SUPPLEMENTARY FILES

The structure of transcription termination factor Nrd1 reveals an original mode for GUAA recognition

Elsa Franco-Echevarría, Noelia González-Polo, Silvia Zorrilla, Santiago Martínez-Lumbreras, Clara M Santiveri, Ramón Campos-Olivas, Mar Sánchez, Olga Calvo, Beatriz González*, José Manuel Pérez-Cañadillas*.

TABLE OF CONTENTS

Supplementary Table S1: Plasmids and protein constructs used in this work. Keywords: NMR=Nuclear Magnetic Resonance, X-ray= X-ray crystallography, F.A.= Fluorescence anisotropy, ITC= Isothermal titration calorimetry, Fun. stu.= Functional studies in *S. cerevisiae* (drop tests and northern block). pET28_txAHTEV is a pET28-derived plasmid in which the 6xHis-Trombin_site has been replaced by trxA-6xHis-TEVsite

Plasmid/vector	Expression host/strain	Use	Source
pET28-txAHTEV-Nrd1 301-489	*E. coli* BL21(DE3)	X-ray; NMR:F.A.	This work
pET28-txAHTEV-Nrd1 265-489	*E. coli* BL21(DE3)	NMR	This work
pET28-txAHTEV-Nrd1 290-468	*E. coli* BL21(DE3)	X-ray; NMR;F.A., ITC.	This work
pRS316-*NRD1*	*Mat a ura3-52, leu2-3,112, trp1-1 , his3-11,15, ade2-1, met2Δ1, lys2Δ2, can1-100, nrd1ΔHIS3 [pRS316-NRD1]*	Fun. stu.	(1-3)
pRS415-*NRD1*	*Mat a ura3-52, leu2-3,112, trp1-1 , his3-11,15, ade2-1, met2Δ1, lys2Δ2, can1-100, nrd1ΔHIS3 [pRS415-NRD1]*	Fun. stu.	This work
pRS415-nrd1−Δ301-306	*Mat a ura3-52, leu2-3,112, trp1-1 , his3-11,15, ade2-1, met2Δ1, lys2Δ2, can1-100, nrd1ΔHIS3*	Fun. stu.	This work
pRS415-nrd1−Δ301-336	*Mat a ura3-52, leu2-3,112, trp1-1 , his3-11,15, ade2-1, met2Δ1, lys2Δ2, can1-100, nrd1ΔHIS3*	Fun. stu.	This work
pRS415-nrd1−Δ412-463	*Mat a ura3-52, leu2-3,112, trp1-1 , his3-11,15, ade2-1, met2Δ1, lys2Δ2, can1-100, nrd1ΔHIS3*	Fun. stu.	This work
pRS415-nrd1−Δ301-336,Δ412-463	*Mat a ura3-52, leu2-3,112, trp1-1 , his3-11,15, ade2-1, met2Δ1, lys2Δ2, can1-100, nrd1ΔHIS3*	Fun. stu.	This work
pRS415-nrd1-V467stop	*Mat a ura3-52, leu2-3,112, trp1-1 , his3-11,15, ade2-1, met2Δ1, lys2Δ2, can1-100, nrd1ΔHIS3*	Fun. stu.	This work
pRS415-nrd1-K490stop	*Mat a ura3-52, leu2-3,112, trp1-1 , his3-11,15, ade2-1, met2Δ1, lys2Δ2, can1-100, nrd1ΔHIS3*	Fun. stu.	This work
pRS415-nrd1-L558stop	*Mat a ura3-52, leu2-3,112, trp1-1 , his3-11,15, ade2-1, met2Δ1, lys2Δ2, can1-100, nrd1ΔHIS3*	Fun. stu.	This work
pRS415-nrd1-Q567stop	*Mat a ura3-52, leu2-3,112, trp1-1 , his3-11,15, ade2-1, met2Δ1, lys2Δ2, can1-100, nrd1ΔHIS3*	Fun. stu.	This work
pRS415-nrd1-K335E	*Mat a ura3-52, leu2-3,112, trp1-1 , his3-11,15, ade2-1, met2Δ1, lys2Δ2, can1-100, nrd1ΔHIS3*	Fun. stu.	This work
pRS415-nrd1-K335R	*Mat a ura3-52, leu2-3,112, trp1-1 , his3-11,15, ade2-1, met2Δ1, lys2Δ2, can1-100, nrd1ΔHIS3*	Fun. stu.	This work
pRS415-nrd1-K335M	*Mat a ura3-52, leu2-3,112, trp1-1 , his3-11,15, ade2-1, met2Δ1, lys2Δ2, can1-100, nrd1ΔHIS3*	Fun. stu.	This work

Supplementary Table S1 (cont.)

