RESEARCH ARTICLE

# Dynamical comparison between Drosha and Dicer reveals functional motion similarities and dissimilarities

Rotem Aharoni[1], Dror Tobi [1,2]*

**1** Department of Molecular Biology, Ariel University, Ariel, Israel, **2** Department of Computer Sciences, Ariel University, Ariel, Israel

* drorto@ariel.ac.il

## Abstract

Drosha and Dicer are RNase III family members of classes II and III, respectively, which play a major role in the maturation of micro-RNAs. The two proteins share similar domain arrangement and overall fold despite no apparent sequence homology. The overall structural and catalytic reaction similarity of both proteins, on the one hand, and differences in the substrate and its binding mechanisms, on the other, suggest that both proteins also share dynamic similarities and dissimilarities. Since dynamics is essential for protein function, a comparison at their dynamics level is fundamental for a complete understanding of the overall relations between these proteins. In this study, we present a dynamical comparison between human Drosha and Giardia Dicer. Gaussian Network Model and Anisotropic Network Model modes of motion of the proteins are calculated. Dynamical comparison is performed using global and local dynamic programming algorithms for aligning modes of motion. These algorithms were recently developed based on the commonly used Needleman-Wunsch and Smith-Waterman algorithms for global and local sequence alignment. The slowest mode of Drosha is different from that of Dicer due to its more bended posture and allow the motion of the double-stranded RNA-binding domain toward and away from its substrate. Among the five slowest modes dynamics similarity exists only for the second slow mode of motion of Drosha and Dicer. In addition, high local dynamics similarity is observed at the catalytic domains, in the vicinity of the catalytic residues. The results suggest that the proteins exert a similar catalytic mechanism using similar motions, especially at the catalytic sites.

## Introduction

Sequence and structure alignment algorithms are long standing techniques to study proteins. Sequence alignment algorithms [1–3] identify residue conservation, while structure-based classification algorithms like CATH [4], SCOP [5], and DALI [6] provide a good overview of the entire protein structure universe. However, protein structures are dynamic rather than static, and understanding the relation between protein function and dynamics is fundamental

for comprehending the protein structure–dynamics–function relationship. Such an understanding can stem from a comparison of the dynamics of related proteins or the same protein in different states. Traditionally, dynamical comparison of proteins was limited to conserved residues between related proteins [7–9]. In recent years, we [10–13] and others [14–17] have developed several tools and techniques for comparison of protein dynamics that are sequence and structure independent, contributing to the development of the field of comparative dynamics. The dynamical similarity between human Drosha and Giardia Dicer is presented in this study.

Dynamical comparison relies on analysis of low-frequency normal modes from coarse-grained elastic network models (ENM) such as the Gaussian Network Model (GNM) [18] or the Anisotropic Network Model (ANM) [19, 20]. This analysis was proven useful in unraveling the collective modes and, in particular, those at the low frequency end of the mode spectrum that underlie protein equilibrium dynamics [21]. With increasing availability of structural data for well-studied proteins in different forms (liganded, complexed, or free), there is increasing evidence in support of the correspondence between functional changes in structures observed in experiments and the global motions predicted by these coarse-grained analyses [22]. The low frequency modes (usually first 20) are concerned with functional motion and the similarity tend to decrease with higher energy modes. Comparison was made between principal component analysis modes obtained from micro- to milli- second full atomic molecular dynamics (MD) simulations [23, 24] and modes obtained from ANM. Close overlap was found between the principal modes of these two techniques, reinforcing normal mode analysis as a tool for exploring protein dynamics [25].

The Ribonuclease III (RNase III) family is a group of endoribonucleases that specifically degrades double-stranded RNA (dsRNA) at selected target sites, producing typical staggered ends containing 2-nt 3′-overhangs [26]. The cleavage is executed by a conserved catalytic site, which is placed in a unique endonuclease domain fold called a RNase III domain (RIIID) [27–29]. RNase III family members are divided into three classes based on domain organization. Drosha and Dicer are RNase III family members of classes II and III, respectively, which play a major role in the maturation of micro-RNAs (miRNAs) [30, 31]. In the nucleus, Drosha together with its cofactor DiGeorge Syndrome Critical Region 8 (DGCR8) constitute a complex known as Microprocessor. The complex excises a long primary transcript (pri-miRNA) to release hairpin shaped precursor-miRNA (pre-miRNA) of ~70 nucleotides in length. The pre-miRNA is subsequently cleaved at the cytoplasm by Dicer, yielding a miRNA duplex of ~22nt in length, when one strand of this duplex remains as a mature miRNA, while the other strand is degraded [26, 30–34].

Recent studies have found that human Drosha and Giardia Dicer (considered as the 'minimal' Dicer) share a similar domain arrangement and overall fold, besides the C-terminal component, despite no sequence homology [35, 36]. Both proteins have an elongated structure (Fig 1), with two upper RNase III Domains, RIIIDa and RIIIDb, that form a single processing center with two catalytic sites, one on each domain [33, 35, 37]. The RIIIDa is linked to a long α-helix called a Connector, which is encircled by the N-terminal residues that form the Platform domain. The Connector is followed by a bottom globular domain called PAZ for Dicer. The electron density of this region for Drosha was unclear due to structural flexibility, although the strong electron density blobs and predicted secondary structure imply a globular region that may adopt a PAZ-like fold (hence, referred to as the PAZ-like domain) [35]. The lower domains (Connector, Platform, and PAZ/PAZ-like) are collectively referred to as the central domain (CED). In addition to the similar domains, Drosha and Dicer have distinct domains. The two RIIIDs of Dicer are connected by a large helical domain called the Bridging domain, which is missing in Drosha. Another difference is that unlike Dicer, the RIIIDb of
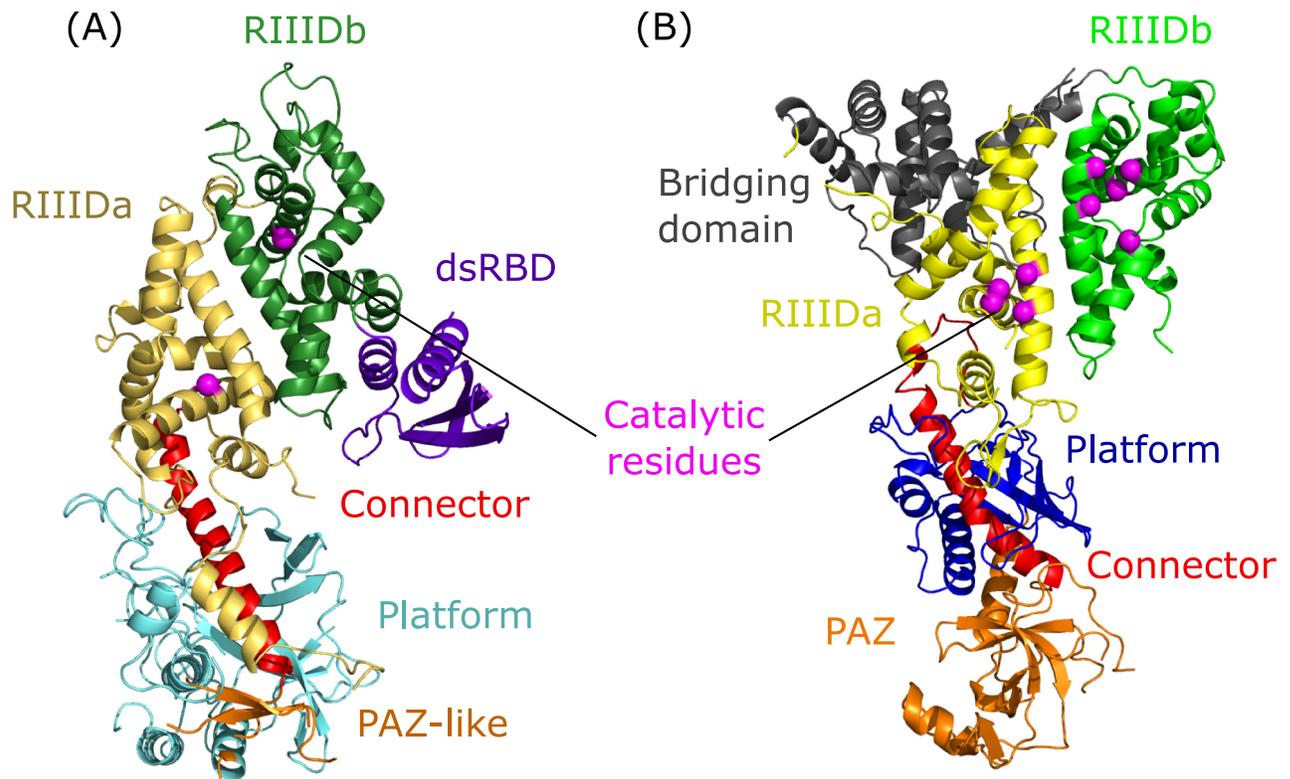
**Fig 1. Structures of human Drosha and Giardia Dicer.** Front view ribbon representation of the structure of (A) Drosha and (B) Dicer. The proteins are colored with similar colors for similar domains. Domains colors are as follow: N-terminal Platform domain (Drosha: cyan, Dicer: dark blue), putative PAZ-like and PAZ domains (Drosha and Dicer: orange), Connector (red), RIIIDa (Drosha: gold, Dicer: yellow), RIIIDb (Drosha: green-forest, Dicer: light green), Dicer Bridging domain (gray), Drosha dsRBD domain (purple). The catalytic residues of both RIIID domains are represented as magenta spheres.

