Supporting Information

Lin et al. 10.1073/pnas.0813255106

```
       M  V  W  L   T  S  I   A  F  I   V  T  L  L   G  A  Q   Y  D  I
   1   ATGGTTTGGC TGACGAGTAT CGCGTTTATT GTGACTCTTT TAGGAGCACA ATACGACATA
       V  T  A  Q   A  I  S   V  A  T   S  V  P  S   V  F  S   S  P  S
  61   GTGACTGCGC AGGCAATTTC AGTTGCAACT TCTGTCCCAT CAGTGTTCAG TAGCCCTAGC
       L  A  S  G   F  L  G   C  L  T   T  G  I  G   Q  S  P   D  F  P
 121   CTTGCAAGTG GTTTCCTTGG ATGTCTCACA ACTGGTATTG GACAATCTCC AGATTTCCCG
       F  Q  E  Q   Q  D  L   D  D  L   A  Q  V  I   L  S  A   V  T  S
 181   TTTCAAGAAC AACAAGATTT AGATGACTTA GCACAAGTAA TTCTTTCGGC AGTAACCAGT
       N  T  D  T   S  K  S   A  R  A   Q  A  L  S   T  A  L   A  S  S
 241   AATACTGACA CCTCAAAGTC AGCGAGAGCA CAAGCCTTGA GCACTGCATT AGCATCTTCC
       L  A  D  L   L  I  S   E  S  S   G  S  S  Y   Q  T  Q   I  S  A
 301   TTAGCCGACC TACTGATTTC CGAATCAAGT GGAAGCAGCT ACCAAACTCA AATATCTGCC
       L  T  N  I   L  S  D   C  F  V   T  T  T  G   S  N  N   P  A  F
 361   CTCACTAATA TCCTATCCGA TTGTTTTGTC ACAACAACTG GATCAAACAA TCCTGCATTT
       V  S  R  V   Q  T  L   I  A  V   L  S  Q  S   S  S  N   A  I  S
 421   GTATCAAGAG TTCAAACACT TATAGCAGTG CTTTCTCAAA GCAGCAGTAA TGCAATTTCA
       G  A  T  G   G  S  A   F  A  Q   S  Q  A  F   Q  Q  S   A  S  Q
 481   GGCGCAACAG GTGGCTCCGC ATTCGCACAA AGTCAGGCGT TCCAACAAAG CGCATCACAA
       S  A  G  L   S  A  S   R  A  G   S  T  S  S   S  T  T   T  T  T
 541   AGTGCTGGCC TAAGTGCTTC CCGAGCAGGA TCCACATCCT CTTCCACCAC TACCACCACC
       S  G  A  T   S  Q  A   A  S  Q   S  A  S  S   S  Y  S   S  A  F
 601   TCGGGAGCAA CAAGTCAAGC AGCAAGCCAG AGTGCAAGCA GTTCTTATAG CTCTGCATTT
       A  Q  A  A   S  S  A   L  A  T   S  S  A  I   S  R  A   F  A  S
 661   GCACAAGCGG CCTCTTCTGC CCTTGCGACC TCCAGTGCTA TCAGCAGAGC CTTCGCTTCT
       V  S  S  A   S  A  A   S  S  L   A  Y  N  I   G  L  S   A  A  R
 721   GTCAGTTCCG CCTCTGCCGC CTCCAGCCTT GCTTATAACA TAGGCTTATC CGCCGCACGG
       S  L  G  I   A  S  D   T  A  L   A  G  A  L   A  Q  A   V  G  G
 781   TCCCTCGGAA TAGCCTCTGA CACAGCCCTC GCTGGTGCCT TAGCTCAAGC TGTGGGTGGA
       V  G  A  G   A  S  A   S  A  Y   A  N  A  I   A  R  A   A  G  Q
 841   GTAGGCGCGG GAGCCTCTGC ATCTGCCTAC GCTAATGCTA TTGCACGTGC CGCTGGACAA
       F  L  A  T   Q  G  V   L  N  A   G  N  A  S   A  L  A   G  S  F
 901   TTCTTAGCCA CTCAGGGTGT TTTGAACGCA GGCAATGCAT CTGCCCTAGC AGGCAGCTTC
       A  R  A  L   S  A  S   A  E  S   Q  S  F  A   Q  S  Q   A  Y  Q
 961   GCCAGAGCCC TCTCGGCCTC AGCAGAATCC CAGTCATTCG CACAGAGTCA AGCCTACCAA
       Q  A  S
1021   CAGGCATCG
```

