Supporting Information

van Roon *et al.* 10.1073/pnas.0802494105

SI Text

Cloning, Expression, and Purification of Rds3p. The Rds3 coding gene was amplified by PCR, using yeast genomic DNA. The Rds3 gene was cloned by using BamHI and HindIII restriction sites into a pRK172 vector, in which hexa-histidine tagged GST and a tobacco etch virus protease cleavage site had been cloned between NdeI and BamHI restriction sites immediately upstream of the protein coding region. The pRK172-GST-Rds3 plasmid was transformed into BL21(DE3)RIL CodonPlus cells (Stratagene). The protein was either expressed in rich medium or in M9 minimal medium supplemented with $^{15}NH_4Cl$ or $^{15}NH_4Cl$ and [$^{13}C_6$]-glucose to obtain ^{15}N or ^{15}N-^{13}C isotope labeled protein, respectively, for NMR experiments. The GST-Rds3p fusion protein was expressed overnight at 15°C after induction with 0.5 mM IPTG. Upon induction the medium was supplemented with 0.1 mM $ZnCl_2$. Harvested cells were resuspended in glutathione-binding buffer [20 mM Tris·HCl (pH 7.4), 500 mM Urea, 500 mM NaCl, 1 μM $ZnSO_4$, and 10 mM β-mercaptoethanol], containing 1 mM phenylmethylsulfonyl fluoride and complete protease inhibitor mixture (Roche) and lysed by sonication. The lysate was cleared by centrifugation, the supernatant loaded on glutathione-Sepharose (GE-Healthcare) equilibrated with glutathione-binding buffer and the protein was eluted by a 0–20 mM linear gradient of glutathione. Peak fractions were incubated overnight with His-tagged tobacco etch virus protease at room temperature while dialysing against Ni-NTA binding buffer [20 mM Tris·HCl (pH 7.4), 500 mM Urea, 500 mM NaCl, 25 mM imidazole, 1 μM $ZnSO_4$ and 10 mM β-mercaptoethanol]. After complete cleavage the sample was loaded on Ni-NTA agarose (Qiagen) to remove His-tagged tobacco etch virus protease, His-tagged GST and minor protein contaminants. Finally, Rds3p was dialyzed against Q-buffer [20 mM Tris·HCl (pH 8.0), 200 mM NaCl, 1 μM $ZnSO_4$, 10 mM β-mercaptoethanol], loaded onto high-trap Q-Sepharose (GE-Healthcare) and collected in the flowthrough separated from other remaining contaminants. Rds3p was concentrated and buffer exchanged by using an Amicon-Ultra-15 concentrator (Millipore) with an exclusion size of 5 kDa.

Analytical gelfiltration was performed on a Superdex S200 HR 10/300 column with 20 mM Tris·HCl (pH 7.4), 200 mM NaCl and 1 mM DTT as eluent.

NMR Experiments. NMR samples comprised 0.3–0.4 mM (^{15}N-labeled) or 0.6–0.8 mM (^{15}N, ^{13}C-labeled) solutions of Rds3p (with additional vector-derived residues Gly-Gly-Ser- at the N terminus in place of the native residue Met 1) in NMR buffer [20 mM Tris·HCl (pH 7.0), 1 mM [$^2H_6$] DTT, 200 mM NaCl]. Data were acquired at 27°C on Bruker Avance 800, DMX600 and DRX500 spectrometers. Resonances were assigned by using a standard suite of triple resonance NMR experiments, and ^1H, ^{15}N and ^{13}C chemical shifts were calibrated by using sodium 3,3,3-trimethylsilylpropionate (TSP) as an external ^1H reference (1).