Drosha is bound through a broad and mostly hydrophobic interaction surface to a double-stranded RNA-binding domain (dsRBD). Drosha also has a specific conserved insertion of 67 residues that form two helical segments within its RIIIDa. The first helix (Bump helix) is half buried at a cleft formed by Platform and Connector. The other helix (mobile-basic helix or MB helix) is not shown in the structure, due to missing electron density [35, 36]. In addition, Drosha has at its N-terminal two domains: proline-rich (P-rich) and arginine/serine-rich (R/S-rich), which are dispensable for pri-miRNA processing activity in vitro [38]. Structurally, Drosha has a more bended posture compared with Dicer, which is primarily caused by the former's extensive interaction surface between Connector and RIIIDa that is highly conserved and hydrophobic [35].

Drosha and Dicer act as "rulers" and "scissors" since they measure and cut at a fixed distance of each substrate [26, 36]. Drosha and Dicer catalytic domains (RIIIDs) are highly conserved among the RNase family and share a cleavage mechanism that involves two metal ions [29, 30, 33, 39]. In the single processing center formed by the intramolecularly dimerized RIIIDs, each catalytic site on each of the domains, RIIIDa and RIIIDb, cleaves one strand of an RNA duplex, the 3' and 5' strands, respectively. The Dicer action mechanism is better studied than Drosha's. The 2-nt, 3'-overhangs of a pre-miRNA are recognized and anchored to a specific pocket in the PAZ domain, while the 5' lies adjacent to the specific PAZ domain loop [30, 40]. The pre-

miRNA twist approximately two helical turns towards the catalytic valley of the processing center that is located between the two RIIIDs. The distance between the PAZ and RIIIDs domains, in the case of Dicer, is 65Å and corresponds the length of 25bp, the length of the miRNA duplex. At the processing center, the pre-miRNA is cleaved, liberating a mature miRNA duplex [32, 41–43]. The ligand binding mechanism of Drosha is slightly different than that of Dicer. Drosha binds asymmetrically to the C-terminal tail (CTT) domain of two DGCR8 molecules, one on each RIIID. The two DGCR8 dimerize with the apical stem of pri-miRNA and so augment the substrate binding affinity of Drosha, and enhance Microprocessor fidelity [36, 38, 39, 44, 45]. Drosha interacts with the basal part. A recent study by Kwon et al. [35] presents an RNA-binding model of human Drosha with pri-miRNA. According to this model, the last base pair (bp) before the basal junction of pri-miRNA is adjacent to the Bump helix. The basal junction is where the dsRNA of the lower stem bifurcates to basal segments of single stranded RNA (ssRNA). The basal ssRNA segments of pri-miRNA are located near a positively charged conserved surface on the CED and extends to the PAZ-like domain, rather than the RIIIDs, due to a blockage against the dsRNA stem. The blockage is caused by bending of the Connector and the core of the Platform, toward the dsRNA, and due to the Bump helix with its surrounding loop. The pri-miRNA extends up from the basal junction (adjacent to the Bump helix) about one helical turn of ~28Å [38], which matches the length spanned by ~11bp, toward the catalytic site at RIIIDa. The dsRBD of Drosha interacts transiently with the dsRNA stem and helps to locate the processing centers ~11bp from the basal junction [39]. The pri-miRNA is then cleaved, yielding a 2-nt 3′-overhangs pre-miRNA, which is later processed by Dicer.

The overall structural and enzymatic reaction similarity of both proteins, on the one hand, and differences in the substrate and its binding mechanisms, on the other, suggests that both proteins share dynamics similarity and dissimilarities related to their common and unique functions. In this paper, we examine the dynamics similarity and dissimilarities of human Drosha and Giardia Dicer (hereafter referred to as Drosha and Dicer, respectively) by comparing their GNM and ANM modes of motions, using recently developed global [13] and local [46] mode alignment algorithms. We show that there is an overall low global dynamical similarity between the two proteins, while higher local dynamical similarity co-localizes with the catalytic domains.

## Methods

### Protein structures

Protein data bank (PDB) [47, 48] entries 5b16 [35] and 2ffl [40] of Drosha and Dicer, respectively, were used for dynamical comparison. The structures were superimposed using the Iterative magic fit tool of Swiss PDB viewer (SPDBV) [49] and their ANM modes of motion were calculated, as previously reported [20, 50, 51]. Each residue was represented by a single node positioned at its $C^\alpha$ atom and a cutoff distance of 15Å was used [52]. GNM modes were calculated as previously described [11, 18, 53] using $C^\alpha$ atom cutoff distance of 7.3Å. The first six ANM modes (and one for GNM modes) are so-called trivial modes with zero frequency (eigenvalue) and correspond to rigid-body rotation and translation. Therefore, first (slowest) mode refers to the first mode with non-zero eigenvalue.

The Drosha PDB entry includes two short helices of DGCR8 CTT, one on each RIIID, that are sufficient to interact with and stabilize the protein. In addition, some unknown (UNK) residues of the Platform and PAZ-like domains of Drosha are defined as hetero-atoms (HETATM). Both DGCR8 helices were found to have a negligible effect on the calculated modes of motion. Thus, the presented results were calculated without the two DGCR8 CTTs.

## GNM mode analysis

The motions along different GNM modes are found by eigenvalue decomposition of the connectivity matrix $\Gamma = U \Lambda U^{-1}$ [18]. U is the orthogonal matrix of eigenvectors $u_k$ of $\Gamma$ and $\Lambda$ is the diagonal matrix of the eigenvalues ($\lambda_k$), $1 \leq k \leq N$. The eigenvalues represent the frequencies of the $N$-1 non-zero GNM modes, and are organized in ascending order such that $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_{N-1}$ and $\lambda_N = 0$. The $i$th element $(u_k)_i$ of the $k$th eigenvector describes the fluctuation (deformation) of residue $i$ from its equilibrium position along the $k$th principal coordinate. The MSF of residue $i$ can be rewritten as a weighted sum of the square fluctuations driven by all modes as:

$$\langle \Delta \overrightarrow{R}_i^2 \rangle = \sum_k [\Delta \overrightarrow{R}_i^2]_k = 3k_B T/\gamma \sum_k [\lambda_k^{-1}(\overrightarrow{u}_k)_i^2] \tag{1}$$

GNM enables us to predict the relative *sizes* of motions accessed by different modes, not their *directions, the GNM fluctuations being isotropic by definition*. High fluctuations in absolute values correspond to protein structural regions with high mobility and vice versa. Regions with opposite GNM sign move in anti-correlated manners. Hinge is defined as the region where the fluctuations change sign. The directions of collective motions can be characterized by the Anisotropic Network Model (ANM) [20].

The fluctuations predicted by the GNM are isotropic and hence there is no information on the 'directions' of motion in the calculated modes, just their sizes and sign. In each mode, residues with opposite signs are moving in an anti-correlated manner. High (low) absolute fluctuation values indicate high (low) mobility. Hinge regions are defined as residues where the mode signs change. Regions where no more than three residues were of opposite signs were not considered as hinges.

