Structure of the dimerization domain of DiGeorge Critical Region 8

Rachel Senturia,¹ Michael Faller,¹ Sheng Yin,² Joseph A. Loo,^{1,2,3,4}
Duilio Cascio,³ Michael R. Sawaya,³ Daniel Hwang,¹ Robert T. Clubb,^{2,3,4}
and Feng Guo^{1,4*}

¹Department of Biological Chemistry in David Geffen School of Medicine, University of California, Los Angeles, California 90095

²Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90095

³Department of Energy Institute of Genomics and Proteomics, University of California, Los Angeles, California 90095

⁴Molecular Biology Institute, University of California, Los Angeles, California 90095

Received 30 March 2010; Revised 23 April 2010; Accepted 27 April 2010

DOI: 10.1002/pro.414

Published online 7 May 2010 proteinscience.org

Abstract: Maturation of microRNAs (miRNAs, ~22nt) from long primary transcripts [primary miRNAs (pri-miRNAs)] is regulated during development and is altered in diseases such as cancer. The first processing step is a cleavage mediated by the Microprocessor complex containing the Drosha nuclease and the RNA-binding protein DiGeorge critical region 8 (DGCR8). We previously reported that dimeric DGCR8 binds heme and that the heme-bound DGCR8 is more active than the heme-free form. Here, we identified a conserved dimerization domain in DGCR8. Our crystal structure of this domain (residues 298–352) at 1.7 Å resolution demonstrates a previously unknown use of a WW motif as a platform for extensive dimerization interactions. The dimerization domain of DGCR8 is embedded in an independently folded heme-binding domain and directly contributes to association with heme. Heme-binding-deficient DGCR8 mutants have reduced pri-miRNA processing activity *in vitro*. Our study provides structural and biochemical bases for understanding how dimerization and heme binding of DGCR8 may contribute to regulation of miRNA biogenesis.

Keywords: dimerization; WW motif; X-ray crystallography; 3D domain swapping; heme; RNA-binding protein; microRNA processing

Introduction

microRNAs (miRNAs) regulate gene expression by targeting specific messenger RNAs for translational repression or degradation.^{1,2} More than 700 human miRNAs have been cloned.³ They have important functions such as developmental timing and pattern-

ing, hematopoiesis, and apoptosis.^{4,5} Some miRNAs have been shown to be oncogenic or act as tumor suppressors.⁶ miRNAs have also been implicated in neurodegenerative diseases.⁷

miRNAs are transcribed as long primary miRNAs (pri-miRNAs), which undergo sequential cleavage steps in the nucleus and cytoplasm.^{8–10} In the first step, a pri-miRNA is cleaved into a ~65-nt intermediate, termed precursor miRNAs (pre-miRNAs), by the combined action of an RNase III enzyme Drosha¹¹ and an RNA-binding protein DiGeorge critical region 8 (DGCR8; the name in mammals; its fly and worm homologs are called Pasha and Pash-1, respectively).^{11–15} DGCR8 and Drosha form a complex termed the Microprocessor. The pre-miRNA is then transported to the cytoplasm and is further cleaved by Dicer, another RNase III

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: Jonsson Comprehensive Cancer Center at UCLA; Grant sponsor: Basil O'Connor Starter Scholar Research Award (March of Dimes Foundation); Grant number: 5-FY06-592; Grant sponsor: NIH; Grant numbers: GM080563, RR20004, AI52217.

*Correspondence to: Feng Guo, Department of Biological Chemistry, David Geffen School of Medicine, University of California, Los Angeles, CA 90095. E-mail: fguo@mbi.ucla.edu

enzyme, to give a duplex. The miRNA duplex is incorporated into the effector complex, RNA-induced silencing complex, and is unwound into mature single-stranded forms.

DGCR8 is essential for miRNA maturation.^{12,13,16} It independently binds to pri-miRNAs^{17–19} and makes a major contribution to their recognition by the processing machinery. In addition, proper control of DGCR8 expression and activity appears to be important for normal miRNA biogenesis. Downregulation of DGCR8 expression using RNA interference globally represses miRNA maturation and enhances cellular transformation and tumorigenesis,²⁰ consistent with clinical observations linking expression levels of miRNA processing factors and outcomes of cancers in patients.²¹ DGCR8 is among the ~30 genes that are heterozygously deleted in patients with DiGeorge syndrome.²² Haploinsufficiency of DGCR8 in mice causes abnormal miRNA processing and induces behavioral and neuronal deficits similar to a subset of symptoms observed in DiGeorge syndrome.²³

The C-terminal portion of the 773-residue DGCR8 protein (residues 499–751) contains two double-stranded RNA-binding domains (dsRBDs) and a conserved C-terminal tail (CTT) that are sufficient for pri-miRNA processing.^{18,24} The N-terminal portion of DGCR8 (residues 1–275) is required for nuclear localization.²⁴ The only recognizable motif in the central region of DGCR8 is a WW motif (residues 307–329), which is a widely distributed protein motif characterized by two tryptophan residues.²⁵ WW motifs are often used for interacting with proline-containing sequences. However, the function of the WW motif in DGCR8 has not been identified.