Plasmid/vector	Expression host/strain	Use	Source
pRS415-nrd1-T340A	*Mat a ura3-52, leu2-3,112, trp1-1 , his3-11,15, ade2-1, met2Δ1, lys2Δ2, can1-100, nrd1ΔHIS3*	Fun. stu.	This work
pRS415-nrd1-W353A	*Mat a ura3-52, leu2-3,112, trp1-1 , his3-11,15, ade2-1, met2Δ1, lys2Δ2, can1-100, nrd1ΔHIS3*	Fun. stu.	This work
pRS415-nrd1-R374A	*Mat a ura3-52, leu2-3,112, trp1-1 , his3-11,15, ade2-1, met2Δ1, lys2Δ2, can1-100, nrd1ΔHIS3*	Fun. stu.	This work
pRS415-nrd1-K380A	*Mat a ura3-52, leu2-3,112, trp1-1 , his3-11,15, ade2-1, met2Δ1, lys2Δ2, can1-100, nrd1ΔHIS3*	Fun. stu.	This work
pRS415-nrd1-R413G	*Mat a ura3-52, leu2-3,112, trp1-1 , his3-11,15, ade2-1, met2Δ1, lys2Δ2, can1-100, nrd1ΔHIS3*	Fun. stu.	This work
pRS415-nrd1-Y418A	*Mat a ura3-52, leu2-3,112, trp1-1 , his3-11,15, ade2-1, met2Δ1, lys2Δ2, can1-100, nrd1ΔHIS3*	Fun. stu.	This work
pRS415-nrd1-W437A	*Mat a ura3-52, leu2-3,112, trp1-1 , his3-11,15, ade2-1, met2Δ1, lys2Δ2, can1-100, nrd1ΔHIS3*	Fun. stu.	This work
pRS415-nrd1- K335E,Y418A	*Mat a ura3-52, leu2-3,112, trp1-1 , his3-11,15, ade2-1, met2Δ1, lys2Δ2, can1-100, nrd1ΔHIS3*	Fun. stu.	This work
pRS415-nrd1-C415S,C416S	*Mat a ura3-52, leu2-3,112, trp1-1 , his3-11,15, ade2-1, met2Δ1, lys2Δ2, can1-100, nrd1ΔHIS3*	Fun. stu.	This work
pET28-txAHTEV-Nrd1 290-468 K335E	*E. coli* BL21(DE3)	F.A.	This work
pET28-txAHTEV-Nrd1 290-468 K335M	*E. coli* BL21(DE3)	F.A.	This work
pET28-txAHTEV-Nrd1 290-468 K335R	*E. coli* BL21(DE3)	F.A.	This work
pET28-txAHTEV-Nrd1 290-468 T430A	*E. coli* BL21(DE3)	F.A.	This work
pET28-txAHTEV-Nrd1 290-468 W353A	*E. coli* BL21(DE3)	F.A.	This work
pET28-txAHTEV-Nrd1 290-468 R374A	*E. coli* BL21(DE3)	F.A.	This work
pET28-txAHTEV-Nrd1 290-468 K380A	*E. coli* BL21(DE3)	F.A.	This work
pET28-txAHTEV-Nrd1 290-468 R405A	*E. coli* BL21(DE3)	F.A.	This work
pET28-txAHTEV-Nrd1 290-468 R413G	*E. coli* BL21(DE3)	F.A.	This work
pET28-txAHTEV-Nrd1 290-468 Y418A	*E. coli* BL21(DE3)	F.A.	This work
pET28-txAHTEV-Nrd1 290-468 W437A	*E. coli* BL21(DE3)	F.A.	This work
pET28-txAHTEV-Nrd1 290-468 C415S/C416S	*E. coli* BL21(DE3)	F.A.	This work

Supplementary Table S2: Crystallographic data statistics and refinement.

