# **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 ( $\alpha$ S20C) was prepared against the peptide Ac-CFGQNRNDMQKNNYQPNLQ-amide. The following antibodies were obtained from commercial sources: polyclonal  $\alpha$ -FLAG antibody (Sigma), monoclonal anti-tubulin DM1A (Sigma), monoclonal  $\alpha$ -Nop1p (EnCor Biotech), and monoclonal  $\alpha$ -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- $20^{SUZ-C}$  (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  $\mu$ l 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  $\mu$ l of the same buffer. A fraction of each supernatant and bead prep was suspended in 2X sample buffer and 10  $\mu$ l of each sample fractionated on a NuPAGE Bis-Tris Gel (Invitrogen). The gel was blotted to nitrocellulose and the membrane probed with the  $\alpha$ -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 RNAbinding 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 highthroughput 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). Posttranscriptional 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	IVALV <mark>V</mark> ADS <mark>WEDA</mark> DADPVKEL	32
CG31908 Dmel 24582402	2	SHGEDVLDNWEEIDEDGLSMT	22
LOC661564_Tcas_91088557	5	QQEVE <mark>VLESWEEI</mark> EETDVLDQ	25
LOC551727 Amel 66551291	2	STMDDVLESWEE1EESEVLNK	22
AAEL007556 Aaeg 108876516	8	QQQDLVDDSWEE1DEDRLAAR	28
PM20_Hsap_27526566	6	MEDEE <mark>V</mark> AESWEEAADSGKSKS	26
109926 Drer 71834312	1	MDDEEVAESWEEAADSGEMER	21
MGC76116_Xtro_45361483	1	MEEDE <mark>VAESWEEA</mark> ADSGEIDR	21
RGD1560286 Rnor 62649944	1	MEDEEVAESWEEAADSGEIDR	21
Nvec1000024849	7	EDDED <mark>I</mark> GDSWEDMADSGELER	27
Bflo100000327	3	AGNDD <mark>V</mark> FDSWEDLADSGELER	23
Cint0100133317	1	MADDS IWDS WEDF ADSGDIDK	21
consensus/100%		<mark>l</mark> s <mark>W</mark> <mark>h</mark> s	

#### B. SUZ

Secondary structure Szy-20\_Cele\_32565720 CBG03223\_Cbri\_39597541 GA16555\_Dpse\_125987257 PM20\_Hsap\_27526566 v1g244736\_Nvec\_156375029 AaeL\_AAEL007556\_Aaeg\_108876516 CC1G\_06969\_Ccin\_116501212 LOC551727\_Amel\_66551291 LOC100113898\_Nvit\_156544936 Loc100113898 Nvit\_156544936 zgc:109926\_Drer\_71834312 Loc61564\_Tcas\_91089557 Loc419462\_Ggal\_71894853 MGC82291\_X1ae\_I47902077 GSTEN:00027356:G:001\_Tnig\_47229659 GSTEN:00020589:G:001\_Tnig\_47222758 CNBM0160\_Cneo\_134092253 UM00917.1\_Umay\_71004796 PFB0855c\_Pfal\_124801414 PVX\_002503\_Dviv\_155093653 PVX\_002630\_Pviv\_156093653 PY03794\_Pyoe\_82539177 LOC580614\_Spur\_115629162 R3hdm1\_Rnor\_109497817 GSTEN:00019622:G:001\_Tnig\_47216735 GSTEN:00019622:0:001\_HIIg\_47216735 GSTEN:00031493:G:001\_Thig\_47216135 LOC711448\_Mnul\_109097378 LOC470786\_Ptro\_114585947 LOC564134\_Drer\_125845449 enc\_Dmel\_24656987 enc\_Dme1\_24056987 GA10596\_Dpse\_125979031 AaeL\_AAEL009189\_Aaeg\_108874747 v1g244887\_Nvec\_156374018 OsI\_031367\_Osat\_125564755 Os05g0411600\_Osat\_115463877 OsI\_000006\_05at\_125524045 osolg0100600\_osat\_115433958 VITISV\_041705\_vvin\_147772646 VITISV\_007445\_vvin\_147820324 AT3656680\_Atha\_15228989 AC183371g11v1\_Mtru\_124361205 AT2G40960\_Atha\_18405560 DIP1\_Zmay\_72536048 AtDIP1\_Atha\_10178049 AT3G10770\_Atha\_18398995 MGG 11709 Mgri 145605289 An08g00940 Anig 145239131 HIGG 03832 Bfuc 154316829 FG10074.1 Gzea 46137117 UM03778.1 Umay 71019389 YALIOD12430g Ylip 50550543 CC16 00906 Ccin 116507627 consensus/85%

