Structure
Ways & Means

DOLORS: Versatile Strategy for Internal Labeling and Domain Localization in Electron Microscopy

Pick-Wei Lau,[1,2] Clinton S. Potter,[1,3] Bridget Carragher,[1,3,*] and Ian J. MacRae[2,*]
[1]National Resource for Automated Molecular Microscopy
[2]Department of Molecular Biology
[3]Department of Cell Biology
The Scripps Research Institute, La Jolla, CA 92037, USA
*Correspondence: bcarr@scripps.edu (B.C.), macrae@scripps.edu (I.J.M.)
http://dx.doi.org/10.1016/j.str.2012.10.019

SUMMARY

Single-particle electron microscopy (EM) is a powerful tool for studying the structures of large biological molecules. However, the achievable resolution does not always allow for direct recognition of individual protein domains. Labels that can be visualized by EM have been developed for protein termini, but tagging internal domains remains a challenge. We describe a robust strategy for determining the position of internal sites within EM maps, termed domain localization by RCT sampling (DOLORS). DOLORS uses monovalent streptavidin added posttranslationally to tagged sites in the target protein. Internal labels generally display less conformational flexibility than terminal labels, providing more precise positional information. Automated methods are used to rapidly generate assemblies of unique 3D models allowing the attachment sites of labeled domains to be accurately identified and thus provide an overall architectural map of the molecule.

INTRODUCTION

Single-particle electron microscopy (EM) has been established as an important tool in macromolecular structure determination. Progress in recent years has allowed the determination of subnanometer (Wiedenheft et al., 2011) and even atomic resolution EM structures (Zhang et al., 2008). However, many single-particle EM structures cannot be reconstructed to a resolution that allows unambiguous identification of secondary structures or discrete domains (Lau et al., 2009), particularly in the case of small, flexible macromolecules with low symmetry. Nevertheless, even at moderate resolution, it is still possible to gain mechanistic insights into the structure if functional components or domains within the macromolecule can be located within the EM map (Chittuluru et al., 2011; Lees et al., 2010; Shanks et al., 2010; Yip et al., 2010).

Tagging discrete domains or subunits of a complex with a recognizable label can provide their location within EM maps. Antibodies can be useful labeling tools because they are easily recognized by EM (Chittuluru et al., 2011), but monoclonal antibodies are costly and time consuming to develop and must have a high enough affinity to remain associated with the target particle under the dilute conditions used for EM. To overcome these issues, a variety of fusion protein-based tags have been developed, in which the protein of interest is expressed in frame with a recognizable protein tag. Successful examples include green fluorescent protein (GFP) (Choy et al., 2009), actin polymer (Stroupe et al., 2009), and the dynein light-chain-interacting domain (DID) (Flemming et al., 2010). Protein fusion tags are typically appended to either the N or C terminus of the studied protein, allowing the identification of the beginning or end of the polypeptide chain in the EM map. This approach is best suited for macromolecular complexes composed of multiple small polypeptide subunits. In such cases, identifying the termini of individual subunits within an EM reconstruction may provide sufficient positional information to draw meaningful mechanistic insights. In contrast, for a large protein composed of many domains within a single polypeptide chain, knowing only the positions of the N and C termini may be insufficient to establish the functional architecture of the particle. Furthermore, most methods for labeling termini are not well suited for labeling internal sites because the labels are necessarily large globular domains (e.g., GFP) and need to fold independently of the protein in which they are inserted without disrupting the overall tertiary/quaternary structure of the subject particle.

An additional limitation of most labeling methods is that terminal labels tend to be flexible (Lees et al., 2010; Shanks et al., 2010; Yip et al., 2010), and thus, the attachment site can be difficult to precisely locate by EM. Flexibility stems from unstructured elements commonly located at protein termini and also because the label is tethered to only one point in the subject protein. Terminal labels thus often experience a large degree of conformational freedom and adopt a variety of positions relative to the subject particle, which consequently limits accuracy in locating the position of the label relative to the particle of interest. Attempts to visualize flexible labels using EM typically result in 2D class averages displaying a blur of densities corresponding to the label, and deriving 3D maps from these structures is challenging (Chittuluru et al., 2011; Lees et al., 2010; Shanks et al., 2010; Yip et al., 2010).

To overcome these limitations, we have devised a strategy termed domain localization by RCT sampling (DOLORS) for labeling and localizing internal domains and positions within