In a previous study, we found that a truncated form of DGCR8 (called NC1, amino acids 276–751) overexpressed in *E. coli* binds heme.¹⁸ Heme-bound NC1 is a dimeric protein associated with one heme, and it is more active than the heme-free form in reconstituted pri-miRNA processing assays. In addition, our parallel investigation revealed that caspase-mediated cleavage of DGCR8 produces a C-terminal fragment called DGCR8^{C2} with its N-terminus located around residue 406. DGCR8^{C2} does not bind heme, is monomeric, and is inactive in pri-miRNA processing because of the presence of its autoinhibition domain (residues 406–498) [Fig. 1(A)]. These observations suggested that dimerization of DGCR8 serves a regulatory function and heme is probably involved in miRNA processing through association with DGCR8.

Here, we present the identification and crystal structure of the dimerization domain of DGCR8. This structure shows an unexpected use of WW motif for dimerization and uncovers an extensive dimerization interface that includes highly conserved residues. The dimerization domain is embed-

ded in an independently folded heme-binding domain (HBD). The second tryptophan in the WW motif (residue 329) is required for association of DGCR8 with heme. Mutant DGCR8 proteins that fail to bind heme have reduced miRNA processing activity *in vitro*. Potential biological significance of dimerization and heme binding of DGCR8 in miRNA maturation is discussed.

Results

DGCR8 contains a conserved dimerization domain

We found that DGCR8^{276–412}-His₆ (the numbers in superscript represent the range of residues), which included the WW motif, existed as a dimer when expressed in *E. coli* [Fig. 1(B)]. The DGCR8^{276–412}-His₆ protein elutes from a size-exclusion column as a single peak, indicating a single oligomerization state. Using the elution volume of DGCR8^{276–412}-His₆ and a standard curve and assuming that this protein adopts a globular shape, its molecular weight is estimated to be 42 kDa, two to three times the molecular weight of each protomer (16.5 kDa). This result is consistent with the expectation that DGCR8^{276–412}-His₆ is a dimer, especially considering that nearly all DGCR8 proteins we have characterized elute in size-exclusion chromatography (SEC) as if they were with slightly higher molecular weights.

We further examined the oligomerization state of DGCR8^{276–412}-His₆ using mass spectrometry. Between residues 276 and 412 of DGCR8, there is a single cysteine residue at position 352. To avoid complications from potential nonspecific disulfides and to address the question if the dimerization is mediated by a disulfide between the two subunits, we mutated Cys352 to a serine in the context of DGCR8^{276–412}-His₆. This mutant eluted in SEC at the same volume as the wild type [Fig. 1(C)], indicating that a disulfide is not required for maintaining its oligomerization state. Finally, the DGCR8^{276–412}-His₆ C352S was characterized using electrospray ionization mass spectrometry [ESI-MS; Fig. 1(D)].²⁶ In relatively mild “native” conditions, some dimer was observed, whereas in denaturing conditions, only monomeric protein could be detected. The differences between the mass spectra in native and denaturing conditions confirm that DGCR8^{276–412}-His₆ is a dimer.

Nuclear magnetic resonance (NMR) characterization of ¹⁵N-labeled DGCR8^{276–412} revealed only ~50 peaks (~138 expected) in the backbone amide region of the Heteronuclear Single Quantum Coherence (HSQC) spectrum, suggesting that a large portion of this protein was unstructured and/or aggregating. Thereafter, we further narrowed the

Figure 1. Identification of a dimerization domain in DGCR8. A: Schematics of the DGCR8 constructs used in this article. Affinity tags are not shown. B: The SEC chromatogram of DGCR8²⁷⁶⁻⁴¹²-His₆, as the last step of purification, with a SDS gel of the peak fraction indicated on the right. “D” represents dimeric proteins (either covalently linked through a disulfide or noncovalently associated) sometimes observed on SDS gels. C: Same as (A), except that the DGCR8²⁷⁶⁻⁴¹²-His₆ C352S mutant was used. D: ESI mass spectra of the DGCR8²⁷⁶⁻⁴¹²-His₆ C352S mutant protein in both native (upper panel) and denaturing (lower panel) conditions. The dimer (●●; 33,264 Da) and monomer (●; 16,653 Da) peaks are indicated. (E) ESI mass spectra of DGCR8²⁷⁶⁻³⁵³. The dimer (●●; 17,745 Da) and monomer (●; 8,872 Da) species are shown.

dimerization domain to residues 276–353, as demonstrated by ESI-MS [Fig. 1(E)].

The sequences of DGCR8 homologs are conserved in the dimerization domain region, especially

the WW motifs, with that of Pash-1 from *C. elegans* showing the largest deviation. We found that a truncated Pash-1 (residues 149–266) from *C. elegans*, equivalent to the human DGCR8²⁷⁶⁻⁴¹², was also a

Figure 2. Pash-1 from *C. elegans* forms a noncovalent dimer in solution. ESI mass spectra of the Pash-1^{149–266} C224S/C244S mutant protein containing a C-terminal His₆ tag, in both native (upper panel) and denaturing (lower panel) conditions. The dimer species (●●; 27,209 Da) was apparent in native conditions, whereas monomer (●; 13,603 Da) was observed as the dominant species in denaturing conditions. The Cys-to-Ser mutations avoided the potential complications from covalent dimers that might form during manipulation of the protein samples.

heme-free dimer as indicated by ESI-MS (Fig. 2). Mutations of the only two cysteine residues in Pash-1^{149–266}, C224 and C244, to serine did not affect the

dimerization. Therefore, dimerization appears to be a conserved function of this domain in the DGCR8 family of RNA-binding proteins.

