Published online 9 October 2020 *Nucleic Acids Research, 2020, Vol. 48, No. 19* **11097–11112**
doi: 10.1093/nar/gkaa827

ERH facilitates microRNA maturation through the interaction with the N-terminus of DGCR8

S. Chul Kwon[1,2,3,†], Harim Jang[1,2,†], Siyuan Shen[4,†], S. Chan Baek[1,2], Kijun Kim[1,2], Jihye Yang[1,2], Jeesoo Kim[1,2], Jong-Seo Kim[1,2], Suman Wang[4], Yunyu Shi[4], Fudong Li[4,*] and V. Narry Kim[1,2,*]

[1]Center for RNA Research, Institute for Basic Science, Seoul 08826, Korea, [2]School of Biological Sciences, Seoul National University, Seoul 08826, Korea, [3]School of Biomedical Sciences, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China and [4]Hefei National Laboratory for Physical Sciences at Microscale, School of Life Sciences, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, Anhui 230026, China

Received May 13, 2020; Revised September 14, 2020; Editorial Decision September 15, 2020; Accepted October 02, 2020

ABSTRACT

The microprocessor complex cleaves the primary transcript of microRNA (pri-miRNA) to initiate miRNA maturation. Microprocessor is known to consist of RNase III DROSHA and dsRNA-binding DGCR8. Here, we identify Enhancer of Rudimentary Homolog (ERH) as a new component of Microprocessor. Through a crystal structure and biochemical experiments, we reveal that ERH uses its hydrophobic groove to bind to a conserved region in the N-terminus of DGCR8, in a 2:2 stoichiometry. Knock-down of ERH or deletion of the DGCR8 N-terminus results in a reduced processing of suboptimal pri-miRNAs in polycistronic miRNA clusters. ERH increases the processing of suboptimal pri-miR-451 in a manner dependent on its neighboring pri-miR-144. Thus, the ERH dimer may mediate 'cluster assistance' in which Microprocessor is loaded onto a poor substrate with help from a high-affinity substrate in the same cluster. Our study reveals a role of ERH in the miRNA biogenesis pathway.

INTRODUCTION

The canonical maturation pathway of microRNA (miRNA) begins with the transcription of a miRNA gene by RNA polymerase II. The primary transcript (also known as pri-miRNA) is cleaved by an RNase III complex microprocessor. After nuclear export, the cleaved hairpin product (called precursor miRNA or pre-miRNA) is further processed by another RNase III DICER. The resulting short duplex RNA is then loaded onto the argonaute (AGO) protein, and one strand of the duplex is discarded, producing the mature miRNA–AGO complex (referred to as RNA-induced silencing complex or RISC) (1).

Microprocessor is known to consist of two proteins, RNase III DROSHA and dsRNA-binding protein DGCR8 (2–8). DROSHA and DGCR8 are necessary and sufficient to cleave many pri-miRNAs tested *in vitro* (2,5). Several other proteins have been reported to associate with microprocessor and participate in pri-miRNA processing. For example, DEAD-box RNA helicases DDX5 and DDX17 were reported as DROSHA-interacting proteins (5). hnRNPA1 and KSRP were shown to regulate the processing of a subset of pri-miRNAs (9–11). SRSF3 binds to a *cis*-acting element (CNNC motif) to enhance the efficiency and accuracy of processing (12,13). However, except for DROSHA and DGCR8, none of these proteins are considered as the Microprocessor components because they do not form a stable complex with DROSHA or DGCR8.

A canonical pri-miRNA has characteristic features that are recognized by Microprocessor, including a ∼35-bp double-stranded stem, an apical loop (>10 bp), and basal single-stranded regions (12,14,15). In addition, four sequence motifs (basal UG, apical UGU, mismatched GHG and flanking CNNC motifs) enhance the processing efficiency and cleavage specificity (12,13,15,16).

Recent biochemical and structural studies enabled a clear understanding of the processing mechanism. Single-molecular measurement of the Microprocessor complex showed that the complex is a heterotrimer with one molecule of DROSHA and two molecules of DGCR8 (2,8). DROSHA, together with DGCR8, forms an elongated shape and spans the entire hairpin to measure its stem length (∼35-bp) (15,17–19). The RNA-binding affin-

*To whom correspondence should be addressed. Tel: +82 2 880 9120; Fax: +82 2 887 0244; Email: narrykim@snu.ac.kr
Correspondence may also be addressed to Fudong Li. Email: lifudong@ustc.edu.cn
†The authors wish it to be known that, in their opinion, the first three authors should be regarded as Joint First Authors.

ity mainly comes from DGCR8 rather than DROSHA (2). However, DROSHA constitutes the core of microprocessor. The central domain of DROSHA fits into the basal part of pri-miRNA while the double-stranded RNA-binding domain (dsRBD) of DROSHA binds to the mismatched GHG motif to fix the position of pri-miRNA within microprocessor (16,18,19). Two RNase III domains form an intramolecular dimer to create the processing center that cleaves pri-miRNA (6).

Here, we sought to identify the interactors of microprocessor, and found a small globular protein, Enhancer of Rudimentary Homolog (ERH), as a stable Microprocessor component. Our study reveals that ERH binds to the N-terminus of DGCR8 and enables efficient processing of suboptimal hairpins in polycistronic pri-miRNAs, providing new insights into the pri-miRNA processing.

MATERIALS AND METHODS

Cell lines

HEK293E cells (human embryonic kidney 293 EBNA1; authenticated by ATCC STR profiling) were grown in DMEM (Welgene) supplemented with 9% fetal bovine serum (Welgene). K562 cells (Korean Cell Line Bank; authenticated by ATCC STR profiling) were grown in RPMI1640 (Welgene) supplemented with 9% fetal bovine serum (Welgene). HCT116 cells (authenticated by ATCC STR profiling) and HCT116 DROSHA knock-out cells (whose parental cells were authenticated by ATCC STR profiling) (20) were cultured in McCoy's 5A medium (Welgene) supplemented with 9% fetal bovine serum (Welgene).

Preparation of knock-in cell line by CRISPR/Cas9

CRISPR-Cas9 experiments were designed and conducted based on previous papers (21,22). Briefly, an oligo duplex encoding the sgRNA sequence was cloned at the Bbs I site of pX458-pSpCas9 (BB)-2A-GFP plasmid (Addgene #48138). To construct a donor vector, the ~600 bp homology sequence to the target site was amplified by overlapping PCR and inserted into a Sal I and Not I digested pGL3-Basic plasmid (Promega). Cas9 and donor plasmids were co-transfected using Lipofectamine 2000 (Thermo Fisher Scientific) into HEK293E cells which were seeded in a six-well plate one day before transfection. Two days after transfection, cells were diluted to a final concentration of 0.5 cells per 100 μl and plated in two 96-well plates. After 2 weeks, single cell-derived colonies were picked and cultured in a 96-well plate for further experiments. To detect the insertion of the tag sequence by PCR analysis, the genomic DNA of each clone was extracted by using QuickExtract solution (Epicentre) and PCR was carried out. Genomic amplicons of the target region were cloned into a TopBlunt V2 vector (Enzynomics) and verified by Sanger sequencing. Subsequently, the expression of the 3xFlag-2xStrep tag was confirmed by immunoprecipitation and Western blotting.

Immunoprecipitation for mass analysis

Approximately 1.3E9 cells were resuspended in the Lysis buffer (20 mM Tris pH 8.0, 150 mM NaCl, 2 mM β-mercaptoethanol, 100 ng/μl RNase A) and lysed by sonication. After centrifugation at 35 000 g for 1 h at 4°C, lysates were incubated with Strep-Tactin beads (IBA Lifesciences) at 4°C overnight. Beads were washed with the Wash buffer (20 mM Tris pH 8.0, 150 mM NaCl, 2 mM beta-mercaptoethanol) for 5 times and incubated with the 50 mM biotin-containing Wash buffer for 1 h at 4°C. Biotin eluates were further incubated with anti-Flag M2 agarose beads for 3 h at 4°C. After washing, proteins were eluted with 500 ng/μl of the 3× Flag peptide-containing Wash buffer.

LC–MS/MS analysis and protein identification

The immuno-precipitated protein samples were first reduced and alkylated in 8 M urea with 50 mM ammonium bicarbonate (ABC) buffer. After 10-fold dilution with 50 mM ABC buffer, the protein samples were digested with 2% (w/w) trypsin at 37°C overnight. The resulting peptide samples were subject to C18 clean-up and loaded to the in-house packed trap column (3 cm × 150 μm i.d) and capillary analytical column (100 cm × 75 μm i.d.) with 3 μm Jupiter C18 particles (Phenomenex) for peptide separation. A flow rate of 300 nl/min and a linear gradient ranging from 95% solvent A (water with 0.1% formic acid) to 40% of solvent B (acetonitrile with 0.1% formic acid) for 100 min were applied on nanoACQUITY UPLC (Waters) coupled with Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific), which was operated at sensitive mode using the following parameters: m/z 350–1800 of precursor scan range, 1.4 Th of precursor isolation window, 30% of normalized collision energy (NCE) for higher-energy collisional dissociation (HCD), 30 s of dynamic exclusion duration, 120k or 30k resolution at m/z 200 and 100 or 54 ms of maximum injection time for full MS or MS/MS scan, respectively.

MS raw data files were processed with MaxQuant (version 1.5.3.30) against the human SwissProt database at default settings (20 ppm or 6 ppm of precursor ion mass tolerances for initial or main search, respectively, and 0.5 Da for fragment ion masses). Enzyme specificity was set to trypsin/P and a maximum of two missed cleavages were allowed. Cysteine carbamidomethylation and methionine oxidation were selected as fixed and variable modifications, respectively. A 1% false discovery rate was required at both the protein- and the peptide-level.

Sequence analysis

Protein secondary structure prediction was carried out by using PSIPRED 4.0 (23), and disordered regions were predicted with IsUnstruct v2.02 (24). Conservation scores were calculated by using the Scorecons server (25,26) with the input sequences listed in Supplementary Figure S2.

Immunoprecipitation of endogenous proteins

For endogenous protein immunoprecipitation, HEK293E cells grown on 100 mm dishes were lysed in the IP buffer (20 mM Tris pH 7.5, 100 mM KCl, 0.2 mM EDTA, 0.2% NP-40, 10% glycerol, 0.4 ng/μl RNase A and protease inhibitor cocktails (Calbiochem)) and sonicated. Clear lysates

were collected by centrifugation for 10 min at 4°C. The antibody was pre-incubated with 20 μl of prewashed protein G Sepharose (GE healthcare) at 4°C for 1 h. Antibody-conjugated beads were incubated with lysates at 4°C for 1–2 h and washed with the Wash buffer (20 mM Tris pH 7.5, 100 mM KCl, 0.2 mM EDTA, 0.2% NP-40, 10% glycerol) for 5 times.

Plasmids

Primers used for cloning are provided in Supplementary Table S5.

