Supplemental Data

The Conserved Protein SZY-20 Opposes the Plk4-Related Kinase ZYG-1 to Limit Centrosome Size

Mi Hye Song, L. Aravind, Thomas Müller-Reichert, and Kevin F. O'Connell

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Protein Structure Analysis

HMM searches were carried out using the hmm_search program of the HMMER package, after they were optimized with the hmm_caliberate program (Eddy 1998). Multiple alignments were constructed using MUSCLE and KALIGN programs followed by manual adjustments based on PSI-BLAST results (Edgar 2004; Lassmann and Sonnhammer 2006). Protein secondary structure was predicted using a multiple alignment as the input for the JPRED program, with information extracted from a PSSM, HMM and the seed alignment itself (Cuff and Barton 2000). The domain architectures were determined using a panel of PSSM profiles and HMMs developed for sensitive and accurate detection of known proteins domains (Schaffer et al. 1999; Bateman et al. 2002). These were run using the RPS-BLAST program (Schaffer et al. 1999) or the hmm_search program (Eddy 1998) respectively. The networks were then constructed using custom PERL scripts and rendered using the PAJEK program (http://vlado.fmf.uni-lj.si/pub/networks/pajek/).

Antibodies

An affinity-purified anti-SZY-20 rabbit polyclonal antibody (αS20C) was prepared against the peptide Ac-CFGQNRNDMQKNNYQPNLQ-amide. The following antibodies were obtained from commercial sources: polyclonal α-FLAG antibody (Sigma), monoclonal anti-tubulin DM1A (Sigma), monoclonal α-Nop1p (EnCor Biotech), and monoclonal α-GST (Invitrogen). All antibodies were used at a 1:500-1:2,000 dilution .

Cloning and Protein Expression

The following constructs were expressed in *E.coli* and purified using Glutathione Sepharose 4 Fast Flow (GE Healthcare) according to the manufacturer's instructions: GST-SZY-20 (pMS3.2), GST-SZY-20^{SUZ} (pMS3.8), GST-SZY-20$^{SUZ\text{-}C}$ (pMS3.9), GST-SZY-20^{dm} (pMS3.10), GST (pGEX-6P-1) (GE Healthcare). GST expression vectors were constructed using Gateway technology to insert *szy-20* sequences into the destination vector pDEST-15.

RNA-Binding Assays

RNA-binding assays were performed as described (Audhya et al. 2005). Purified GST-tagged proteins were dialyzed into a buffer of 25 mM Tris-HCl, pH 7.4; 150 mM NaCl; 1% Triton X-100; one Complete EDTA-free protease inhibitor tablet (Roche)/50 ml. Purified proteins were incubated with 50 µl of 50% poly(U)-Sepharose beads (GE) for 30 min with gentle agitation at 4°C. The beads were spun in a microfuge for 2 min at 3000 rpm. The supernatants were reserved and the beads washed ten times in 1 ml of ice-cold buffer before final suspension in 50 µl of the same buffer. A fraction of each supernatant and bead prep was suspended in 2X sample buffer and 10 µl of each sample fractionated on a NuPAGE Bis-Tris Gel (Invitrogen). The gel was blotted to nitrocellulose and the membrane probed with the α-GST antibody. Purified GST was

used as negative control and GST-RGG (gift of Anjon Audhya and Karen Oegema) as a positive control.

SUPPLEMENTAL RESULTS

Protein Sequence Analysis Indicates that SZY-20 Is Involved in RNA Metabolism

To investigate the evolutionary relationships and possible functions of the three conserved blocks of SZY-20 orthologs, we carried out iterative searches using a sequence profile or hidden Markov model constructed from the conserved alignment blocks. We noticed that regions homologous to the central domain, the SUZ domain, and the C-terminal domain, the SUZ-C domain, were detected in several proteins (e-value <.001) independently of each other. These proteins were from a wide range of eukaryotes including animals, fungi, *Dictyostelium*, plants, and the apicomplexan *Plasmodium falciparum*. This suggests that they represent two distinct ancient conserved domains that have combined with each other as well as other modules, in different lineages in the course of evolution. The combination of SUZ and SUZ-C is so far observed only in the animal SZY-20 homologs. Several systematic studies on domain architectures have shown that they have important implications for protein function (Aravind 2000).