## ANM mode alignment matrix

The ANM analysis describes motions of the proteins in Cartesian space and hence include information on the direction and the size of the motion. The commonly used Needleman–Wunsch [1] and Smith–Waterman [2] algorithms for global and local sequence alignment were modified to align ANM modes with few modifications. ANM mode analysis results in a set of vectors $\{\overrightarrow{U}\}$ describing the deformation of residues from their equilibrium position (native structure) in Cartesian space. Let $\overrightarrow{U_i^k}$ be the deformation vector of residue $i$ in mode $k$ of one protein and $\overrightarrow{V_j^l}$ the deformation vector of residue $j$ in mode $l$ of another protein. The score for residues $i$ and $j$ used for calculation of the alignment matrix of modes $k$ and $l$ is defined as:

$$S_{ij} = \frac{\overrightarrow{U_i^k} \cdot \overrightarrow{V_j^l}}{|\overrightarrow{U_i^k}||\overrightarrow{V_j^l}|} - C \ where \ 0 \leq C \leq 1 \tag{2}$$

$S_{ij}$ will be positive if their cosine value is greater than C that is the two deformation vectors pointing in the same direction and negative if their cosine is smaller than C. Here we used C = 0.7 radians (~40˚) to define the threshold for vector similarity. In case of alignment of homologous or identical proteins, it is possible to guide the algorithm to prefer the matching of spatially close residues by applying distance constraints. Distance constraints were applied in the present work by modifying the alignment score $S_{ij}$ as follows:

$$S_{ij} = \begin{cases} \dfrac{\overrightarrow{U_i^k} \cdot \overrightarrow{V_j^l}}{|\overrightarrow{U_i^k}||\overrightarrow{V_j^l}|} - C, \ r_{ij} \leq R_c \\ -1 \qquad\qquad , r_{ij} > R_c \end{cases} \tag{3}$$

where $r_{ij}$ is the $C^\alpha$ distance between residues $i$ and $j$ and $R_c$ is the cutoff distance set here to 20Å. Since the sign of the fluctuations in each mode is arbitrary, the alignment of two modes, $a$ and $b$, is done twice: Once between the two original modes ($a$ and $b$) and once between mode $a$ and the negative of the second mode–$b$, with the best alignment (highest score) being used.

## ANM global mode alignment

The algorithm enables alignment of two modes even if their length is different as it creates insertions/deletions along the alignment in order to obtain the optimal alignment. Gap opening and extension penalties were set to 1.0 and 0.1. After the alignment process, in order to obtain a unified alignment score that is not dependent on the alignment length, an average alignment score (AAS) was calculated. The AAS is calculated as the average $S_{ij}$ along the alignment with the exception of gap position and distance constraint regions that received a score of zero. The AAS range from zero (no match) to one (full match). Alignment was performed for the 40 slowest modes. This algorithm is an extension of our previous GNM modes alignment [10] and was reported previously in [13].

## ANM local mode alignment

For each pair from the $n$ slowest aligned modes, an alignment matrix is created and the best non-overlapping (up to 200) gapless matches with minimal length of seven residues (gapless alignment of a single mode pair) are kept. The top $2n$ matches (best scores) are selected and best $S_{ij}$ is kept, for residues $i$ and $j$ of the first and second aligned proteins, for each mode combination [46]. The final residue dynamical similarity score (RDSS) of each residue is the sum of all its best (kept) $S_{ij}$. The RDSS are then normalized to the range [0,99] to fit the B-factor scale of the protein structures. The PDB file B-factor values are replaced by the RDSS and best dynamically matching areas are visualized by the B-factor coloring option of the used viewer, Pymol [54]. The ability of the current algorithm to identify local dynamics similarity is demonstrated in our recent papers, where we present a detailed dynamical comparison between myoglobin and hemoglobin [46] and show its ability to perform dynamics-based clustering [55].

## Results

### Structural similarity of Drosha and Dicer

Human Drosha and Giardia Dicer are not apparent homologous proteins (global and local sequence similarity of 26.0% and 31.9%, respectively). However, recent studies [35, 36] found that both proteins share very similar domain arrangement and overall similar fold, apart from the C-terminal part (Fig 1).

Superimposing of Drosha and Dicer structures with SPDBV is depicted in Fig 2. Only 225 residues, mostly of the top part, out of 722 and 732 of Drosha and Dicer, respectively, were well superposed with RMSD = 1.31 Å. The top parts of the two proteins, containing the two RIIID domains are well superimposed (213 residues out of 421 and 444 of Drosha and Dicer, respectively), including residues at catalytic sites (see also S1 Fig. panel A, Supporting Information). Starting from the C-terminus of the Connector helix down to the PAZ/PAZ-like domain, Drosha has more stooped posture of the RIIID domains relative to the CED as compared with Dicer [35]. The distance between the N-terminal of both Connector domains (S271 of Drosha, I251 of Dicer) is 20.21Å, and the bending angle is ~30˚ as noted by Kwon et al. [35]. Nevertheless, the lower parts (CED) of the two proteins are also superimposable (S1 Fig. panel B), including the Connectors and the kink-inducing prolines (P864 and P266 for Drosha and Dicer, respectively) with high overlap degree (RMSD = 1.66 Å, 53 superposed residues out
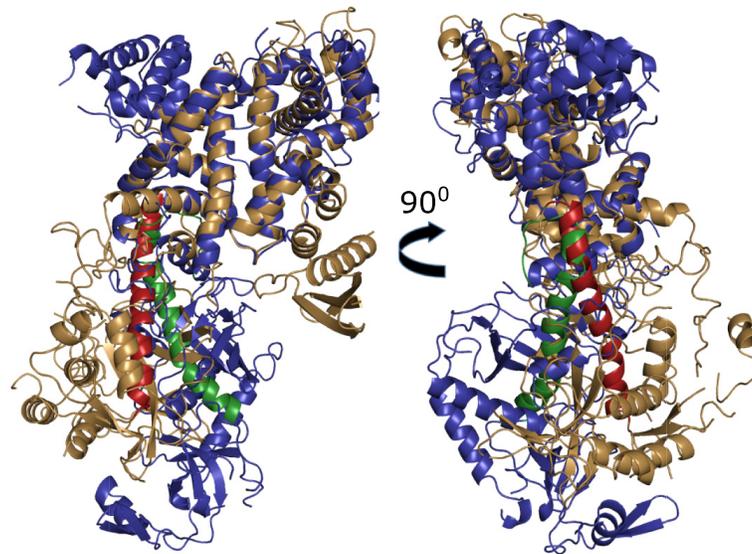
**Fig 2. Superimposition of Drosha and Dicer.** Front (left) and side (right) views of superimposition of Drosha (sand) upon Dicer (deep blue). Connector domain of Drosha is colored firebrick and of Dicer green-forest. Starting at the C-terminus of the Connector helix Drosha has more stooped posture compared with Dicer. The distance between the N-terminal of both Connector domains is 20.21Å, and the bending angle is ~30˚.

of 301 and 288 of Drosha and Dicer, respectively). Drosha's Connector is shorter than Dicer's (30 vs 38 residues), but its Platform is much bigger than Dicer's Platform (293 vs 124 residues). However, both Connector and Platform domains have similar orientation and show the same fold topology of both proteins. These general structural and folding similarities, despite no apparent homology suggest they share dynamical similarities.