Immunoprecipitation of ectopically expressed proteins

For ectopic expression of the tagged protein (Flag, HA, myc), 10 μg of 1:1 mixture of branched polyethylenimine (PEI) (Sigma) and linear PEI (Polysciences) were diluted in 400 μl OPTI-MEM with total 5 μg of the expression plasmids. The mixture was incubated for 5 min before adding to 50% confluent HEK293E cells in a 100 mm dish. After 2 days, cells were harvested and lysed in the Lysis buffer (20 mM Tris pH 7.5, 150 mM NaCl, 0.1 ng/μl RNase A, and protease inhibitor cocktails (Calbiochem)) by sonication. The cell lysates were collected by full speed centrifugation for 10 min at 4°C. For Flag or HA-tagged protein immunoprecipitation, anti-Flag M2 affinity gel (Sigma-Aldrich) or anti-HA agarose affinity gel (Sigma-Aldrich) was pre-washed three times with 1 ml of the Wash buffer (20 mM Tris pH 7.5, 150 mM NaCl). In the case of anti-myc immunoprecipitation, the antibody was pre-incubated with 20 μl of prewashed protein G sepharose (GE healthcare) at 4°C for 1 h. Lysates and beads were incubated at 4°C for 1–2 h. Incubated beads were washed three times with the Wash buffer. Remaining supernatants were removed using a 30 gauge needle. Beads were resuspended in 1× SDS protein sample buffer (50 mM Tris pH 6.8, 10% Glycerol, 2% SDS, 100 mM DTT, 0.01% bromophenol blue) for western blotting.

Western blotting

For western blotting, protein samples were boiled at 95°C for 5 min before separating on Novex WedgeWell 8–16% Tris–Glycine Mini Gels (Thermo Fisher Scientific). Protein samples were transferred to polyvinylidene fluoride membrane (GE Healthcare). The signals were detected using the Odyssey infrared imaging system (Odyssey Sa, LI-COR) or using SuperSignal West Pico PLUS Chemiluminescent Substrate (Thermo Fisher Scientific).

Protein expression and purification

The full length protein coding sequence of human ERH (residues 1–104) was amplified by PCR from a human brain cDNA library and cloned into a modified pET28a (Novagen) vector without a thrombin protease cleavage site (termed p28a). SUMO tagged DGCR8$^{96–126}$ was also amplified by PCR from the human brain cDNA library and cloned to a modified pET28a vector with a SUMO protein fused at the N-terminus following the His_6 tag. All the Mutants were generated using the Takara MutanBEST Kit.

All proteins were expressed in Escherichia coli BL21(DE3) cells. Cells were grown in Luria-Bertani medium at 37°C until the OD_{600} reached about 0.8. Protein expression was induced with 0.3 mM β-D-1-thiogalactopyranoside (IPTG) for 24 h at 16°C. The His_6-tagged proteins, as well as SUMO tagged DGCR8 peptides, were first purified by Ni-chelating resin (Qiagen) in buffer A (20 mM Tris–HCl (pH 7.8), 500 mM NaCl). Proteins were further purified by size-exclusion chromatography on a Hiload 16/60 Superdex 75 column (GE healthcare) in buffer B (20 mM Tris–HCl (pH 7.8), 150 mM NaCl).

For preparation of the ERH-DGCR8 complex, the purified ERH protein was mixed with His_6-SUMO tagged DGCR8$^{96–126}$ protein at a 1:1.5 molar ratio and then the His_6-SUMO tag was cleaved by incubation with ULP1 (ubiquitin-like-specific protease 1) at 16°C overnight. After the digestion, the ERH-DGCR8$^{96–126}$ complex was further purified by size-exclusion chromatography on a Hiload 16/60 Superdex 75 column (GE healthcare) in buffer B.

Crystallization, data collection and structure determination

The ERH-DGCR8$^{96–126}$ complex crystals were grown from a solution containing 0.2 M imidazole malate pH 6, 8% (w/v) PEG4000 using the hanging drop vapor diffusion method at 20°C. The crystals were soaked in the mother liquor supplemented with 25% (v/v) glycerol and flash-frozen in liquid nitrogen. X-ray diffraction data were collected on beamline 19U1 of the Shanghai Synchrotron Radiation Facility (SSRF). The diffraction data were processed using the HKL2000 software (HKL Research) (27). The structures of the complex were determined by molecular replacement with the MOLREP program using the structure of human ERH (PDB ID: 1W9G) as the search model (28,29). The model was further built and refined using Coot (30) and Phenix.refine (31), respectively. Crystal diffraction data and refinement statistics are shown in Supplementary Table S4.

Isothermal titration calorimetry

All the ITC experiments were carried out on a MicroCal PEAQ-ITC calorimeter (Malvern panalytical) at 293 K. We conducted the ITC experiments using SUMO-tagged DGCR8 peptides for titration into ERH proteins. The titration protocol consisted of a single initial injection of 1 μl, followed by 19 injections of 2 μl His_6-SUMO-tagged DGCR8 peptides into the sample cell containing ERH protein. Thermodynamic data were analyzed with a single-site binding model using MicroCal PEAQ-ITC Analysis Software provided by the manufacturer.

Small RNA sequencing

Sequencing libraries were generated as previously described (16,32). Briefly, HEK293E cells were tandemly transfected with siRNAs using Lipofectamine 2000 on days 0 and 3. On day 5, total RNA was purified using TRIzol (Thermo Fisher Scientific), and 10 μg of total RNA was used for library preparation. K562 cells were resuspended in OPTI-

MEM (Thermo Fisher Scientific) and transfected with siRNAs using Neon (Thermo Fisher Scientific) with the condition of 1450 V, 10 ms, 3 pulses. Four days after transfection, total RNA was purified with TRIzol. Most of the sequencing analysis steps were done as previously described (16). In this paper, all the genome mapped reads (including ambiguously mapped ones) were used for counting miRNA reads. The normalization was done by RPM (reads per million) or DESeq2 (33). Small RNA-seq results from previous papers were downloaded from GEO (GSE141098 (34), GSE116303 (35)) and analyzed through the same pipeline except for handling different adaptors. To filter out lowly expressed miRNAs for reliable analysis, baseMean values from DESeq2 were used for cutoff: 800 (siERH), 100 (DGCR8 Δex2), 50 (SAFB DKO, HEK293T), 100 (SAFB DKO, Ramos). Linear regression models were drawn using seaborn.regplot (size of the confidence interval = 95%).

In vitro high-throughput microprocessor assay

We prepared a DNA template library for human pri-miRNAs (miRBase v21; (36)) by using massive parallel synthesis technology (Cellemics). The template consists of 125-nt minimal pri-miRNAs and 18-nt adapter sequences at both ends for PCR amplification. Pri-miRNA substrates were generated by T7 *in vitro* transcription and gel purification. Microprocessor complex was prepared from HEK293E cells overexpressing DROSHA-Flag and DGCR8-HA proteins. The enzyme was purified through Flag-IP and 3xFlag-peptide elution. With those substrates and enzymes, we conducted in vitro processing for 1 h at 37°C. Input and processed RNAs were separately purified and sequenced.

Sorting suboptimal pri-miRNAs

From the sequence results, we defined 'Cleavage Productivity' and 'Cleavage Specificity' to sort out suboptimal pri-miRNAs. Cleavage Productivity was defined as the product of cleavage efficiency and productive cleavage ratio that represents the ratio of productive cleavage over inverted (or abortive) cleavage. Cleavage Specificity was calculated as the cleavage proportion for the primary cleavage site. By taking mirtrons (37) as negative controls and DROSHA-dependent miRNAs (20) as positive controls, we postulated cutoffs for both metrics and achieved >82% of true positive rate and >98% of true negative rate. For those pri-miRNAs that failed the cutoffs for either Cleavage Productivity or Cleavage Specificity, we sort them as suboptimal pri-miRNAs. More detailed explanation for cleavage metrics will be reported elsewhere.

Global impacts of ERH, SAFB/SAFB2, or DGCR8 exon 2

After DESeq2 analysis, we further selected pri-miRNAs with sufficient depths for robust analysis. Due to the depth imbalances among libraries, we applied separate cutoffs to balance the number of pri-miRNAs in final sets (baseMean parameter over 100 for SAFB/SAFB2 Ramos and DGCR8 exon 2, 50 for SAFB/SAFB2 HEK293T and 800 for ERH library). Next, we classified pri-miRNAs into four subgroups based on genomic organization and optimality to see the impacts separately: 'clustered & optimal', 'clustered & suboptimal', 'stand-alone & optimal' and 'stand-alone & suboptimal'. We assigned a pri-miRNA as clustered if it colocalize with another pri-miRNA within 1500 bp in the genome.

ERH knock-down and qRT-PCR

To validate the role of ERH in the polycistronic suboptimal miRNA processing, siERH targeting 3′ UTR of ERH was transfected into 50% confluent HEK293E cells using Lipofectamine 3000 (Thermo Fisher Scientific) following manufacturer's instructions. After 48 h, the cells were trypsinized and counted for reverse transfection. For reverse transfection, 40 pmol of siRNA and 1 μg of pri-miR-144–451a expression vector were co-transfected into 4E5 cells in a six-well culture plate using 3 μl of Lipofectamine 3000. After additional 48 hours, RNA was harvested with TRIzol.

Relative accumulation of miRNA was calculated by two separate qPCR results. The level of mature miRNAs was measured with the TaqMan MicroRNA Assays (Thermo Fisher Scientifics, #001973 for U6 snRNA, #002676 for miR-144, #001105 for miR-451a) as manufacturer's instruction. The level of ectopic pri-miRNA transcripts was measured with qRT-PCR using SYBR green MIX (Thermo Fisher Scientific). The cDNAs for the qPCR were synthesized from 500 ng of total RNA using random hexamer primers (Thermo Fisher Scientific) and RevertAid Reverse Transcriptase (Thermo Fisher Scientific). U6 snRNA was used as the internal control for both qPCR. The pri-miRNA expression vector contains an SV40 artificial intron, in which the pri-miRNA sequence is inserted, and the sfGFP coding sequence. Transfection efficiencies between samples were normalized using sfGFP. Primers used in this experiment are provided in Supplementary Table S5.

ERH rescue experiment

ERH rescue experiment was performed through two repeated transfections. Endogenous ERH was knocked down with shERH targeting 3′ UTR during the first two days. Polycistronic miRNA expression plasmids and shERH plasmids were co-expressed during the next two days.

shERH plasmids were diluted in 500 μl of OPTI-MEM with 15 μl of Fugene HD (Promega) and mixed well with rapid pipetting. The mixture was incubated for 5 min before adding into 50% confluent HEK293E cells. After 48 h, the cells were treated with trypsin and then collected and counted. Total 2.5 μg of shERH plasmids, pri-miRNA expression plasmids, wild-type or homodimerization mutants ERH plasmids (1:2:2 ratio) were co-transfected into 4E5 cells in six-well culture plates using 5 μl of Fugene HD. After 48 h, the total RNA was harvested.

RESULTS

ERH associates with DGCR8

To examine if there is an unknown factor involved in the processing of pri-miRNA, we identified Microprocessor-

associated proteins using a recently developed strategy for human protein complexes (21). Briefly, we first knocked-in a segment encoding three consecutive Flag and two Strep tags (3xFlag-2xStrep) downstream of the start codon in the DGCR8 locus using CRISPR-Cas9. The DGCR8 complex expressed at the endogenous level was then isolated by Strep pull-down and biotin elution, followed by immunoprecipitation (IP) with anti-Flag antibody and elution with 3xFlag peptide (Figure 1A, Supplementary Figure S1A and B). This Strep-Flag tandem affinity purification allowed highly stringent purification of the complex. Only three proteins, DGCR8, DROSHA and ERH, were specifically detected with high LFQ intensities (Figure 1B, C and Supplementary Table S1). Because we treated RNase A during purification, ERH seems to bind to DGCR8 (or DROSHA) in an RNA-independent manner.