Both domains show a strong association with other previously known RNA-binding domains (Figure S3) (Anantharaman et al. 2002): SUZ is fused with R3H and RRM domains and RGG repeats in previously characterized multidomain RNA-binding proteins. These include the Encore protein from *D. melanogaster* (Hawkins et al. 1997), the Dip1 protein from *Zea mays* (Saleh et al. 2006) and R3hdm1protein from mammals. On the other hand, SUZ-C was found in other RNA-binding proteins fused to the S1-like domain in the animal CSDE1 (Grosset et al. 2000; Dormoy-Raclet et al. 2007), the LA and RRM domains in proteins typified by *Xenopus* Acheron (Klein et al. 2002), and to R3H and RRM domains in various uncharacterized proteins from fungi and *Dictyostelium*. SZY-20 was also found to be part of a complex identified in high-throughput interaction studies containing various RNA-binding proteins with KH, CCCH, RRM and RPR domains (Li et al. 2004). Based on the "syntax" of the domain architectures and the physical association, we predicted that the SUZ and SUZ-C domains function in connection to RNA-protein complexes. The SUZ domain enrichment in positively charged residues, including four universally conserved positively charged positions (Figure S2B), supports a role in direct RNA-binding. While this might be possible for the SUZ-C domain as well, it should be noted that it is a very small domain that is always found in one or more copies at the C-terminal end of all polypeptides, which additionally possess other N-terminal RNA binding domains. Therefore the SUZ-C domain might potentially function as a tag or a signal that helps in localizing the polypeptide to specific complexes.

SUPPLEMENTAL REFERENCES

Anantharaman, V., Koonin, E.V., and Aravind, L. (2002). Comparative genomics and evolution of proteins involved in RNA metabolism. Nucleic Acids Res *30*, 1427-1464.

Aravind, L. (2000). Guilt by association: contextual information in genome analysis. Genome Res *10*, 1074-1077.

Audhya, A., Hyndman, F., McLeod, I.X., Maddox, A.S., Yates, J.R., 3rd, Desai, A., and Oegema, K. (2005). A complex containing the Sm protein CAR-1 and the RNA helicase CGH-1 is required for embryonic cytokinesis in Caenorhabditis elegans. J Cell Biol *171*, 267-279.

Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M., and Sonnhammer, E.L. (2002). The Pfam protein families database. Nucleic Acids Res *30*, 276-280.

Cuff, J.A. and Barton, G.J. (2000). Application of multiple sequence alignment profiles to improve protein secondary structure prediction. Proteins *40*, 502-511.

Dormoy-Raclet, V., Markovits, J., Malato, Y., Huet, S., Lagarde, P., Montaudon, D., Jacquemin-Sablon, A., and Jacquemin-Sablon, H. (2007). Unr, a cytoplasmic RNA-binding protein with cold-shock domains, is involved in control of apoptosis in ES and HuH7 cells. Oncogene *26*, 2595-2605.

Eddy, S.R. (1998). Profile hidden Markov models. Bioinformatics *14*, 755-763.

Edgar, R.C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics *5*, 113.

Grosset, C., Chen, C.Y., Xu, N., Sonenberg, N., Jacquemin-Sablon, H., and Shyu, A.B. (2000). A mechanism for translationally coupled mRNA turnover: interaction between the poly(A) tail and a c-fos RNA coding determinant via a protein complex. Cell *103*, 29-40.

Hawkins, N.C., Van Buskirk, C., Grossniklaus, U., and Schupbach, T. (1997). Post-transcriptional regulation of gurken by encore is required for axis determination in Drosophila. Development *124*, 4801-4810.

Klein, S.L., Strausberg, R.L., Wagner, L., Pontius, J., Clifton, S.W., and Richardson, P. (2002). Genetic and genomic tools for Xenopus research: The NIH Xenopus initiative. Dev Dyn *225*, 384-391.

Lassmann, T. and Sonnhammer, E.L. (2006). Kalign, Kalignvu and Mumsa: web servers for multiple sequence alignment. Nucleic Acids Res *34*, W596-599.

Li, S., Armstrong, C.M., Bertin, N., Ge, H., Milstein, S., Boxem, M., Vidalain, P.O., Han, J.D., Chesneau, A., Hao, T., Goldberg, D.S., et al. (2004). A map of the interactome network of the metazoan C. elegans. Science *303*, 540-543.

Newton, K., Petfalski, E., Tollervey, D., and Caceres, J.F. (2003). Fibrillarin is essential for early development and required for accumulation of an intron-encoded small nucleolar RNA in the mouse. Mol Cell Biol *23*, 8519-8527.