## GNM slow modes of Drosha and Dicer

GNM analysis was used to calculate the slowest (global) motions of Drosha and Dicer. The two slowest modes of motions are presented in Fig 3 with the structures on the right colored according to the residue's mode signs. The hinge sites of the first modes are at the crossover between residues C891-L892, T898-H899, L1018-D1019 and C1031-R1032 for Drosha and H328-T329 for Dicer. For both proteins, these hinge residues are located at the bottom of RIIIDa, the C-terminus of the Connector helix, and close to the interface between the Platform and RIIID domains. These findings suggest that the top and the bottom parts of each protein move in an anti-correlated manner. Drosha and Dicer structures are usually divided into two parts: a top part that includes the catalytic domains (RIIIDs) and a bottom part that includes the CED domains. The GNM slowest mode correlates with this division and indicates its dynamical role. The second GNM mode shows two hinge regions for both proteins. The first hinge of Drosha is located slightly below the axis of the hinge region shown by the first mode of GNM, and includes the lower half of the Connector, Platform, and PAZ-like domains (crossover between residues D532-G533, E555-E556, H869-I870 and N905-F906), whereas the second hinge separates the RIIIDs and dsRBD (P1247-R1248). As for Dicer, one hinge is located between the Platform and PAZ domains (L133-M134, L256-I257), and the other hinge resides between the Bridging domain and RIIIDa (A349-R350, L379-V380, I420-Y421, G543-F544, Q572-C573, and L593-A594). It should be noted that in both first GNM modes, residues D929 and T958 of Drosha have opposite signs. However, residues 930–957 are
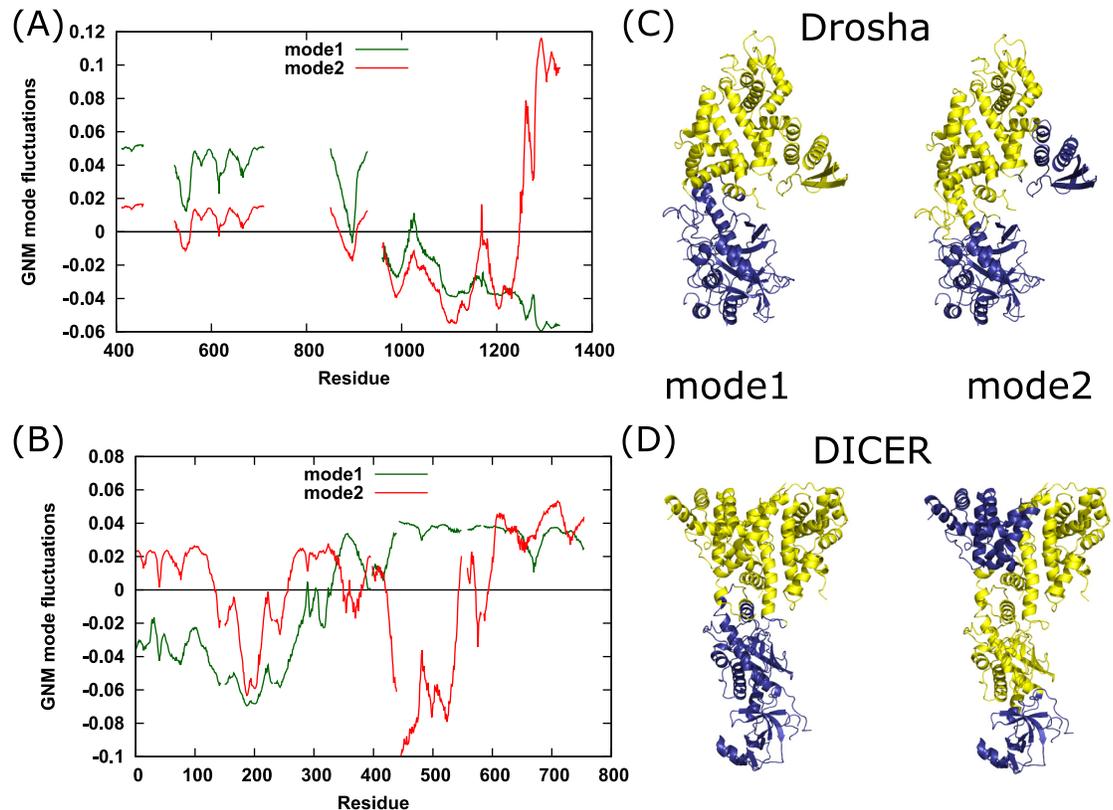
**Fig 3. GNM first and second slow modes.** Fluctuation values of the GNM first (green) and second (red) slowest modes of (A) Drosha and (B) Dicer. Residues with opposite signs are moving in anti-correlated manner and high (low) absolute fluctuation values indicate high (low) mobility. Residues at the crossover between positive and negative fluctuations act as hinges in the movement of the proteins. (C) Drosha and (D) Dicer structures colored according to the positive and negative signs of the first and second GNM slowest modes.

missing and belong to the MB-helix. Thus, the hinge region is located within this segment. Similarly, the PAZ-like domain is absent from the analysis.

## Drosha and Dicer mean square fluctuations profile

The above analysis enabled us to identify slow modes hinge regions. A complementary analysis of mean square fluctuations (MSF) indicates the mobility of each residue as calculated from all modes [18]. The MSF profile curves for Drosha and Dicer and the structures colored according to the residue mobility are depicted in Fig 4 and the catalytic residues are marked as red dots. The average MSF of RIIIDa is lower than that of RIIIDb (0.28 vs. 0.33 for Drosha, and 0.24 vs. 0.28 for Dicer), that is, RIIIDb is more mobile than RIIIDa. The catalytic residues tend to occupy the minima in the MSF curve, and thus they are relatively immobile. Dynamically, Drosha and Dicer exhibit very similar fluctuation profiles. Both catalytic domains have lower average mobility (0.39 for Drosha, and 0.33 for Dicer) relative to the non-catalytic domains (nonCD), while the lowest mobility is observed in the processing center cleft, including the two catalytic sites. Low mobility, though higher than that of catalytic domains, is also shown at the Connectors' C-terminus and Platforms' β-sheets of Dicer, compared with most of the Connector and Platform domains of Drosha. Higher mobility observed for PAZ than PAZ-like domains, probably resulting from the fact that most of the Paz-like residues are missing from the structure due to high thermal fluctuations.
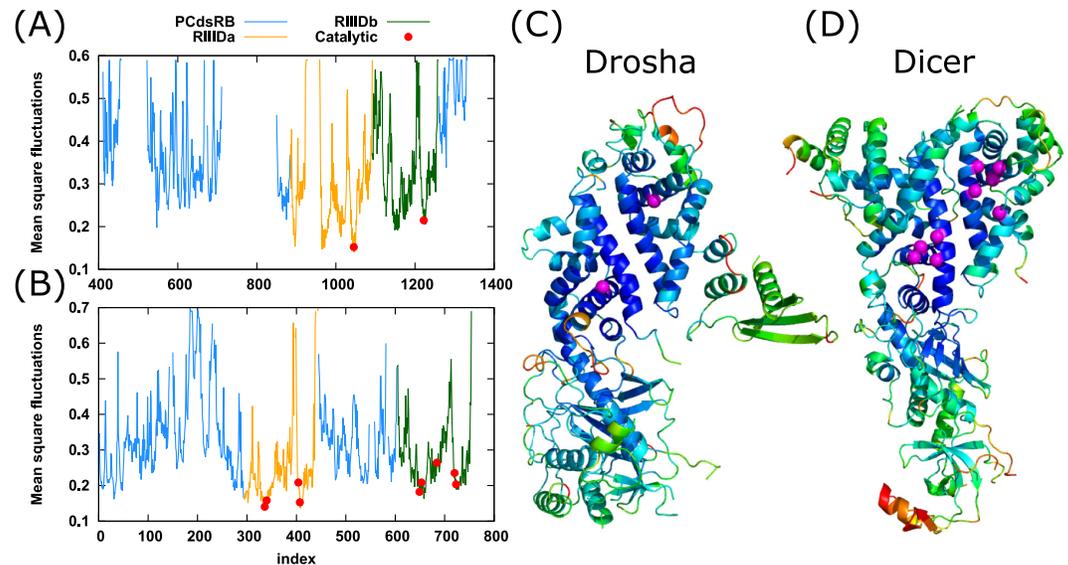
**Fig 4. GNM mean-square fluctuations profile.** Distribution of mean-square fluctuations (ordinate) as a function of residue index (abscissa) for (A) Drosha and (B) Dicer. Peaks represent the most mobile residues and minima the low mobility regions. Fragmented areas in the curve are UNK or missing residues in the PDB files. Curve colors are as follow: non catalytic domains (light blue; PCdsRB; platform, connector, dsRBD, bridging domain), RIIIDa (orange), RIIIDb (green). Catalytic residues are marked by filled red circles. (C) Drosha and (D) Dicer ribbon diagram is colored according to their MSF values. Warm and cold colors (red and blue scales, respectively) indicate most mobile and constrained regions, respectively. PCdsR designate: platform, connector, and dsRBD domains.

https://doi.org/10.1371/journal.pone.0226147.g004

## Drosha and Dicer global mode alignment

The above calculation was performed using modes predicted by GNM, which are isotropic. More detailed dynamical information can be obtained by ANM modes of motion, which are anisotropic. ANM modes of the proteins were calculated and global mode alignment between the forty slowest modes of the two proteins were performed. The AAS matrix for alignments of the five slowest modes is presented in S2 Fig. Only the second mode of Drosha and Dicer show low to moderate similarity, with an AAS score of 0.33 while the rest of the modes are globally dissimilar. Even if global similarities are not observed, the protein's motion for at least the first five slowest modes is characterized by an anti-correlated movement of the upper domains and the CED, suggesting its functional importance. The motion is a semi-circular twisting and differs between the different modes as well as between both proteins in the rotation axis and direction (Fig 5). The change of the movement direction occurs at the hinge region between the RIIIDs and the CED, close to the lower catalytic site, in correlation with the location of the GNM identified hinge regions.