For ^{15}N-labeled protein samples, the following data were acquired: 2D datasets: [^{15}N-^1H] HSQC; pseudo3D datasets: ^{15}N T_1, ^{15}N T_2, and ^{15}N-^1H heteronuclear NOE. For ^{15}N/^{13}C-labeled protein samples, the following data were acquired: 2D: [^{15}N-^1H] HSQC, [^{13}C-^1H] HSQC covering the full ^{13}C spectral width, constant-time [^{13}C-^1H] HSQC covering only the aliphatic ^{13}C region, constant-time [^{13}C-^1H] HSQC covering only the aromatic ^{13}C region, [^1H-^1H] NOESY experiments (without heteronuclear filtering; τ_m = 50 and 150 ms), [^1H-^1H] NOESY experiments filtered to remove ^{15}N-coupled signals in F_2 only (τ_m = 150 ms); 3D datasets: CBCANH, CBCACONH, HNCO, HBHANH, HBHACONH, [^1H-^{13}C-^1H] HCCH-TOCSY, [^{13}C-^{13}C-^1H] HCCH-TOCSY, HNHB, HACAHB-COSY, ^{15}N NOESY-HSQC (τ_m = 150 ms), ^{13}C NOESY-HSQC (τ_m = 150 ms), separate datasets acquired for ^{13}C aliphatic and aromatic spectral regions.

Samples for ^{113}Cd-Zn exchange experiments were prepared by adding a fivefold excess of ≈30 mM ^{113}Cd-EDTA solution to a 0.3 mM solution of ^{15}N-labeled Rds3p in NMR buffer.

Structure Calculations. NOE distance restraints were derived from analysis of all of the data from NOE-based experiments. Cross peaks were picked, integrated, and chemical-shift matched by using the program Analysis developed by the CCPN Project (2). Stereospecific assignments for 32 $C^\beta H_2$ groups were made by analyzing HNHB and HACAHB-COSY spectra (3).

Structures were calculated by using ARIA 1.1.2 (4), which makes use of CNS 1.1 (5) for simulated annealing calculations. Initial calculations made use of ARIA in an unmodified form to derive the structure of the protein without defining any zinc atoms or using any interligand constraints, and without including any NOE cross-peak assignments in the input data. These calculations established the binding connectivity for the zinc ions unambiguously as shown in Fig. S3 and Table S1 and gave the same knotted fold as the later calculations including zinc.

Once the binding sites for the zinc atoms were identified from these initial structures, the ARIA code was modified to include zinc atoms in the simulated annealing protocol and unambiguous distance and dihedral angle constraints were added to define the geometry of the zinc binding sites. These modifications followed an essentially similar protocol to that used for disulphide bridges in ARIA conventionally. Each time that a stage of torsion angle dynamics is initiated, three of four of the Zn-ligand bonds are removed (so as to leave the zinc attached to the protein but to eliminate all rings) and subsequently each time Cartesian dynamics or minimization is initiated these three zinc-ligand bonds are reformed. In addition, the number of structures calculated in iteration eight of the ARIA protocol was increased from 20 to 100, and the number in the water refinement stage was increased from 10 to 50. Of the 50 water refined structures, the 20 lowest-energy structures were included in the final ensemble for deposition.

Additional independent calculations to verify the fold were carried out by using manual assignment of NOE cross peaks or using the semiautomatic program ATNOS/CANDID (6, 7). In each case the data converged well on the knotted fold even in early cycles of refinement, showing that the fold was well determined by the data; for the 10 cycles of ARIA calculations (cycles 0–8 followed by water refinement) the numbers of accepted conformers that had the correct knotted fold in successive cycles [as determined by using the protein knot server at http://knots.mit.edu (8)] were 5/10, 9/10, 9/10, 9/10, 9/10, 10/10, 10/10, 10/10, 49/50, and 20/20 respectively, whereas for the 7 cycles of ATNOS/CANDID calculations the corresponding numbers were 9/10, 10/10, 20/20, 20/20, 20/20, 20/20, and 20/20. The few outliers differed from the correct structure by having different chain crossings, but the nature of these differences followed no pattern across different outliers.