ERH was initially identified as a recessive enhancer of a weak wing phenotype of the rudimentary mutant in *Drosophila melanogaster* (38). Diverse functions of ERH have been described in multiple organisms. In *Xenopus*, ERH was reported as a cell type-specific transcriptional repressor (39) while in *Schizosaccharomyces pombe*, Erh1 promotes meiotic mRNA decay during vegetative growth (40). Human ERH was shown to be involved in chromosome segregation during mitosis (41,42).

The structure of ERH may provide an insight into how a conserved ERH protein family can perform multiple functions. ERH is a small globular protein (∼12 kDa) that forms a homodimer (43–45). In *S. pombe*, a dimerized Erh1 binds and connects two molecules of Mmi1, a YTH domain-containing RNA-binding protein, and the dimerization is required for the in vivo activity of Mmi1 (46,47). Thus, the functions of ERH may derive from its binding partners.

We validated the interaction between DGCR8 and ERH by co-IP (in the presence of RNase A) and Western blotting. First, when endogenous DGCR8 was pulled down, we could detect endogenous ERH from the precipitant (Figure 1D). Conversely, when endogenous ERH was immuno-precipitated, endogenous DGCR8 was co-precipitated (Figure 1E). In this condition, hnRNPA1, an abundant nuclear RNA-binding protein, was not precipitated with either DGCR8 or ERH (Figure 1D and E). Also, it is noted that endogenous DROSHA was detected in the ERH immuno-purified sample (Figure 1E), implying that ERH associates with Microprocessor.

To determine which component of Microprocessor ERH directly binds to, we examined the interaction between over-expressed proteins. Ectopic DROSHA was co-expressed with the C-terminal tail (CTT) of DGCR8 which constitutes the DROSHA-binding motif and competes out endogenous DGCR8 so the DROSHA protein is devoid of full-length DGCR8 (2,17). In this experimental condition, we could not detect meaningful interaction between DROSHA and ERH (Figure 1F), indicating that ERH binds to DGCR8 rather than DROSHA. To confirm this interaction unambiguously, we performed co-IP using the DROSHA knock-out cell line (20). Notably, the interaction between ERH and DGCR8 was not affected by the absence of DROSHA, indicating that ERH binds to DGCR8, independently of DROSHA (Figure 1G).

ERH binds to the N-terminus of DGCR8

The molecular function of each domain of DGCR8 has been extensively studied. RNA-binding heme domain (RHED) mediates homodimerization and recognizes the apical loop of pri-miRNA (48,49). Two double-stranded RNA-binding domains (dsRBDs) bind to the upper stem of pri-miRNA and are mainly responsible for RNA-binding affinity of Microprocessor (2,19). However, the function of the N-terminal region of DGCR8 is still unknown. The N-terminus contains regions that are conserved across vertebrates, suggesting its important contribution (Figure 2A and Supplementary Figure S2A).

To find the ERH-binding site of DGCR8, we tested the interaction between ERH and serial deletion mutants of DGCR8. Deletion of the N-terminus of DGCR8 (1–275 aa) abrogated the interaction (Figure 2B). We narrowed down the interface between DGCR8 and ERH using additional N-terminal deletion mutants and found that the region between I87 and A146 is necessary (Figure 2C).

The region between P96 and G139 is highly conserved among vertebrates (Figure 2A, highlighted in yellow and Supplementary Figure S2A). We further prepared several DGCR8 mutants that do not disrupt the overall protein folding based on the predicted secondary structure. Initially, we focused on the region between P96 and L115 because there were some conserved hydrophobic residues that might be used for protein docking (Supplementary Figure S2A). We deleted the region by generating DG115–773 (where 1–114 aa of DGCR8 was deleted) and DGΔ96–115 (where 96–115 aa of DGCR8 was removed). These two mutants could not pull down ERH, confirming the importance of this region in the interaction with ERH (Figure 2D). We also tested additional mutants in which five hydrophobic residues were replaced with alanine or serine (DG_Ala mut: V105A/F108A/L112A/L114A/L115A; DG_Ser mut: V105S/F105S/L112S/L114S/L115S). These two DGCR8 mutants were impaired in ERH binding, indicating that the hydrophobic residues may participate in the interaction (Figure 2D and Supplementary Figure S2B).

To identify a necessary and sufficient binding region, we prepared three short DGCR8 fragments fused to the C-terminus of GFP (Figure 2E). DG103–126 that contains the five hydrophobic residues we tested, did not bind to ERH, but DG96–126 was able to interact with ERH (Figure 2E and Supplementary Figure S2C). Further, DG96–139 binds more avidly to the ERH protein compared to DG96–126 (Figure 2E and Supplementary Figure S2C), indicating that this conserved block of DGCR8 (P96–G139) serves as the ERH-binding domain.

ERH interacts with DGCR8 in a 2:2 stoichiometry

Recent structure of *S. pombe* Erh1-Mmi1$^{95-122}$ complex showed that an Erh1 homodimer binds to Mmi1 in a 2:2 stoichiometry, mainly through a conserved dimer interface characteristic of ERH family proteins (46), which raises the question of whether similar mechanisms are used in the ERH-DGCR8 interaction. First, we performed isothermal titration calorimetry (ITC) assays to determine the binding affinity between ERH and the DGCR8$^{96-126}$ peptide. The result showed that ERH binds to DGCR8$^{96-126}$

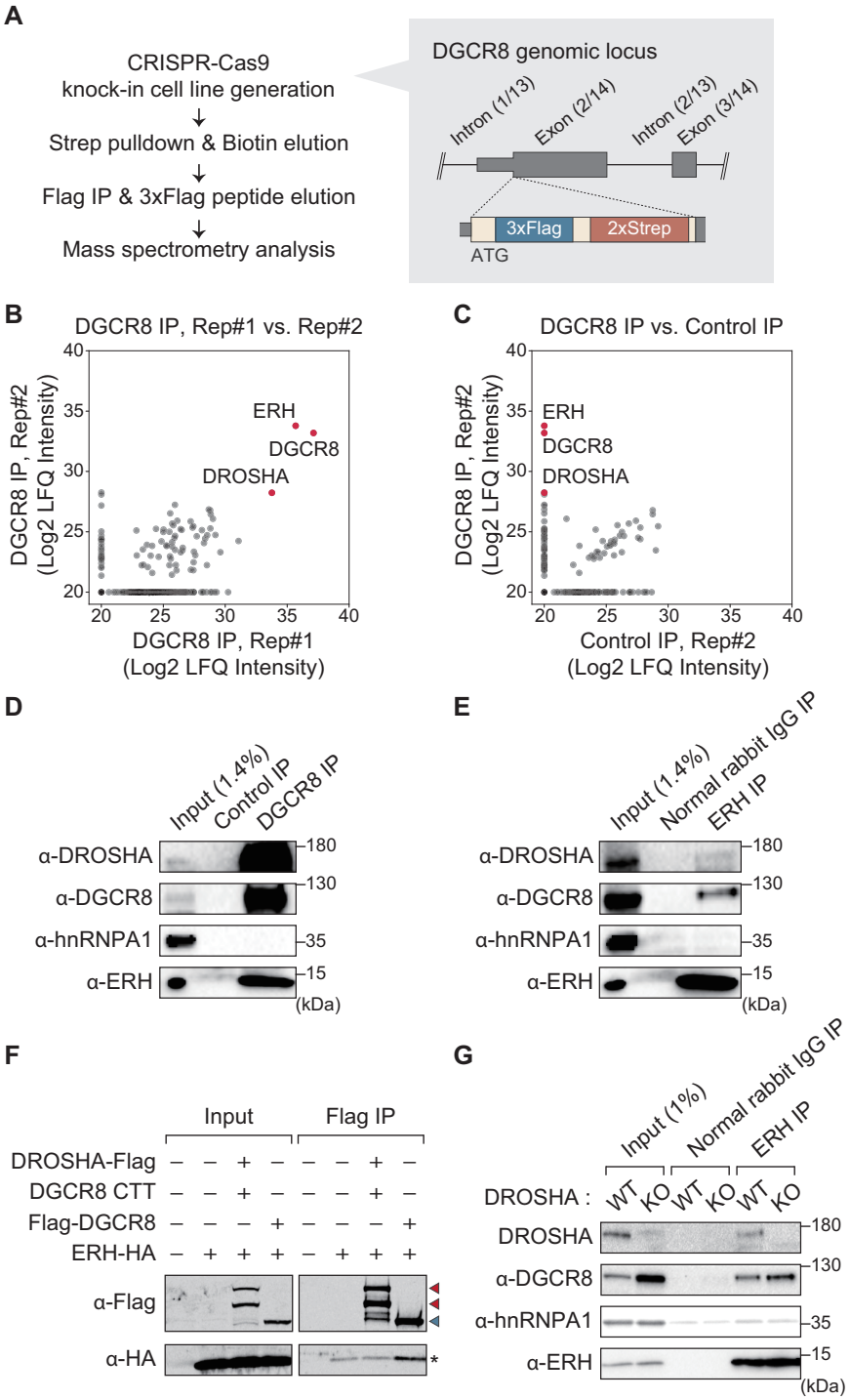

Figure 1. ERH binds to DGCR8. (**A**) Scheme of 3xFlag-2xStrep-DGCR8 knock-in cell generation, immunoprecipitation, and mass analysis. (**B, C**) Mass spectrometry results from DGCR8 IP (using the 3xFlag-2xStrep-DGCR8 cell line) and control IP (using the parental cell line). (**D, E**) Western blots showing the interaction between endogenous DGCR8 and ERH using HEK293E. The ascites fluid from the SP2/0 fusion partner cell was used for Control IP in D. (**F**) Western blots with overexpressed DROSHA (red triangle) or DGCR8 (blue triangle). DROSHA-Flag was co-transfected with the C-terminal tail of DGCR8 (see the text). Some overexpressed ERH proteins were nonspecifically pulled down by anti-Flag agarose beads and denoted with an asterisk (*). (**G**) Western blots using HCT116 wild-type or DROSHA knock-out cells (20).