The first two or three slowest modes usually describe global functional motions that are to great biological interests [22]. Therefore, the dynamical difference of the slowest mode between the two proteins is of great importance. The slowest (first) modes of Drosha and Dicer show anti-correlated rotation of the top part against the CED (Fig 5A). However, the rotation axis of the Drosha's top part is located between the RIIIDb and dsRBD, while the rotation center of Dicer's top part is located above RIIIDa. The rotation axis of the CED is similar for Drosha and Dicer and located at the center of their domains. Movies capturing the slowest mode motion of Drosha and Dicer are provided in S1 and S2 Files, respectively. Drosha's slowest mode results in a motion of the dsRBD toward the lower cleavage site at the binding cleft and a counter motion of the lower part of RIIIDa and CED domains. At the same time, RIIIDb
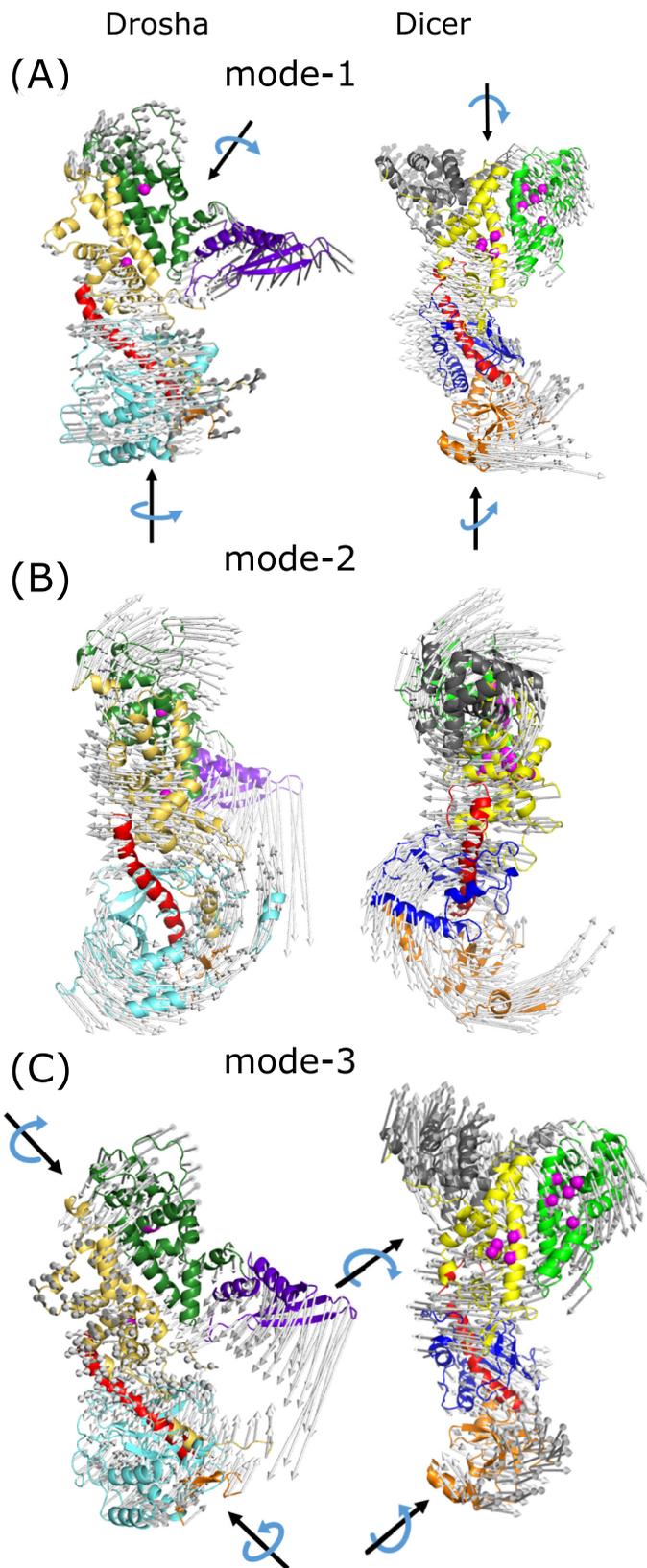
Drosha    Dicer

(A)    mode-1



(B)    mode-2



(C)    mode-3

**Fig 5. Porcupine plots of three slowest modes of Drosha and Dicer.** Drosha (left) and Dicer (right) (A) first, (B) second, (C) and third ANM modes. The motion of the first and third modes is dissimilar for the two proteins, while the second mode shows reasonable similarity. Protein ribbon representation is colored according to the domain colors (see Fig 1 legend) and the arrows displaying the movement direction are colored light gray. Black arrows depict the rotation axis, blue arrows show the motion direction. Upper and lower parts of both proteins are moving in an anti-correlated circular shape. Porcupine plots were created using Modevectors [56].

and the top part of RIIIDa move backwards, away from dsRBD direction. The slowest mode of Dicer results in a sidewise anti-correlated twisting motion of the top part and the CED.

Drosha and Dicer's motion according to their second mode is similar (Fig 5B). Despite existent differences between the motions, the global similarity of the two modes is pronounced. Additional viewpoints of this mode are shown in S2 Fig. panels B-E. The motion mainly describes the semi-circular, front-wise bending motion of the upper domains towards the lower domains (Platform and PAZ/PAZ-like). The hinge axis of both proteins is located between the RIIIDs and the CED, at the C-terminus of the Connector helix. The third mode of Drosha and Dicer is also dissimilar, due to differences in the rotation axis and motion direction of the top part and the CED domains (Fig 5C). Drosha's upper rotation axis is at the interface between the two RIIIDs, at the catalytic cleft, and the lower rotation axis is located at around PAZ-like domain and the Bump helix of RIIIDa. Whereas Dicer upper and lower rotation axes are at the interface between RIIIDa and the Bridging domain, and around the PAZ domain, respectively.

### Drosha and Dicer local mode alignment

Global alignments revealed dynamical similarity between Drosha and Dicer of only one slow mode of each protein, although local similarities may still exist. Local mode alignments of the three, twelve, and forty slowest modes of the two proteins were performed. The higher the mode number, the higher the energetic cost of the protein to move along this mode, and the motions become more local and less global. Fig 6 shows the protein structures colored according to the RDSS. Warm colors (high RDSS) represent high local dynamic similarity (also referred here to as dynamical conservation) and cold colors (low RDSS) represent low dynamic similarity. In all three local mode alignments performed, moderate to high dynamic conservation (high RDSS) is observed at both RIIIDs and the Connector domain. RIIIDb is a little more mobile relative to RIIIDa (as predicted by MSF) and also tends to show higher conservation.

Each of the examined local mode alignments has a slightly different local dynamic similarity profile, although the three alignments show dynamical conservation at the same domains (two RIIIDs and Connector, as mentioned). In the first three modes, very high dynamic similarity is observed at the top part of the RIIIDb domain, while the RIIIDa helix located at the cleft between the two catalytic domains and, in the case of Drosha, the following helix, have very low dynamic similarity. At the twelve modes alignment, the dynamic similarity level of the top part of RIIIDb slightly declined, whereas the dynamic similarity of its lower part as well as the RIIIDa helix, that is in the cleft between the two RIIIDs, increased. Local mode alignment of forty modes shows that the highest RDSS residues concentrate at the areas containing the catalytic residues at each of the RIIIDs, parallel to rising of the dynamic similarity level of RIIIDa. Despite the few differences in the dynamic conservation profile of the three alignments, all three alignments show higher conservation in the vicinity of the catalytic residues for each catalytic domain at the processing center.