Dataset	Nrd1 $_{301-489}$ SeMet	Nrd1 $_{301-489}$	Nrd1 $_{290-468}$	Nrd1 $_{290-468}$	Nrd1 $_{290-468}$	Nrd1 $_{290-468}$
Data collection						
RNA	-	-	-	GUAA	CGUAAA	UUAGUAAUCC
RNA built	-	-	-	GUAA	CGUAAA	AGUAAU
Space group	$P4_32_12$	$P4_32_12$	$P6_5$	$P4_32_12$	$P4_32_12$	$P4_32_12$
Unit cell, *a,b,c* (Å) γ (Å)	61.05, 61.05, 157.60	60.99, 60.99, 157.96	58.39, 58.39, 92.45 120.0	64.94, 64.94, 158.39	64.53, 64.53, 157.03	66.16, 66.16, 156.91
Synchroton Beamline/Detector	XALOC (ALBA) PILATUS 6M	ID23-2 (ESRF) PILATUS 2M	XALOC (ALBA) PILATUS 6M	XALOC (ALBA) PILATUS 6M	XALOC (ALBA) PILATUS 6M	XALOC (ALBA) PILATUS 6M
Temperature (°C)	-173	-173	-173	-173	-173	-173
Wavelength (Å)	0.979330	0.976250	0.97930	0.979260	0.979290	0.979260
Resolution (Å)	48.26 - 2.20 (2.27-2.20)	48.14 - 2.30 (2.38-2.30)	92.45 - 1.60 (1.63-1.60)	158.39 - 2.45 (2.55 -2.45)	45.63 - 3.40 (3.67 – 3.40)	156-91- 3.53 (3.87 - 3.53)
Data processing						
Total reflections	400454 (34541)	335798 (34968)	476523 (24279)	211854(25842)	120281 (24903)	45676 (10425)
Unique reflections	15954 (1360)	13708 (1307)	23634 (1197)	12717(1460)	5025 (989)	4716 (1077)
Multiplicity	25.1 (25.4)	24.5 (26.8)	20.2 (20.3)	16.7 (17.7)	23.9 (25.2)	9.7 (9.7)
Completeness (%)	100.0 (100.0)	99.1 (98.5)	100.0 (100.0)	96.1 (100.0)	100.0 (100.0)	99.5 (99.1)
Mean I/σ (I)	43.6 (6.9)	32.1 (11.1)	26.8 (6.1)	28.2 (3.4)	22.2 (8.5)	9.7 (2.4)
R_{merge}[a] (%)	5.8 (59.8)	7.0 (38.5)	8.1 (60.1)	4.1 (63.6)	15.7 (57.8)	8.1 (43.5)
R_{pim}[b] (%)	1.2 (12.2)	1.5 (7.6)	1.8 (13.6)	10 (15.4)	3.3 (11.6)	2.7 (14.6)
CC1/2	1.00 (0.98)	0.99 (0.99)	0.99 (0.95)	1.00 (0.97)	0.99 (0.99)	0.99 (0.99)
Asymetric unit	1	1	1	1	1	1
Wilson B factor (Å 2)	39.07	41.06	14.92	50.41	65.80	98.53

Supplementary Table S2: (Cont).

Crystal data	Nrd1 $_{301-489}$ SeMet	Nrd1 $_{301-489}$	Nrd1 $_{290-468}$	Nrd1 $_{290-468}$	Nrd1 $_{290-468}$	Nrd1 $_{290-468}$
Refinement						
R_{work} / R_{free}[c] (%)		20.68/23.30	17.76/20.80	19.67/21.72	18.11/21.59	22.51/26.86
Nº of atoms/Mean B factors (Å^2)		1373/59.95	1587/21.67	1448/62.97	1437/89.51	1447/128.46
Protein		1327/60.00	1389/20.20	1310/61.71	1310/85.70	1303/125.05
RNA		-	-	82/75.74	125/129.08	136/159.90
Ligand		4/74.00	60/28.66	16/107.47	-	8/169.30
Water molecules		42/57.70	138/33.22	38/54.51	1/43.95	2/69.17
Ramachandran plot (%) Favoured / Outliers		99.0 / 0	99.0 / 0	98.0 / 0	98.0 / 0	96.0/ 1
R.m.s. Bonds (Å)/Angles (°)		0.010/1.31	0.007/1.25	0.009/1.36	0.007/1.22	0.007/1.16
PDB codes		5O1W	5O1X	5O1Y	5O1Z	5O20
Missing residues in pdb		301-305/472-485	290-301/464-468	290-301/465-468	290-301/465-468	290-301/464-486

