W., Arendall,W.B., Snoeyink,J., Richardson,J.S. *et al.* (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.*, **35**, 375–383.
46. McCoy,A.J., Grosse-Kunstleve,R.W., Storoni,L.C. and Read,R.J. (2005) Likelihood-enhanced fast translation functions. *Acta Crystallogr. D*, **61**, 458–464.
47. Smart,O.S., Womack,T.O., Flensburg,C., Keller,P., Paciorek,W., Sharff,A., Vonrhein,C. and Bricogne,G. (2012) Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta Crystallogr. D*, **68**, 368–380.
48. Zheng,G., Lu,X.J. and Olson,W.K. (2009) Web 3DNA – A web server for the analysis, reconstruction and visualization of three-dimensional nucleic-acid structures. *Nucleic Acids Res.*, **37**, W240–W246.
49. Krissinel,E. and Henrick,K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D*, **60**, 2256–2268.
50. Ishida,T. and Kinoshita,K. (2007) PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res.*, **35**, 460–464.
51. Buchan,D.W.A., Minneci,F., Nugent,T.C.O., Bryson,K. and Jones,D.T. (2013) Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res.*, **41**, W349–W357.
52. Bamford,S., Dawson,E., Forbes,S., Clements,J., Pettett,R., Dogan,A., Flanagan,A., Teague,J., Futreal,P.A., Stratton,M.R. *et al.* (2004) The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *Brit J. Cancer*, **91**, 355–358.
53. Forbes,S.A., Beare,D., Gunasekaran,P., Leung,K., Bindal,N., Boutselakis,H., Ding,M.J., Bamford,S., Cole,C., Ward,S. *et al.* (2015) COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res.*, **43**, D805–D811.
54. Nagaich,A.K., Appella,E. and Harrington,R.E. (1997) DNA bending is essential for the site-specific recognition of DNA response elements by the DNA binding domain of the tumor suppressor protein p53. *J. Biol. Chem.*, **272**, 14842–14849.
55. Balagurumoorthy,P., Lindsay,S.M. and Harrington,R.E. (2002) Atomic force microscopy reveals kinks in the p53 response element DNA. *Biophys. J.*, **101**, 611–623.
56. Beno,I., Rosenthal,K., Levitine,M., Shaulov,L. and Haran,T.E. (2011) Sequence-dependent cooperative binding of p53 to DNA targets and its relationship to the structural properties of the DNA targets. *Nucleic Acids Res.*, **39**, 1919–1932.
57. Jordan,J.J., Menendez,D., Sharav,J., Beno,I., Rosenthal,K., Resnick,M.A. and Haran,T.E. (2012) Low-level p53 expression changes transactivation rules and reveals superactivating sequences. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 14387–14392.
58. Aue,A., Hinze,C., Walentin,K., Ruffert,J., Yurtdas,Y., Werth,M., Chen,W., Rabien,A., Kilic,E., Schulzke,J.D. *et al.* (2015) A grainyhead-like 2/Ovo-like 2 pathway regulates renal epithelial barrier function and lumen expansion. *J. Am. Soc. Nephrol.*, **26**, 2704–2715.
59. Kokoszynska,K., Ostrowski,J., Rychlewski,L. and Wyrwicz,L.S. (2008) The fold recognition of CP2 transcription factors gives new insights into the function and evolution of tumor suppressor protein p53. *Cell Cycle*, **7**, 2907–2915.
60. Ho,W.C., Fitzgerald,M.X. and Marmorstein,R. (2006) Structure of the p53 core domain dimer bound to DNA. *J. Biol. Chem.*, **281**, 20494–20502.
61. Rudolph,M.J. and Gergen,J.P. (2001) DNA-binding by Ig-fold proteins. *Nat. Struct. Mol. Biol.*, **8**, 384–386.
62. Rifat,Y., Parekh,V., Wilanowski,T., Hislop,N.R., Auden,A., Ting,S.B., Cunningham,J.M. and Jane,S.M. (2010) Regional neural tube closure defined by the Grainy head-like transcription factors. *Dev. Biol.*, **345**, 237–245.
63. Boglev,Y., Wilanowski,T., Caddy,J., Parekh,V., Auden,A., Darido,C., Hislop,N.R., Cangkrama,M., Ting,S.B. and Jane,S.M. (2011) The unique and cooperative roles of the Grainy head-like transcription factors in epidermal development reflect unexpected target gene specificity. *Dev. Biol.*, **349**, 512–522.
64. Rohs,R., Jin,X., West,S.M., Joshi,R., Honig,B. and Mann,R.S. (2010) Origin of specificity in protein–DNA recognition. *Annu. Rev. Biochem.*, **79**, 233–269.
65. Harrison,C.J., Bohm,A.A. and Nelson,H.C.M. (1994) Crystal-structure of the DNA-binding domain of the heat-shock transcription factor. *Science*, **263**, 224–227.
66. Nikolov,D.B. and Burley,S.K. (1994) 2.1-Angstrom resolution refined structure of a TATA box-binding protein (Tbp). *Nat. Struct. Mol. Biol.*, **1**, 621–637.