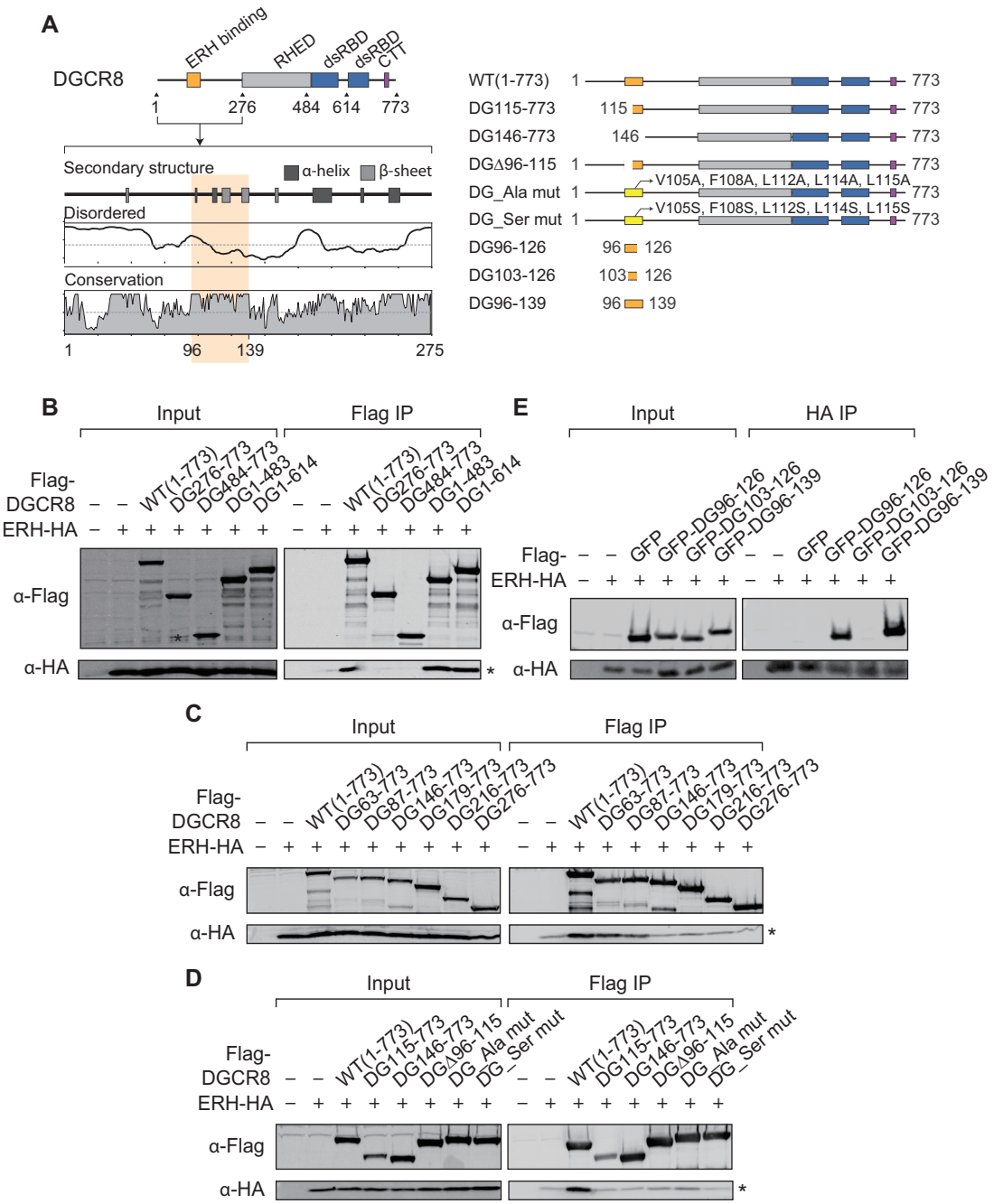

Figure 2. ERH binds to the N-terminus of DGCR8. (**A**) Sequence analysis of the N-terminus of human DGCR8. (**B–D**) Western blots with deletion or point mutants of DGCR8. Some overexpressed ERH proteins were nonspecifically pulled down by anti-Flag agarose beads and denoted with an asterisk (*). (**E**) Western blots with GFP-fused DGCR8 fragments.

with a dissociation constant (K_D) of about 19.6 μM and an N value of ∼1 (0.965) (Figure 3A and Supplementary Table S3). Next, we assessed the stoichiometry of the ERH-DGCR8$^{96–126}$ complex by analytical size-exclusion chromatography (SEC). Similar to Erh1, the ERH protein eluted as a single peak at 11.94 mL, representing an ERH homodimer in solution. We further purified the ERH-DGCR8$^{96–126}$ complex by incubating the ERH protein with His$_6$-SUMO tagged DGCR8$^{96–126}$, followed by cleavage of the His$_6$-SUMO tag by the ULP1 enzyme. The ERH-DGCR8$^{96–126}$ complex eluted as a single peak at 11.68 mL, indicating that the association of DGCR8$^{96–126}$ did not perturb the ERH homodimer (Figure 3B). In combination with the *N* value of the ITC results, we concluded that ERH and DGCR8$^{96–126}$ form a stable 2:2 stoichiometry complex similar to the Erh1-Mmi1$^{95–122}$ complex. However, the binding affinity of ERH with DGCR8$^{96–126}$ (∼19.7 μM) is apparently lower than that of Erh1-Mmi1$^{95–122}$ (∼0.83 μM), sug-

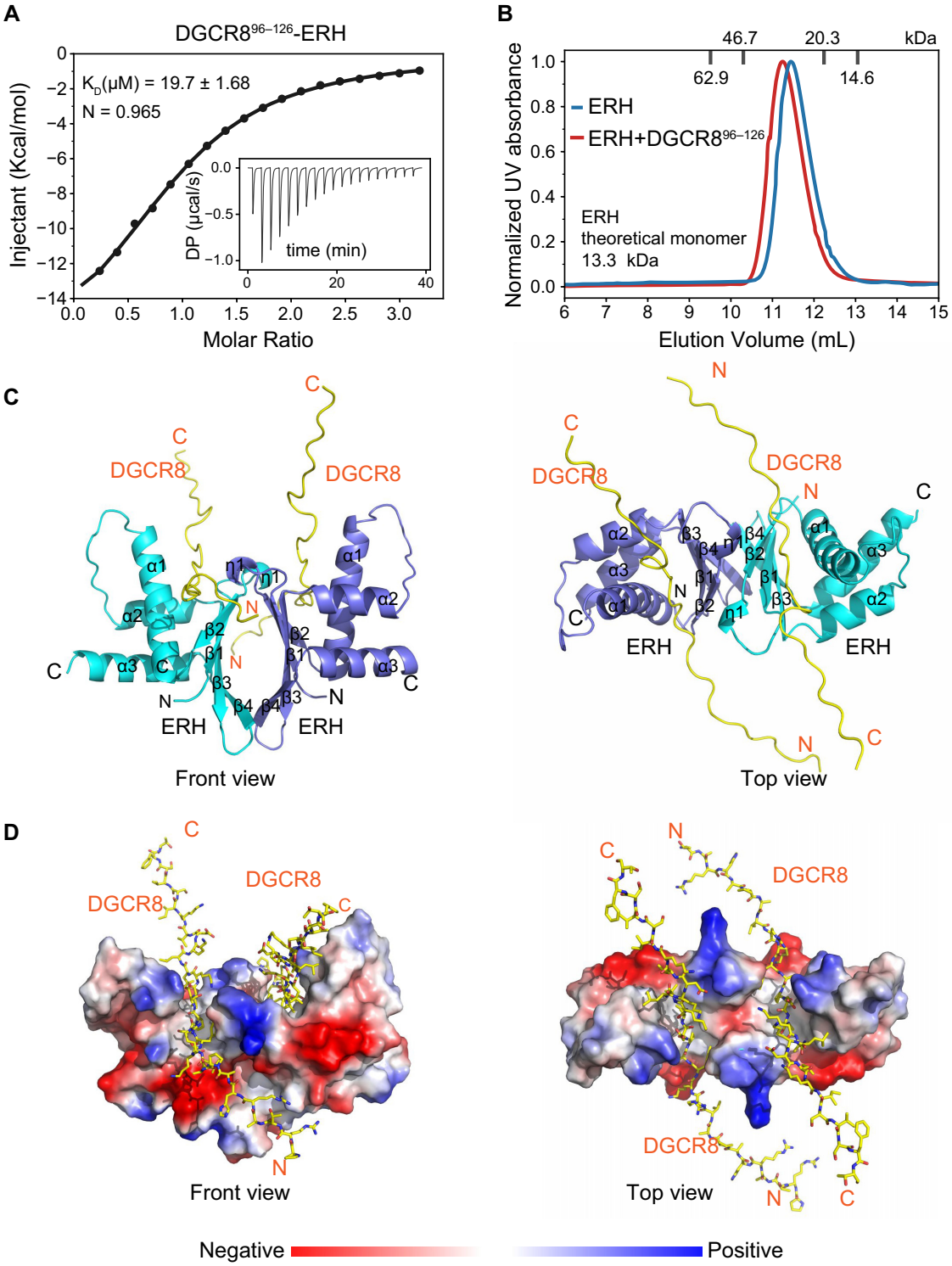

Figure 3. Structure of ERH in complex with the DGCR8$^{96–126}$ peptide. (**A**) The raw ITC titration data of ERH with SUMO-tagged DGCR8$^{96–126}$ and its fitting curve are shown. K_D, dissociation constant; DP, differential power; N, binding stoichiometry. (**B**) Size-exclusion analysis of ERH (blue) and ERH-DGCR8$^{96–126}$ (red). Marker sizes are 62.9 kDa (Albumin), 46.7 kDa (Ovalbumin), 20.3 kDa (Chymotrypsinogen), 14.6 kDa (Ribonuclease A). Also shown are theoretical monomer sizes (kDa) of ERH. (**C**) Ribbon representations of ERH bound to the DGCR8$^{96–126}$ peptide. One monomer of the ERH homodimer is colored in slate, while the other is colored in cyan. The two bound DGCR8$^{96–126}$ peptides are colored in yellow. (**D**) The two DGCR8$^{96–126}$ peptides are represented as sticks (yellow) on the molecular face of the ERH homodimer. Red and blue colors denote negative and positive surface charge, respectively.

gesting a different interaction manner which requires further structural investigations.

Structure of the ERH-DGCR8$^{96–126}$ Complex

We crystalized the ERH-DGCR8$^{96–126}$ complex (PDB ID: 7CNC) in space group C2, containing one ERH monomer and one DGCR8$^{96–126}$ peptide in each crystallographic asymmetric unit, and determined the structure at 1.6 Å resolution (Supplementary Table S4). Consistently with the 2:2 stoichiometry model based on the results from the SEC and ITC experiments, two neighboring ERH-DGCR8$^{96–126}$ complexes constitute a heterotetramer (Figure 3C). In the final model, most of the residues of the ERH protein (residues 1–100) and the DGCR8$^{96–126}$ peptide (residues 96–125) could be unambiguously built, with the electron densities of only small parts (the N-terminal His_6 tag and a few C-terminal residues of ERH and the DGCR8 peptide) being invisible, probably due to flexibility. The $2F_o - F_c$ electron density map for residues 96–125 of DGCR8 is shown in Supplementary Figure S3. Similar to the apo ERH structure (PDB ID: 1W9G), each ERH monomer folds as an α-β protein composed of a four-stranded antiparallel β sheet (β1–β4) with three amphipathic α helices (α1–α3) situated on one face of the β sheet. The outside faces of the β-sheets of two monomers constitute the homodimer interface and form a pseudo-β-barrel, mainly via hydrophobic interactions among several highly conserved residues. Comparison of our structure and the apo ERH structure do not show any significant structural changes upon DGCR8$^{96–126}$ binding (Supplementary Figure S4A–C). The all atom root-mean-square deviation (rmsd) is 0.22 Å.

In the structure, two DGCR8 peptides bind on the surface of the ERH homodimer, with each of them adopting an extended conformation. Although most residues of the DGCR8$^{96–126}$ peptide could be clearly traced, only a small portion of this peptide was found to interact with ERH. The interacting portion of peptide (residues 108–114) is located in the hydrophobic groove between α-helix bundle (mainly α1) and the β-sheet (termed 'α-β groove'), with the other portions of the peptides either lying outside the ERH homodimer or making few nonspecific contacts with other neighboring ERH proteins due to crystal packing (Figure 3C and D). Unexpectedly, we could not find any contacts between the DGCR8 peptide and the homodimer interface of ERH, while the conserved dimer interface of *S. pombe* Erh1 plays dominant roles in Mmi1 binding (46). The dimer interface and the α-β groove of Erh1 interact with the C-terminal and N-terminal portions of the Mmi1$^{95–122}$ peptide, respectively. Hence, >2500 Å^2 of the solvent exposed surface of Erh1 is buried by Mmi1, while only ∼1190 Å^2 of the solvent-exposed surface of ERH is buried by the two DGCR8 peptides. These observations also explain why the binding between ERH and DGCR8 is apparently weaker than the yeast Erh1-Mmi1 complex, which implies that the DGCR8-ERH interaction may be reversible and controllable.