Crystal structure of the dimerization domain of DGCR8

We determined a crystal structure of the dimerization domain, human DGCR8^{276–353}, at 1.7 Å resolution using the single-wavelength anomalous dispersion (SAD) method (Table I). High-quality electron density maps allowed the structure of residues 298–352 to be traced with no ambiguity [Fig. 3(A)], while the remaining 23 terminal residues appeared to be disordered. There was one DGCR8 polypeptide chain in the asymmetric unit. Consistent with our biochemical studies, DGCR8^{276–353} appeared to be a dimer with the two subunits related by a crystallographic twofold symmetry [Fig. 3(B)].

The structure of each subunit is composed of four short (three to four residues each) β-strands surrounded by structured loops [Fig. 3(B,C)]. The three N-terminal β-strands (β1–β3) from one subunit form a twisted antiparallel β-sheet with the downstream strand from the other subunit (β4′) in a dimer. The β1–β3 strands are formed by residues in the WW motif. The relatively large number of Pro

Table I. Crystallographic Statistics

Data collection	Native	Se-Met
Space group	P4 ₃ 2 ₁ 2	P4 ₃ 2 ₁ 2
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	39.95, 39.95, 80.62	39.81, 39.81, 80.42
α, β, γ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	60–1.7 (1.76–1.7)	60–2.5 (2.59–2.5)
Observed reflections	68,401	30,318
Unique reflections	7,721	2,527
Wavelength (Å)	0.97918	0.97918
<i>R</i> _{sym}	0.076 (0.478)	0.074 (0.125)
<i>I</i> /σ(<i>I</i>)	24.6 (5.0)	28.4 (17.9)
Completeness (%)	99.9 (100.0)	99.8 (99.6)
Redundancy	8.9 (9.1)	12 (9.7)
Refinement		
Resolution (Å)	28.3–1.7	
No. of reflections used	7,435	
<i>R</i> _{work} (<i>R</i> _{free})	0.1742 (0.2052)	
Average <i>B</i> factor (Å ²)		
Protein/water	25.68/29.70	
Root mean square deviation		
Bond length (Å)/angles (°)	0.016/1.726	
SAD phasing		
No. of selenium sites found		1
Figure of merit (after density modification)		0.538 (0.785)
Content of asymmetric unit		
No. of protein molecules	1	
No. of protein		
Residues/atoms	55/440	
No. of solvent atoms	42	
Ramachandran statistics		
Allowed/generous/disallowed (%)	100/0/0	

$R_{\text{sym}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$. $R_{\text{work}} = \sum |F_o - F_c| / \sum F_o$. $R_{\text{free}} = \sum |F_o - F_c| / \sum F_o$, calculated using a random set containing 5% reflections that were not included throughout refinement.

Figure 3. Crystal structure of DGCR8 dimerization domain. A: The σ -weighted $2F_o - F_c$ electron density map of the dimerization domain in stereo and contoured at 1.0σ cutoff. B: Stereo diagram of the structure, viewed along a crystallographic twofold axis (represented by the black oval). The two subunits are drawn in green and cyan. Secondary structures and residues from the cyan subunit are labeled with a prime, both here and below. C: Sequence alignment of the DGCR8 homologs. The secondary structures of human DGCR8, determined by the crystal structure, are shown at the bottom in green. The two Trp residues, outlining the WW motif, are highlighted in yellow background. Residues directly involved in dimerization interface interactions are illustrated by “*.” Residues required for heme binding are indicated by bars.

residues (9 Pro of 55 ordered residues in the structure) contributes to the formation of loop structures. However, these Pro residues do not directly interact with the WW motif, unlike the proline-rich ligands that bind to other WW domains.²⁵

The interface between the two subunits of the dimerization domain is very extensive, with $\sim 2600 \text{ \AA}^2$ of solvent-accessible surface buried [Fig. 3(B)]. Each subunit has 29% of its total surface area buried in the dimer interface. The twofold symmetric dimer interface is composed of extensive hydrogen bonds and hydrophobic interactions. At the center of the dimer, the side chains of Ser330 from both subunits are hydrogen bonded to each other. Two hydro-

gen bonds are formed between main chains of strands $\beta 3$ and $\beta 4'$ in geometry typical of an antiparallel β -sheet. Surrounding these main-chain-main-chain interactions, Pro304, Trp307, Val326, Val327, Thr328, and Trp329 from one subunit form a hydrophobic patch complementing the surface contributed by residues Pro332, Tyr333, Leu335, Ile340, Pro345, Ile350, and Pro351 from the neighboring subunit [Fig. 4(A)]. In peripheral loop regions of the dimer interface, the side chain of Arg325 forms two hydrogen bonds with the main-chain carbonyls of Gly336 and Gly338, respectively; and the side chain of His312 hydrogen bonds with the main-chain carbonyl of His343 [Fig. 4(B)]. This dimerization