Structures were visualized by using the program pymol (www.pymol.org), and electrostatic surfaces were calculated by using the programs pdb2pqr (9) (slightly modified to allow

inclusion of the zinc ions) and APBS (10), setting the contour potentials at ±15kT/e.

Splicing Assay. Splicing extracts were prepared as described in ref. 11 from a *GAL1*::*rds3–1* strain grown either at 30°C with galactose or at 37°C with glucose to metabolically deplete Rds3p (12). One hundred nanograms of TAP-purified SF3b complex (13), 300 ng of Rds3p or 1.5 μl of buffer D was incubated with the inactivated extract at 4°C for 1 h before addition of *in vitro* transcribed radioactively labeled actin pre-mRNA and ATP. Extract from yeast grown at 30°C was incubated similarly with buffer D, SF3b complex or Rds3p as a control. The pre-mRNA and spliced products were resolved on a 6.5% acrylamide gel with 8 M urea run at 22 W for 65 min at room temperature. The gels were visualized with a Typhoon phosphorimager (Amersham Pharmacia).

Electromobility Shift Assay. Rds3p (20–250 μM) was incubated with various DNA and RNA constructs (2–5 μM) in binding buffer [20 mM Hepes (pH 7.4), 150 mM NaCl, 10 mM β-mercaptoethanol] with and without 5 mM $MgCl_2$ for 30 min at room temperature. The reaction mixtures were resolved on a 6% acrylamide gel run for 1 h at 90 V in a cold room (4°C) and a 1% agarose gel run for 45 min at 25 mA in a cold room in Tris·borate buffer and stained with toluidine blue and ethidium bromide, respectively.

1. Wishart DS, *et al.* (1995) 1H, 13C and 15N chemical shift referencing in biomolecular NMR. *J Biomol NMR* 6:135–140.
2. Vranken WF, *et al.* (2005) The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins* 59:687–696.
3. Grzesiek S, Kuboniwa H, Hinck AP, Bax A (1995) Multiple-quantum line narrowing for measurement of H-alpha-H-beta J-couplings in isotopically enriched proteins. *J Am Chem Soc* 117:5312–5315.
4. Linge JP, O'Donoghue SI, Nilges M (2001) Automated assignment of ambiguous nuclear overhauser effects with ARIA. *Methods Enzymol* 339:71–90.
5. Brünger AT, *et al.* (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 54:905–921.
6. Herrmann T, Güntert P, Wüthrich K (2002) Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. *J Biomol NMR* 24:171–189.
7. Herrmann T, Güntert P, Wüthrich K (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J Mol Biol* 319:209–227.
8. Kolesov G, Virnau P, Kardar M, Mirny LA (2007) Protein knot server: Detection of knots in protein structures. *Nucleic Acids Res* 35:W425–W428.
9. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA (2004) PDB2PQR: An automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res* 32:W665–W667.
10. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc Natl Acad Sci USA* 98:10037–10041.
11. Lin RJ, Newman AJ, Cheng SC, Abelson J (1985) Yeast mRNA splicing in vitro. *J Biol Chem* 260:14780–14792.
12. Wang Q, Rymond BC (2003) Rds3p is required for stable U2 snRNP recruitment to the splicing apparatus. *Mol Cell Biol* 23:7339–7349.
13. Dziembowski A, *et al.* (2004) Proteomic analysis identifies a new complex required for nuclear pre-mRNA retention and splicing. *EMBO J* 23:4847–4856.