Saleh, A., Lumbreras, V., Lopez, C., Dominguez-Puigjaner, E., Kizis, D., and Pages, M. (2006). Maize DBF1-interactor protein 1 containing an R3H domain is a potential regulator of DBF1 activity in stress responses. Plant J *46*, 747-757.

Schaffer, A.A., Wolf, Y.I., Ponting, C.P., Koonin, E.V., Aravind, L., and Altschul, S.F. (1999). IMPALA: matching a protein sequence against a collection of PSI-BLAST-constructed position-specific score matrices. Bioinformatics *15*, 1000-1011.

Wootton, J.C. (1994). Non-globular domains in protein sequences: automated segmentation using complexity measures. Comput Chem *18*, 269-285.

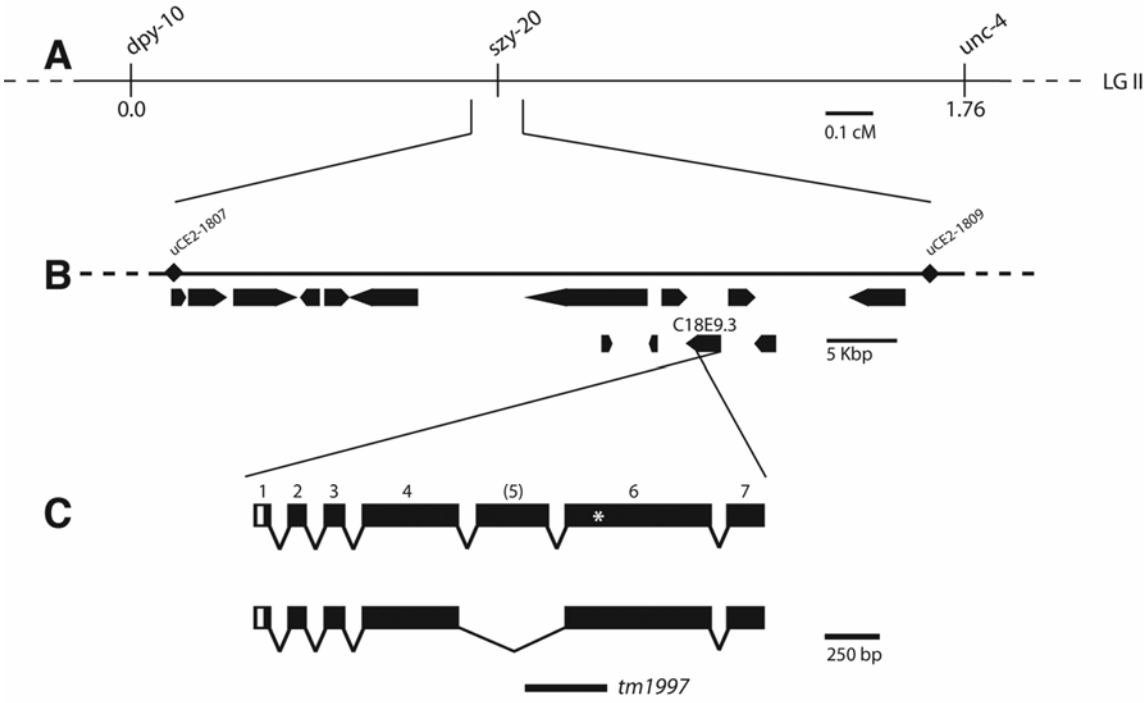

Figure S1. Cloning and Structure of the *szy-20* Locus

(A) The genetic interval containing the *szy-20* gene.
(B) The physical interval containing *szy-20* as defined by SNP mapping. The positions and orientations of the fourteen predicted ORFs in this interval are indicated.
(C) Shown are the intron-exon structures of the two major *szy-20* transcripts that are produced by alternative splicing of exon 5. The positions of the szy-20(*bs52)* mutation (asterisk) and the 370 bp deletion *szy-20(tm1997)* are also shown.