Figure 4. The extensive dimerization interface of DGCR8. A: The hydrophobic surface of a subunit of DGCR8^{276–353}. The cyan subunit of the dimerization domain in Figure 2(B) is viewed from the top. The surface composed of hydrophobic residues at the interface is drawn in yellow and those of hydrophilic residues in orange. The crystallographic twofold axis is shown by the dashed line. B: The C-terminal loop of the dimerization domain (residues 336–346, green) interacts with the surface of WW motif ($\beta 1'$ – $\beta 3'$) from the neighboring subunit (cyan).

interface can be summarized as the WW motif from one subunit interacting with the C-terminal portion of the other subunit [residues 332–352, containing strand $\beta 4'$; Fig. 4(B)]. Therefore, the dimerization domain of DGCR8 appears to be a three-dimensional (3D) domain-swapped dimer.²⁷ The residues that form the dimerization interface in human DGCR8 are highly conserved among the DGCR8 homologs [Fig. 3(C)].

A novel use of the WW motif for dimerization

There are five classes of WW domains that are distinguished by the proline-containing peptide ligands to which they bind.²⁵ The DGCR8 WW motif in each subunit is structurally superimposable to the other classes of WW domains [examples are shown in Fig. 5(A)]. The RMSD values of their superimposed $C\alpha$ atoms are around 1 Å. The largest structural variations occur in the loop between the $\beta 1$ and $\beta 2$ strands.

Based on available three-dimensional structures, we aligned the DGCR8 sequences with all known classes of WW motifs [Fig. 5(B)]. DGCR8 contains all the most conserved residues among all WW motifs: Trp307, Tyr319, Trp329, and Pro332. Within

each subunit in our structure, as well as in other WW domain structures, Pro332 stacks with Trp307 and forms a hydrophobic core with Tyr319. These interactions have been shown to be important for folding and structural stability of other WW domains.³⁰ Similarly, mutation of either Trp307 or Pro332 to alanine in DGCR8 in the context of NC1 caused a great reduction of structural stability, as indicated by elevated susceptibility to proteolysis during bacterial expression and subsequent purification. However, the NC1 Y319A mutant remains soluble in solution and is active in pri-miRNA processing assays (data not shown), indicating some unique folding properties of DGCR8. Overall, the sequence and structure of the WW motif in DGCR8 are similar to the well-characterized WW domains.

The WW motif embedded in the dimerization domain of DGCR8 is distinct from most other WW motifs in that it directly mediates dimerization. Most of the characterized WW domains are monomeric.²⁵ The only other known WW homodimer is located in mouse Sav1 protein.²⁸ The NMR structure of this WW dimer reveals that the β -sheets from both subunits directly contact each other through hydrophobic interactions, forming a β -barrel together. DGCR8 dimerizes through a different strategy, namely domain swapping between the WW domain and the immediate C-terminal region from the partner subunit [Fig. 3(B)]. The surface that a classic WW domain uses to interact with proline-

Figure 5. Comparison of the WW motifs in DGCR8 and other WW domains. A: The WW motif of DGCR8 is structurally superimposable to other WW domain structures. The main chains of DGCR8 (green), mSav1 (purple, PDB ID 2dvv),²⁸ and dystrophin (yellow, PDB ID 1EG3) are shown.²⁹ (B) Structure-based sequence alignment of WW families. Black rectangles delineate the most conserved residues among all WW domains. The secondary structures of human DGCR8 assigned based on our crystal structure are indicated at the bottom.

containing ligands is adopted by DGCR8 as a part of the dimerization interface [Fig. 4(B)]. Thus, it is unlikely that the WW motif in DGCR8 binds proline-rich ligands.

The dimerization domain is embedded in an independently folded heme-binding domain

Our previous study suggests that residues 276–498 may constitute a heme-binding domain [Fig. 1(A)]. Indeed, overexpression of this region with a C-terminal His₆ tag (HBD-His₆) produced an independently folded, soluble heme-binding domain [Fig. 6(A)] that interacted with heme in a way similar to the NC1 protein.¹⁸ The purified HBD-His₆ was absorbed at peak wavelengths of 450, 367, and 556 nm [Fig. 6(B)] and was a dimer bound by one heme molecule, as shown using ESI-MS [Fig. 6(C)]. Thus, the dimerization domain of DGCR8 is embedded in the HBD.

The dimerization domain provides a surface for interaction with heme

We next examined the contribution of the DGCR8 dimerization domain to the interaction with heme using site-directed mutagenesis. In these experiments, the ability of a mutant to bind heme was tested through overexpression of these proteins in *E. coli*, followed by purification using ion exchange and SEC. The overexpression experiments were performed both in the presence and in the absence of δ -aminolevulinic acid (δ -ALA), a heme biosynthesis precursor. We previously showed that the wild-type NC1 protein was fully occupied by heme when δ -ALA was supplied in the bacterial cell culture at the time of induction, whereas about half of the total NC1 protein was present in a heme-free state when δ -ALA was not added.¹⁸ In the same study, we also found that mutation of Cys352 to an Ala, Ser, or His residue completely abolished heme binding to the NC1 protein expressed in bacteria.