Fig. S1. N-terminal cDNA sequence of TuSp1. The translated amino acids are indicated above the DNA sequence. The predicted signal peptide is highlighted in bold. NTD (solid line) and RP1 (dashed line) are underlined.

```
RP1  SASRAGSTSSSTTTTTSGATSQAASQSASSSYSSAFAQAASSALATSSAISRAFASVSSA 60
RP2  SASRAGSTSSSTTTTTSGATSQAASQSASSSYSSAFAQAASSSLATSSAISRAFASVSSA 60
NTD  QAISVATSVPSVFSSPSLASGFLGCLTTGIGQSPDFPFQEQQDLDDLAQVILSAVTSNTD 60
CTD  ---GISVGVPGYLRTPS-----------STILAPSNAQIISLGLQTTLAPVLSSSGLSSA 46
          .   ..   :.*           .   :.   .  .  *        :     .:

RP1  SAASSLAYNIGLSAARSLGIASDTALAGALAQAVGGVGAGASASAYANAIARAAGQFLAT 120
RP2  SAASSLAYNIGLSAARSLGIASDTALAGALAQAVGGVGAGASASAYANAIARAAGQFLAT 120
NTD  TSKS--------ARAQALSTALASSLADLLISESSGSSYQTQISALTNILS---DCFVTT 109
CTD  SASAR-----------------VSSLAQSLASALSTSRGTLSLSTFLNLLSSISSEIRAS 89
     :: :                   ::**  * .  .      . *:  * ::   . : ::

RP1  QGVL-NAGNASALAGSFARALSASAESQSFAQSQAYQQASA-FQQAAAQSAAQ- 171
RP2  QGVL-NAVNASSLGSALANALSDSAANS--AVSGNYLGVSQNFGRIAPVTGGTA 171
NTD  TGSN-NPAFVSRVQTLIAVLSQSSSNAISGATGGSAFAQSQAFQQSASQSAGL- 161
CTD  TSLDGTQATVEVLLEALAALLQVINGAQITDVNVSSVPSVNAALVSALVA---- 139
      .   .   .. :   :*   .          .             *  :
```

Fig. S2. Internal aa sequence alignment of TuSp1 domains. The sequence identity among NTD, RP1/RP2, and CTD ranges from 11% to 19%, whereas RP1 and RP2 share 78% sequence identity. The structural regions are underlined and highlighted in bold.

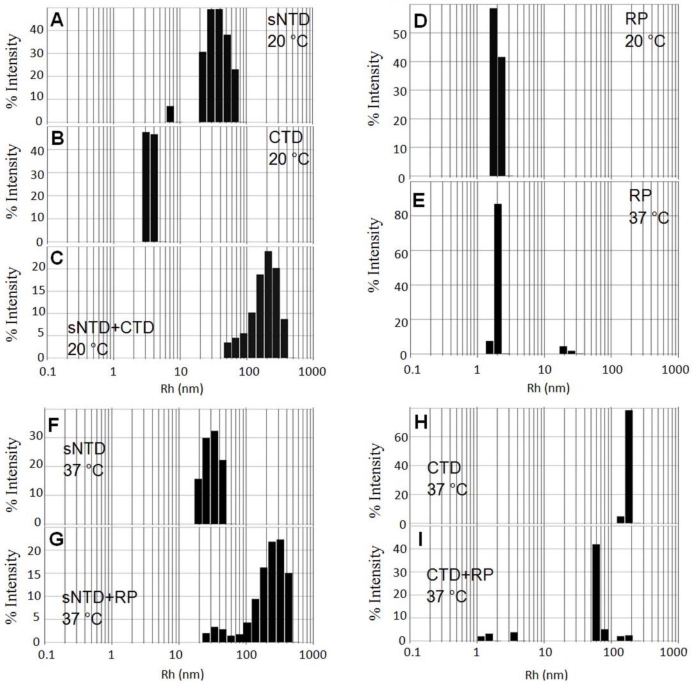

Fig. S3. DLS of TuSp1 domains. (*A*–*C*) terminal domains at 20 °C: sNTD (*A*), CTD (*B*), sNTD + CTD (*C*). (*D* and *E*) RP domain (RP1/RP2) at 20 °C (*D*) and at 37 °C (*E*). RP1 and RP2 gave rise to the same results, and only RP1 is shown here. (*F* and *G*) sNTD and RP domain at 37 °C. (*F*) sNTD. (*G*) sNTD + RP. (*H* and *I*) CTD and RP domain at 37 °C. (*H*) CTD. (*I*) CTD + RP. Each protein concentration was adjusted to ≈50 μM in 10 mM Tris buffer (pH 7.0).