The hydrophobic α-β groove of ERH consists of bulky side chains of L6, V8, P10, T18, M35, R42, F60 and L64. The side chains of F108, P110, L112 and L114 of DGCR8 project into this hydrophobic groove, participating in the ERH-DGCR8 interaction (Figure 4A and B). Notably, the phenyl group of F108 of DGCR8 is located between two proline residues, one of which comes from P110 of DGCR8 and the other comes from P14 of another ERH monomer (Figure 4B). Such Pro-Phe-Pro packing triad seems to play indispensable role in the ERH-DGCR8 interaction since either DGCR8^{F108A} or DGCR8^{P110A} is impaired significantly in binding to wild-type ERH (Figure 4C and Supplementary Table S3). Furthermore, the importance of two Leu side chains (L112 and L114) was confirmed by a double-mutant (DGCR8$^{L112A/L114A}$) which severely reduced affinity to ERH (Figure 4C and Supplementary Table S3).

We also measured the binding affinities of ERH mutants to DGCR8$^{96–126}$. Alanine substitutions of the ERH residues at the hydrophobic α-β groove, including ERH^{T18A}, ERH^{M35A}, ERH^{F60A} and ERH^{L64A}, affected the interaction (Figure 4D and Supplementary Table S3). When we simultaneously mutated F60 and L64 into alanine in ERH, the affinity between ERH$^{F60A/L64A}$ and DGCR8 was disrupted markedly, emphasizing the importance of this hydrophobic groove (Figure 4D and Supplementary Table S3). In addition to these hydrophobic contacts, the side chain of D111 of DGCR8 forms two hydrogen bonds with the side chain of R42, possibly providing further binding affinity and specificity (Figure 4B).

Suboptimal clustered pri-miRNAs are regulated by ERH and the N-terminus of DGCR8

We next investigated whether ERH affects the biogenesis of miRNAs. We performed small RNA-seq after ERH knockdown in HEK293E and K562 cells. Some miRNAs were significantly altered (Figure 5A, Supplementary Figure S5A and B). For validation, we carried out qRT-PCR and compared the results with small RNA-seq. Upon ERH depletion, the levels of pri-miR-196a-1 and pri-miR-197 modestly decreased while their mature counterparts slightly increased (Figure 5B). On the contrary, pri-miR-181b-1 and pri-miR-425 were markedly up-regulated while their mature miRNAs were down-regulated substantially in ERH depleted cells (Figure 5B). These data suggest that ERH plays a role in the pri-miRNA processing of a subset of miRNAs.

Recently, the Guang-Hui Liu group generated a cell line with a deletion of the DGCR8 N-terminal 240 aa (DGCR8 Δex2; deletion of exon 2) in human mesenchymal stem cells (35). The deletion resulted in differential expression of certain miRNAs as determined by small RNA-seq (35). We noticed that the deleted region of DGCR8 (1–240 aa) encompasses the ERH-binding site (96–139 aa). Therefore, we hypothesized that the miRNA expression profiles might be comparable. Indeed, even though the cell types are different, there was a correlation between commonly expressed miRNAs (Figure 5C). Notably, many commonly down-regulated miRNAs are polycistronic pri-miRNAs (Figure 5C, colored in red). It is also noted that among the miRNAs that we validated (Figure 5B), pri-miR-181b-1 and pri-miR-425 whose maturation was strongly regulated by ERH are polycistronic miRNAs, while pri-miR-196a-1 and pri-miR-197 are single-hairpin miRNAs. Taken together, these data

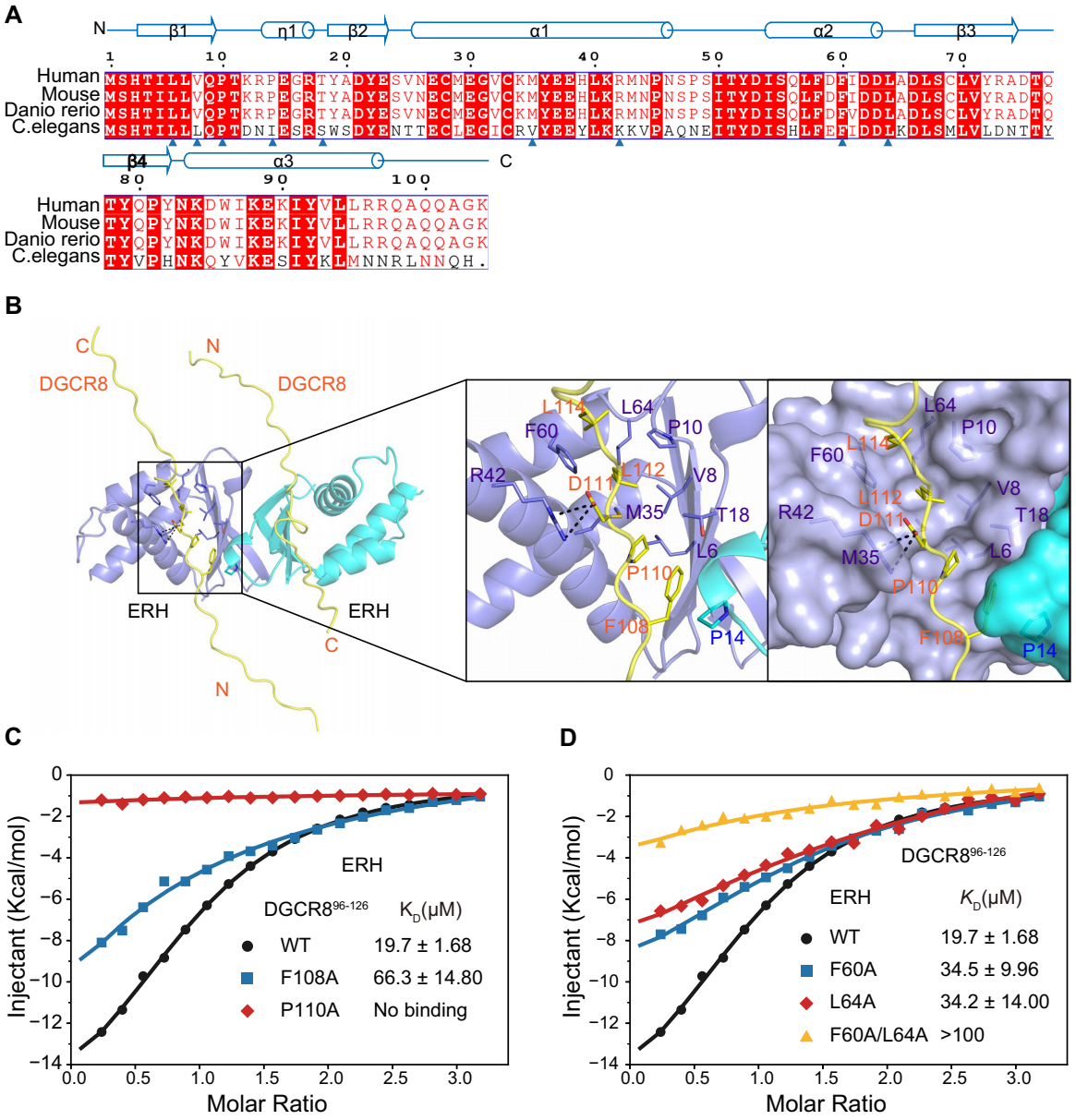

Figure 4. Molecular interface between ERH and DGCR8$^{96–126}$. (**A**) Sequence alignment of ERH proteins from *Homo sapiens* (human), *Mus musculus* (mouse), *Danio rerio,* and *C. elegans*. The alignment was generated by ESPript3 with CLUSTALW. The secondary structures of ERH, as determined by DSSP, are shown above the sequences. Red squares, identical residues; blue triangles, ERH key residues involved in the DGCR8$^{96–126}$ interaction. (**B**) Close-up views of the interactions between ERH and the DGCR8$^{96–126}$ peptide. (Left) ERH (slate and cyan) and DGCR8$^{96–126}$ (yellow) are shown as ribbons with selected side-chains as sticks. Hydrogen bonds are shown as black dashed lines. (Right) The Van der Waals surface views of ERH-DGCR8$^{96–126}$. The Van der Waals surface of the ERH is depicted as a semitransparent skin (slate). The DGCR8 peptide is represented as a stick diagram (yellow). (**C**) ITC fitting curves of ERH^{WT} using DGCR8$^{96–126}$ (black), DGCR8^{F108A} (blue), and DGCR8^{P110A} (red) are shown. (**D**) ITC fitting curves of ERH^{WT} (black), ERH^{F60A} (blue), ERH^{L64A} (red), ERH$^{F60A/L64A}$ (orange) using DGCR8$^{96–126}$ are shown.

implicate that ERH and the N-terminus of DGCR8 may act together in the processing of certain clustered pri-miRNAs.

About one-third of human miRNAs reside in 'polycistronic clusters' with more than one hairpins in a single nascent transcript (50,51). Intriguingly, the processing of some hairpins depends on that of their neighboring hairpins (52,53), indicating that there is an interplay between clustered hairpins. In a recent study, the Lai group and our laboratory revealed that the processing of pri-miR-451a, a suboptimal substrate for Microprocessor, is enhanced by the neighboring optimal hairpin, pri-miR-144 (54). The Bartel group also independently made a similar observation and showed that ERH facilitates cluster assistance (55). Moreover, before these two papers were published, a preprint from the Herzog group reported that pri-miR-15a could be processed by the help of the neighbor pri-miR-16-1 and a trans-acting protein SAFB2 (34). Among the miRNAs that we validated, pri-miR-181b-1 and pri-miR-425 are polycistronic miRNAs, while pri-miR-196a-1 and pri-miR-197 are single-hairpin miRNAs (Figure 5B).