A. N-terminal extension

```
Szy20_Cele_32565720          65  NVSLVVADSWDDADADPVKEL 85
CBG03223_Cbri_39597541       12  IVALVVADSWEDADADPVKEL 32
CG31908_Dmel_24582402         2  SHGEDVLDNWEEIDEDGLSMT 22
LOC661564_Tcas_91088557       5  QQEVEVLESWEEIEETDVLDQ 25
LOC551727_Amel_66551291       2  STMDDVLESWEEIEESEVLNK 22
AAEL007556_Aaeg_108876516     8  QQQDLVDDSWEEIDEDRLAAR 28
PM20_Hsap_27526566            6  MEDEEVAESWEEAADSGKSKS 26
109926_Drer_71834312          1  MDDEEVAESWEEAADSGEMER 21
MGC76116_Xtro_45361483        1  MEEDEVAESWEEAADSGEIDR 21
RGD1560286_Rnor_62649944      1  MEDEEVAESWEEAADSGEIDR 21
Nvec1000024849                7  EDDEDIGDSWEDMADSGELER 27
Bflo1000000327                3  AGNDDVFDSWEDLADSGELER 23
Cint0100133317                1  MADDSIWDSWEDFADSGDIDK 21
consensus/100%                   .....l.-sW--h..s.....
```