--ННННННННННН ---EEE---------EEEEE------102 KEAFFEKVKAEES 17 EEPKRVFLRRPKDG 40 PAPTYEERQAAYQAARNRILGTEY 209 38 PAQTYEERQAAYQAARNRILGSEY 154 EEPKRVFLRRPKDG 49 KEAYFEKVKAEES 17 INQTVKILRRPAQA PIKTLKOREQEYAEARLRILGAAK 78 GEPVLMVLHKSTN 11 13 152 PPPQIRILKRPTSN 42 KVPIVIQDDSLPA 14 PVKSLAQREAEYAEARKRILGSAS 107 1 14 VHKTLAERQAQYAEARARILGSAC 115 YKPQLR**ILKRP**DSG 50 KTTVVIOEDTNRT 1 QRPQMQILRRPQSQ 15 QIKSLOORKQEYAEARLRILGSAH 60 SLPAMLEEELRPR 3 128 QQPTMRILKRPTAN 17 TMEQLKEREARYQAARERIFGSGG 156 77 PMPDLILSRNSSL 12 50 STRMIMLGEDGMR PKPTVKILKRPSRD PKPTVKILKRPDRN 17 PIKSLKOREQEYAEARKRILGEEK 122 15 PIKSLKOREQEYAEARRRILGEEK 116 5 47 DTRIIILGDETMR 4 PPPQIR ILKRPSSN 14 QVKSLAQREAEYAEARKRILGSAT 46 RTAIVIQDDSLPA 111 46 SVPIKILLTGDDA PEPTVKILKRPS--14 -KKTLQQRKQEYAEARLRILGETE 114 42 KVPIVIQDDSVPS PPPQIR ILKRPATN 14 PVKSLAQREAEYAEARKRILGSAS 107 14 PVKSLAQREAEYAEARKRILGSAS 107 PPPQIRILKRPTSN 42 RAPVVIQDDSLPS 27 RTAIVIQDDSLPA PPPQIRILKRPTNN 14 QMKTLAQREAEYAEARRRILGSAS 92 PPPQIR**ILKRPASN** 14 QVKSLAQREAEYAEARKRILGSAC NTAMVIRDDSLAA 1 100 RTPSTTIIAPQAL PTPTVQLLRRTPNS 16 KGKSMEERAEEYRLARERIFGAEA 166 19 EHKTVEQRQHEYRLARERIFGAPS ARAPPRILORPKEA 128 ALPSKSGITI---194 47 SEKTLEOREKEYNKIRARIFSNFN 481 NNNNININNNNIN LKKKITLLKRNDIK 586 PKKKITLLKRNEMK 26 SQKTLEQREREYKKIRARIFSNFN 399 310 IKADILTSANPPS 13 60 GEKTLEQREQEYNKIRARIFSNFN 195 86 DSPNISMKNSMC-AKKKIVLLKRNEIK ESTAKKILKRDTAS 18 RSKSFEERHEEYEKARARIFNQEV 402 331 RIPEOKFREHIKE 3 DFQKRYILKRDNSS 16 RSKSIEEREEEYQRARDRIFSQDS 302 232 RIPDQKFNEHIKD 4 16 RSKSMEEREEEYQRARERIFAHDV 450 12 QNKSVEEREEEYKRARERIFSQET 157 23 FSSSSHKRRQIFRGNREGLSRTSS 258 DFQKRYILKRDNSS 380 RIPDOKFSEHIKD 4 90 RILEQRFLDLVHK 5 ISHWKS**ILKRD**GSL EFQQRFILKRDDAS 181 RIPEORFSEHIKD 4 RIPEQRFCEHLKD ESQKRFILKRDNSS 27 ICNETYKKRQLFRGNRDGSGRTSG MEERRLILRRDASQ 187 RIPEHRFAEHVQE 2 16 RSRSIEEREEEYQRVRERIFSQDA 255 17 KAKSFEEREEDYDRARSRIFSRTG 576 510 RIPEIRFQSLVR-DDARKSILKRDTHS 19 KAKSFEEREEDYDRARSRIFSRTG 544 19 KAKSFEEREEEYDKVKRRIFKNRE 364 476 RIPEIRFOSLVR-DDVRKS**ILKRDTHS** 292 RIPEIRFKTLINE EEPRKSILKRDTHS 20 RSKSIEEREEEYVKARQRIFNNPN 346 15 LQKSVEERKEEYNKARARIFNNNS 205 267 RIPEORFADLVCL GAMPRSILRRONTP 9 SIAKVAIKORPOKH 133 RLPAVRLSDIOVN EKLKFVICPRPKAF 12 AARTVEERIEEYNKARARIFNGSI 180 109 KLPVIALSEVPSK DTSTSVVVKRKDTD 35 PAPSLKEREAAYRAARERIFSAHD 202 IPPVLVSDVLWEY 3 290 35 PAPSLKEREAAYRAARERIFSAHD 221 15 NSKSVEERKEEYNRARARIFNSSS 138 133 IPPVLVSDVLWEY DTSTSVVVKRKDTD 3 GVIKVAIKORPOKR 67 RLPLIRLADIPVS 6 13 SVRTVEERKEEYDKARRIFSTA 175 13 PLRSVEERKEEYDKARRIFSGLT 179 13 PLRSVEERKEEYDKARRIFSGLT 174 EHIKIAIRRPNKT 105 RCPAICLSEIPAK 108 KFPAVRLSEIPVA ESRKVSIKTRPSKG 8 104 KYPVVRLSEIPAK ELKKIVIOSRPHKT 105 REPAARLSEIPVK EHMKVVIKPRPTKG 11 LLKSVEERKEDYDRARARIFNGLV 173 EKLKFVIYQRPKAF 12 VPKTVDERIEEYNKARARIFNGSI 193 123 RYPAIFYQKFLAS 13 LLRSVEERKEEYDKARARIFNSPS 167 13 LLRSVEERKEEYDKARARIFNSPS 168 96 RYPYVCLSEIPVK 8 EGFKIAIKPRPNRG 98 REPYVCLSETPVK 7 EGEKTATEPREKEG 343 RVPPSLQSIAAAN AMLPRKIMRRGEEG 28 QKLTREEREEAYNKARLRIFGSSA 430 342 RLPTPLSVLHAAT AMPAMKIMRRTDSE 39 LTLTREEREAKYQEARERIFRDFT 437 25 EKLSREEREAVYLAARERIFGKED 445 26 OKLTREEREELYKLARERIFGSSE 426 362 RIPPSLTSISNPP NLPAMKIMRRGGDG 8 10 AVLPKKIMRRGODS 340 RVPTSLALINPSP 22 KDMTLEEREASYKAARARIFGDMA 286 TKPAFKIMHRDPSS 214 PIPAVTSASASSA 457 SIPMEELAWAGAA 14 NGPPLRLMRRNSPR 28 TSTSLEEREQAYQKARERIFKDVD 549 95 RIPERRICELVPP 9 IMQRSAQDRRNKSH 36 KHMTIQEREAAYNEARSRIFMDFE 190 .....lbpRs... ...boh.pRc...Y...AR.RIhs... p.s.....