### Discussion

Drosha and Dicer are classes II and III members, respectively, of the RNase III protein family that play a key role in miRNA biogenesis [30, 31]. Functionally, both proteins act as a "ruler"
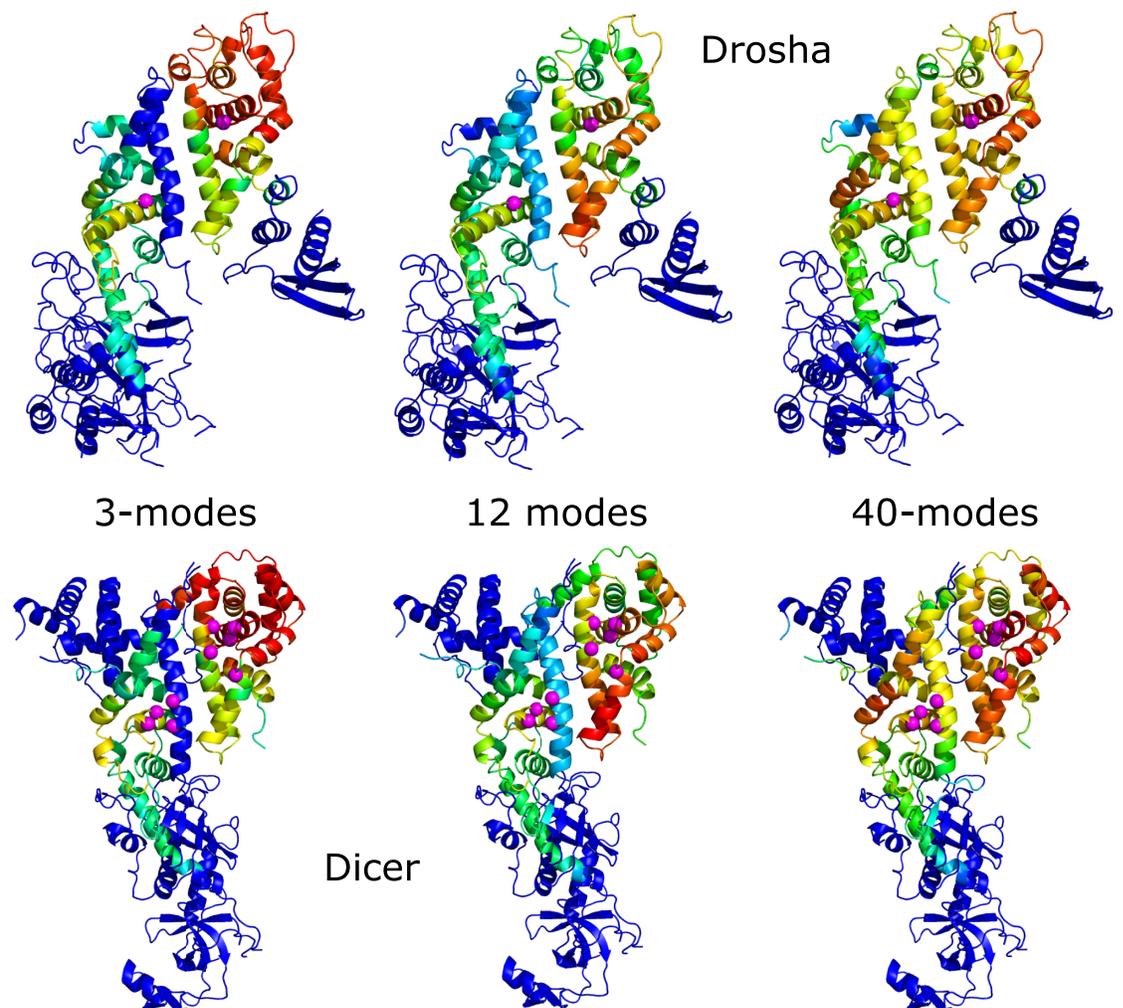
**Fig 6. Local ANM modes alignment of Drosha and Dicer.** Protein structures of Drosha (top) and Dicer (bottom) colored according to the RDSS calculated from local alignment of the three (left), twelve (middle), and forty (right) ANM slowest modes. Catalytic residues are represented as magenta spheres. Warm colors (high RDSS) represent high local dynamic similarity and cold colors (low RDSS) represent low dynamic similarity.

and "scissors" as they measure and cut at a fixed distance of each substrate [26, 36]. The dsRNA substrate cleavage is carried out at the single processing center cleft containing two catalytic sites, generated by two intramolecularly dimerized strictly conserved catalytic domains (RIIIDa, RIIIDb). Previous studies report high similarity of domain arrangement and overall folding, despite few clear differences [35, 36]. The high structural resemblance of both proteins together with their similar function suggest they also share related dynamical similarities. However, since there are also differences between the proteins' substrate and its binding mechanism, the dynamical comparison is expected to show dissimilarities related to unique functions. Therefore, comparison between Drosha and Dicer at their dynamics level is necessary to deepen the understanding of the relations between these proteins. To test this hypothesis, we analyzed the overall motion of each protein and dynamical comparison was performed using

three methods: (i) GNM analysis, (ii) Global ANM modes alignment, and (iii) local ANM modes alignment. The last two methodologies were recently developed by us based on the commonly used sequence comparison algorithms, Needleman-Wunsch [1] and Smith-Waterman [2].

The GNM analysis shows that the least mobile region for both proteins is the substrate binding cleft (Fig 4). In addition, the catalytic residues are relatively immobile and tend to occupy the minima in the GNM MSF profile. These results are consistent with a previous study by Yang et al. [57] who showed that most of the catalytic residues in enzymes tend to have significantly smaller mobility compared to the average residues, probably to protect the delicate arrangement and orientation of these functional groups. A study by Macrae et al. [41] reported two flexible hinge regions at the Dicer structure. The first hinge is located between the Platform domain and RIIID domains. The induced positional shift is diffused along the Platform loop and the loop following the Connector helix, leading to bending of the RIIID domains towards the Platform, in addition to 5° rotation of the RIIID region around the long axis of the protein. The second hinge resides between the Platform and PAZ domains and is traced to strictly conserved Pro266. This flexible hinge induces a kink at the Connector helix, resulting in a shift of 5Å of the PAZ domain, and aids in directing the helix towards RIIIDa [40]. This kink-inducing proline is also found to be conserved in Drosha homologs (Pro864 in Drosha) [35]. The hinge found at the first GNM mode together with the lower hinge received at the second GNM mode for Dicer are consistent with those reported by Macrae et al. [40, 41]. The results also show a resemblance of the hinge locations of Drosha and Dicer, as well as the anti-correlated motion of the upper and lower parts.

GNM modes comparison can be viewed as a low-resolution dynamical comparison, since these modes provide information on the size of the fluctuations but not their directions. ANM modes provide both the direction and size of the motion, hence, can be viewed as a high-resolution dynamical comparison. Furthermore, each mode vectors $\{\vec{U}\}$ (see Methods) describing the deformation of residues from their equilibrium position of each mode are independent of each other. Therefore, both global and local alignment algorithms can be used to compare these modes. The results show partial dynamical similarity between Drosha and Dicer. Global mode alignments show a moderate similarity only of the second mode of Drosha and Dicer among the 40 slowest modes. Despite the existing similarities between Drosha and Dicer, there are significant differences between them, due to their distinct domains of each protein and the bent posture of Drosha. In addition, the substrates properties and their binding nature are likely to necessitate unique global dynamics for cleavage. The typical pri-miRNA cleaved by Drosha contains a dsRNA stem of ~11bp that is bound to the top part and flanked by single-stranded basal segments that are apparently bound to the CED domain [35]. However, the substrate of Dicer is a ~25bp dsRNA bound to the top part and the PAZ domain. Moreover, the substrate binding and cleavage mechanism of Drosha and Dicer are slightly different, with Drosha functioning as a part of the microprocessor complex [33, 39] while Dicer acts as an independent unit. Taking all the above together explains the dissimilar dynamics at the global level. Plausible explanation for difference in the global (first mode) dynamics of the two proteins is that since Dicer's substrate is more rigid, bending motion of the first and second modes will result in separation of the cleaved dsRNA. However, since the ssRNA bound to Drosha's CED domain are more flexible the slowest mode of Drosha evolve a stretching away movement of the CED domain with inward movement of the dsRBD as a more efficient motion that can separate the cleaved RNA products.

Although no significant global similarity was found, both Drosha and Dicer exhibit similar global motion trend, of bending the upper and lower domains toward each other, while

moving in a semi-circular manner. Since this kind of movement is observed in the different modes of the two proteins it is probably of a functional significance of assisting the cleavage of each substrate. Rotating the substrate can impair its stability by generating high torsional stress or by stretching chemical bonds. Bending of the upper and lower domains towards each other together with the substrate can physically assist the cleavage mechanism.

As members of RNase III family, Drosha and Dicer catalytic domains are highly conserved and function with the same mechanism of dsRNA cleavage, although they have different substrates (pri- and pre- miRNA). Therefore, it is expected that the two proteins will have similar but not identical local dynamics. Local mode alignments show higher dynamic conservation at both RIIIDs, in the vicinity of the catalytic residues and at the Connector domain. The high dynamical conservation of the processing center indicates that Drosha and Dicer use similar local dynamics for the catalytic mechanism.