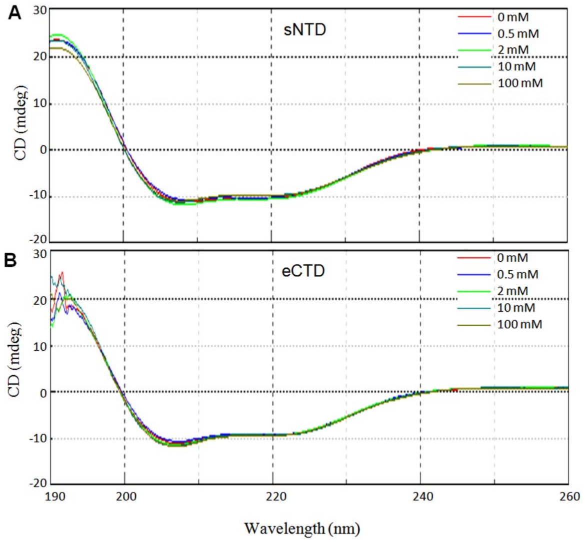

Fig. S4. CD spectra of sNTD (*A*) and eCTD (*B*) of TuSp1 in a series of DPC concentrations at 20 °C. The sNTD/eCTD concentration was $\approx$10 μM, and DPC concentration varied from 0 mM to 100 mM in 10 mM phosphate (pH 6.5).