#### C. SUZ-C

Secondary Struct. LOC703901\_Mmul\_109080056 ENSANGG00000019760\_Agam\_118785968 AaeL AAEL010018 Aaeg\_108873813 dCSDE1\_Dme1\_47026413 LOC100026945\_Mdom\_126277631 LARP6\_Gga1\_118095675 GSTEN:00022434:G:001\_Tnig\_47220914 zgc:66107\_Drer\_41054017 MGC130966\_X1ae\_148233564 Larp6\_Rnor\_109483591 Loc420093\_Ggal\_50761032 Loc510025081\_Mdom\_126323256 Loc551727\_Amel\_66551291 Loc664329\_Tcas\_91087187 Loc100113898\_Nvit\_156544936 zgc:109926\_Drer\_71834312 LOC584092\_Spur\_115733035 LOC655517\_Tcas\_91079915 LOC100123206\_Nvit\_156537407 LOC100088086\_Oana\_149577832 MGC76116\_Xtro\_45361483 v1g244736\_Nvec\_156375029 Loc100012018 Mdom 126340373 GSTEN:00020562:G:001\_Tnig\_47222740 csde1\_Xtro\_71895719 LOC558349 Drer 125823502 DDBDRAFT 0218799 Ddis 66808135 LOC414014 Amel\_66566223 LOC414014 Anne1 000001 CSDE1 Ptro 114558794 SNOG 09039 Pnod 111062111 A0090003001325 Aory\_83768247 ACC690005001252 ACT9 03768247 AaeL AAEL007556 Aaeg 108876516 LOC661564 Tcas 31088557 CG31908 Dmel 24582402 Bflo1000000327 Bflo Bflo1000000327 Bflo1000000327 RGD1560286\_predicted\_Rnor\_62649944 PM20\_Hsap\_27526566 CBG03223\_Cbri\_39597541 Szy-20\_Cele\_32565720 consensus/85%