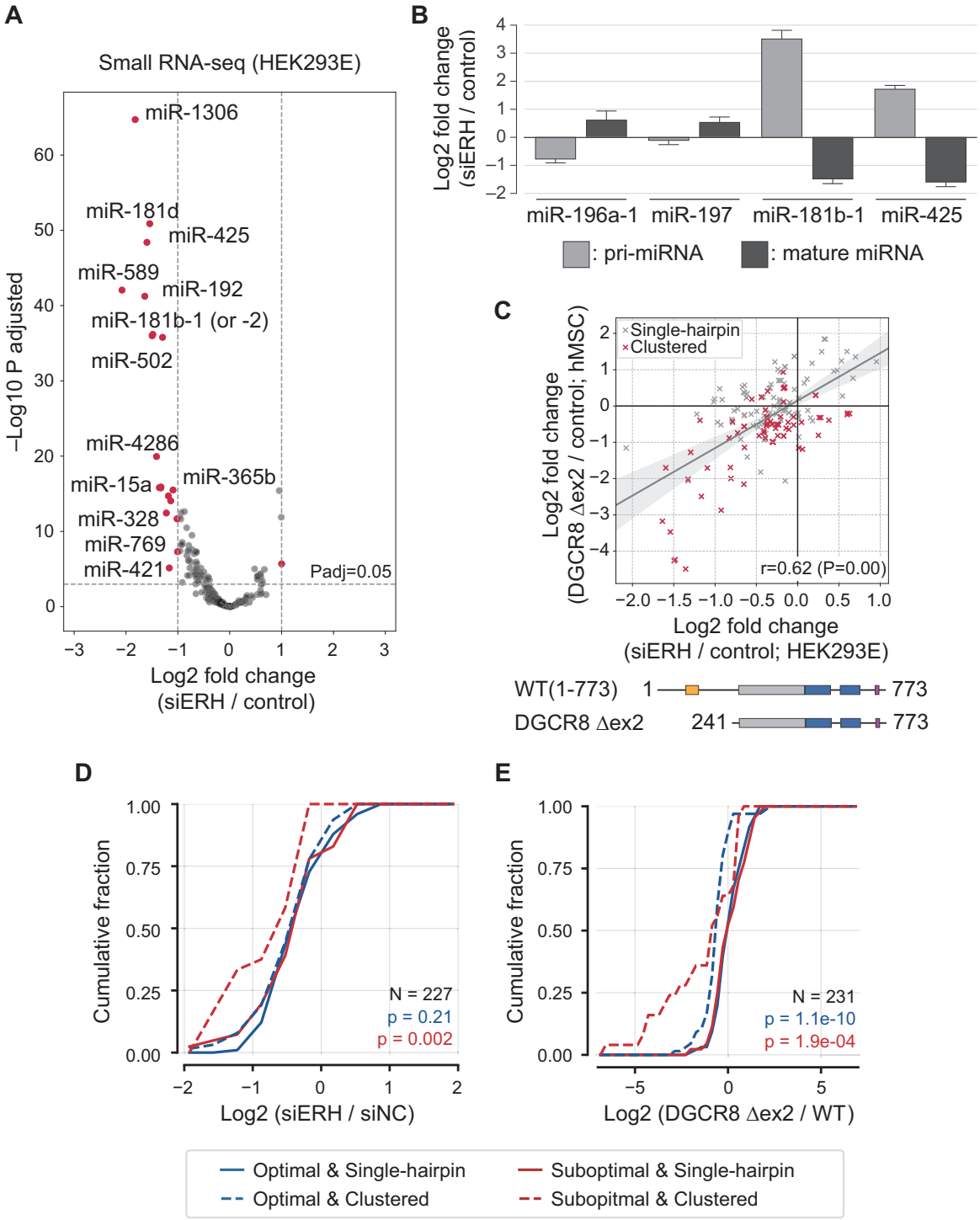

Figure 5. Suboptimal and clustered pri-miRNAs are regulated by ERH and the N-terminus of DGCR8. (**A**) Volcano plot showing differentially expressed miRNAs from ERH-depleted HEK293E cells. (**B**) qRT-PCR for validation of small RNA-seq. The expression levels of pri-miRNAs were detected by SYBR Green-based qRT-PCR after knocking down ERH for 3 days, and those of mature miRNAs were deduced from small RNA-seq. Error bars indicate standard error of the mean ($n = 3$) (**C**) Scatter plots of $\log_2$ fold changes of siERH (this study) versus DGCR8 Δex2 (35). A miRNA is defined as 'clustered' if another miRNA gene is found within ±1,500 bp in the genome. r is Pearson's correlation coefficient. (**D**) Global impacts of ERH on miRNA abundances. miRNAs were classified by their genomic organizations and optimal features for processing: 'Optimal & Single-hairpin' (blue line), 'Optimal & Clustered' (blue dashed line), 'Suboptimal & Single-hairpin' (red line) and 'Optimal & Clustered' (red dashed line). Fold-changes of 'Optimal & Clustered' or 'Suboptimal & Clustered' pri-miRNAs were compared against all stand-alone pri-miRNAs through one-sided Mann–Whitney *U* test. (**E**) Same analysis using DGCR8 Δex2 cells (35).

To systematically investigate the characteristics of the miRNAs affected by ERH, we utilized a high-throughput *in vitro* processing data which experimentally determined the relative processing efficiency and accuracy of all annotated human pri-miRNAs (K.K., S.C.B. and V.N.K., manuscript in preparation). 'Suboptimal' miRNAs were defined as those that are not efficiently or accurately processed in vitro (see Materials and Methods for details). 'Clustered' miRNAs were classified based on the genomic distance from the closest neighboring miRNA locus (within 1500 bp). Interestingly, the ERH knockdown affected the 'suboptimal and clustered' miRNAs more strongly than those in the other groups (Figure 5D). Likewise, the suboptimal and clustered miRNAs tended to be down-regulated more markedly than the other miRNAs in the DGCR8 Δex2 cell lines (Figure 5E).

ERH is required for the cluster assistance of pri-miR-144–451a

To experimentally verify the role of ERH in the processing of clustered suboptimal pri-miRNAs, we chose the pri-miR-144–451a cluster (54,55) because pri-miR-451a was affected substantially in our small RNA-seq data from K562 cells (Supplementary Figure S5B). We generated a construct containing both hairpins or those with a single hairpin (Figure 6A and B). siERH or control siRNA was transfected along with the miRNA constructs into HEK293E cells where endogenous pri-miR-144∼451a is barely detected. As expected from earlier findings (54,55), a higher level of mature miR-451a is expressed from the miR-144–451a constructs than the stand-alone miR-451a construct (Figure 6B). Importantly, in ERH-depleted cells, the mature miR-451a level from the miR-144∼451a construct was drastically reduced (Figure 6B). Moreover, when ERH was ectopically expressed in ERH-depleted cells, ERH increased the expression of miR-451a without significantly altering miR-144 expression (Figure 6C). Taken together, our data demonstrates a role of ERH in cluster assistance of miR-451a.

DISCUSSION

This study introduces ERH as the third core component of the Microprocessor complex. Stringent tandem affinity purification of DGCR8 from the endogenous genomic context yielded only three proteins: DGCR8 (the bait), DROSHA (the well-known partner of DGCR8), and ERH (Figure 1B and C). From literature search, we noticed that the interaction between DGCR8 and ERH has been detected in several previous proteomic experiments even though the results were not validated (56,57). Moreover, the theoretical molecular weight of the complex is close to the experimentally measured weight of the Microprocessor complex when ERH is considered. Previously, we estimated the mass of an endogenous Microprocessor as 364 (±25) kDa using a sedimentation coefficient and a Stokes radius (2). The theoretical weight of the complex is 355 kDa if the complex consists of one molecule of DROSHA (159 kDa), two molecules of DGCR8 (86 + 86 kDa), and a dimer of ERH (12 + 12 kDa). Thus, the full microprocessor complex is likely to be a heteropentamer.

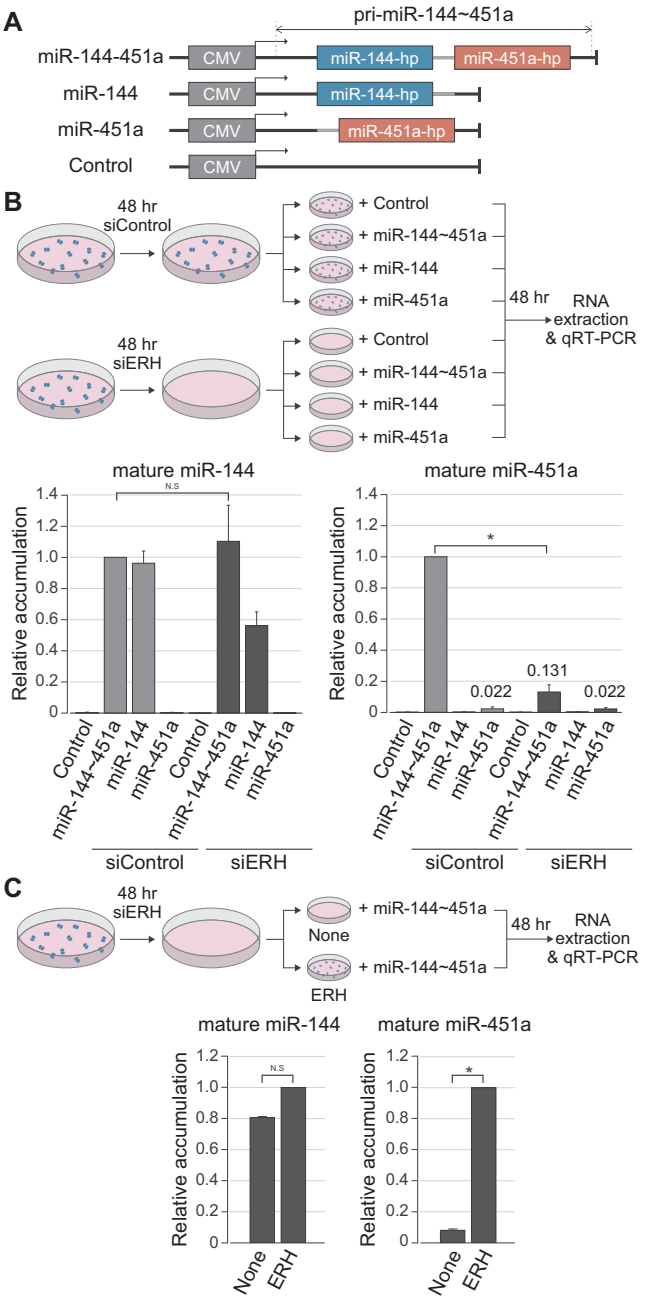

Figure 6. ERH is required for the processing of suboptimal clustered pri-miR-451a. (**A**) Scheme of ectopic pri-miR-144–451a constructs. (**B, C**) qRT-PCR showing the ERH dependency of pri-miR-144–451a processing. Error bars indicate standard error of the mean ($n = 3$). *$P < 0.01$; N.S., not significant.

We note that while we were still investigating the action mechanism of ERH, the Bartel group reported that ERH is co-purified with Microprocessor and is involved in cluster assistance (55). In our study, we not only independently discover ERH as a co-purified protein but further reveal that ERH directly binds to the N-terminus of DGCR8. Furthermore, our structural and biochemical data indicate that ERH and DGCR8 form a complex in a 2:2 stoichiometry (Figure 3). The high-resolution structure ERH-