67. Andrabi,M., Mizuguchi,K. and Ahmad,S. (2014) Conformational changes in DNA-binding proteins: Relationships with precomplex features and contributions to specificity and stability. *Proteins: Struct. Funct. Bioinf.*, **82**, 841–857.
68. Ma,B. and Levine,A.J. (2007) Probing potential binding modes of the p53 tetramer to DNA based on the symmetries encoded in p53 response elements. *Nucleic Acids Res.*, **35**, 7733–7747.
69. Chene,P. (2001) The role of tetramerization in p53 function. *Oncogene*, **20**, 2611–2617.
70. Pham,N., Lucumi,A., Cheung,N. and Viadiu,H. (2012) The tetramer of p53 in the absence of DNA forms a relaxed quaternary state. *Biochemistry*, **51**, 8053–8055.
71. Darido,C., Georgy,S.R., Wilanowski,T., Dworkin,S., Auden,A., Zhao,Q., Rank,G., Srivastava,S., Finlay,M.J., Papenfuss,A.T. *et al.* (2011) Targeting of the tumor suppressor GRHL3 by a miR-21-dependent proto-oncogenic network results in PTEN loss and tumorigenesis. *Cancer Cell*, **20**, 635–648.
72. Nagaich,A.K., Bhattacharyya,D., Brahmachari,S.K. and Bansal,M. (1994) CA/TG sequence at the 5′ end of oligo(A)-tracts strongly modulates DNA curvature. *J. Biol. Chem.*, **269**, 7824–7833.
73. Muller,P.A.J. and Vousden,K.H. (2013) p53 mutations in cancer. *Nat. Cell Biol.*, **15**, 2–8.
74. Petrof,G., Nanda,A., Howden,J., Takeichi,T., McMillan,J.R., Aristodemou,S., Ozoemena,L., Liu,L., South,A.P., Pourreyron,C. *et al.* (2014) Mutations in GRHL2 result in an autosomal-recessive ectodermal Dysplasia syndrome. *Am. J. Hum. Genet.*, **95**, 308–314.
75. Peyrard-Janvid,M., Leslie,E.J., Kousa,Y.A., Smith,T.L., Dunnwald,M., Magnusson,M., Lentz,B.A., Unneberg,P., Fransson,I., Koillinen,H.K. *et al.* (2014) Dominant mutations in GRHL3 cause Van der Woude Syndrome and disrupt oral periderm development. *Am. J. Hum. Genet.*, **94**, 23–32.
76. Leslie,E.J., Liu,H., Carlson,J.C., Shaffer,J.R., Feingold,E., Wehby,G., Laurie,C.A., Jain,D., Laurie,C.C., Doheny,K.F. *et al.* (2016) A genome-wide association study of nonsyndromic cleft palate identifies an etiologic missense variant in GRHL3. *Am. J. Hum. Genet.*, **98**, 744–754.
77. Joerger,A.C. and Fersht,A.R. (2016) The p53 pathway: origins, inactivation in cancer, and emerging therapeutic approaches. *Annu. Rev. Biochem.*, **85**, 375–404.
78. Brandt,T., Kaar,J.L., Fersht,A.R. and Veprintsev,D.B. (2012) Stability of p53 homologs. *PLoS ONE*, **7**, e47889.
79. Shirra,M.K., Zhu,Q., Huang,H.C., Pallas,D. and Hansen,U. (1994) One exon of the human LSF gene includes conserved regions involved in novel DNA-binding and dimerization motifs. *Mol. Cell Biol.*, **14**, 5076–5087.
80. Coull,J.J., Romerio,F., Sun,J.M., Volker,J.L., Galvin,K.M., Davie,J.R., Shi,Y., Hansen,U. and Margolis,D.M. (2000) The human factors YY1 and LSF repress the human immunodeficiency virus type 1 long terminal repeat via recruitment of histone deacetylase 1. *J. Virol.*, **74**, 6790–6799.
81. Yoo,B.K., Emdad,L., Gredler,R., Fuller,C., Dumur,C.I., Jones,K.H., Jackson-Cook,C., Su,Z.Z., Chen,D., Saxena,U.H. *et al.* (2010) Transcription factor Late SV40 Factor (LSF) functions as an oncogene in hepatocellular carcinoma. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 8357–8362.
82. Grant,T.J., Bishop,J.A., Christadore,L.M., Barot,G., Chin,H.G., Woodson,S., Kavouris,J., Siddiq,A., Gredler,R., Shen,X.N. *et al.* (2012) Antiproliferative small-molecule inhibitors of transcription factor LSF reveal oncogene addiction to LSF in hepatocellular carcinoma. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 4503–4508.
83. Rajasekaran,D., Siddiq,A., Willoughby,J.L.S., Biagi,J.M., Christadore,L.M., Yunes,S.A., Gredler,R., Jariwala,N., Robertson,C.L., Akiel,M.A. *et al.* (2015) Small molecule inhibitors of Late SV40 Factor (LSF) abrogate hepatocellular carcinoma (HCC): evaluation using an endogenous HCC model. *Oncotarget*, **6**, 26266–26277.