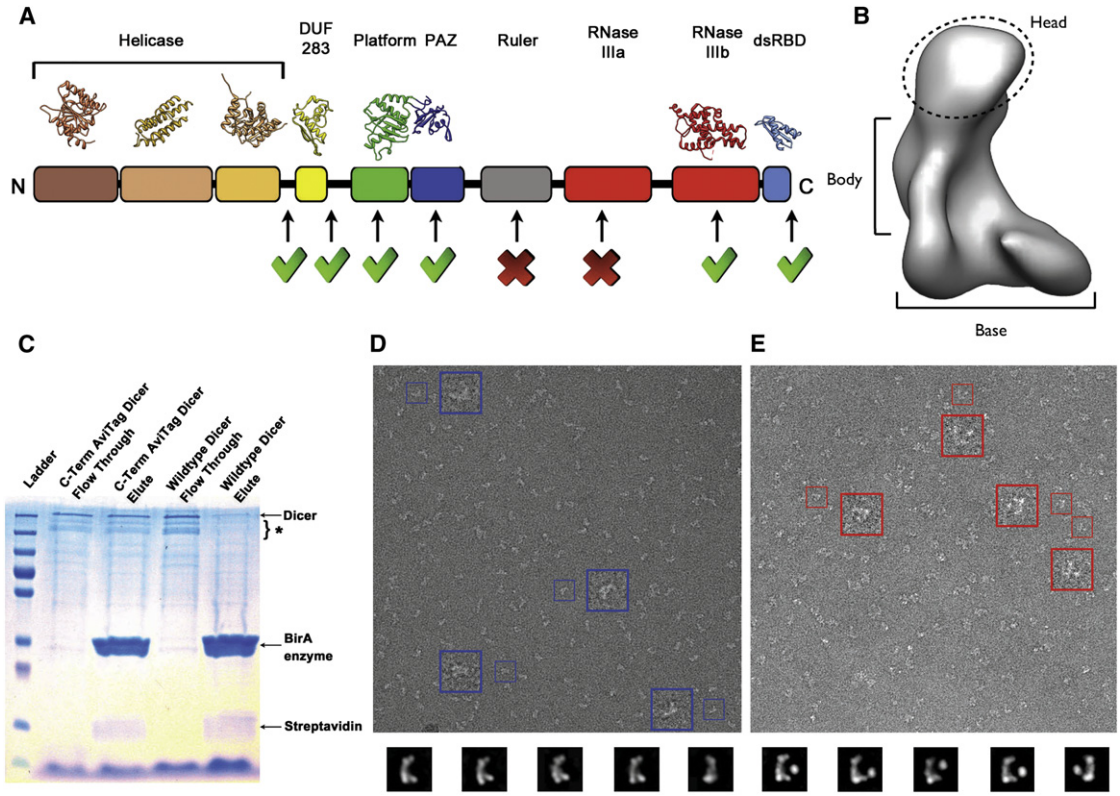

Figure 1. Labeling and Visualizing Dicer with Streptavidin Tag

(A) Schematic showing the functional domains of human Dicer. Crystal structures representative of most functional domains are known. Green checks indicate successful localization of the indicated positions using DOLORS. Red crosses indicate failed attempts. N, N terminal; C, C terminal.

(B) A representative view of the full-length Dicer EM map with major features labeled.

(C) C-terminal (C-term) AviTag Dicer and control wild-type Dicer were incubated with BirA in a biotinylation reaction. The streptavidin-labeled Dicer coeluted with His_6-streptavidin from a HisTrap column. The BirA enzyme, containing a His_6 tag, coeluted from the column. The asterisk (*) indicates Dicer containing a protease-sensitive loop in the RNase III domains that when cleaved does not significantly alter the structure of the protein.

(D and E) EM micrographs of (D) unlabeled human Dicer and (E) C-terminal streptavidin-labeled Dicer. The lower panels display representative class averages from the two data sets. Boxes show magnifications of individual particles.

See also Table S1.

proteins. The labeling strategy involves site-specific biotinylation of the subject protein, followed by posttranslational tagging with streptavidin and subsequent visualization and 3D reconstruction by EM. This approach minimizes perturbation of the subject protein molecule, leading to a high success rate for labeling internal domains. Moreover, internal labels are usually less flexible than terminal labels, which, when combined with high-throughput image acquisition and data processing to generate multiple unique reconstructions, leads to remarkably accurate domain localization in EM maps. We originally developed DOLORS to map the domain architecture of human Dicer, a specialized ribonuclease that cleaves long double-stranded RNA (dsRNA) into small fragments of discrete length as part of the RNA interference pathway (Bernstein et al., 2001). Dicer is composed of a large (220 kDa) polypeptide chain that contains multiple discrete functional domains. High-resolution crystal structures of many of these domains (or homologous domains) have been established (Figure 1A) (Du et al., 2008; Qin et al., 2010; Takeshita et al., 2007). However, due to its size and lack of symmetry, single-particle EM reconstructions of Dicer yield only a moderate resolution of 20 Å (Lau et al., 2009; Wang et al., 2009), and thus, the positions of individual domains remained ambiguous (Figure 1B).

Here, we describe the DOLORS method in detail, the strengths and weaknesses of this approach, and how it was used to successfully determine the domain organization of Dicer and gain structural insight into its mechanism of action. We found that short (approximately five residue) surface loops tended to be the best biotinylation sites, in that these led to samples with low conformational heterogeneity, with minimal disruption to the structure of the target protein. We further show that conformational heterogeneity in label position can be reduced by reducing the length of the labeled loop. We also show that random conical tilt (RCT) reconstruction of labeled particles allows unbiased determination of the label position and that by comparing multiple reconstructions, the precision of the positional information can be assessed. The combined findings provide a method for determining domain topology in EM maps of moderate resolution.

RESULTS

General Approach

Labeling in DOLORS exploits the AviTag (Beckett et al., 1999), a flexible 15-amino acid sequence that is recognized as a substrate by the biotin ligase enzyme BirA (Cronan, 1989). The AviTag is engineered into an internal loop in the subject protein. The recombinant subject protein is then expressed, purified, and biotinylated in vitro using BirA. After removal of excess biotin, His_6-tagged, monovalent streptavidin (Howarth et al., 2006) is bound to the subject protein at the site of biotinylation. The population of labeled particles can be enriched by selecting for the His_6-tagged streptavidin. Following this selection step, free streptavidin is separated from the labeled target particles using size exclusion chromatography.

The sample can then be prepared for EM, using either negative stain or vitrification. Streptavidin is a compact, 60 kDa protein that is easily visualized when attached to a larger subject particle. Here, we used negative-stain EM preservation to provide improved contrast for Dicer, which is both small and asymmetric. Automated data acquisition software (Suloway et al., 2005) and a streamlined data analysis pipeline (Lander et al., 2009) were used to acquire and process tilted pairs of images. Particles extracted from the untilted images were aligned and classified to form 2D class averages, which were subsequently converted to 3D maps using particles from the corresponding tilted images by RCT reconstruction (Radermacher et al., 1986). In most cases, the streptavidin label was readily visualized in the 2D class averages, typically adopting a variety of conformations, most likely due to some flexibility in the attachment. Using the RCT strategy, multiple conformations were converted into 3D maps, and the attachment site of the tag was located on the surface of the structure. Sampling of the streptavidin label positions from multiple 3D maps provided higher confidence in localizing the labeled site.

Engineering Biotinylation Sites

To generate streptavidin-labeled proteins, DNA encoding a variation of the 15-amino acid AviTag (LNDILEAQ**K**IEWHEG) sequence (Beckett et al., 1999) is cloned into the coding sequence of internal surface accessible loops (or the one of the termini) in a cDNA clone of the subject protein. The AviTag contains a lysine residue that is specifically recognized as a substrate for the Biotin Ligase enzyme BirA (Cronan, 1989). In the case of human Dicer, we tested a total of 12 different AviTagged protein constructs (Table S1 available online).