	E	SEEEE	CEE			
83	PRPEO	LVSPLKNITLDDANAPRLM	LHOPRO	PDNSM-	GCGAER	124
67	PRPER	LISRIKLNSVDDS-GPRLT	IRAPK	PDGTK-	GFHASA	907
31	VRAER	LSRIKLNSVDDS-GPRLT	TRAPKO	PDGTK-	GFHISV	971
08	DRPDR	LISRLKLNG-DDT-VPRLI	IRAPK	PQG-K-	GFSVLA	1046
66	PAASP	LLTR-KVQNADGSPVG	LRLPRO	PDDTR-	GFHGGY	503
41	PVSSP	LAR-KIQNADGLPVG	LRLPRO	PDGTK-	GFHSGC	478
84	PVGSP	LAR-KIQNADGLPPG	VRLPR	PDGTR-	GFHA	419
37	PVGSP	LAR-KIQNADGLPPG	VRLPRO	PDGTR-	GFHCPP	474
43	PAASP	LGR-KIQNADGLPPG	LRLPR	PDGTK-	<b>GF</b> HNGG	480
67	PGASP	LLSR-KMQTADGLPVG	LRLPRO	PDDTR-	<b>GF</b> HGGH	604
93	PPRTP	LAVKPKSR	LRLPH	PDGTR-	GFYNSI	423
00	VTHSW:	SWGPNPVPPLPRLD	SRLPH	PDGTR-	GFHHRV	436
30	QEINK	QPKSTTPSSSGLPSN	LRMPIC	PDGTR-	GFNVRR	168
10	SQSSD	SNPRSRSNSTA-FQNGEN	VRMPRO	PDGSK-	GFRQRP	450
24	PEVNR	QGKPLVAGGNNLPTN	TRMPT	PDGTK-	<b>GF</b> NVRR	162
27	GSSHT	SEENRPGNH	VRQPA	PDGTQ-	<b>GF</b> HHQR	160
85	PRPER	RRLQSLPNDEKGLRLI	LRQPIC	PKSAEY	<b>GF</b> KIER	825
59	ARPER	LISRLRTVSVDDA-GPKLT	LRQPK	PDGSK-	GFSSEA	899
43	QRPER	LISRLRTVSLEDS-GPKLT	VRQPK	PDGTR-	GFSQER	883
37	PGPDR	LVNRLKSITLDDASAPRLM	LRQPK	PDNSK-	<b>gf</b> gaer	678
11	EQEKP	ADRPARINQAEEIRQPNN	/IRQPLC	PDGSQ-	GFRQRR	152
38	RDLKP	VLQASSEEKPDNI	IRQPLO	RANH	GDVPQH	172
47	EQEKP	LDRPTRISQPEDTRQPNN	/IRQPLS	PDGSQ-	<b>gf</b> kqhr	88
09	PRPER	LVKRLKSVTLDDANAPRLI	LRQPRO	PDSTKV	PLRPSA	751
49	PRPDR	LVSRLKSITLDDASAPRLM	LRQPRO	PDNTK-	<b>GF</b> GAER	790
86	PRPDR	LVNRLKSITLDDSNAPRLV	IRQPRO	PDNSK-	GFSVER	827
63	SSPNN	LISNNSINFSGNSNSFNSP	LRQPRO	PDGTK-	<b>GF</b> SDNY	904
40	QRPER	LISRLRTTSLEDT-GPKLT	VRQPRO	PDGTR-	<b>gf</b> sqer	880
06	PRPDR	VNRLKNITLDDASAPRLM	LRQPRO	PDNSM-	<b>GF</b> GAER	847
27	ENPGP	GGHRAFSNAYDDR	SRQPRO	PLPERGQ-	<b>GF</b> GRQR	565
33	SSAGP	GSNRSIGVGFDGQSQERVI	PIRQPRO	PLPEKGP-	GFRRQN	576
90	GTMGP	PAYGGPLGHNSNNN	LRTPAC	PDGSH-	<b>GF</b> AMRR	327
20	KYGIS	VACKARSSDLDN	IRLPK	PDGSK-	GFNILD	155
57	HQAQQ	GQHPHPFNDL <mark>A</mark>	LRLPK	PCPNGSI-	<b>GF</b> QMRR	392
17	QEERP	HLVQQLEELKVEDGVN	VRQPRO	PDGSA-	GFQLQR	156
11	EQEKP	LDRPTRISQPEDSRQPSN	IRQPLO	PDGSQ-	GFKQRR	152
98	EQEKP	LDRPTRISQPEDSRQPNN	IRQPLO	PDGSQ-	GFKQRR	139
73	RQVNA	REQQAGNMPPNTGAGI	GAHPMI	TNQWP-	<b>TL</b> SQTR	512
79	RNVNA	REQHANNGSPNTGAGI	GPHPM	ISA-SQWP-	<b>AL</b> QQNR	619
	p.	<b>1</b>	.R.Pb	PCSS	GF	

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

5

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; ccharged (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  $\alpha$ -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<sup>SUZ</sup> mutation, the protein sequence R138-N139-R140 was changed to G138-N139-G140, and in the SZY-20<sup>SUZ-C</sup> 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- $\alpha$ -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 ( $\alpha$ S20C, 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 ( $\alpha$ S20C, 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  $\mu$ m (A) and 5  $\mu$ m (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,  $10 \mu m$ .



# 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.