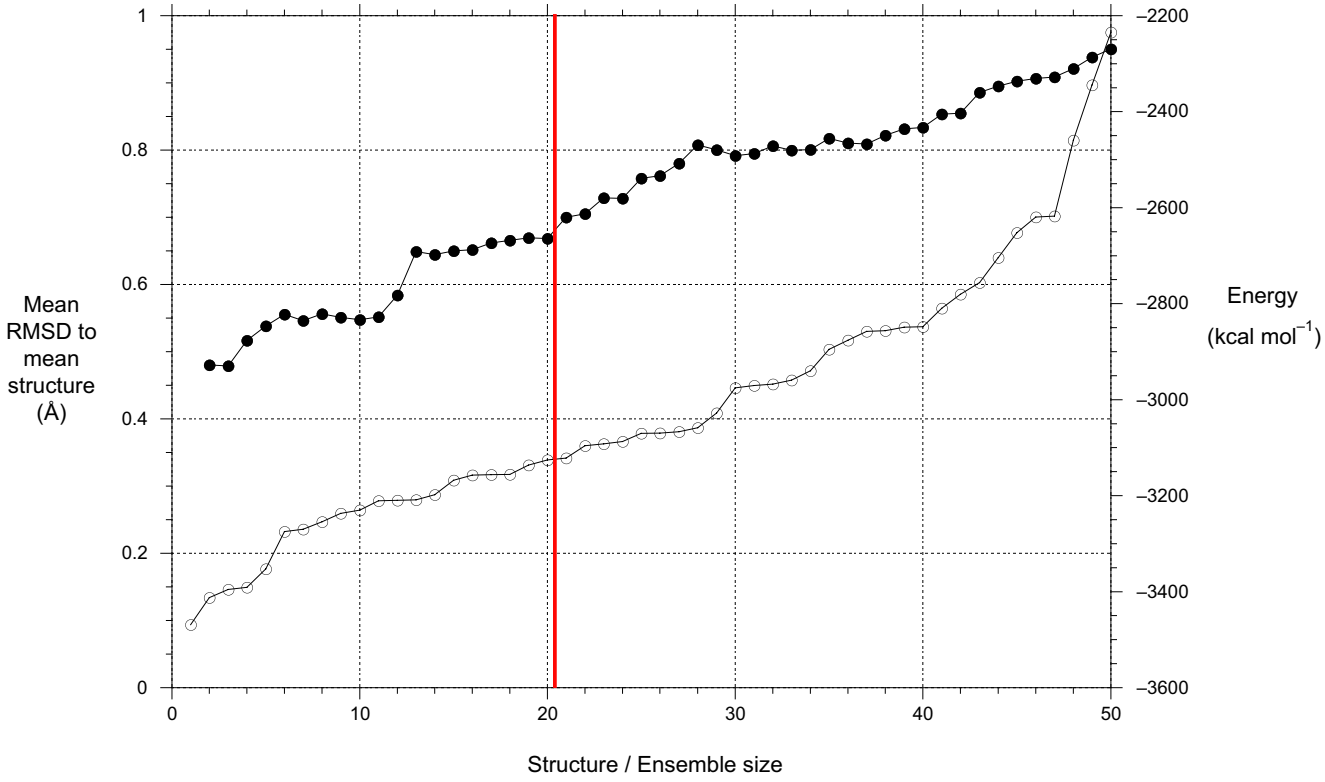

Fig. S1. RMSD and energy profiles of structures from the water refinement stage of the Aria protocol. RMSD values (filled circles) are independently calculated for each ensemble size, using the program CLUSTERPOSE [Diamond R (1995) Coordinate-based cluster analysis. *Acta Crystallogr D* 51:127–135], adding successive structures in order of increasing CNS total energy. Open circles represent the CNS total energy. Only structures to the left of the vertical red line were included when calculating the structural statistics.