Internal AviTags work best for labeling when inserted into short surface loops that do not contribute to the structure of the target protein. In the case of Dicer, several approaches were taken for selecting such sites. For the PAZ, Platform, and RNase IIIb domains, identification of AviTag insertion sites was facilitated by inspection of crystal structures of these domains (Figure 1A). The most efficiently labeled and best-behaved target sites were short (approximately five amino acid) surface loops that were disordered in the crystal structures and thus did not appear to contribute to the structural stability of the protein. For selecting label sites in the absence of a crystal structure, we relied on structural prediction algorithms to identify candidate loops with similar properties. The I-TASSER server (Roy et al., 2010) was used to predict protein structure and identify likely exposed loops. We also used a combination of secondary structure prediction and sequence alignment between Dicer sequences from related organisms, with the idea that poorly conserved loops are less likely to contribute to protein structure. In the case of DUF283 labeling, we analyzed amino acids flanking the DUF283 domains with the DISOPRED2 Prediction of Protein Disorder Server (Ward et al., 2004). Charged or hydrophilic amino acids with the highest disorder score were chosen as sites for AviTag insertion.

Results from 12 different constructs indicate that DOLORS is most successful when a crystal structure of the domain being studied is available for selecting the AviTag insertion site. All four structure-guided constructs led to well-behaved proteins that were successfully labeled and visualized by EM (Table S1). Constructs designed by structure prediction worked with a lower frequency. Only two out of seven proteins gave reliable structural information. However, half of the failures were well-behaved proteins that labeled efficiently but did not give rise to RCT reconstructions for which streptavidin density could clearly be distinguished from Dicer. This is likely due to the fact that alignment of Dicer particles is greatly facilitated by its "L" shape (Figure 1B). Labels that happen to obscure the L shape may interfere with accurate particle alignment and thus may give rise to poor reconstructions. Of the 12 constructs made, only 2 produced Dicer proteins that aggregated, indicating that in general, well-chosen AviTag insertions do not dramatically disrupt protein structure or stability.

Preparation of Labeled Particles

Subject particles bearing the AviTag insertion were purified from cell lysate prior to in vitro biotinylation. In the case of Dicer, recombinant proteins initially contained an N-terminal His_6 tag, which was used for purification. His_6 tags were removed by proteolysis before biotinylation of the AviTag by incubating the purified protein with BirA. After biotinylation, Dicer samples were dialyzed extensively to remove excess biotin and then added slowly to an excess of monovalent streptavidin (Howarth et al., 2006), bearing a His_6 tag. Streptavidin-Dicer complexes were separated from nonlabeled proteins by Ni-chelate chromatography. Typically, about 60% of each Dicer sample was retained on the column, suggesting that the labeling efficiency is usually $\geq$60% (Figure 1C). Streptavidin-Dicer complexes were finally separated from free streptavidin and BirA by size exclusion chromatography.

Identifying Label Position by RCT Sampling

The 3D position of the streptavidin on labeled particles is identified by comparing multiple RCT reconstructions of the labeled particles to the EM map of the unlabeled protein. For Dicer, purified samples were negatively stained using 2% uranyl acetate. Automated EM software, Leginon (Yoshioka et al., 2007), was used to acquire large sets of tilted pairs of images suitable for RCT reconstruction. In general, more than 500 micrograph tilt pairs were acquired in a single unattended overnight session. Analysis of large EM data sets was greatly facilitated by Appion (Lander et al., 2009), an efficient image-processing pipeline that tracks all metadata and allows easy transitions through the steps of data analysis from particle picking to volume reconstruction. Appion facilitates streamlined EM data processing and analysis by using a relational database to track all processing jobs and results, together with all the generated metadata,

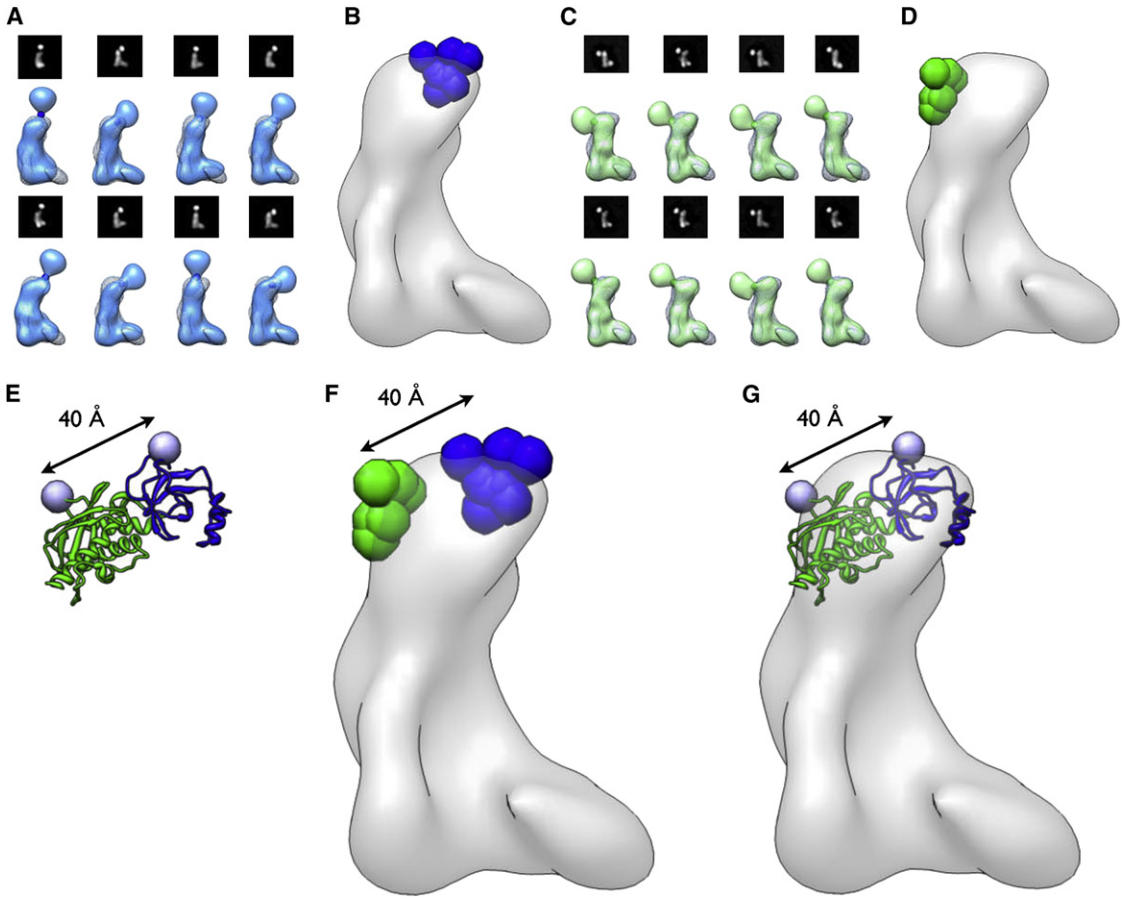

Figure 2. Identifying Internal Domains in an EM Map

(A–D) Streptavidin was attached to a loop in either (A) PAZ or (C) Platform domain of Dicer. 2D class averages of labeled proteins are shown with corresponding 3D RCT reconstructions. Each map is overlaid onto the unlabeled Dicer EM map (shown in gray). The estimated attachment sites for streptavidin are indicated by spheres (PAZ is in blue; Platform is in green). The estimated attachment sites for (B) PAZ-labeled and (D) Platform-labeled Dicer from eight independent RCT reconstructions mapped onto the unlabeled Dicer EM map.