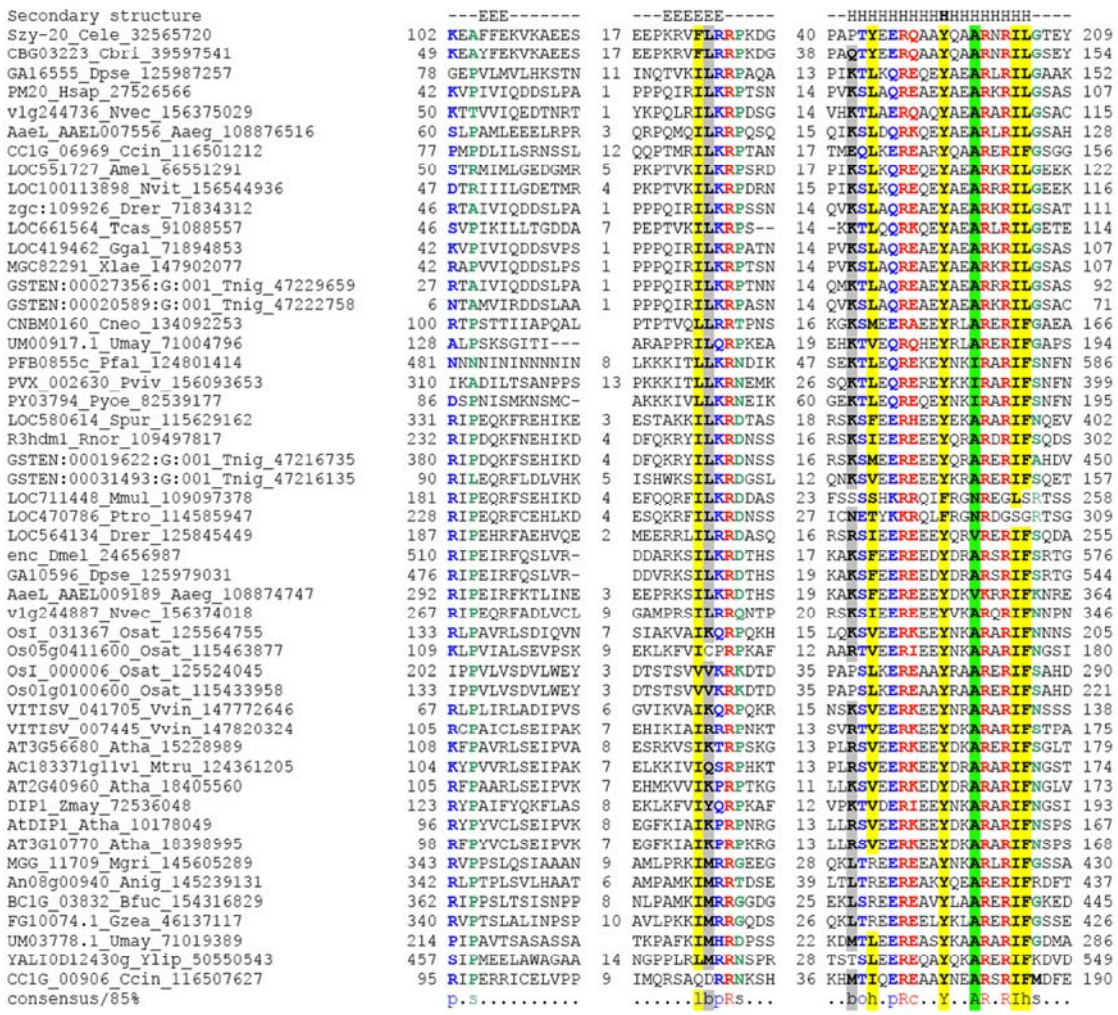

Figure S2. Alignments of the N-Terminal Extension, SUZ, and SUZ-C Domains

The predicted secondary structure using the JPRED program (Wootton, 1994) is indicated above the alignment. E is for an extended conformation or a beta strand and H is for a helical conformation. The columns are colored coded as follows: p-polar (KRHEDNQST) blue; s-small (GASTVHND) green; l-aliphatic (ILVMA) yellow highlight; h-hydrophobic (LIVMAWYFC) yellow highlight; o-alcoholic (ST) blue; c-charged (DERK) red; b-bulky (KREIFYWLQM) grey highlight.
(A) N-terminal stretch: *Aaeg : Aedes aegypti; Amel : Apis mellifera; Cbri : Caenorhabditis briggsae; Cele : Caenorhabditis elegans; Dmel : Drosophila melanogaster; Drer : Danio rerio; Hsap : Homo sapiens; Nvec : Nematostella vectensis; Rnor : Rattus norvegicus; Tcas : Tribolium castaneum; Bflo : Branchiostoma floridae; Xtro : Xenopus tropicalis.; Cint : Ciona intestinalis.*
(B) SUZ: *Aaeg : Aedes aegypti; Amel : Apis mellifera; Anig : Aspergillus niger; Atha : Arabidopsis thaliana; Bfuc : Botryotinia fuckeliana; Cbri : Caenorhabditis briggsae; Ccin : Coprinopsis cinerea; Cele : Caenorhabditis elegans; Cneo : Cryptococcus neoformans; Dmel : Drosophila melanogaster; Dpse : Drosophila pseudoobscura; Drer : Danio rerio; Ggal : Gallus gallus; Gzea : Gibberella zeae; Hsap : Homo sapiens; Mgri : Magnaporthe grisea; Mmul : Macaca mulatta; Mtru : Medicago truncatula; Nvec : Nematostella vectensis; Nvit : Nasonia vitripennis; Osat : Oryza sativa; Pfal : Plasmodium falciparum; Ptro : Pan troglodytes; Pviv : Plasmodium vivax; Pyoe : Plasmodium yoelii; Rnor : Rattus norvegicus; Spur : Strongylocentrotus purpuratus; Tcas : Tribolium castaneum; Tnig : Tetraodon nigroviridis; Umay : Ustilago maydis; Vvin : Vitis vinifera; Xlae : Xenopus laevis; Ylip : Yarrowia lipolytica; Zmay : Zea mays.*
(C) SUZ-C: *Aaeg : Aedes aegypti; Agam : Anopheles gambiae; Amel : Apis mellifera; Aory : Aspergillus oryzae; Bflo : Branchiostoma floridae; Cbri : Caenorhabditis briggsae; Cele : Caenorhabditis elegans; Ddis : Dictyostelium discoideum; Dmel : Drosophila melanogaster; Drer :Danio rerio; Ggal : Gallus gallus; Hsap : Homo sapiens; Mdom : Monodelphis domestica; Mmul : Macaca mulatta; Nvec : Nematostella vectensis; Nvit : Nasonia vitripennis; Oana : Ornithorhynchus anatinus; Pnod : Phaeosphaeria nodorum; Ptro : Pan troglodytes; Rnor : Rattus norvegicus; Spur : Strongylocentrotus purpuratus; Tcas : Tribolium castaneum; Tnig : Tetraodon nigroviridis; Xlae : Xenopus laevis; Xtro : Xenopus tropicalis.*