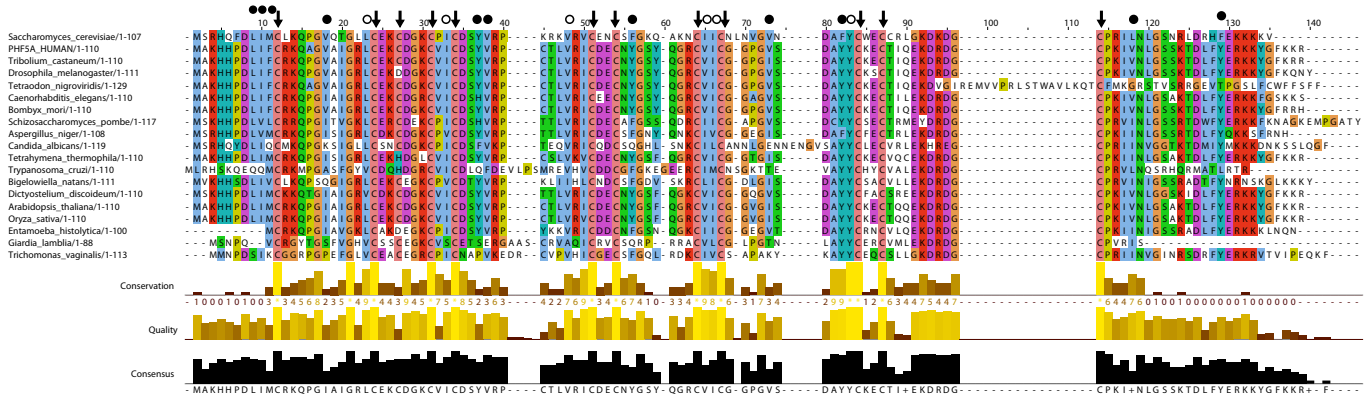

Fig. S2. Sequence alignment of Rds3p. Rds3p is highly conserved throughout all eukaryotic species from yeast to plants, insects and vertebrates. The human sequence also represents *Rattus norvegicus*, *Mus musculus*, *Danio rerio*, *Bos Taurus*, *Xenopus laevis*, *Gallus gallus*, *Canis familiaris*, *Pan troglodytes*, *Monodelphis domestica*, and *Equus caballus*, which all have 100% sequence identity. The zinc binding cysteines (arrows) are highly conserved in all species. In a few species the second cysteine of the first CXXC motif is replaced by aspartic acid or histidine, but these residues can also coordinate zinc, so the zinc finger is presumably not disrupted (*cf.* refs. 1 and 2). The buried hydrophobic residues (open circles) stabilizing the core of the protein between the β-triangle and the zinc fingers are very highly conserved also. The closed circles denote the conserved surface exposed hydrophobic residues. The multiple sequence alignment was generated with ClustalW (3).

1. Lipski BP, Beals CR, Staunton DE (1998) Leupaxin is a novel LIM domain protein that forms a complex with PYK2. *J Biol Chem* 273:11709–11713.
2. Michelsen JW, *et al*. (1994) Mutational analysis of the metal sites in LIM domain. *J Biol Chem* 269:11108–11113.
3. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680.