(E) Crystal structures of PAZ (blue) and Platform (green) domains with the AviTag insertion sites indicated by spheres. The distance between the two sites is 40 Å.

(F) The estimated attachment sites for PAZ and Platform are separated by an average distance of 40 Å.

(G) PAZ-Platform crystal structure modeled into the EM map based on labeling results.

See also Figure S1.

and provide connections between disparate processing steps. For RCT data set analysis, the Appion pipeline allows users to select any aligned class average generated from an untilted set of images and immediately generate a 3D volume from the particles in the associated tilted images, essentially providing a "one-click" RCT reconstruction (Voss et al., 2010) in a straightforward and transparent manner. Note that the method does not require a preferred particle alignment but only that particles in any particular relative orientation can be aligned to form class averages. Unsupervised reference-free alignment and classification of labeled particles typically resulted in many class averages that contained a distinct additional density adjacent to the "L"-shaped Dicer (Figures 1D and 1E). 3D models from selected 2D class averages were generated and overlaid onto the unlabeled Dicer EM map (Lander et al., 2009). To localize the position of the label, six to eight independent RCT reconstructions of each labeled Dicer were overlaid onto the unlabeled Dicer map. RCT reconstructions were chosen based on the quality and resolution of the RCT and cross-correlation to the unlabeled Dicer map. For each overlaid RCT map, the streptavidin attachment site was estimated as the shortest distance connecting Dicer and streptavidin densities (Figures 2, 3, and 4).

Mapping Domain Positions with DOLORS

Comparing multiple independent RCT reconstructions allows assessment of the precision with which the streptavidin attachment sites have been estimated. For example, in Dicer particles labeled in the PAZ domain, eight estimated attachment sites fell within a radius of 10 Å in the front of the head region of Dicer (Figures 2A and 2B). Similarly, attachment sites for Dicer particles labeled in the Platform domain fell within a 10 Å radius in the back of the head region (Figures 2C and 2D). The precision in identifying both streptavidin attachment sites was sufficient to allow clear distinction between the PAZ and Platform labels.

In the crystal structure of *Giardia* Dicer, the PAZ and Platform domains are tightly associated with each other and function as an intact structural module (Macrae et al., 2006). Based on the

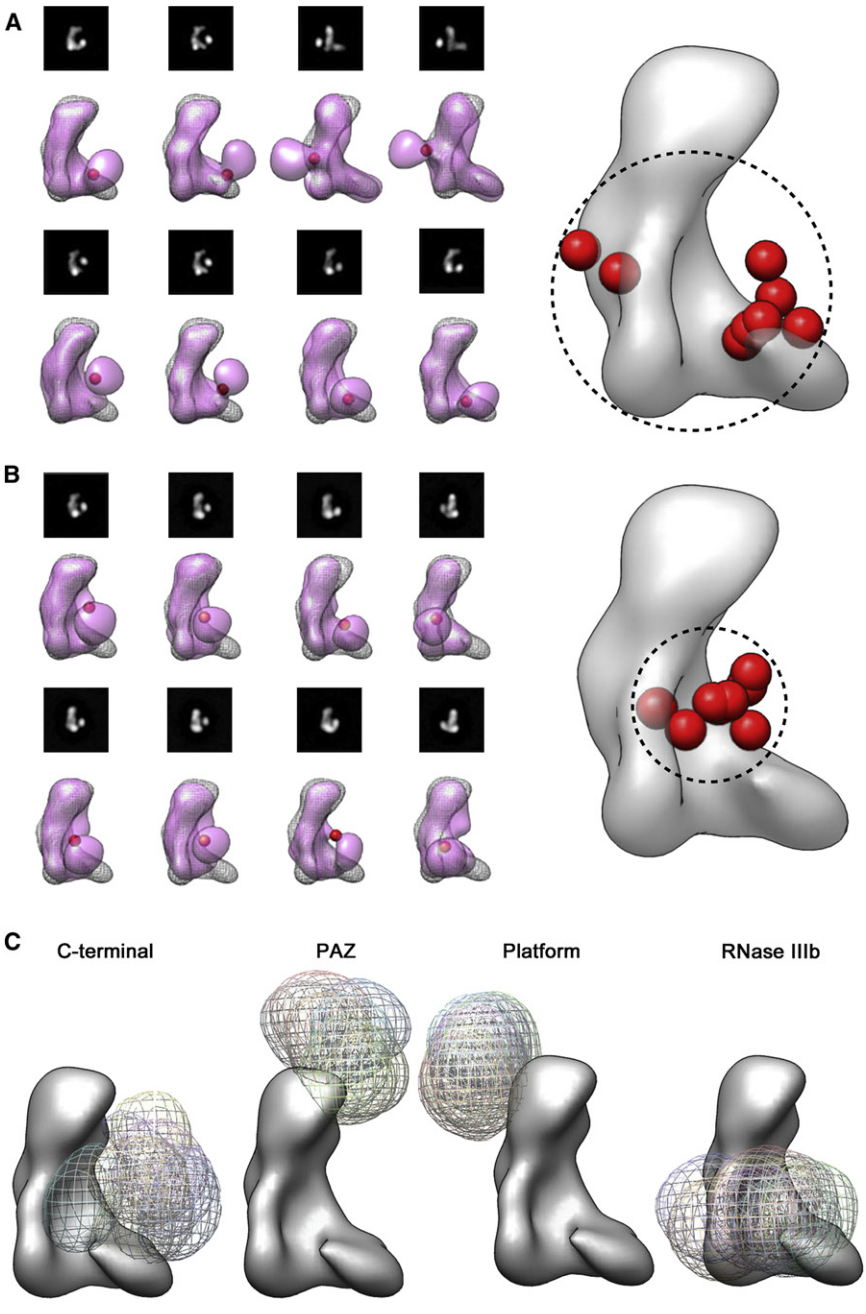

Figure 3. Comparing and Reducing Label Flexibility

(A and B) The RNase IIIb domain was labeled by attaching streptavidin to either (A) a 45-amino acid residue surface loop or (B) the same loop truncated to 25 residues. 2D class averages are shown with corresponding 3D RCT reconstructions overlaid on the unlabeled Dicer EM map, shown in gray (left panels). The estimated streptavidin attachment sites are indicated by red spheres. Estimated attachment sites from eight RCT reconstructions mapped on the Dicer EM map (right panels). Dashed circles indicate the smallest sphere in which all sites fall.