Here, we report an additional residue, Trp329, as required for heme binding. Even in the overexpression condition with δ -ALA, the purified NC1 W329A and W329H mutant proteins did not associate with heme at all, as indicated by the lack of Soret band in the UV-vis absorption spectra of these proteins [Fig. 7(A,B)]. In contrast, residues located in other regions of the DGCR8 dimerization domain, such as His312, are not required for binding heme regardless if δ -ALA was used (data not shown). The mutants described here are unlikely to be misfolded: they remain soluble in solution, behave similarly to the wild-type protein during purification, and bind pri-miRNAs with affinity similar to the wild type [Fig. 8(B); and data not shown].

Although the dimerization domain of DGCR8 does not bind heme when overexpressed in *E. coli*, it contains both Cys352 and Trp329. Strikingly, these residues from both DGCR8 subunits are clustered

Figure 6. DGCR8 contains a dimeric heme-binding domain. A: The SEC chromatogram of HBD-His₆ in the last step of purification, with a SDS gel of the peak fraction indicated in the inset. B: UV-visible spectrum of HBD-His₆ showed the absorption peaks at the same wavelengths as those of the heme-bound NC1 protein. C: ESI mass spectrum of HBD-His₆. The major complex observed is the dimer/heme species (●; 53,894 ± 4 Da), with minor contributions from the monomer/heme (▲; 27,252 ± 1 Da) and monomer (△; 26,637 ± 1 Da) species.

on a common face of the dimeric structure, strongly suggesting that this subdomain of HBD directly contacts heme [Fig. 7(C)]. In this structural architecture, dimerization is directly linked to the ability to bind heme. The heme-binding surface outlined by our mutagenesis data is almost 20 Å in the longest dimension (between the furthest atoms of Trp329), long enough to accommodate a heme molecule,

Figure 7. The dimerization domain contributes a surface for association with heme. A and B: Trp329 is required for heme binding. The NC1 W329A (A) and W329H (B) mutants were overexpressed in *E. coli* both in the presence of 1 mM δ -ALA. The SEC chromatograms obtained in the last step of purification are shown, with the 280- and 450-nm absorption curves drawn in black and red, respectively. Asterisks indicate an impurity (mostly nucleic acids) from the bacterial extract. The UV-vis spectra of the SEC peak fractions (indicated by bars above the SEC peaks) are shown below. The results for these mutants expressed in the absence of δ -ALA are similar (not shown). C: Surface drawing of the dimerization domain structure with the residues required for heme binding (Trp329 and Cys352) shown in red. The other residues in the two subunits in the dimer are shown in green and cyan, respectively.

Figure 8. The DGCR8 W329A and W329H mutants have reduced activity. A: In reconstituted pri-miRNA processing assays, a pri-miR-30a fragment was uniformly labeled with ^{32}P and was incubated with purified recombinant Drosha and 50 nM of wild-type or mutant proteins (in the context of NC1) for 45 min. The reactions were analyzed using a 15% denaturing polyacrylamide gel and autoradiography; M, size markers. B: Filter-binding assays. The upper panel indicates the data fitted using a cooperative trimer model, in which three NC1 molecules binds one RNA cooperatively. The low panel shows the Hill plot of the RNA-binding data. Only the region in the binding transition is shown.

which has edges of 11–13 Å and a diagonal distance of 13–15 Å.

Heme-binding-deficient DGCR8 mutants have reduced pri-miRNA processing activity

The NC1 W329A and W329H mutant proteins were tested in the reconstituted pri-miRNA processing assays. When compared with the heme-bound NC1, both W329A and W329H mutants (heme-free dimers) showed reduced pri-miRNA processing activity [Fig. 8(A)], similar to the NC1 C352A mutant characterized previously.¹⁸ We previously showed using filter-binding assays that the heme-bound wild-type NC1 proteins bind to the pri-miR-30a RNA with high cooperativity (Hill coefficient of 2.9).¹⁸ Here, using the same assay, we show that the binding affinity and cooperativity of the W329A dimer to pri-miR-30a are similar to those of the wild-type NC1 [Fig. 8(B)]. These results are consistent with a role of heme in regulating miRNA maturation.

Discussion

Here, we identified a dimerization domain in DGCR8 that is highly conserved in this family of RNA-binding proteins. Our crystal structure of this domain and mutagenesis data indicate that the WW motif of DGCR8 serves as a major component of the domain-swapped dimer and directly contributes to the association with heme. Mutations of key residues in the dimerization domain reduce the pri-miRNA processing activity of DGCR8 *in vitro*, suggesting a regulatory function of the dimerization domain, as well as the larger heme-binding domain, in miRNA maturation.

Kim and coworkers¹⁴ previously showed that the Microprocessor complex contains more than one copy of the DGCR8 protomer. In their experiments, they expressed a mixture of DGCR8 with either V5 or FLAG tag in HeLa cells and found that the V5- and FLAG-tagged DGCR8 coimmunoprecipitated with each other. The direct interactions we observed between the dimerization domains may be responsible for the coimmunoprecipitation of DGCR8 molecules. Together, our data suggest that dimeric DGCR8 may be a functional form of this protein in cells.