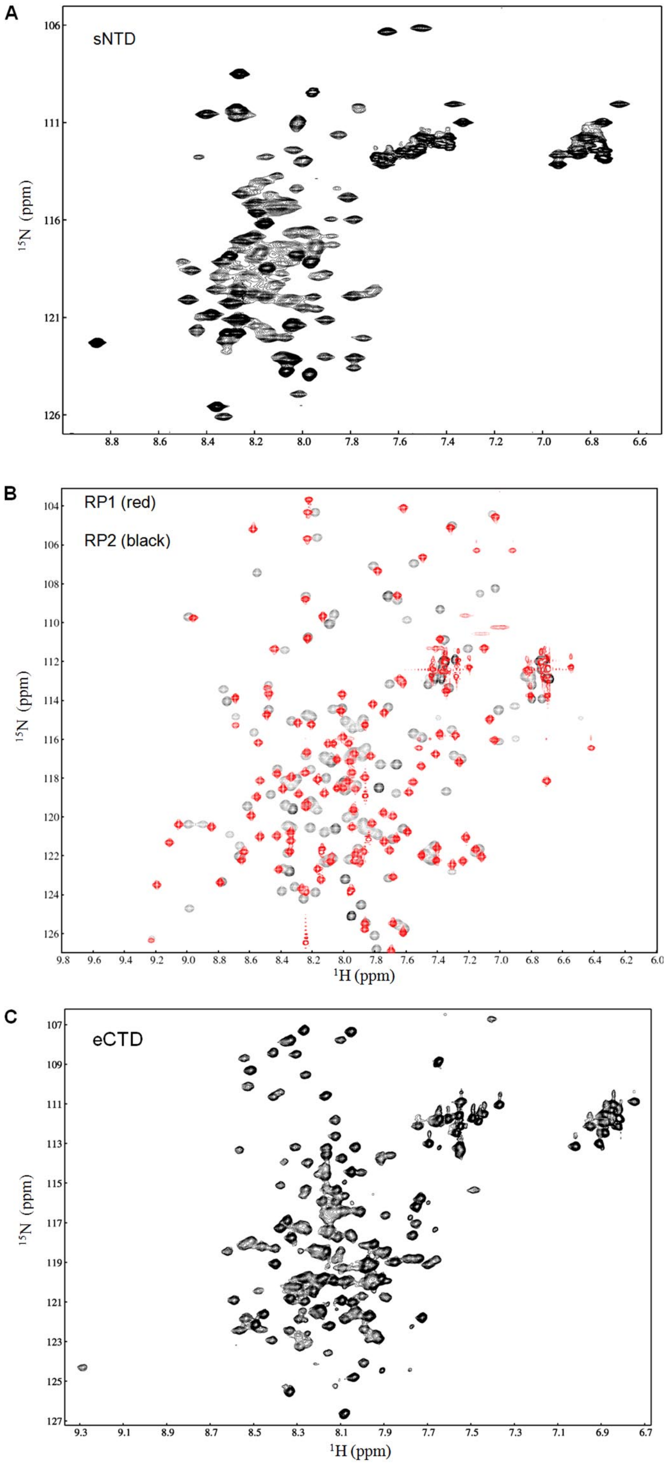

Fig. S5. [^{1}H,^{15}N]HSQC spectra of TuSp1 domains. (*A*) Spectrum of sNTD acquired at 500 MHz, 30 °C, pH 6.0, and 5 mM phosphate, 100 mM DPC. (*B*) Overlay of RP1 (red) and RP2 (black) spectra recorded at 800 MHz, 17 °C, pH 6.5, 50 mM phosphate, and 100 mM NaCl. (*C*) Spectrum of eCTD acquired at 800 MHz, 35 °C, pH 6.5, 10 mM phosphate, 100 mM DPC. For all of the samples, the protein concentration was ≈1 mM.

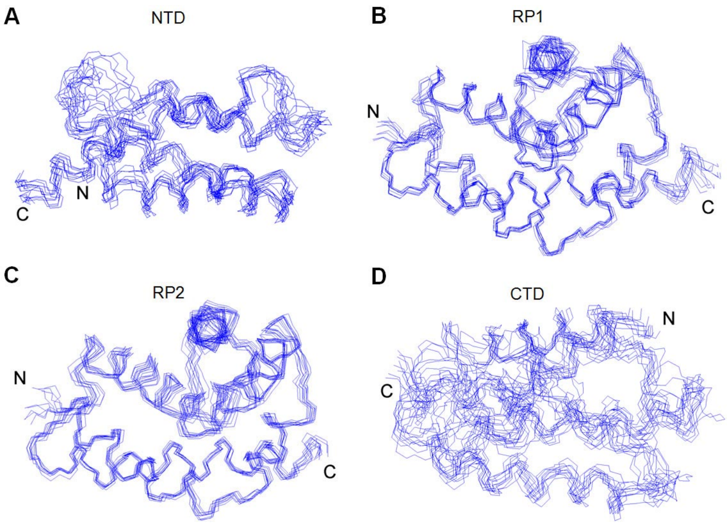

Fig. S6. Overlay of the 10 lowest energy structures of NTD (*A*), RP1 (*B*), RP2 (*C*), and CTD (*D*). Unstructured N- and C-terminal regions are not shown.

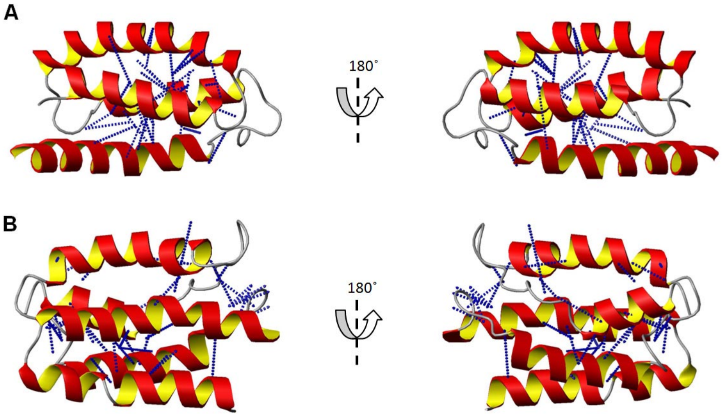

Fig. S7. Detected long-range NOEs (blue dashed lines) in the structural region of NTD (*A*) and CTD (*B*).