(C) The streptavidin density (colored mesh) from eight independent RCT reconstructions from C-terminal, PAZ, Platform, and RNase III-labeled Dicer, overlaid onto the EM map of Dicer (solid gray). The C-terminal attachment sites are diffuse and difficult to localize. The PAZ and Platform attachment sites are tightly clustered, allowing a positional estimate within a 10 Å radius. The RNase IIIb attachment site, whereas more flexible than those of PAZ and Platform, still allowed a positional estimate within a 20 Å radius.

See also Figure S1.

crystal structure, our PAZ and Platform AviTag insertions were placed about 40 Å apart (Figure 2E). Importantly, the central positions of the streptavidin attachment sites in the PAZ and Platform domains are separated by a distance of 40 Å, closely matching the distance between the two sites in the crystal structure (Figure 2F). Based on the positions of the labels, the PAZ-Platform crystal structure can be docked into the head region of Dicer (Figure 2G).

Label Flexibility Influences Mapping Precision

Streptavidin labels on short disordered loops tend to give the most tightly clustered estimated attachment sites and thus the highest confidence in label position. For example, the AviTags in the Dicer PAZ and Platform domains, which gave tight clustering (Figure 2), were both inserted into loops approximately five amino acids long. In contrast, tagging a longer loop in the RNase IIIb domain resulted in more heterogeneous results (Figure 3A). In this case, the AviTag was inserted in the middle of a large, 30-residue loop. The resulting loop, after AviTag insertion, was therefore 45 amino acid residues long. This loop was initially chosen because it is disordered in the crystal structure (Takeshita et al., 2007), suggesting that it does not contribute to structural stability of Dicer and would likely be an accessible site for BirA. Indeed, labeling was efficient, and RCT reconstruction was successful. However, streptavidin attachment sites estimated from eight RCT maps sampled a spherical volume with an ~45 Å radius (Figure 3B). To increase our precision, we reduced the flexibility of the streptavidin bound within the RNase IIIb domain by shortening the loop to which it was attached. Instead of simply inserting the AviTag sequence into the loop, we replaced residues N1780 to E1800 of the loop with the AviTag. This reduced the length of the loop from 45 to 25 amino acid residues. Reduced flexibility of the streptavidin was readily observed in 2D class averages and 3D RCT reconstructions (Figure 3A). All the mapped attachment sites fell within a 20 Å radius sphere, allowing a much more accurate localization of the streptavidin attachment site (Figure 3B). Thus, reduced loop length correlated with increased precision in positional information.

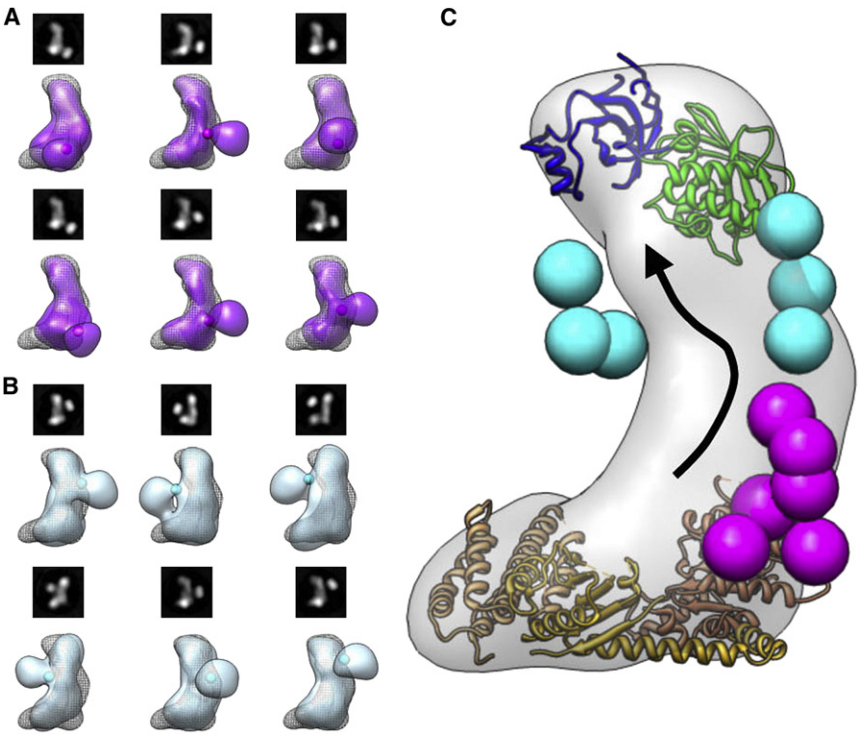

Figure 4. Topology Mapping Using DOLORS
(A and B) Dicer was labeled within the loop (A) DUF-prior and (B) DUF-after. 2D class averages are shown with corresponding 3D RCT reconstructions. Each map is overlaid onto the unlabeled Dicer EM map (shown in gray). The estimated attachment sites for streptavidin are indicated by spheres (DUF-prior is in pink; DUF-after is in cyan). (C) Positions of streptavidin attachment sites for DUF-prior and DUF-after follow the proposed path of the polypeptide, from helicase to PAZ/Platform, within Dicer EM map.
See also Figure S1.