Previous studies showed that (a) the dsRBDs and CTT regions of DGCR8 (residues 499–751) are monomeric in the absence of RNAs and are sufficient for pri-miRNA processing^{18,24}; (b) the DGCR8^{406–751} is also monomeric but is inactive, and therefore, the immediate N-terminal neighboring region (residues 406–498) of dsRBD1 is an autoinhibition domain (Gong *et al.* manuscript provided with this submission); and (c) NC1 (residues 276–751), containing the dimerization and autoinhibition domains (they form the HBD), dsRBDs and CTT, can form heme-bound dimer that is active in pri-

miRNA processing.¹⁸ Thus, the dimerization domain we characterized in this study can activate miRNA maturation through releasing the autoinhibition. The mechanism of activation by the dimerization domain likely involves its interactions with the autoinhibition domain and probably with a heme cofactor.

The dimerization domain of DGCR8 is stabilized through 3D domain swapping between the WW domain of one subunit and the immediate downstream region of its partner subunit (Fig. 2). We speculate that “unswapping” would occur if the linker between the $\beta 3$ and $\beta 4$ strands (residues 329–332) loops back so that $\beta 4$ occupies the position of $\beta 4'$ in a dimer. In this unswapped monomeric conformation, most interactions between neighboring subunits in the dimer are preserved. Indeed, some wild-type NC1 protein adopts a monomeric conformation in the absence of heme.¹⁸ Thus, 3D domain swapping could be a structural mechanism through which DGCR8 forms the heme binding pocket, because an unswapped monomer would not contain a complete heme-binding surface.

The association with heme could alter the distribution of DGCR8 in different oligomerization states. We note that the heme-free NC1 C352A mutant adopts mostly monomeric conformation,¹⁸ whereas the heme-free W329A and W329H mutants mainly exist as dimers [Fig. 7(A,B)]. Mutation of either C352 or W329 causes defects in pri-miRNA processing *in vitro*. Therefore, both dimerization and heme binding of DGCR8 are likely to be important for its functions.

Cys352 is completely conserved in the DGCR8 family proteins [Fig. 3(C)]. Trp329 is conserved in all vertebrate DGCR8 homologs, but is substituted by an alanine in the DGCR8 homolog (Pasha) in flies and by a histidine in the Pash-1 protein in worms, respectively.¹² Mutation of Trp329 to either alanine or histidine in human DGCR8 abolished its interaction with heme in the bacteria expression system [Fig. 7(A,B)]. Thus, the fly and worm homologs of DGCR8 either do not bind heme or use another residue to serve the same function of Trp329. It is intriguing to note that *C. elegans* does not synthesize its own heme.³¹

The molecular mechanism involving heme-DGCR8 interaction may influence miRNA maturation in several aspects. First, it has been reported that, during development and in tumors, the processing efficiency of pri-miRNAs is decreased.³² However, the mRNA levels of Drosha and DGCR8 are not induced concordantly with the production of mature miRNAs. Post-translational regulation of DGCR8 activity mediated by dimerization and heme binding may explain this apparent discrepancy. Second, recent studies show that the stability of DGCR8 messenger RNA is regulated through cleavage of its pri-miRNA-like hairpin structures by the

Microprocessor complex.^{33–35} Thus, DGCR8-heme association may be part of the homeostatic control of miRNA biogenesis. Finally, signals that modulate DGCR8 activity likely work in combination with other mechanisms to control the expression and activity of miRNAs. For example, several other RNA binding proteins, such as Lin-28, have been shown to regulate the processing of specific miRNAs or miRNA families.^{36,37} Further investigation is needed to completely elucidate the complex regulatory network of miRNA processing. This study provides a structural basis for understanding the regulatory functions of DGCR8 in miRNA maturation in animals.

Materials and Methods

Plasmids

The DGCR8^{276–353} sequence was cloned into pHis-Sumo, a pET-28a vector containing an N-terminal His₆ tag followed by a SUMO moiety.³⁸ The HBD-His₆ expression plasmid contains coding sequences for residues 276–498 of DGCR8 inserted into the pET-24a+ vector between the NdeI and NotI sites. Site-directed mutagenesis was carried out using the standard 4-primer PCR method. The coding sequences of all plasmids were verified by sequencing.

Protein expression, purification, and characterization

Native and SeMet-labeled DGCR8^{276–353} was expressed in *E. coli* and was purified using a Ni affinity chromatography followed by SEC. The His₆-SUMO moiety of the fusion protein was subsequently cleaved off using the Ulp1 protease and was removed using a Ni column. The untagged DGCR8^{276–353} was concentrated and buffer exchanged into 20 mM Tris-HCl (pH 8.0), 80 mM NaCl, and 20 mM dithiothreitol (DTT). HBD-His₆ was expressed in *E. coli* with δ -ALA added to 1 mM at the time of induction. HBD-His₆ was purified by affinity chromatography followed by SEC. The SEC buffer contained 20 mM Tris-HCl (pH 8.0), 400 mM NaCl, and 1.0 mM DTT. The purification was performed in open air, but the chromatography buffers were degassed. The UV-vis spectrum of the purified HBD-His₆ shown in Figure 6(B) was acquired in the SEC buffer without additional treatment to the protein. The expression, purification, and biochemical characterization of the wild-type and mutant NC1 proteins were performed using the procedures described previously.¹⁸ Detailed procedures are described in Supporting Information.