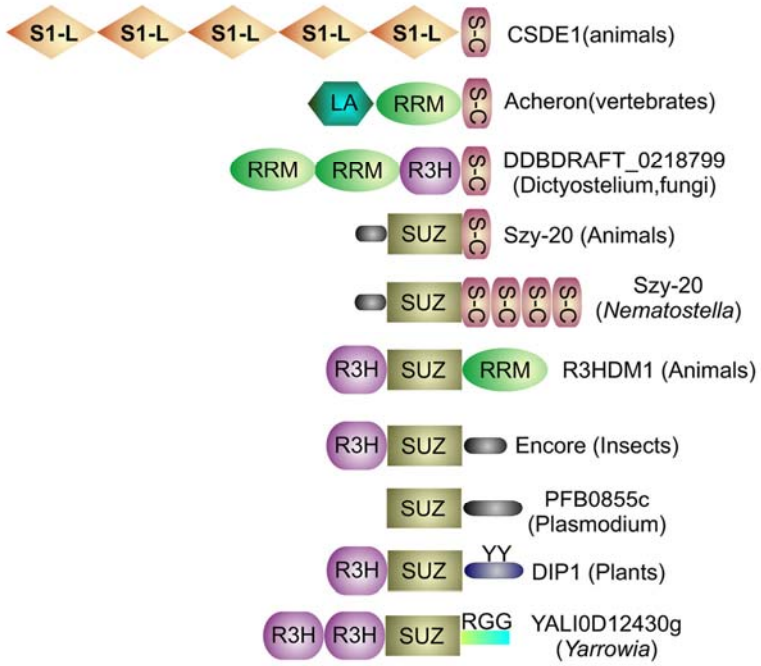

Figure S3. Protein Sequence Analysis Indicates that SZY-20 Is Involved in RNA Binding
The phyletic patterns indicate the organisms in which orthologs of a protein are found. The domain abbreviations are: RGG- RGG repeats; S-C: SUZ-C; S1-L: S1-like OB fold

domain; YY: a conserved region with conserved tyrosines. Note that only the globular domains are shown approximately to scale. The rest of the low complexity sequence in the proteins is not displayed in the schematic domain architectures. The unlabeled grey modules represent regions of extended conservation that are uniquely found in orthologs of the indicated proteins.

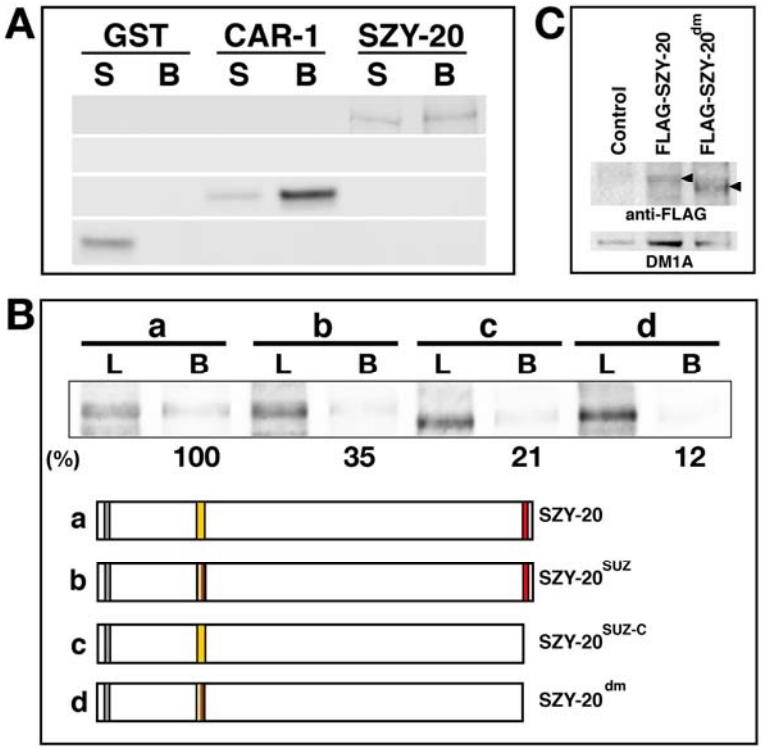

Figure S4. SZY-20 Binds RNA via the SUZ and SUZ-C Domains

(A and B) Wild-type and mutant versions of GST-SZY-20 were incubated with poly-U beads. Supernatant (S) represents the unbound fraction and washed beads (B) represent the bound fraction. Proteins were detected by immunoblotting using the α-GST antibody. A GST-tagged version of the CAR-1 RGG domain served as a positive control and GST as a negative control.

(B) SUZ and SUZ-C domains possess RNA-binding activity. The relative RNA-binding activity is shown beneath the blot, and was calculated by normalizing the fraction bound (B) to the total amount of input protein and then expressing this as a percentage of wild-type RNA-binding activity. Twenty percent of the input protein was loaded (L) on the gel. The structure of each protein assayed is illustrated in a-d. In the SZY-20^{SUZ} mutation, the protein sequence R138-N139-R140 was changed to G138-N139-G140, and in the SZY-20^{SUZ-C} mutation, the C-terminal 20 amino acids, including the entire SUZ-C domain, were deleted.

(C) Immunoblot showing expression of exogenous FLAG-tagged SZY-20 proteins in embryos. Blots were probed with an anti-FLAG antibody and an anti-α-tubulin (DM1A) antibody as a loading control.