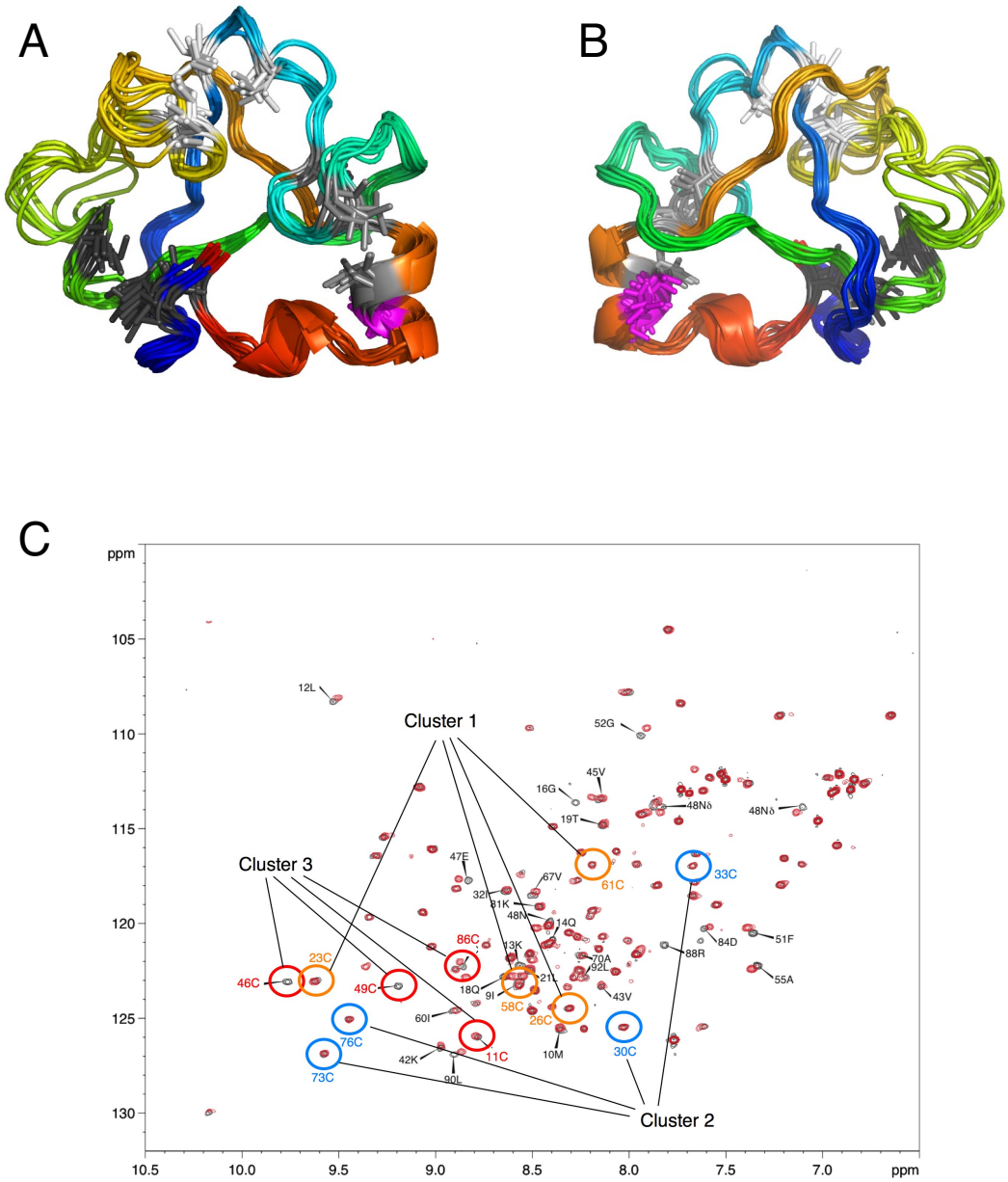

Fig. S3. Zinc cysteine connectivities. (*A* and *B*) Two views (related by a 180° rotation) of the 10 lowest energy structures from an ensemble of 20 calculated with no zinc-ligand or inter-ligand constraints, and no NOE cross-peak assignments in the input data. The smoothed protein backbone is shown in chainbow colouring from residue Met10 (blue) to residue Leu90 (red), and the cysteines involved in metal clusters are colored white for cluster 1 (Cys23, 26, 58 and 61), light grey for cluster 2 (Cys30, 33, 73 and 76) and dark grey for cluster 3 (Cys11, 46, 49 and 86). The clusters are most clearly seen in *A*; *B* shows that Cys77 (magenta), although sequentially adjacent to cluster 2, projects away from it. (*C*) Superposition of [^{15}N,^{1}H] HSQC spectra of normal Rds3p (containing zinc; black spectrum) and 1 h after addition of a fivefold excess of ^{113}Cd EDTA complex (red spectrum). Backbone amide correlations corresponding to the metal-binding cysteins are indicated with orange, blue and red circles for clusters 1, 2, and 3, respectively. Only for the peaks of cluster 3 are there significant chemical shift changes or disappearances, strongly suggesting that it is undergoing metal exchange whereas clusters 1 and 2 remain substantially unaffected. Some further assignments for peaks that shift are indicated in black; almost all are close in space to cluster 3. At later times, the protein partially precipitated and the spectra became somewhat more complicated and much less intense, rendering the intended application of [^{113}Cd, ^{1}H] correlation experiments essentially impossible. The final calculations, in which explicit zinc binding connectivities and geometrical constraints