ESI mass spectrometry

A nanoESI source and Au/Pd-coated borosilicate glass capillaries (Proxeon Biosystems, Cambridge, MA) were coupled to a Synapt HD mass spectrom-

eter (Waters, Milford, MA) to acquire positive ion ESI-MS spectra. All protein samples were desalted into a buffer containing 100 mM ammonium acetate (pH 6.8) and 1 mM DTT (10 mM was used for DGCR8^{276–353}), and the samples were concentrated using Amicon centrifugal filter devices (10-kDa MW cutoff).

Crystallization, data collection, structure determination, refinement, and analysis

Crystals of DGCR8^{276–353} were obtained using the hanging drop vapor diffusion method. The protein solution (~6 mg/mL) was mixed at 1:1 ratio (v/v) with the well solution, which contained 0.2M ammonium acetate, 100 mM Bis-Tris pH 5.5, 25% polyethylene glycol 3350, and 10 mM nicotinamide adenine dinucleotide (NAD). Crystals were grown at 18°C in 5 days. They were transferred to a cryoprotection solution containing the same components of the well solution and 20% glycerol for 10 min before flash freezing in liquid nitrogen. X-ray diffraction data were collected at the Advanced Photon Source. Reflections were integrated and scaled using HKL2000.³⁹ SAD phasing was performed using SHELX.⁴⁰ The model was built using ARPWARP⁴¹ and COOT and was refined using REFMAC.⁴² Simulated annealing was performed in PHENIX.⁴³ The structure was validated using the programs PROCHECK (version 3.5.4),⁴⁴ ERRAT,⁴⁵ and VERIFY_3D.⁴⁶ Although addition of NAD to the crystallization solution improved the size and quality of the crystals, no clear electron density was observed for NAD. Solvent-accessible surface areas were calculated using AREAIMOL⁴⁷ with a probe radius of 1.4 Å. Structural superposition were performed using the program MAPS.⁴⁸ Coordinates and structure factor amplitudes have been deposited in the PDB with accession code [3LE4](#).

Acknowledgments

The authors thank the UCLA-DOE X-ray Crystallography Core Facility, which is supported by DOE Grant DE-FC02-02ER63421. They also thank M. Capel, K. Rajashankar, F. Murphy, J. Schuermann, and I. Kourinov at NE-CAT beamline 24-ID-C. NE-CAT is supported by Grant RR-15301 from the NCCR, and APS is supported by DOE under Contract DE-AC02-06CH11357.

References

1. Bartel DP (2009) MicroRNAs: target recognition and regulatory functions. *Cell* 136:215–233.
2. Filipowicz W, Bhattacharyya SN, Sonenberg N (2008) Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat Rev Genet* 9:102–114.

3. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ (2008) miRBase: tools for microRNA genomics. *Nucleic Acids Res* 36:D154–D158.
4. Ambros V (2004) The functions of animal microRNAs. *Nature* 431:350–355.
5. Bushati N, Cohen SM (2007) microRNA functions. *Annu Rev Cell Dev Biol* 23:175–205.
6. Croce CM (2009) Causes and consequences of microRNA dysregulation in cancer. *Nat Rev Genet* 10:704–714.
7. Nelson PT, Wang WX, Rajeev BW (2008) MicroRNAs (miRNAs) in neurodegenerative diseases. *Brain Pathol* 18:130–138.
8. Kim VN, Han J, Siomi MC (2009) Biogenesis of small RNAs in animals. *Nat Rev Mol Cell Biol* 10:126–139.
9. Faller M, Guo F (2008) MicroRNA biogenesis: there's more than one way to skin a cat. *Biochim Biophys Acta* 1779:663–667.
10. Jinek M, Doudna JA (2009) A three-dimensional view of the molecular machinery of RNA interference. *Nature* 457:405–412.
11. Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Radmark O, Kim S, Kim VN (2003) The nuclear RNase III Drosha initiates microRNA processing. *Nature* 425:415–419.
12. Denli AM, Tops BB, Plasterk RH, Ketting RF, Hannon GJ (2004) Processing of primary microRNAs by the Microprocessor complex. *Nature* 432:231–235.
13. Gregory RI, Yan KP, Amuthan G, Chendrimada T, Doratotaj B, Cooch N, Shiekhattar R (2004) The Microprocessor complex mediates the genesis of microRNAs. *Nature* 432:235–240.
14. Han J, Lee Y, Yeom KH, Kim YK, Jin H, Kim VN (2004) The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev* 18:3016–3027.
15. Landthaler M, Yalcin A, 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.
16. Wang Y, Medvid R, Melton C, Jaenisch R, Blelloch R (2007) DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. *Nat Genet* 39:380–385.
17. Han J, Lee Y, Yeom KH, Nam JW, Heo I, Rhee JK, Sohn SY, Cho Y, Zhang BT, Kim VN (2006) Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell* 125:887–901.
18. Faller M, Matsunaga M, Yin S, Loo JA, Guo F (2007) Heme is involved in microRNA processing. *Nat Struct Mol Biol* 14:23–29.
19. Sohn SY, Bae WJ, Kim JJ, Yeom KH, Kim VN, Cho Y (2007) Crystal structure of human DGCR8 core. *Nat Struct Mol Biol* 14:847–853.
20. Kumar MS, Lu J, Mercer KL, Golub TR, Jacks T (2007) Impaired microRNA processing enhances cellular transformation and tumorigenesis. *Nat Genet* 39:673–677.
21. Merritt WM, Lin YG, Han LY, Kamat AA, Spannuth WA, Schmandt R, Urbauer D, Pennacchio LA, Cheng JF, Nick AM, Deavers MT, Mourad-Zeidan A, Wang H, Mueller P, Lenburg ME, Gray JW, Mok S, Birrer MJ, Lopez-Berestein G, Coleman RL, Bar-Eli M, Sood AK (2008) Dicer, Drosha, and outcomes in patients with ovarian cancer. *N Engl J Med* 359:2641–2650.
22. Shiohama A, Sasaki T, Noda S, Minoshima S, Shimizu N (2003) Molecular cloning and expression analysis of a novel gene DGCR8 located in the DiGeorge syndrome chromosomal region. *Biochem Biophys Res Commun* 304:184–190.
23. Stark KL, Xu B, Bagchi A, Lai WS, Liu H, Hsu R, Wan X, Pavlidis P, Mills AA, Karayiorgou M, Gogos JA (2008) Altered brain microRNA biogenesis contributes to phenotypic deficits in a 22q11-deletion mouse model. *Nat Genet* 40:751–760.
24. Yeom KH, Lee Y, Han J, Suh MR, Kim VN (2006) Characterization of DGCR8/Pasha, the essential cofactor for Drosha in primary miRNA processing. *Nucleic Acids Res* 34:4622–4629.
25. Ilsley JL, Sudol M, Winder SJ (2002) The WW domain: linking cell signalling to the membrane cytoskeleton. *Cell Signal* 14:183–189.
26. Yin S, Loo JA (2009) Mass spectrometry detection and characterization of noncovalent protein complexes. *Methods Mol Biol* 492:273–282.
27. Liu Y, Eisenberg D (2002) 3D domain swapping: as domains continue to swap. *Protein Sci* 11:1285–1299.
28. Ohnishi S, Guntert P, Koshiba S, Tomizawa T, Akasaka R, Tochio N, Sato M, Inoue M, Harada T, Watanabe S, Tanaka A, Shirouzu M, Kigawa T, Yokoyama S (2007) Solution structure of an atypical WW domain in a novel beta-clam-like dimeric form. *FEBS Lett* 581:462–468.
29. Huang X, Poy F, Zhang R, Joachimiak A, Sudol M, Eck MJ (2000) Structure of a WW domain containing fragment of dystrophin in complex with beta-dystroglycan. *Nat Struct Biol* 7:634–638.
30. Jager M, Dendle M, Kelly JW (2009) Sequence determinants of thermodynamic stability in a WW domain—an all-beta-sheet protein. *Protein Sci* 18:1806–1813.
31. Rao AU, Carta LK, Lesuisse E, Hamza I (2005) Lack of heme synthesis in a free-living eukaryote. *Proc Natl Acad Sci USA* 102:4270–4275.
32. Thomson JM, Newman M, Parker JS, Morin-Kensicki EM, Wright T, Hammond SM (2006) Extensive post-transcriptional regulation of microRNAs and its implications for cancer. *Genes Dev* 20:2202–2207.
33. Han J, Pedersen JS, Kwon SC, Belair CD, Kim YK, Yeom KH, Yang WY, Haussler D, Blelloch R, Kim VN (2009) Posttranscriptional crossregulation between Drosha and DGCR8. *Cell* 136:75–84.
34. Triboulet R, Chang HM, Lapierre RJ, Gregory RI (2009) Post-transcriptional control of DGCR8 expression by the Microprocessor. *RNA* 15:1005–1011.
35. Shenoy A, Blelloch R (2009) Genomic analysis suggests that mRNA destabilization by the microprocessor is specialized for the auto-regulation of Dgcr8. *PLoS One* 4:e6971.
36. Viswanathan SR, Daley GQ, Gregory RI (2008) Selective blockade of microRNA processing by Lin28. *Science* 320:97–100.
37. Heo I, Joo C, Cho J, Ha M, Han J, Kim VN (2008) Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA. *Mol Cell* 32:276–284.
38. Reverter D, Lima CD (2004) A basis for SUMO protease specificity provided by analysis of human Senp2 and a Senp2-SUMO complex. *Structure* 12:1519–1531.
39. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 276:307–326.
40. Sheldrick GM (2008) A short history of SHELX. *Acta Crystallogr A* 64:112–122.
41. Perrakis A, Morris R, Lamzin VS (1999) Automated protein model building combined with iterative structure refinement. *Nat Struct Biol* 6:458–463.
42. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53:240–255.

43. Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* 58:1948–1954.
44. Laskowski RA, Moss DS, Thornton JM (1993) Main-chain bond lengths and bond angles in protein structures. *J Mol Biol* 231:1049–1067.
45. Colovos C, Yeates TO (1993) Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci* 2:1511–1519.
46. Luthy R, Bowie JU, Eisenberg D (1992) Assessment of protein models with three-dimensional profiles. *Nature* 356:83–85.
47. Collaborative Computational Project N (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50:760–763.
48. Zhang Z, Lindstrom M, Unge J, Peterson C, Lu G (2003) Potential for dramatic improvement in sequence alignment against structures of remote homologous proteins by extracting structural information from multiple structure alignment. *J Mol Biol* 332:127–142.