We also found that streptavidin attached to the C terminus of Dicer gave rise to heterogeneous RCT reconstructions and imprecise positional information. Although the streptavidin was easily distinguishable within individual class averages, the positions of the streptavidin in each RCT reconstruction in relation to Dicer were too diverse to allow an accurate determination of the position of the C terminus of Dicer (Figure 3C). Taking the average position of all streptavidins from individual RCT reconstruction suggests that the C terminus lies within the body of Dicer. Although the C terminus only contains six disordered amino acid residues (Du et al., 2008), the streptavidin label appears to be able to take on a wide range of positions relative to Dicer. This observation is consistent with the idea that terminal labels will often be more flexible and conformationally heterogeneous than internal labels because they are tethered to only one point of the target protein and thus experience a higher degree of freedom. We also note that the positions of labels in the body of Dicer tended to be less precisely identified than those extending from the head (Figure 3C).

Topology Mapping Using DOLORS

By tagging sequential positions in a protein sequence, it is possible to establish the topology of a polypeptide chain in an EM map using DOLORS. In the case of Dicer, deletion studies had established that the N-terminal helicase domains reside in the base of Dicer L (Lau et al., 2012; Wang et al., 2009). Intriguingly, DOLORS mapped the Platform/PAZ module, which comes shortly after the helicase in the primary sequence, to the opposite end L (Figures 1A and 4). These data indicate that the 160 amino acid residues between the helicase and the Platform/PAZ module span the entire length of the Dicer protein. To test this model, we used DOLORS to map the polypeptide chain at two positions between helicase and Platform/PAZ. The residues intervening between these positions also contain a domain named DUF283, which adopts a double-stranded RNA binding domain (dsRBD) fold (Dlakić, 2006; Qin et al., 2010). We inserted the AviTag between residues D620 and G621, which are in a predicted loop N terminal to DUF283 (DUF-prior). We generated a second construct with the AviTag between residues E731 and E732, which are in a predicted loop after DUF283 in the primary sequence (DUF-after). For the DUF-prior streptavidin-labeled Dicer, DOLORS mapped the position of the loop close to the base of the Dicer L, adjacent to the helicase (Figure 4A). For DUF-after, the position of the loop mapped to the head region of Dicer, adjacent to the Platform domain (Figure 4B). These results support the model in which DUF283 connects the helicase to the Platform/PAZ module by spanning the length of the molecule (Figure 4C). The position of the DUF283 also suggests the possibility of its dimerization with the C-terminal dsRBD of Dicer, which is shown to also reside within the body of Dicer (Lau et al., 2012). Most importantly, these results show that DOLORS can also be used for domain tracing within a moderate-resolution EM map.

DISCUSSION

DOLORS is a valuable approach for EM structure determination in that it can provide information for the underlying architecture of a molecular complex at moderate resolution. The method is especially useful for large proteins for which crystal structures of isolated domains have been established, although we believe DOLORS could be applied to any molecular complex in which recombinant components can be incorporated. Although a single polypeptide was studied here, we expect that DOLORS will also be useful in determining structures of multisubunit molecular complexes where the localization of individual subunits is particularly challenging even with the availability of crystal structures. Using the small 15-amino acid AviTag insertion as a biotinylation target to which streptavidin is bound posttranslationally leads to remarkably high success rates when generating labeled protein samples. The AviTag also allows the use of internal labels, which tend to be more conformationally homogenous than terminal labels and thus provide

more precise positional information. The rigid shape, structural stability, and size of streptavidin facilitate unambiguous identification of the marker in relation to the protein of interest. Moreover, the subpicomolar affinity of streptavidin for the biotinylated target protein allows the labeled complex to remain intact at dilute concentrations and in heavy metal stain. In our experience, streptavidin labels are usually clearly observed in 2D class averages, and thus, even 2D projection analysis may be useful in identifying the location of the label. However, the use of automated data acquisition allows the generation of many RCT reconstructions for each labeled sample with relative ease. This approach facilitates RCT sampling, which provides a direct indication of how precisely the position of a label is known and ultimately more reliable structural information.

EXPERIMENTAL PROCEDURES

Generation of AviTag Expression Constructs

Desired sites of AviTag insertion were first mutated by QuikChange (Stratagene) to add a restriction site unique to the target protein expression plasmid. Complementary DNA oligonucleotides encoding the AviTag sequence (LNDILEAQ**K**IEWHEG), containing the appropriate sticky ends when annealed, were cloned into the unique restriction site. Dicer constructs with AviTags inserted in the PAZ, Platform, and the truncated RNase IIIb loop have been described previously by Lander et al. (2009). In DUF-prior, DUF-after, and RNase IIIb constructs, AviTags were inserted between residues D620 and G621, E731 and E732, and R1791 and S1792, respectively. Cloning details are described in the Supplemental Experimental Procedures.

Preparation of Streptavidin-Labeled Target Proteins

All Dicer proteins were produced in Sf9 cells using the Bac-to-Bac Baculovirus Expression System (Invitrogen) as described previously (MacRae et al., 2008). Recombinant Dicer proteins were first purified via an N-terminal His_6 tag, which was subsequently removed by treatment with TEV protease. Protein samples were then passed through a 5 ml HisTrap FF (GE Healthcare Life Sciences) column to remove additional contaminating proteins. The flow-through was collected, concentrated, and exchanged into a buffer containing 250 mM potassium glutamate with 25 mM Tris (pH 8.0) to remove sodium chloride, which inhibits BirA, prior to biotinylation. Biotinylation reactions were carried out in a reaction volume of about 0.5 ml. One-tenth the volume of a 10X BirA reaction buffer, containing 100 mM ATP, 100 mM magnesium acetate, and 500 μM d-biotin (pH 8), was added to the concentrated sample. The biotinylation reaction was started with the addition of 20 μl of 5 mg/ml purified BirA enzyme and incubated for 1 hr at 37°C.

Following the biotinylation reaction, samples were dialyzed extensively, against 0.3 M NaCl, 10 mM imidazole, 0.5 mM TCEP, 50 mM Na_2HPO_4 (pH 8), to remove excess biotin. The biotinylated protein samples were then bound to purified, His_6-tagged monovalent streptavidin (Howarth et al., 2006). The streptavidin-tagged samples were purified from untagged proteins by applying the sample to a 1 ml HisTrap FF column (untagged proteins lacked a His_6 tag at this stage of the preparation and thus did not bind the HisTrap column). The bound protein was eluted by increasing imidazole concentration to 300 mM. The Dicer-Streptavidin complex was then separated from free streptavidin and BirA by size exclusion chromatography using a HiLoad 16/60 Superdex 200 column (GE Healthcare Life Sciences).