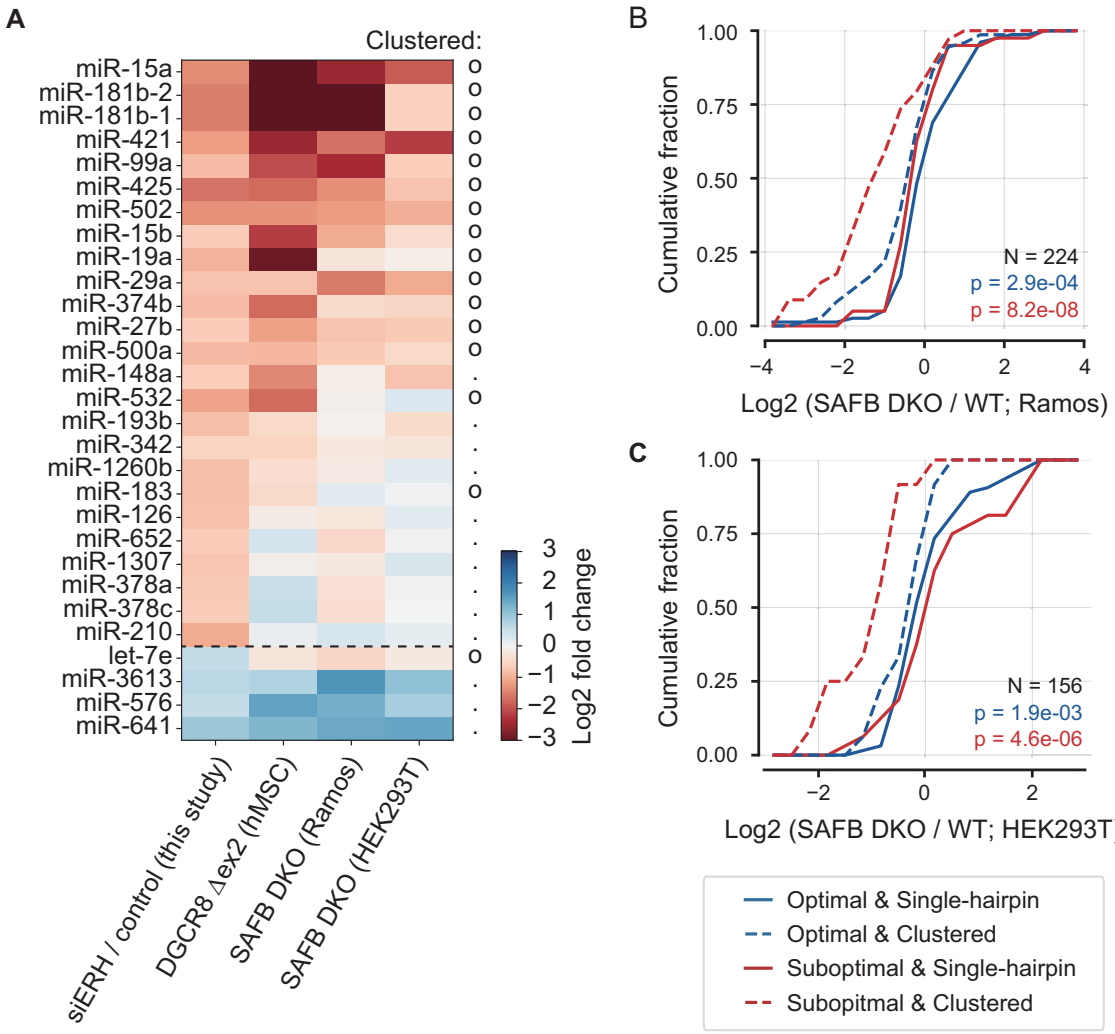

Figure 7. ERH, SAFB2, and the N-terminus of DGCR8 play a role together in the cluster-assisted processing of suboptimal hairpins. (**A**) Heatmap for comparing small RNA-seq results from siERH, DGCR8 exon 2 deletion, and SAFB/SAFB2 double knock-out (DKO). Commonly expressed miRNAs in all samples were first selected, and then differentially expressed miRNAs in the siERH sample (log_2 fold change >0.5 or <–0.5) were shown. (B, C) Same analysis of Figure 5D and E using SAFB/SAFB2 double knock-out in Ramos (**B**) or HEK293T (**C**).

DGCR8$^{96–126}$ uncovers the atomic basis of the interaction along the hydrophobic α–β grooves of ERH dimer (Figure 4).

This study confirms and extends the finding that ERH facilitates the processing of suboptimal polycistronic miRNAs. Our high-throughput in vitro processing data combined with small RNA sequencing data systematically revealed the association between the ERH-dependence, processing efficiency, and genomic organization (Figure 5). It is noted, however, that there are certain exceptions. For example, miR-589 and miR-328 are affected by ERH knockdown, but they appear to be stand-alone miRNAs with the nearest miRNA loci at 215 923 bp and 1 031 031 bp away, respectively (Figure 5A and Supplementary Table S2). Therefore, it is yet to be examined how the specificity of ERH regulation is determined and if there are any additional unknown mechanisms.

Recently, the Herzog group identified SAFB2 as a factor involved in cluster assistance through genome-wide CRISPR screen (34). Interestingly, among the unvalidated positive hits from the screen (34), we could find ERH, implying that both ERH and SAFB2 are required for cluster assistance. To test this possibility, we compared the miRNA alteration patterns of siERH-treated HEK293E cells (this study), the SAFB/SAFB2 double knock-out (DKO) Ramos or HEK293T cells (34). The overall alteration patterns were notably similar, and many of the commonly down-regulated pri-miRNAs reside in clusters (Figure 7A and Supplementary Figure S6). Furthermore, suboptimal miRNAs in clusters were mainly affected by SAFB/SAFB2 double knock-out (Figure 7B and C). Together, these data suggest that ERH, SAFB2, and the N-terminus of DGCR8 act together in the cluster-assisted processing of suboptimal hairpins. It is likely that the large SAFB2 protein may encompass the whole Microprocessor units by interacting with both DROSHA and ERH, because the interaction between ERH and SAFB has been previously reported (58). SAFB2, however, does not seem

to be a stable component of Microprocessor as it was not detected in our isolation experiment (Figure 1). The detailed action mechanism of DROSHA, DGCR8, ERH and SAFB2 needs further investigation using in vitro reconstitution with recombinant proteins.

Other candidate genes from the CRISPR screening (34) provide additional insight into cluster assistance. First, there were positive hits on the miR-1306 hairpin which resides in the second exon of the DGCR8 gene. Since Cas9 induces a short deletion to a target gene, it is likely that the DGCR8 protein was still expressed with a small deletion, and that ERH cannot bind to this DGCR8 as in the DGCR8 Δex2 (Figure 5C). Second, the screening also found UROD (also known as uroporphyrinogen decarboxylase), which is an essential enzyme involved in the heme biosynthetic process. Because heme is a cofactor for the proper dimerization of RHED domain of DGCR8 (48,49), the defect of UROD may affect DGCR8 dimerization. Combining these results and our data, we envision that proper dimerization of DGCR8 may be important for cluster assistance. The suboptimal hairpin may not be recognized easily by an ERH-lacking Microprocessor because two DGCR8 molecules may not be accurately aligned in the absence of ERH.

ERH is a highly conserved protein identified in most eukaryotes and modulates the function of partner proteins (57,59). There may be multiple ERH-containing complexes, and the Microprocessor seems to be one of them. Notably, the ERH-binding site of DGCR8 is only present in vertebrate homologs of DGCR8. Thus, it is likely that ERH has recently joined the miRNA biogenesis pathway and contributed to the emergence of new polycistronic miRNAs in vertebrates through cluster assistance. Our study provides an example of how the conserved protein ERH plays a unique role in a specific phylogenetic lineage.

Additional Note: While our manuscript was under revision, the bioRxiv manuscript from the Herzog group was published with additional data which experimentally validated the role of ERH in the processing of a suboptimal and clustered miRNA, pri-miR-15a (60).

DATA AVAILABILITY

The atomic coordinates and structure factors for the ERH-DGCR8 complex structure have been deposited to the Protein Data Bank (PDB) under the accession code PDB: 7CNC. The small RNA-seq data in this study have been deposited in the NCBI Gene Expression Omnibus under the accession number: GSE149421. Data and codes for in vitro Microprocessor processing experiments have not been included in this submission because the complete dataset will be reported separately. The other codes are available at https://github.com/jimkwon/nar2020.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Eunji Kim and Daeun Choi for the technical help, and Young-suk Lee for the critical reading of the manuscript. We also thank the staff of BL19U1 beamline at National Center for Protein Science Shanghai and Shanghai Synchrotron Radiation Facility, Shanghai, People's Republic of China, for assistance during data collection.
Author contributions: S.C.K. and H.J. performed biochemical experiments. S.S. performed protein expression, crystallographic experiments, and ITC experiments with the help of S.W. S.C.K. and S.C.B. carried out bioinformatics. J.Y. generated knock-in cell lines. J.K. and J.-S.K. performed mass spectrometry experiments and analysis. K.K. shared unpublished results. Y.S. supervised and helped structural experiments. F.L. designed the structural studies and determined the structure. S.C.K., H.J., S.S., S.C.B., F.L. and V.N.K. analyzed the data and wrote the manuscript.

FUNDING

Institute for Basic Science from the Ministry of Science and ICT of Korea [IBS-R008-D1 to S.C.K., H.J., J.Y., J.K., S.C.B., K.K., J.-S.K., V.N.K.]; BK21 Research Fellowships from the Ministry of Education of Korea [to H.J., S.C.B., K.K.]; Ministry of Science and Technology of China [2016YFA0500700, 2019YFA0508403 to S.S., S.W., Y.S., F.L]; Strategic Priority Research Program of the Chinese Academy of Sciences [XDB39000000 to S.S., S.W., Y.S., F.L.]; Chinese National Natural Science Foundation [U1932122 to S.S., S.W., Y.S., F.L]. Funding for open access charge: Institute for Basic Science; Ministry of Science and Technology of China; Strategic Priority Research Program of the Chinese Academy of Sciences; Chinese National Natural Science Foundation.
Conflict of interest statement. None declared.

REFERENCES

1. Treiber,T., Treiber,N. and Meister,G. (2019) Regulation of microRNA biogenesis and its crosstalk with other cellular pathways. *Nat. Rev. Mol. Cell Biol.*, **20**, 5–20.
2. Nguyen,T.A., Jo,M.H., Choi,Y.-G., Park,J., Kwon,S.C., Hohng,S., Kim,V.N. and Woo,J.-S. (2015) Functional anatomy of the human microprocessor. *Cell*, **161**, 1374–1387.
3. Lee,Y., Ahn,C., Han,J., Choi,H., Kim,J., Yim,J., Lee,J., Provost,P., Rådmark,O., Kim,S. *et al.* (2003) The nuclear RNase III Drosha initiates microRNA processing. *Nature*, **425**, 415–419.
4. Denli,A.M., Tops,B.B.J., Plasterk,R.H.a., Ketting,R.F. and Hannon,G.J. (2004) Processing of primary microRNAs by the Microprocessor complex. *Nature*, **432**, 231–235.
5. Gregory,R.I., Yan,K.-P., Amuthan,G., Chendrimada,T., Doratotaj,B., Cooch,N. and Shiekhattar,R. (2004) The microprocessor complex mediates the genesis of microRNAs. *Nature*, **432**, 235–240.
6. Han,J., Lee,Y., Yeom,K.-H., Kim,Y.-K., Jin,H. and Kim,V.N. (2004) The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev.*, **18**, 3016–3027.
7. Landthaler,M., Yalcin,A. and Tuschl,T. (2004) The human DiGeorge syndrome critical region gene 8 and its D. melanogaster homolog are required for miRNA biogenesis. *Curr. Biol.*, **14**, 2162–2167.
8. Herbert,K.M., Sarkar,S.K., Mills,M., Delgado De la Herran,H.C., Neuman,K.C. and Steitz,J.A. (2016) A heterotrimer model of the complete microprocessor complex revealed by single-molecule subunit counting. *RNA*, **22**, 175–183.
9. Guil,S. and Cáceres,J.F. (2007) The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a. *Nat. Struct. Mol. Biol.*, **14**, 591–596.
10. Michlewski,G. and Cáceres,J.F. (2010) Antagonistic role of hnRNP A1 and KSRP in the regulation of let-7a biogenesis. *Nat. Struct. Mol. Biol.*, **17**, 1011–1018.