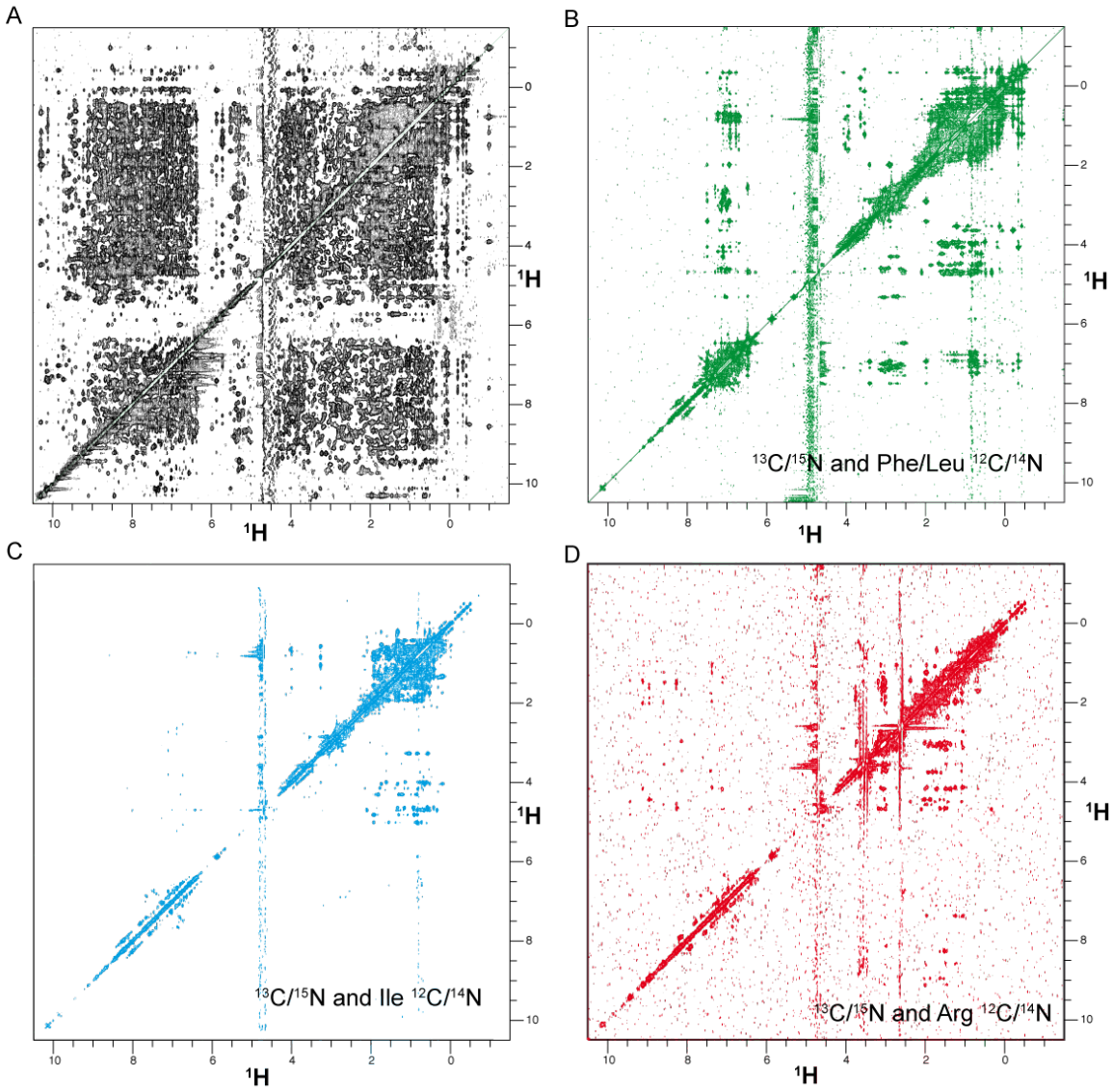
Supplementary Table S3. Summary of NMR restrains and structural calculation statistics for $Nrd1_{290-468}$ solution structure.