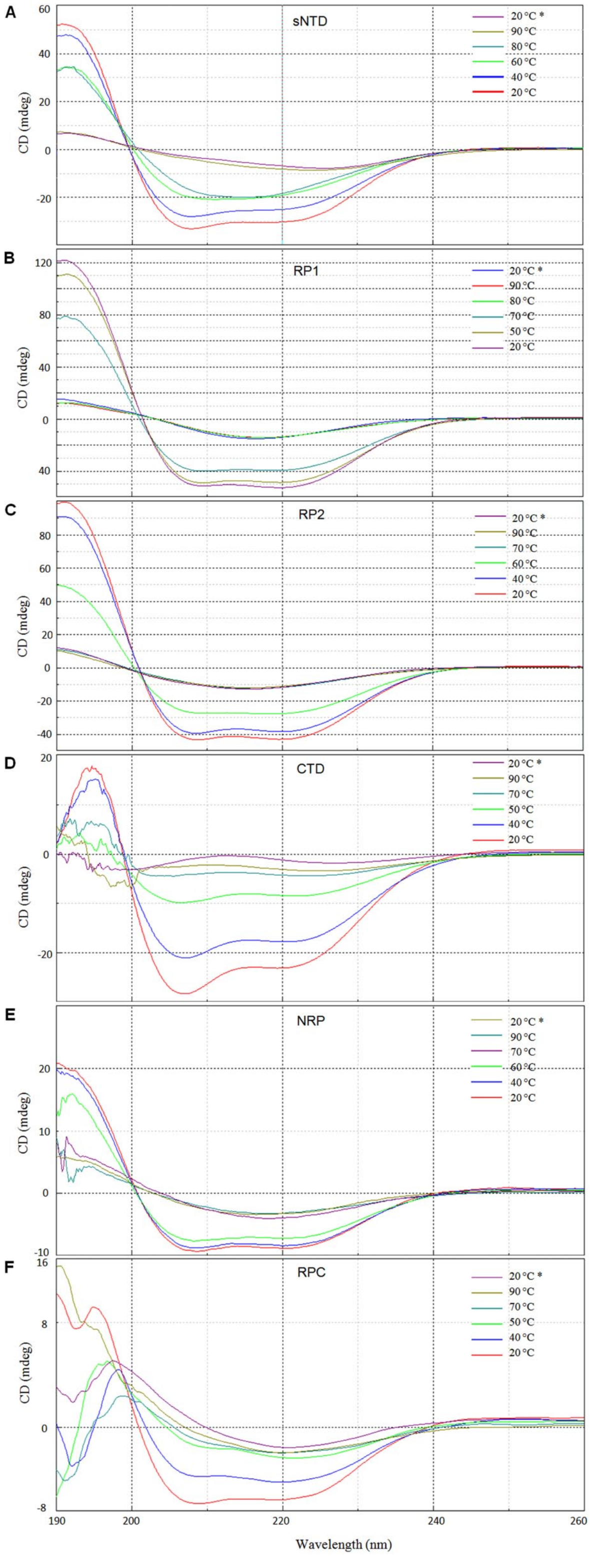

Fig. S8. Temperature-dependent structural transition of TuSp1 domains. (*A*) sNTD. (*B*) RP1. (*C*) RP2. (*D*) CTD. (*E*) NRP. (*F*) RPC. *, CD spectra recorded at 20 °C after cooling down from 90 °C. The protein concentration of each domain was ≈40 μM in 10 mM phosphate buffer (pH 6.5). The protein concentration of NRP/RPC was ≈10 μM in 5 mM Tris buffer (pH 7.0).

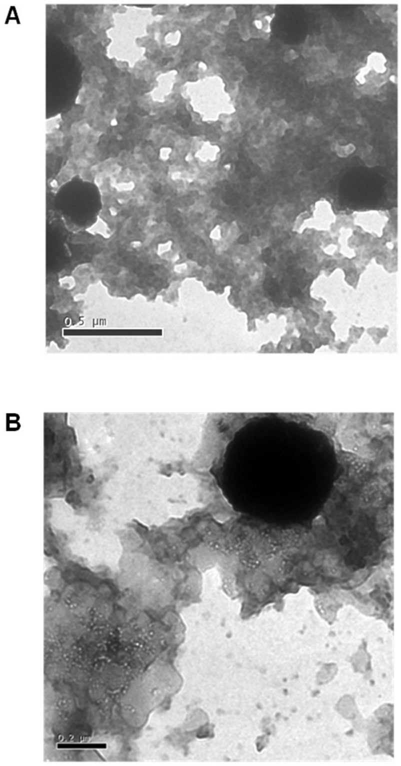

Fig. S9. TEMs of stained sNTD (*A*) and CTD (*B*), illustrating irregular morphology. (Scale bars: (*A*) 500 nm, (*B*) 200 nm.

Table S1. Statistics of pairwise structural comparison of TuSp1 structural domains

Parameters	NTD–RP1	NTD–CTD	RP1–RP2	RP2–CTD
Equivalence residues	82	92	115	96
Structural overlay, %	68	79	95	82
SSAP score	70.81	61.65	90	64.90
Sequence identity, %	9	10	90.21	9
Rms deviation	4.38	10.7	1.58	11.67