11. Trabucchi,M., Briata,P., Garcia-Mayoral,M., Haase,A.D., Filipowicz,W., Ramos,A., Gherzi,R. and Rosenfeld,M.G. (2009) The RNA-binding protein KSRP promotes the biogenesis of a subset of microRNAs. *Nature*, **459**, 1010–1014.
12. Auyeung,V.C., Ulitsky,I., McGeary,S.E. and Bartel,D.P. (2013) Beyond secondary structure: primary-sequence determinants license pri-miRNA hairpins for processing. *Cell*, **152**, 844–858.
13. Kim,K., Nguyen,T.D., Li,S. and Nguyen,T.A. (2018) SRSF3 recruits DROSHA to the basal junction of primary microRNAs. *RNA*, **24**, 892–898.
14. Han,J., Lee,Y., Yeom,K.-H., Nam,J.-W., Heo,I., Rhee,J.-K., Sohn,S.Y., Cho,Y., Zhang,B.-T. and Kim,V.N. (2006) Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell*, **125**, 887–901.
15. Fang,W. and Bartel,D.P. (2015) The menu of features that define primary microRNAs and enable de novo design of microRNA genes. *Mol. Cell*, **60**, 131–145.
16. Kwon,S.C., Baek,S.C., Choi,Y.-G., Yang,J., Lee,Y.-s., Woo,J.-S. and Kim,V.N. (2019) Molecular basis for the single-nucleotide precision of primary microRNA processing. *Mol. Cell*, **73**, 505–518.
17. Kwon,S.C., Nguyen,T.A., Choi,Y.-G., Jo,M.H., Hohng,S., Kim,V.N. and Woo,J.-S. (2016) Structure of human DROSHA. *Cell*, **164**, 81–90.
18. Jin,W., Wang,J., Liu,C.-P., Wang,H.-W. and Xu,R.-M. (2020) Structural basis for pri-miRNA recognition by Drosha. *Mol. Cell*, **78**, 423–433.
19. Partin,A.C., Zhang,K., Jeong,B.-C., Herrell,E., Li,S., Chiu,W. and Nam,Y. (2020) Cryo-EM structures of human Drosha and DGCR8 in complex with primary microRNA. *Mol. Cell*, **78**, 411–422.
20. Kim,Y.-K., Kim,B. and Kim,V.N. (2016) Re-evaluation of the roles of DROSHA, Exportin 5, and DICER in microRNA biogenesis. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, E1881–E1889.
21. Dalvai,M., Loehr,J., Jacquet,K., Huard,C.C., Roques,C., Herst,P., Côté,J. and Doyon,Y. (2015) A scalable genome-editing-based approach for mapping multiprotein Complexes in Human Cells. *Cell Rep.*, **13**, 621–633.
22. Ran,F.A., Hsu,P.D., Wright,J., Agarwala,V., Scott,D.A. and Zhang,F. (2013) Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.*, **8**, 2281–2308.
23. Jones,D.T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.*, **292**, 195–202.
24. Lobanov,M.Y. and Galzitskaya,O.V. (2011) The Ising model for prediction of disordered residues from protein sequence alone. *Phys. Biol.*, **8**, 035004.
25. Valdar,W.S. and Thornton,J.M. (2001) Protein-protein interfaces: analysis of amino acid conservation in homodimers. *Proteins*, **42**, 108–124.
26. Valdar,W.S.J. (2002) Scoring residue conservation. *Proteins*, **48**, 227–241.
27. Otwinowski,Z. and Minor,W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.
28. Navaza,J. and Saludjian,P. (1997) [33]AMoRe: An automated molecular replacement program package. *Methods Enzymol.*, **276**, 581–594.
29. McCoy,A.J., Grosse-Kunstleve,R.W., Adams,P.D., Winn,M.D., Storoni,L.C. and Read,R.J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674.
30. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 486–501.
31. Adams,P.D., Afonine,P.V., Bunkoczi,G., Chen,V.B., Davis,I.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, Biol. Crystallogr.*, **66**, 213–221.
32. Kim,H., Kim,J., Kim,K., Chang,H., You,K.V. and Kim,N. (2019) Bias-minimized quantification of microRNA reveals widespread alternative processing and 3 end modification. *Nucleic Acids Res.*, **47**, 2630–2640.
33. Love,M.I., Huber,W. and Anders,S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.
34. Hutter,K., Lohmüller,M., Jukic,A., Eichin,F., Avci,S., Labi,V., Hoser,S.M., Hüttenhofer,A., Villunger,A. and Herzog,S. (2019) SAFB2 enables the processing of suboptimal stem-loop structures in clustered primary miRNA transcripts. bioRxiv doi: https://doi.org/10.1101/858647, 29 November 2019, preprint: not peer reviewed.
35. Deng,L., Ren,R., Liu,Z., Song,M., Li,J., Wu,Z., Ren,X., Fu,L., Li,W., Zhang,W. *et al.* (2019) Stabilizing heterochromatin by DGCR8 alleviates senescence and osteoarthritis. *Nat. Commun.*, **10**, 3329.
36. Kozomara,A. and Griffiths-Jones,S. (2014) MiRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic. Acids. Res.*, **42**, 68–73.
37. Ladewig,E., Okamura,K., Flynt,A.S., Westholm,J.O. and Lai,E.C. (2012) Discovery of hundreds of mirtrons in mouse and human small RNA data. *Genome Res.*, **22**, 1634–1645.
38. Wojcik,E., Murphy,A.M., Fares,H., Dang-Vu,K. and Tsubota,S.I. (1994) Enhancer of rudimentaryp1, e(r)p1, a highly conserved enhancer of the rudimentary gene. *Genetics*, **138**, 1163–1170.
39. Strandmann,E.P.v., Senkel,S. and Ryffel,G.U. (2001) ERH (enhancer of rudimentary homologue), a conserved factor identical between frog and human, is a transcriptional repressor. *Biol. Chem.*, **382**, 1379–1385.
40. Krzyzanowski,M.K., Kozlowska,E. and Kozlowski,P. (2012) Identification and functional analysis of the erh1+ gene encoding enhancer of rudimentary homolog from the fission yeast Schizosaccharomyces pombe. *PLoS One*, **7**, 14–16.
41. Fujimura,A., Kishimoto,H., Yanagisawa,J. and Kimura,K. (2012) Enhancer of rudimentary homolog (ERH) plays an essential role in the progression of mitosis by promoting mitotic chromosome alignment. *Biochem. Biophys. Res. Commun.*, **423**, 588–592.
42. Weng,M.T., Lee,J.H., Wei,S.C., Li,Q., Shahamatdar,S., Hsu,D., Schetter,A.J., Swatkoski,S., Mannan,P., Garfield,S. *et al.* (2012) Evolutionarily conserved protein ERH controls CENP-E mRNA splicing and is required for the survival of KRAS mutant cancer cells. *Proc. Natl. Acad. Sci. USA*, **109**, E3659–E3667.
43. Arai,R., Kukimoto-Niino,M., Uda-Tochio,H., Morita,S., Uchikubo-Kamo,T., Akasaka,R., Etou,Y., Hayashizaki,Y., Kigawa,T., Terada,T. *et al.* (2005) Crystal structure of an enhancer of rudimentary homolog (ERH) at 2.1 Å resolution. *Protein Sci.*, **14**, 1888–1893.
44. Li,H., Inoue,M., Yabuki,T., Aoki,M., Seki,E., Matsuda,T., Nunokawa,E., Motoda,Y., Kobayashi,A., Terada,T. *et al.* (2005) Solution structure of the mouse enhancer of rudimentary protein reveals a novel fold. *J. Biomol. NMR*, **32**, 329–334.
45. Wan,C., Tempel,W., Liu,Z.J., Wang,B.C. and Rose,R.B. (2005) Structure of the conserved transcriptional repressor enhancer of rudimentary homolog. *Biochemistry*, **44**, 5017–5023.
46. Xie,G., Vo,T.V., Thillainadesan,G., Holla,S., Zhang,B., Jiang,Y., Lv,M., Xu,Z., Wang,C., Balachandran,V. *et al.* (2019) A conserved dimer interface connects ERH and YTH family proteins to promote gene silencing. *Nat. Commun.*, **10**, 251.
47. Hazra,D., Andrić,V., Palancade,B., Rougemaille,M. and Graille,M. (2020) Formation of S. pombe Erh1 homodimer mediates gametogenic gene silencing and meiosis progression. *Sci. Rep.*, **10**, 1034.
48. Quick-Cleveland,J., Jacob,J.P., Weitz,S.H., Shoffner,G., Senturia,R. and Guo,F. (2014) The DGCR8 RNA-binding heme domain recognizes primary microRNAs by clamping the hairpin. *Cell Rep.*, **7**, 1994–2005.
49. Nguyen,T.A., Park,J., Dang,T.L., Choi,Y.-G. and Kim,V.N. (2018) Microprocessor depends on hemin to recognize the apical loop of primary microRNA. *Nucleic Acids Res.*, **46**, 5726–5736.
50. Altuvia,Y., Landgraf,P., Lithwick,G., Elefant,N., Pfeffer,S., Aravin,A., Brownstein,M.J., Tuschl,T. and Margalit,H. (2005) Clustering and conservation patterns of human microRNAs. *Nucleic. Acids. Res.*, **33**, 2697–2706.
51. Lee,Y., Jeon,K., Lee,J.T., Kim,S. and Kim,V.N. (2002) MicroRNA maturation: stepwise processing and subcellular localization. *EMBO J.*, **21**, 4663–4670.
52. Truscott,M., Islam,A.B.M.M.K. and Frolov,M.V. (2016) Novel regulation and functional interaction of polycistronic miRNAs. *RNA*, **22**, 129–138.
53. Lataniotis,L., Albrecht,A., Kok,F.O., Monfries,C.A.L., Benedetti,L., Lawson,N.D., Hughes,S.M., Steinhofel,K., Mayr,M. and Zampetaki,A. (2017) CRISPR/Cas9 editing reveals novel mechanisms of clustered microRNA regulation and function. *Sci. Rep.*, **7**, 8585.

54. Shang,R., Baek,S.C., Kim,K., Kim,B., Kim,V.N. and Lai,E.C. (2020) Genomic clustering facilitates nuclear processing of suboptimal pri-miRNA Loci. *Mol. Cell*, **78**, 303–316.
55. Fang,W. and Bartel,D.P. (2020) MicroRNA clustering assists processing of suboptimal microRNA hairpins through the action of the ERH protein. *Mol. Cell*, **78**, 289–302.
56. Li,S., Wang,L., Fu,B., Berman,M.A., Diallo,A. and Dorf,M.E. (2014) TRIM65 regulates microRNA activity by ubiquitination of TNRC6. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 6970–6975.
57. Kavanaugh,G., Zhao,R., Guo,Y., Mohni,K.N., Glick,G., Lacy,M.E., Hutson,M.S., Ascano,M. and Cortez,D. (2015) Enhancer of rdimentary homolog affects the replication stress response through regulation of RNA processing. *Mol. Cell. Biol.*, **35**, 2979–2990.
58. Drakouli,S., Lyberopoulou,A., Papathanassiou,M., Mylonis,I. and Georgatsou,E. (2017) Enhancer of rudimentary homologue interacts with scaffold attachment factor B at the nuclear matrix to regulate SR protein phosphorylation. *FEBS J.*, **284**, 2482–2500.
59. Weng,M.T. and Luo,J. (2013) The enigmatic ERH protein: its role in cell cycle, RNA splicing and cancer. *Protein and Cell*, **4**, 807–812.
60. Hutter,K., Lohmuller,M., Jukic,A., Eichin,F., Avci,S., Labi,V., Szabo,T.G., Hoser,S.M., Huttenhofer,A., Villunger,A. *et al.* (2020) SAFB2 enables the processing of suboptimal stem-loop structures in clustered primary miRNA transcripts. *Mol. Cell*, **78**, 876–889.