are included, show that, in common with other GATA type zinc-fingers, fingers 1 and 2 of Rds3p have S absolute chirality as defined in terms of sequence priority amongst the zinc-binding residues [see Berg J (1988) Proposed structure for the zinc-binding domains from transcription factor IIIA and related proteins. *Proc Natl Acad Sci USA* 85:99–102]. The formal chirality of finger 3 is R, but this apparent discrepancy arises only because of the chain interruption within the zinc knuckle; were the chain within the knuckle to be continuous then the sequential order of the two cysteines in it would be reversed, restoring the expected S chirality. Another small difference for finger 3 concerns the pattern of rotamers found amongst the zinc-binding cysteines. In common with other regular GATA-type zinc fingers, the cysteines of fingers 1 and 2 both follow the pattern t, g+, g+, g- for the two cysteines of the knuckle and the helix respectively (i.e. their χ^1 angles are approximately 180°, +60°, +60° and −60°). However, the corresponding cysteines of finger 3 show the pattern t, t, g+, g-; again, this difference almost certainly arises because of the interruption of the mainchain within the knuckle.

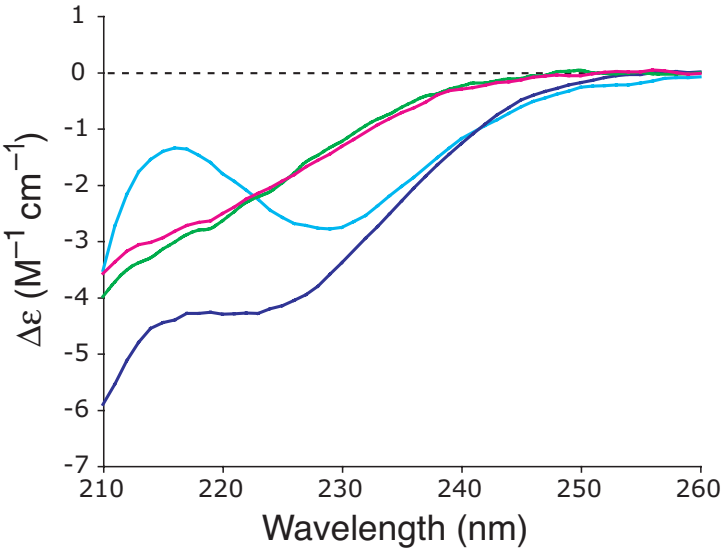

Fig. S4. Overlay of far UV CD spectra of native Rds3p at 20°C (light blue), after heating to 95°C (dark blue), after addition of 10 mM EDTA followed by dialysis to remove the EDTA (green) and after addition of five equivalents of $ZnSO_4$ to the dialyzed sample (pink). The spectrum of Rds3p after removal of the zinc is clearly different from the spectra of the native or heated samples, and addition of $ZnSO_4$ to the EDTA-treated sample did not restore the native spectrum of Rds3p.