Preparation of BirA and Monovalent Streptavidin

The biotin ligase enzyme BirA was expressed in *E. coli* strain BL21(DE3) from the plasmid pET21-BirA (Addgene plasmid 20857) and purified via a C-terminal His_6 tag as described previously by Chen et al. (2005). Monovalent streptavidin was prepared as described by Pettersen et al. (2004). Briefly, active streptavidin (strept-alive) and inactive mutant streptavidin (strept-dead) subunits were expressed separately in BL21(DE3) cells from the plasmids pET21a-Streptavidin-Alive (Addgene plasmid 20860) and pET21a-Streptavidin-Dead (Addgene plasmid 20859), respectively. Streptavidin proteins were purified from inclusion bodies by denaturation and then mixed in a strept-alive:strept-dead ratio of 1:6. Proteins were then refolded by rapid dilution into PBS, followed by concentration via ammonium sulfate precipitation. Streptavidin tetramers containing one His_6 tag (present only on strept-alive subunits) were purified by Ni-chelate chromatography.

Negative Staining and EM

Specimens were negatively stained using either the carbon sandwich (Ohi et al., 2004) or deep staining (Ruiz and Radermacher, 2006) method. The details of the protocol were described previously by Chittuluru et al. (2011) and are included in the Supplemental Experimental Procedures. EM was performed using an FEI Tecnai F20 Twin transmission electron microscope, operating at 120keV. Images were recorded using either a Tietz F415 4K × 4K or a Gatan 4K × 4K pixel CCD camera and using a dose of ~20 $e^-/Å^2$ and a nominal defocus between −1 and −3 μm. Leginon software (Suloway et al., 2005) was used for automated image acquisition at nominal magnifications of 50,000× or 62,000×, corresponding to a pixel size at the specimen level of 0.151 and 0.131 nm, respectively. The RCT node of Leginon (Yoshioka et al., 2007) was used for RCT data collection with image tilt pairs taken at 0° and 50°. Additional details are available in the Supplemental Experimental Procedures.

Image Processing and Model Reconstructions

The Appion software package (Lander et al., 2009) was used for all image processing and model reconstructions. To avoid bias, particles were extracted from the raw micrographs using a reference-free method by Voss et al. (2009). A combination of Xmipp (Scheres et al., 2005) and SPIDER (Frank et al., 1996) protocols was used for alignments and classifications. The 3D reconstruction was performed using SPIDER routines (Frank et al., 1996). Additional details are available in the Supplemental Experimental Procedures.

Model Fitting

Chimera visualization software (Pettersen et al., 2004) was used for all model fitting. Map segmentations were performed using Segger (Pintilie et al., 2010) available as an extension within Chimera. Additional details on modeling are available in the Supplemental Experimental Procedures.

ACCESSION NUMBERS

Representative RCT reconstructions of PAZ, Platform, RNase IIIb, DUF-prior, and DUF-after-labeled Dicers have been deposited at the Electron Microscopy Data Bank (EMDB) and assigned accession codes EMD-2214, EMD-2215, EMD-2216, EMD-2217, and EMD-2218, respectively.

SUPPLEMENTAL INFORMATION

Supplemental Information includes one table, one figure, and Supplemental Experimental Procedures and can be found with this article online at http://dx.doi.org/10.1016/j.str.2012.10.019.

ACKNOWLEDGMENTS

EM imaging and reconstruction were conducted at the National Resource for Automated Molecular Microscopy, which is supported by grants from the National Center for Research Resources (2P41RR017573-11) and the National Institute of General Medical Sciences (9 P41 GM103310-11) from the National Institutes of Health. This work was also supported by NIH grant R01 GM086701 to I.J.M. P.-W.L. is a predoctoral fellow of the American Heart Association. I.J.M. is a Pew Scholar in the Biomedical Sciences. C.S.P., B.C., and I.J.M. conceived of the project. P.-W.L. performed data collection. P.-W.L., C.S.P., B.C., and I.J.M. analyzed the data. P.-W.L., B.C., and I.J.M. wrote the manuscript.

Received: July 23, 2012
Revised: October 11, 2012
Accepted: October 31, 2012
Published: December 4, 2012

REFERENCES

Beckett, D., Kovaleva, E., and Schatz, P.J. (1999). A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. Protein Sci. *8*, 921–929.

Bernstein, E., Caudy, A.A., Hammond, S.M., and Hannon, G.J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature *409*, 363–366.

Chen, I., Howarth, M., Lin, W., and Ting, A.Y. (2005). Site-specific labeling of cell surface proteins with biophysical probes using biotin ligase. Nat. Methods *2*, 99–104.

Chittuluru, J.R., Chaban, Y., Monnet-Saksouk, J., Carrozza, M.J., Sapountzi, V., Selleck, W., Huang, J., Utley, R.T., Cramet, M., Allard, S., et al. (2011). Structure and nucleosome interaction of the yeast NuA4 and Piccolo-NuA4 histone acetyltransferase complexes. Nat. Struct. Mol. Biol. *18*, 1196–1203.

Choy, R.M., Kollman, J.M., Zelter, A., Davis, T.N., and Agard, D.A. (2009). Localization and orientation of the gamma-tubulin small complex components using protein tags as labels for single particle EM. J. Struct. Biol. *168*, 571–574.

Cronan, J.E., Jr. (1989). The *E. coli* bio operon: transcriptional repression by an essential protein modification enzyme. Cell *58*, 427–429.

Dlakić, M. (2006). DUF283 domain of Dicer proteins has a double-stranded RNA-binding fold. Bioinformatics *22*, 2711–2714.

Du, Z., Lee, J.K., Tjhen, R., Stroud, R.M., and James, T.L. (2008). Structural and biochemical insights into the dicing mechanism of mouse Dicer: a conserved lysine is critical for dsRNA cleavage. Proc. Natl. Acad. Sci. USA *105*, 2391–2396.

Flemming, D., Thierbach, K., Stelter, P., Böttcher, B., and Hurt, E. (2010). Precise mapping of subunits in multiprotein complexes by a versatile electron microscopy label. Nat. Struct. Mol. Biol. *17*, 775–778.

Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M., and Leith, A. (1996). SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. J. Struct. Biol. *116*, 190–199.