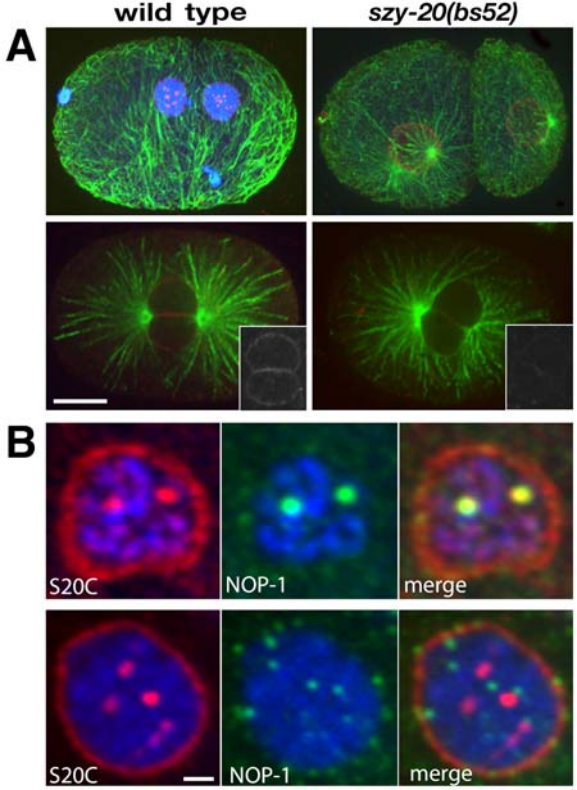

Figure S5. A Fraction of SZY-20 Localizes to the Nucleolus
(A) Embryos were stained for SZY-20 (αS20C, red), tubulin (green), and in some cases, DNA (blue). In wild-type embryos (left), the SZY-20 antibody labels a random number of intranuclear foci as well as the nuclear periphery. The intranuclear foci are absent in *szy-20(bs52)* embryos which lack the epitope recognized by the antibody. Staining of the nuclear periphery is reduced but still present in *szy-20(bs52)* embryos (insets), indicating that nuclear periphery staining is in part due to nonspecific cross-reactivity of the antibody. Shown are Z-projections.
(B) The SZY-20 intranuclear foci partially co-localize with the nucleolar marker NOP-1 (Newton et al. 2003). Embryonic nuclei stained for SZY-20 (αS20C, red), NOP-1 (green) and DNA (blue). Note that in some nuclei, SZY-20 and NOP-1 localize (top) while in others they do not (bottom), indicating temporal differences in the localization of these two proteins to nucleoli. Bars, 10 μm (A) and 1 μm (B).

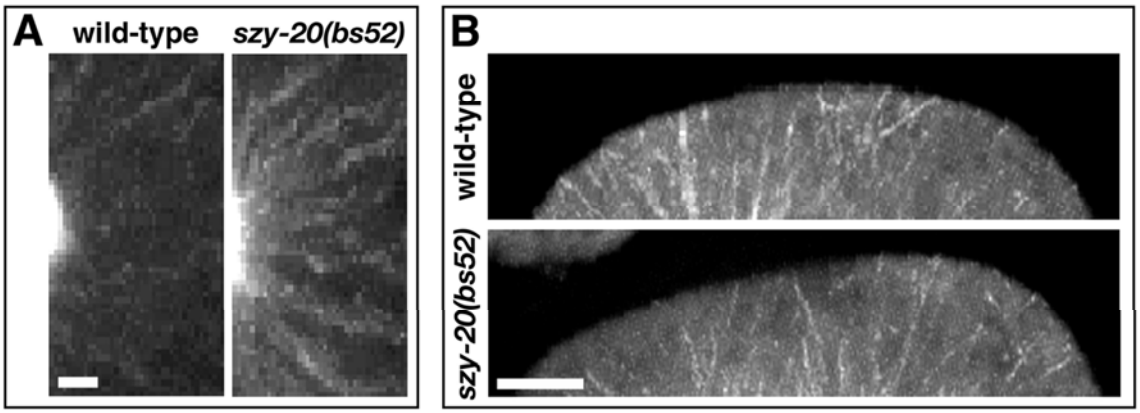

Figure S6. More Microtubules Grow Out from *szy-20* Mutant Centrosomes, but Fewer Microtubules Reach the Cortex
(A and B) Shown are 90-second time-projections of GFP-EBP-2 fluorescence at the centrosome (A) and cortex (B) of embryos at metaphase. Note that more microtubules emerge from *szy-20(bs52)* centrosomes than wild-type centrosomes and that in the mutant relatively fewer microtubules grow long enough to reach the cortex. Bars, 1 μm (A) and 5 μm (B).