Table S1. Cysteine distance matrix

Cβ-Cβ distance	23	26	58	61	30	33	73	76	11	46	49	86
23		**5.68** ***1.02***	**4.41** ***0.54***	**6.85** ***2.23***	13.74 *1.00*	17.47 *1.41*	17.45 *1.83*	19.72 *1.97*	18.28 *1.96*	19.48 *0.96*	15.69 *1.37*	17.18 *0.91*
26	**5.34** ***0.23***		**6.63** ***1.07***	**6.98** ***1.29***	9.43 *1.02*	12.85 *1.23*	13.56 *1.32*	16.08 *1.72*	16.95 *2.85*	19.13 *1.49*	16.12 *1.55*	15.96 *1.66*
58	**4.03** ***0.42***	**5.93** ***0.45***		**5.92** ***1.53***	12.39 *1.01*	15.97 *1.25*	15.69 *2.15*	17.31 *2.20*	14.55 *2.19*	15.28 *1.22*	11.48 *1.51*	13.14 *1.03*
61	**5.48** ***0.19***	**6.73** ***0.27***	**3.82** ***0.25***		12.48 *1.79*	14.41 *2.69*	15.73 *2.04*	17.38 *2.40*	14.68 *2.66*	17.27 *1.57*	13.63 *2.02*	14.30 *1.45*
30	14.00 *0.27*	9.20 *0.28*	12.55 *0.18*	13.55 *0.30*		**5.94** ***0.83***	**4.99** ***1.49***	**7.87** ***1.42***	15.30 *2.35*	17.39 *1.81*	16.56 *2.43*	13.61 *1.68*
33	17.21 *0.20*	12.22 *0.23*	15.64 *0.23*	15.62 *0.20*	**4.78** ***0.13***		**6.77** ***2.48***	**7.85** ***1.95***	16.05 *2.39*	18.89 *2.76*	18.53 *3.18*	14.92 *2.08*
73	18.22 *0.30*	13.35 *0.36*	17.05 *0.22*	18.23 *0.33*	**4.74** ***0.11***	**5.98** ***0.24***		**4.40** ***1.39***	16.01 *3.09*	17.90 *2.87*	17.97 *3.69*	14.23 *2.55*
76	20.03 *0.36*	15.43 *0.36*	18.09 *0.27*	19.11 *0.46*	**6.27** ***0.15***	**5.97** ***0.41***	**4.14** ***0.10***		14.74 *2.59*	16.44 *2.29*	17.52 *2.86*	12.98 *1.47*
11	19.43 *0.60*	18.43 *0.32*	15.47 *0.38*	16.23 *0.68*	15.96 *0.24*	16.97 *0.20*	18.85 *0.34*	16.39 *0.44*		**7.53** ***2.18***	**8.36** ***3.19***	**5.30** ***1.28***
46	20.10 *0.46*	19.56 *0.23*	16.34 *0.34*	18.17 *0.44*	17.16 *0.21*	19.25 *0.30*	19.56 *0.40*	17.42 *0.41*	**5.27** ***0.44***		**5.27** ***1.09***	**4.89** ***1.07***
49	16.49 *0.64*	16.76 *0.39*	12.60 *0.53*	14.08 *0.57*	16.55 *0.29*	18.75 *0.35*	19.96 *0.36*	18.42 *0.45*	**5.23** ***0.66***	**5.02** ***0.29***		**6.39** ***1.28***
86	18.08 *0.37*	16.65 *0.16*	14.39 *0.24*	16.07 *0.42*	13.27 *0.17*	15.23 *0.22*	15.69 *0.34*	13.53 *0.41*	**4.76** ***0.45***	**4.05** ***0.19***	**5.94** ***0.20***	

Distance matrix for Cys Cβ atoms in structures calculated without (top right) or with (bottom left) all zinc binding and inter-ligand distance constraints. For structures calculated without zinc binding or inter-ligand constraints, no prior NOE cross-peak assignment information at all was included in the input data. Distances are grouped to show the clusters that were identified (shown in bold); even when calculated with no inter-ligand or metal binding constraints at all, distances between Cys Cβ atoms within different clusters are all significantly longer than any distance within a single cluster, thereby establishing the metal-binding connectivity pattern unambiguously. In each case, the results show the mean Cys Cβ–Cβ distances (roman) and standard deviations (italic) for the seven lowest-energy structures in an ensemble of 20.