In this study, we present relatively new methodologies for dynamical comparison of proteins and exemplified the use of comparative dynamics as a new layer to study proteins by comparison. Sequence, structure, and dynamical comparison algorithms are three complementary layers to study proteins, and together give a more complete and accurate view of the relation between related proteins.

## Supporting information

**S1 Fig. Superimposition of Drosha and Dicer.**
(PDF)

**S2 Fig. ANM global mode alignment.**
(PDF)

**S1 File. Drosha's slowest mode of motion movie.**
(MP4)

**S2 File. Dicer's slowest mode of motion movie.**
(MP4)

## Author Contributions

**Conceptualization:** Dror Tobi.

**Formal analysis:** Rotem Aharoni.

**Methodology:** Dror Tobi.

**Supervision:** Dror Tobi.

**Writing – original draft:** Rotem Aharoni, Dror Tobi.

## References

1. Needleman SB, Wunsch CD. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol. 1970; 48(3):443–53. Epub 1970/03/01. 0022-2836(70) 90057-4 [pii]. https://doi.org/10.1016/0022-2836(70)90057-4 PMID: 5420325.

2. Smith TF, Waterman MS. Identification of common molecular subsequences. J Mol Biol. 1981; 147 (1):195–7. https://doi.org/10.1016/0022-2836(81)90087-5 PMID: 7265238.

3. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994; 22(22):4673–80. https://doi.org/10.1093/nar/22.22.4673 PMID: 7984417; PubMed Central PMCID: PMC308517.

4. Pearl FM, Bennett CF, Bray JE, Harrison AP, Martin N, Shepherd A, et al. The CATH database: an extended protein family resource for structural and functional genomics. Nucleic Acids Res. 2003; 31 (1):452–5. Epub 2003/01/10. https://doi.org/10.1093/nar/gkg062 PMID: 12520050; PubMed Central PMCID: PMC165509.

5. Andreeva A, Howorth D, Brenner SE, Hubbard TJ, Chothia C, Murzin AG. SCOP database in 2004: refinements integrate structure and sequence family data. Nucleic Acids Res. 2004; 32(Database issue):D226–9. Epub 2003/12/19. https://doi.org/10.1093/nar/gkh039 32/suppl_1/D226 [pii]. PMID: 14681400; PubMed Central PMCID: PMC308773.

6. Holm L, Rosenstrom P. Dali server: conservation mapping in 3D. Nucleic Acids Res. 2010; 38(Web Server issue):W545–9. Epub 2010/05/12. gkq366 [pii] https://doi.org/10.1093/nar/gkq366 PMID: 20457744; PubMed Central PMCID: PMC2896194.

7. Bakan A, Meireles LM, Bahar I. ProDy: protein dynamics inferred from theory and experiments. Bioinformatics. 2011; 27(11):1575–7. Epub 2011/04/08. btr168 [pii] https://doi.org/10.1093/bioinformatics/btr168 PMID: 21471012; PubMed Central PMCID: PMC3102222.

8. Maguid S, Fernandez-Alberti S, Echave J. Evolutionary conservation of protein vibrational dynamics. Gene. 2008; 422(1–2):7–13. Epub 2008/06/26. S0378-1119(08)00228-X [pii] https://doi.org/10.1016/j.gene.2008.06.002 PMID: 18577430.

9. Maguid S, Fernandez-Alberti S, Ferrelli L, Echave J. Exploring the common dynamics of homologous proteins. Application to the globin family. Biophys J. 2005; 89(1):3–13. Epub 2005/03/08. S0006-3495 (05)72652-7 [pii] https://doi.org/10.1529/biophysj.104.053041 PMID: 15749782; PubMed Central PMCID: PMC1366528.

10. Tobi D. Dynamics alignment: comparison of protein dynamics in the SCOP database. Proteins. 2012; 80(4):1167–76. Epub 2012/01/26. https://doi.org/10.1002/prot.24017 PMID: 22275069.

11. Tobi D. Large-scale analysis of the dynamics of enzymes. Proteins. 2013; 81:1910–8. Epub 2013/06/06. https://doi.org/10.1002/prot.24335 PMID: 23737241.

12. Tobi D. Normal Mode Dynamics Comparison of Proteins. Isr J Chem. 2014; https://doi.org/10.1002/ijch.201300142

13. Tobi D. Dynamical differences of hemoglobin and the ionotropic glutamate receptor in different states revealed by a new dynamics alignment method. Proteins. 2017; 85(8):1507–17. https://doi.org/10.1002/prot.25311 PMID: 28459140.

14. Munz M, Lyngso R, Hein J, Biggin PC. Dynamics based alignment of proteins: an alternative approach to quantify dynamic similarity. BMC Bioinformatics. 2010; 11:188. Epub 2010/04/20. 1471-2105-11-188 [pii] https://doi.org/10.1186/1471-2105-11-188 PMID: 20398246; PubMed Central PMCID: PMC2868010.

15. Micheletti C. Comparing proteins by their internal dynamics: exploring structure-function relationships beyond static structural alignments. Phys Life Rev. 2013; 10(1):1–26. Epub 2012/12/04. S1571-0645 (12)00132-7 [pii] https://doi.org/10.1016/j.plrev.2012.10.009 PMID: 23199577.

16. Fuglebakk E, Tiwari SP, Reuter N. Comparing the intrinsic dynamics of multiple protein structures using elastic network models. Biochim Biophys Acta. 2015; 1850(5):911–22. https://doi.org/10.1016/j.bbagen.2014.09.021 PMID: 25267310.

17. Tiwari SP, Reuter N. Conservation of intrinsic dynamics in proteins-what have computational models taught us? Curr Opin Struct Biol. 2017; 50:75–81. https://doi.org/10.1016/j.sbi.2017.12.001 PMID: 29287233.

18. Bahar I, Atilgan AR, Erman B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. Fold Des. 1997; 2(3):173–81. Epub 1997/01/01. https://doi.org/10.1016/S1359-0278(97)00024-2 PMID: 9218955.

19. Eyal E, Chennubhotla C, Yang LW, Bahar I. Anisotropic fluctuations of amino acids in protein structures: insights from X-ray crystallography and elastic network models. Bioinformatics. 2007; 23(13):i175–84. Epub 2007/07/25. 23/13/i175 [pii] https://doi.org/10.1093/bioinformatics/btm186 PMID: 17646294.

20. Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. Biophys J. 2001; 80(1):505–15. Epub 2001/02/13. S0006-3495(01)76033-X [pii] https://doi.org/10.1016/S0006-3495(01)76033-X PMID: 11159421; PubMed Central PMCID: PMC1301252.

21. Kitao A, Go N. Investigating protein dynamics in collective coordinate space. Curr Opin Struct Biol. 1999; 9(2):164–9. Epub 1999/05/14. S0959-440X(99)80023-2 [pii] https://doi.org/10.1016/S0959-440X(99)80023-2 PMID: 10322205.

22. Bahar I, Lezon TR, Yang LW, Eyal E. Global dynamics of proteins: bridging between structure and function. Annu Rev Biophys. 2010; 39:23–42. Epub 2010/03/03. https://doi.org/10.1146/annurev.biophys.093008.131258 PMID: 20192781; PubMed Central PMCID: PMC2938190.

23. Klepeis JL, Lindorff-Larsen K, Dror RO, Shaw DE. Long-timescale molecular dynamics simulations of protein structure and function. Curr Opin Struct Biol. 2009; 19(2):120–7. Epub 2009/04/14. S0959-440X (09)00037-2 [pii] https://doi.org/10.1016/j.sbi.2009.03.004 PMID: 19361980.

24. Shaw DE, Maragakis P, Lindorff-Larsen K, Piana S, Dror RO, Eastwood MP, et al. Atomic-level characterization of the structural dynamics of proteins. Science. 2010; 330(6002):341–6. Epub 2010/10/16. 330/6002/341 [pii] https://doi.org/10.1126/science.1187409 PMID: 20947758.

25. Gur M, Zomot E, Bahar I. Global motions exhibited by proteins in micro- to milliseconds simulations concur with anisotropic network model predictions. J Chem Phys. 2013; 139(12):121912. Epub 2013/10/05. https://doi.org/10.1063/1.4816375 PMID: 24089724; PubMed Central PMCID: PMC3739829.