NMR experimental restrains	
NOE-derived	
Intraresidue	523
Sequential	756
Medium-range (1 < i-j < 4)	405
Long-range (i-j > 4)	1055
Total per residue	5.9
Intra RRM	1301
Intra SD	1284
SD to RRM	153
TALOS+ obtained restrains	
ϕ angle restrains	123
ψ angle restrains	131
Structure statistics	
Mean AMBER energies (kcal/mol ± SD)	
Total	-4338 ± 43
Van der Waals	-1380 ± 20
Restrains (distance + angle)	6 ± 1
Violations	
Distance*	5.3 ± 1.3
Maximum distance violation (Å)	0.25
Angle**	4.2 ± 0.6
Maximum angle violation (º)	18.7
RMSD from ideal geometry	
Bond lengths (Å)	0.010
Bond angles (º)	1.98
Ramachandran Plot analysis	
Residues in most favoured regions	95 ± 1 %
Residues in additionally allowed regions	5 ± 1 %
Residues in disallowed regions	0 ± 0 %
Averages RMSD to mean structure (range)	*(298-463)*
N, CO, C ✔ (Å) (± SD)	0.63 ± 0.11
All heavy (Å) (± SD)	0.97 ± 0.11

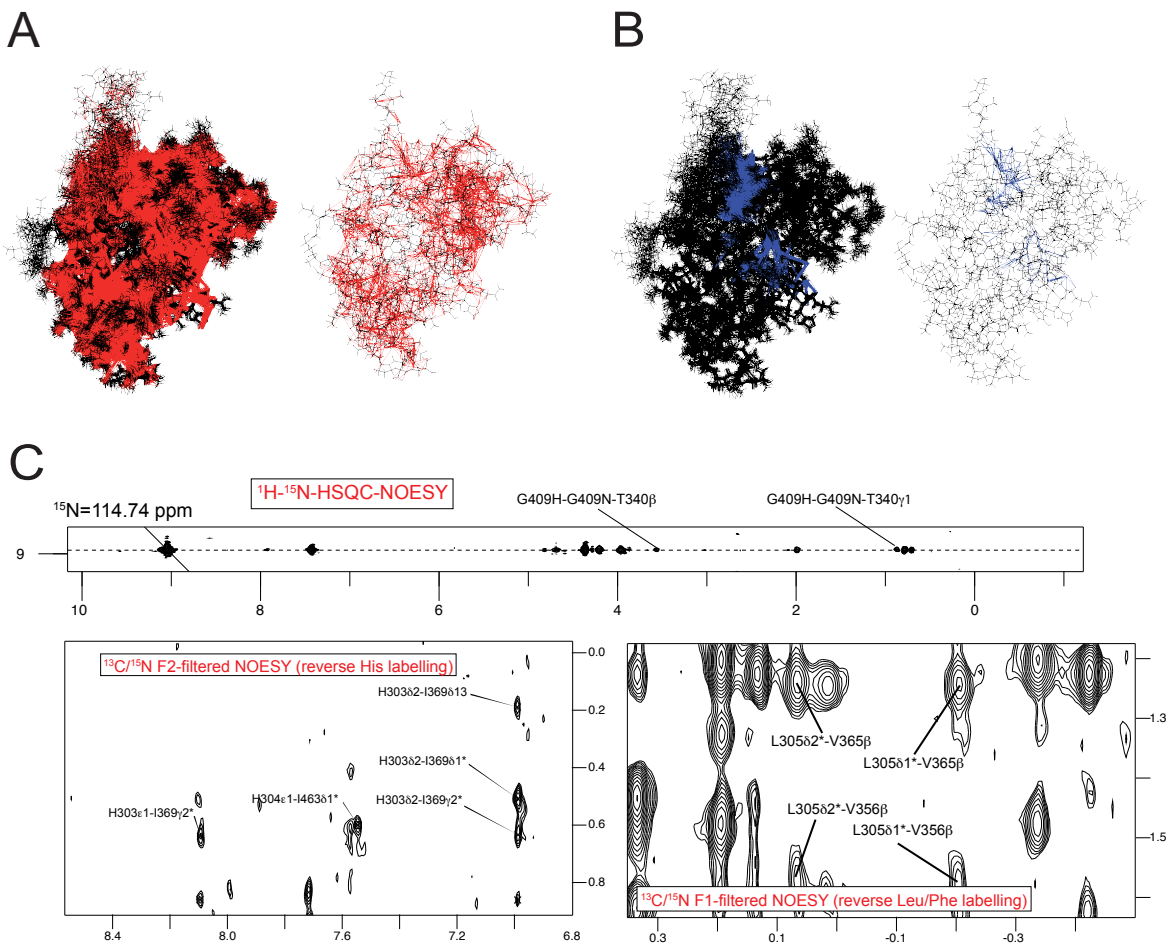
*Averaged value per structure of distance violations > 0.20 Å ± SD.
** Averaged value per structure of total angle violations ± SD.