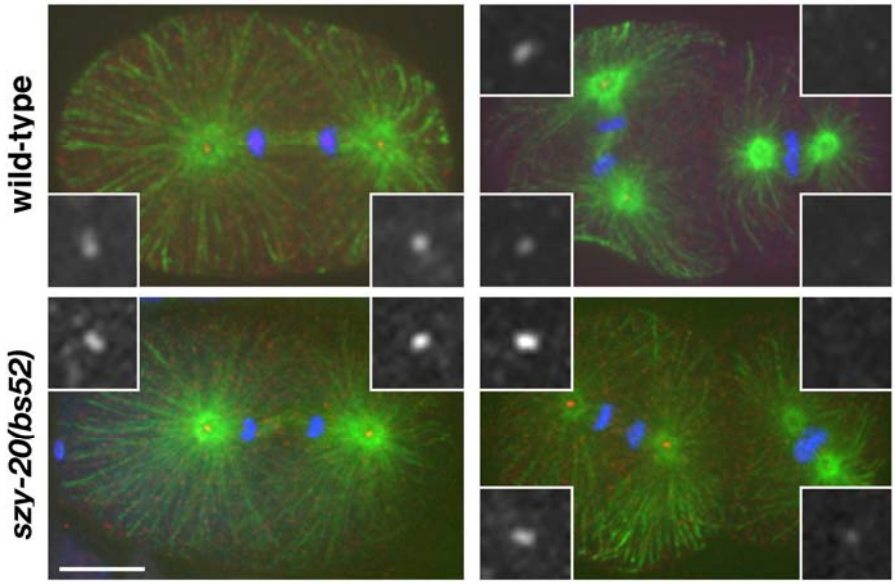

Figure S7. The Level of Centrosome-Associated ZYG-1 in *szy-20(bs52)* Embryos Is Higher Throughout the Cell Cycle

Embryos at anaphase and second mitosis were stained for microtubules (green), ZYG-1 (red) and DNA (blue). Insets are magnified 3-fold. Bar, 10 μm.

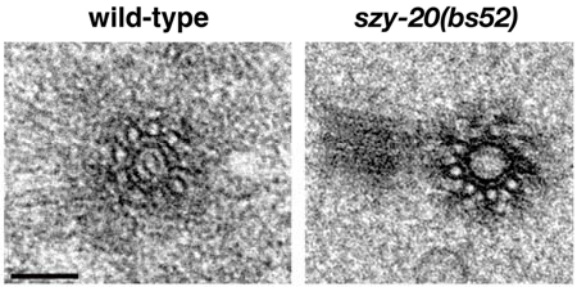

Figure S8. Centriole Structure Is Preserved in *szy-20* Mutants

Shown in cross-section are centrioles from wild-type and *szy-20(bs52)* mutant embryos. Gross defects in the overall structure or 9-fold rotational symmetry are not observed in the mutant. Bar, 100 nm.

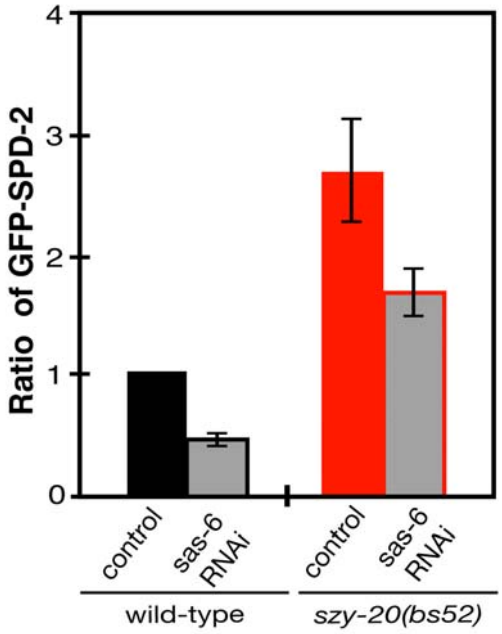

Figure S9. *sas-6 (RNAi)* Reduces the Levels of GFP-SPD-2 at Centrosomes both in Wild-Type and *szy-20* Mutant Embryos

Relative levels of GFP-SPD-2 localization at centrosomes were measured in wild-type and *szy-20(bs52)* embryos that were subjected to *sas-6(RNAi)*. Vertical bars indicate the standard deviation.