Howarth, M., Chinnapen, D.J., Gerrow, K., Dorrestein, P.C., Grandy, M.R., Kelleher, N.L., El-Husseini, A., and Ting, A.Y. (2006). A monovalent streptavidin with a single femtomolar biotin binding site. Nat. Methods *3*, 267–273.

Lander, G.C., Stagg, S.M., Voss, N.R., Cheng, A., Fellmann, D., Pulokas, J., Yoshioka, C., Irving, C., Mulder, A., Lau, P.W., et al. (2009). Appion: an integrated, database-driven pipeline to facilitate EM image processing. J. Struct. Biol. *166*, 95–102.

Lau, P.W., Potter, C.S., Carragher, B., and MacRae, I.J. (2009). Structure of the human Dicer-TRBP complex by electron microscopy. Structure *17*, 1326–1332.

Lau, P.W., Guiley, K.Z., De, N., Potter, C.S., Carragher, B., and MacRae, I.J. (2012). The molecular architecture of human Dicer. Nat. Struct. Mol. Biol. *19*, 436–440.

Lees, J.A., Yip, C.K., Walz, T., and Hughson, F.M. (2010). Molecular organization of the COG vesicle tethering complex. Nat. Struct. Mol. Biol. *17*, 1292–1297.

Macrae, I.J., Zhou, K., Li, F., Repic, A., Brooks, A.N., Cande, W.Z., Adams, P.D., and Doudna, J.A. (2006). Structural basis for double-stranded RNA processing by Dicer. Science *311*, 195–198.

MacRae, I.J., Ma, E., Zhou, M., Robinson, C.V., and Doudna, J.A. (2008). In vitro reconstitution of the human RISC-loading complex. Proc. Natl. Acad. Sci. USA *105*, 512–517.

Ohi, M., Li, Y., Cheng, Y., and Walz, T. (2004). Negative staining and image classification—powerful tools in modern electron microscopy. Biol. Proced. Online *6*, 23–34.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612.

Pintilie, G.D., Zhang, J., Goddard, T.D., Chiu, W., and Gossard, D.C. (2010). Quantitative analysis of cryo-EM density map segmentation by watershed and scale-space filtering, and fitting of structures by alignment to regions. J. Struct. Biol. *170*, 427–438.

Qin, H., Chen, F., Huan, X., Machida, S., Song, J., and Yuan, Y.A. (2010). Structure of the *Arabidopsis thaliana* DCL4 DUF283 domain reveals a noncanonical double-stranded RNA-binding fold for protein-protein interaction. RNA *16*, 474–481.

Radermacher, M., Wagenknecht, T., Verschoor, A., and Frank, J. (1986). A new 3-D reconstruction scheme applied to the 50S ribosomal subunit of *E. coli*. J. Microsc. *141*, RP1–RP2.

Roy, A., Kucukural, A., and Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. Nat. Protoc. *5*, 725–738.

Ruiz, T., and Radermacher, M. (2006). Three-dimensional analysis of single particles by electron microscopy: sample preparation and data acquisition. Methods Mol. Biol. *319*, 403–425.

Scheres, S.H., Valle, M., Nuñez, R., Sorzano, C.O., Marabini, R., Herman, G.T., and Carazo, J.M. (2005). Maximum-likelihood multi-reference refinement for electron microscopy images. J. Mol. Biol. *348*, 139–149.

Shanks, N.F., Maruo, T., Farina, A.N., Ellisman, M.H., and Nakagawa, T. (2010). Contribution of the global subunit structure and stargazin on the maturation of AMPA receptors. J. Neurosci. *30*, 2728–2740.

Stroupe, M.E., Xu, C., Goode, B.L., and Grigorieff, N. (2009). Actin filament labels for localizing protein components in large complexes viewed by electron microscopy. RNA *15*, 244–248.

Suloway, C., Pulokas, J., Fellmann, D., Cheng, A., Guerra, F., Quispe, J., Stagg, S., Potter, C.S., and Carragher, B. (2005). Automated molecular microscopy: the new Leginon system. J. Struct. Biol. *151*, 41–60.

Takeshita, D., Zenno, S., Lee, W.C., Nagata, K., Saigo, K., and Tanokura, M. (2007). Homodimeric structure and double-stranded RNA cleavage activity of the C-terminal RNase III domain of human dicer. J. Mol. Biol. *374*, 106–120.

Voss, N.R., Yoshioka, C.K., Radermacher, M., Potter, C.S., and Carragher, B. (2009). DoG Picker and TiltPicker: software tools to facilitate particle selection in single particle electron microscopy. J. Struct. Biol. *166*, 205–213.

Voss, N.R., Lyumkis, D., Cheng, A., Lau, P.W., Mulder, A., Lander, G.C., Brignole, E.J., Fellmann, D., Irving, C., Jacovetty, E.L., et al. (2010). A toolbox for ab initio 3-D reconstructions in single-particle electron microscopy. J. Struct. Biol. *169*, 389–398.

Wang, H.W., Noland, C., Siridechadilok, B., Taylor, D.W., Ma, E., Felderer, K., Doudna, J.A., and Nogales, E. (2009). Structural insights into RNA processing by the human RISC-loading complex. Nat. Struct. Mol. Biol. *16*, 1148–1153.

Ward, J.J., Sodhi, J.S., McGuffin, L.J., Buxton, B.F., and Jones, D.T. (2004). Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. J. Mol. Biol. *337*, 635–645.

Wiedenheft, B., Lander, G.C., Zhou, K., Jore, M.M., Brouns, S.J., van der Oost, J., Doudna, J.A., and Nogales, E. (2011). Structures of the RNA-guided surveillance complex from a bacterial immune system. Nature *477*, 486–489.

Yip, C.K., Berscheminski, J., and Walz, T. (2010). Molecular architecture of the TRAPPII complex and implications for vesicle tethering. Nat. Struct. Mol. Biol. *17*, 1298–1304.

Yoshioka, C., Pulokas, J., Fellmann, D., Potter, C.S., Milligan, R.A., and Carragher, B. (2007). Automation of random conical tilt and orthogonal tilt data collection using feature-based correlation. J. Struct. Biol. *159*, 335–346.

Zhang, X., Settembre, E., Xu, C., Dormitzer, P.R., Bellamy, R., Harrison, S.C., and Grigorieff, N. (2008). Near-atomic resolution using electron cryomicroscopy and single-particle reconstruction. Proc. Natl. Acad. Sci. USA *105*, 1867–1872.