Supplementary Table S4. NOE-derived distance restraints between RRM and SD domains. The X-ray column shows NMR restraints violations > 0.4 Å found in the x-ray $Nrd1_{290-468}$ structure and reflect minor differences between both structures. Restrains for residues without coordinates in the X-ray are labelled as NA (not available).

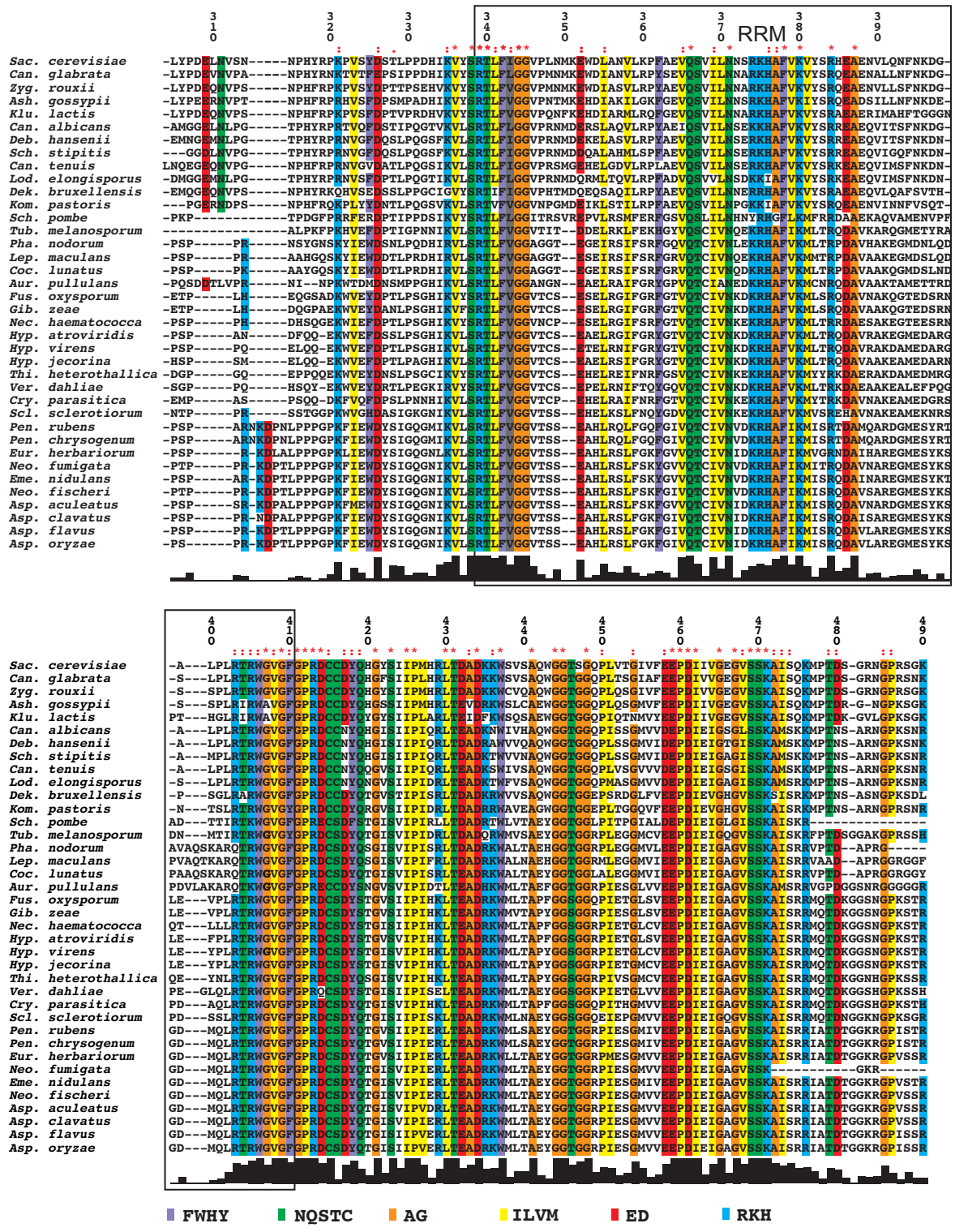
NMR upper limit restraint			X-ray	NMR upper limit restraint			X-ray
294 PRO HA	353 TRP HH2	5.71	NA	305 LEU QQD	352 GLU QB	4.51	
294 PRO HA	353 TRP HZ2	4.77	NA	305 LEU QQD	353 TRP HA	5.06	
294 PRO HB2	353 TRP HH2	4.76	NA	305 LEU QQD	353 TRP HD1	5.36	
294 PRO HB2	353 TRP HZ2	4.66	NA	305 LEU QQD	353 TRP HE1	5.08	
294 PRO HB3	353 TRP HH2	4.76	NA	305 LEU QQD	353 TRP HE3	4.44	0.8
294 PRO HB3	353 TRP HZ2	4.66	NA	305 LEU QQD	353 TRP HH2	4.95	
294 PRO QB	353 TRP HH2	3.96	NA	305 LEU QQD	353 TRP HZ2	4.34	
294 PRO QB	353 TRP HZ2	3.99	NA	305 LEU QQD	353 TRP HZ3	3.98	1.2
294 PRO QG	353 TRP HZ2	5.09	NA	305 LEU QQD	355 LEU QB	5.98	1.2
296 THR HA	353 TRP HE1	4.55	NA	305 LEU QQD	356 ALA H	4.48	
296 THR QG2	353 TRP HE1	5.27	NA	305 LEU QQD	356 ALA QB	3.38	1.1
296 THR QG2	353 TRP HZ2	4.45	NA	305 LEU QQD	365 VAL HB	4.31	0.7
298 PHE HZ	352 GLU H	5.99	NA	305 LEU QQD	365 VAL QQG	3.67	2.0
298 PHE HZ	352 GLU QB	4.68	NA	305 LEU QQD	367 SER H	4.83	
298 PHE HZ	353 TRP HD1	5.06	NA	305 LEU QQD	367 SER HA	4.41	
298 PHE HZ	353 TRP HE1	4.50	NA	305 LEU QQD	368 VAL HB	4.82	
298 PHE QD	353 TRP HE1	4.72	NA	305 LEU QQD	368 VAL QQG	3.44	1.7
298 PHE QD	368 VAL QQG	5.36	NA	306 TYR HA	353 TRP HZ2	5.02	
298 PHE QE	352 GLU QB	3.85	NA	306 TYR QD	366 GLN HA	5.26	
298 PHE QE	352 GLU QG	4.12	NA	306 TYR QE	366 GLN HA	5.20	