26. MacRae IJ, Doudna JA. Ribonuclease revisited: structural insights into ribonuclease III family enzymes. Curr Opin Struct Biol. 2007; 17(1):138–45. https://doi.org/10.1016/j.sbi.2006.12.002 PMID: 17194582.

27. Conrad C, Rauhut R. Ribonuclease III: new sense from nuisance. Int J Biochem Cell Biol. 2002; 34 (2):116–29. https://doi.org/10.1016/s1357-2725(01)00112-1 PMID: 11809414.

28. Filippov V, Solovyev V, Filippova M, Gill SS. A novel type of RNase III family proteins in eukaryotes. Gene. 2000; 245(1):213–21. https://doi.org/10.1016/s0378-1119(99)00571-5 PMID: 10713462.

29. Nicholson AW. Ribonuclease III mechanisms of double-stranded RNA cleavage. Wiley Interdiscip Rev RNA. 2014; 5(1):31–48. https://doi.org/10.1002/wrna.1195 PMID: 24124076; PubMed Central PMCID: PMC3867540.

30. Gan J, Shaw G, Tropea JE, Waugh DS, Court DL, Ji X. A stepwise model for double-stranded RNA processing by ribonuclease III. Mol Microbiol. 2008; 67(1):143–54. https://doi.org/10.1111/j.1365-2958.2007.06032.x PMID: 18047582.

31. Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, et al. The nuclear RNase III Drosha initiates microRNA processing. Nature. 2003; 425(6956):415–9. https://doi.org/10.1038/nature01957 PMID: 14508493.

32. Brodersen P, Voinnet O. The diversity of RNA silencing pathways in plants. Trends Genet. 2006; 22 (5):268–80. https://doi.org/10.1016/j.tig.2006.03.003 PMID: 16567016.

33. Han J, Lee Y, Yeom KH, Kim YK, Jin H, Kim VN. The Drosha-DGCR8 complex in primary microRNA processing. Genes Dev. 2004; 18(24):3016–27. https://doi.org/10.1101/gad.1262504 PMID: 15574589; PubMed Central PMCID: PMC535913.

34. Lee Y, Kim M, Han J, Yeom KH, Lee S, Baek SH, et al. MicroRNA genes are transcribed by RNA polymerase II. EMBO J. 2004; 23(20):4051–60. https://doi.org/10.1038/sj.emboj.7600385 PMID: 15372072; PubMed Central PMCID: PMC524334.

35. Kwon SC, Nguyen TA, Choi YG, Jo MH, Hohng S, Kim VN, et al. Structure of Human DROSHA. Cell. 2016; 164(1–2):81–90. https://doi.org/10.1016/j.cell.2015.12.019 PMID: 26748718.

36. Li S, Patel DJ. Drosha and Dicer: Slicers cut from the same cloth. Cell Res. 2016; 26(5):511–2. https://doi.org/10.1038/cr.2016.19 PMID: 27125999; PubMed Central PMCID: PMC4856758.

37. Zhang H, Kolb FA, Jaskiewicz L, Westhof E, Filipowicz W. Single processing center models for human Dicer and bacterial RNase III. Cell. 2004; 118(1):57–68. https://doi.org/10.1016/j.cell.2004.06.017 PMID: 15242644.

38. Nguyen TA, Jo MH, Choi YG, Park J, Kwon SC, Hohng S, et al. Functional Anatomy of the Human Microprocessor. Cell. 2015; 161(6):1374–87. https://doi.org/10.1016/j.cell.2015.05.010 PMID: 26027739.

39. Han J, Lee Y, Yeom KH, Nam JW, Heo I, Rhee JK, et al. Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. Cell. 2006; 125(5):887–901. https://doi.org/10.1016/j.cell.2006.03.043 PMID: 16751099.

40. Macrae IJ, Zhou K, Li F, Repic A, Brooks AN, Cande WZ, et al. Structural basis for double-stranded RNA processing by Dicer. Science. 2006; 311(5758):195–8. https://doi.org/10.1126/science.1121638 PMID: 16410517.

41. Macrae IJ, Li F, Zhou K, Cande WZ, Doudna JA. Structure of Dicer and mechanistic implications for RNAi. Cold Spring Harb Symp Quant Biol. 2006; 71:73–80. https://doi.org/10.1101/sqb.2006.71.042 PMID: 17381283.

42. Starega-Roslan J, Galka-Marciniak P, Krzyzosiak WJ. Nucleotide sequence of miRNA precursor contributes to cleavage site selection by Dicer. Nucleic Acids Res. 2015; 43(22):10939–51. https://doi.org/10.1093/nar/gkv968 PMID: 26424848; PubMed Central PMCID: PMC4678860.

43. Zamore PD. Thirty-three years later, a glimpse at the ribonuclease III active site. Mol Cell. 2001; 8 (6):1158–60. https://doi.org/10.1016/s1097-2765(01)00418-x PMID: 11885596.

44. Denli AM, Tops BB, Plasterk RH, Ketting RF, Hannon GJ. Processing of primary microRNAs by the Microprocessor complex. Nature. 2004; 432(7014):231–5. https://doi.org/10.1038/nature03049 PMID: 15531879.

**45.** Gregory RI, Yan KP, Amuthan G, Chendrimada T, Doratotaj B, Cooch N, et al. The Microprocessor complex mediates the genesis of microRNAs. Nature. 2004; 432(7014):235–40. https://doi.org/10.1038/nature03120 PMID: 15531877.

**46.** Aharoni R, Tobi D. Dynamical comparison between myoglobin and hemoglobin. Proteins. 2018; 86 (11):1176–83. https://doi.org/10.1002/prot.25598 PMID: 30183107.

**47.** Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic Acids Res. 2000; 28(1):235–42. Epub 1999/12/11. gkd090 [pii]. https://doi.org/10.1093/nar/28.1.235 PMID: 10592235; PubMed Central PMCID: PMC102472.

**48.** Burley SK, Berman HM, Bhikadiya C, Bi C, Chen L, Di Costanzo L, et al. RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. Nucleic Acids Res. 2019; 47(D1):D464–D74. https://doi.org/10.1093/nar/gky1004 PMID: 30357411; PubMed Central PMCID: PMC6324064.

**49.** Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis. 1997; 18(15):2714–23. https://doi.org/10.1002/elps.1150181505 PMID: 9504803.

**50.** Doruker P, Atilgan AR, Bahar I. Dynamics of proteins predicted by molecular dynamics simulations and analytical approaches: application to alpha-amylase inhibitor. Proteins. 2000; 40(3):512–24. Epub 2000/06/22. https://doi.org/10.1002/1097-0134(20000815)40:3<512::AID-PROT180>3.0.CO;2-M [pii]. PMID: 10861943.

**51.** Hensen U, Meyer T, Haas J, Rex R, Vriend G, Grubmuller H. Exploring protein dynamics space: the dynasome as the missing link between protein structure and function. PLoS One. 2012; 7(5):e33931. Epub 2012/05/19. https://doi.org/10.1371/journal.pone.0033931 PONE-D-11-18707 [pii]. PMID: 22606222; PubMed Central PMCID: PMC3350514.

**52.** Eyal E, Lum G, Bahar I. The anisotropic network model web server at 2015 (ANM 2.0). Bioinformatics. 2015; 31(9):1487–9. https://doi.org/10.1093/bioinformatics/btu847 PMID: 25568280; PubMed Central PMCID: PMC4410662.

**53.** Li H, Chang YY, Yang LW, Bahar I. iGNM 2.0: the Gaussian network model database for biomolecular structural dynamics. Nucleic Acids Res. 2016; 44(D1):D415–22. https://doi.org/10.1093/nar/gkv1236 PMID: 26582920; PubMed Central PMCID: PMC4702874.

**54.** The PyMOL Molecular Graphics System, Version 1.8. Schrödinger, LLC.

**55.** Tobi D. Dynamics based clustering of globin family members. PLoS One. 2018; 13(12):e0208465. https://doi.org/10.1371/journal.pone.0208465 PMID: 30513111; PubMed Central PMCID: PMC6279032.

**56.** Law SM. https://pymolwiki.org/index.php/Modevectors. Available from: https://pymolwiki.org/index.php/Modevectors.

**57.** Yang LW, Bahar I. Coupling between catalytic site and collective dynamics: a requirement for mechanochemical activity of enzymes. Structure. 2005; 13(6):893–904. Epub 2005/06/09. S0969-2126(05)00167-X [pii] https://doi.org/10.1016/j.str.2005.03.015 PMID: 15939021; PubMed Central PMCID: PMC1489920.