298 PHE QE	353 TRP HD1	4.84	NA	317 HIS HD2	364 GLU HB2	4.38	
298 PHE QE	353 TRP HE1	4.57	NA	317 HIS HD2	364 GLU HB3	4.38	
298 PHE QE	353 TRP HZ2	4.54	NA	317 HIS HD2	364 GLU QG	4.16	
298 PHE QE	368 VAL QQG	5.29	NA	337 TYR QD	384 ARG QD	4.60	
303 HIS HA	368 VAL H	3.97		338 SER H	384 ARG HH11	4.72	
303 HIS HB2	368 VAL H	4.46		338 SER H	384 ARG HH12	4.72	
303 HIS HB3	368 VAL H	4.46		338 SER H	384 ARG QH1	4.10	
303 HIS HD2	369 ILE H	6.50		338 SER H	406 TRP HH2	5.34	
303 HIS HD2	369 ILE QD1	4.40	1.8	338 SER HA	340 THR H	4.55	
303 HIS HD2	369 ILE QG1	4.50		338 SER HA	406 TRP HE3	5.67	
303 HIS HE1	369 ILE HA	5.49		338 SER HB2	339 ARG H	5.88	
303 HIS HE1	369 ILE HG12	5.65		338 SER HB2	340 THR H	5.10	
303 HIS HE1	369 ILE HG13	5.65		338 SER HB3	339 ARG H	5.88	
303 HIS HE1	369 ILE QD1	5.01		338 SER HB3	340 THR H	5.10	
303 HIS HE1	369 ILE QG1	4.97		338 SER QB	340 THR H	4.29	
303 HIS HE1	369 ILE QG2	5.00		338 SER QB	406 TRP HE3	5.19	
303 HIS QB	368 VAL H	3.86		339 ARG H	459 GLU H	5.34	
303 HIS QB	369 ILE QD1	4.44		339 ARG H	459 GLU HA	5.61	
303 HIS QB	369 ILE QG1	4.59		339 ARG H	459 GLU QB	4.62	
304 HIS HA	368 VAL H	5.30		339 ARG H	459 GLU QG	6.27	
304 HIS HD2	367 SER HB2	4.93		339 ARG QH2	460 PRO HA	5.26	
304 HIS HD2	367 SER HB3	4.93		340 THR H	459 GLU HB2	6.10	
304 HIS HD2	367 SER QB	4.27		340 THR H	459 GLU HB3	6.10	
305 LEU H	367 SER H	4.14		340 THR H	459 GLU QB	5.22	
305 LEU H	367 SER HA	4.75		340 THR HB	408 VAL H	5.09	
305 LEU H	367 SER QB	5.25		340 THR HB	409 GLY H	4.96	
305 LEU HA	353 TRP HE1	6.50		340 THR QG2	407 GLY H	5.19	
305 LEU HA	353 TRP HZ2	4.62		340 THR QG2	408 VAL H	4.65	
305 LEU HG	353 TRP HZ2	4.34	0.9	340 THR QG2	409 GLY H	5.30	
305 LEU QD1	352 GLU HA	5.71	0.8	341 LEU QQD	407 GLY H	4.78	
305 LEU QD1	353 TRP HD1	6.50		378 PHE QD	407 GLY QA	4.66	
305 LEU QD1	353 TRP HE1	6.50		378 PHE QD	408 VAL H	5.24	
305 LEU QD1	353 TRP HE3	5.10		378 PHE QE	407 GLY QA	5.14	1.1
305 LEU QD1	353 TRP HZ2	5.17		378 PHE QE	408 VAL H	5.01	1.4
305 LEU QD1	353 TRP HZ3	4.73		378 PHE QE	408 VAL QQG	5.44	3.3
305 LEU QD1	356 ALA H	5.45		378 PHE QE	418 TYR QE	5.24	3.6
305 LEU QD1	356 ALA QB	4.22		382 TYR HA	443 TRP HD1	4.46	
305 LEU QD1	365 VAL HB	4.96		382 TYR HA	443 TRP HE1	5.89	
305 LEU QD1	367 SER H	5.50		382 TYR QB	443 TRP HD1	4.74	
305 LEU QD1	367 SER HA	5.26		382 TYR QB	443 TRP HE1	5.10	
305 LEU QD1	368 VAL H	5.10		406 TRP H	407 GLY H	5.34	
305 LEU QD1	368 VAL QQG	4.01	1.4	406 TRP H	407 GLY QA	5.21	
305 LEU QD2	352 GLU HA	5.71		406 TRP H	418 TYR HA	5.84	
305 LEU QD2	353 TRP HD1	6.50		406 TRP HD1	418 TYR HA	4.94	
305 LEU QD2	353 TRP HE1	6.50		406 TRP HD1	419 GLN HA	4.50	
305 LEU QD2	353 TRP HE3	5.10	1.0	406 TRP HD1	421 GLY HA2	4.42	
305 LEU QD2	353 TRP HZ2	5.17		406 TRP HD1	421 GLY HA3	4.42	
305 LEU QD2	353 TRP HZ3	4.73	1.6	406 TRP HD1	421 GLY QA	3.84	
305 LEU QD2	356 ALA H	5.45		406 TRP HE1	419 GLN HA	6.36	
305 LEU QD2	356 ALA QB	4.22	1.5	406 TRP HE1	420 HIS HA	4.59	
305 LEU QD2	365 VAL HB	4.96	1.6	406 TRP HE1	421 GLY H	6.30	
305 LEU QD2	367 SER H	5.50		406 TRP HE1	421 GLY QA	5.47	
305 LEU QD2	367 SER HA	5.26		406 TRP HE3	407 GLY H	4.06	
305 LEU QD2	368 VAL H	5.10		406 TRP HE3	421 GLY HA2	5.67	
305 LEU QD2	368 VAL QQG	4.01	1.8	406 TRP HE3	421 GLY HA3	5.67	
305 LEU QQD	352 GLU HA	4.91	1.1	406 TRP HE3	421 GLY QA	4.95	
				406 TRP QB	407 GLY H	4.35	



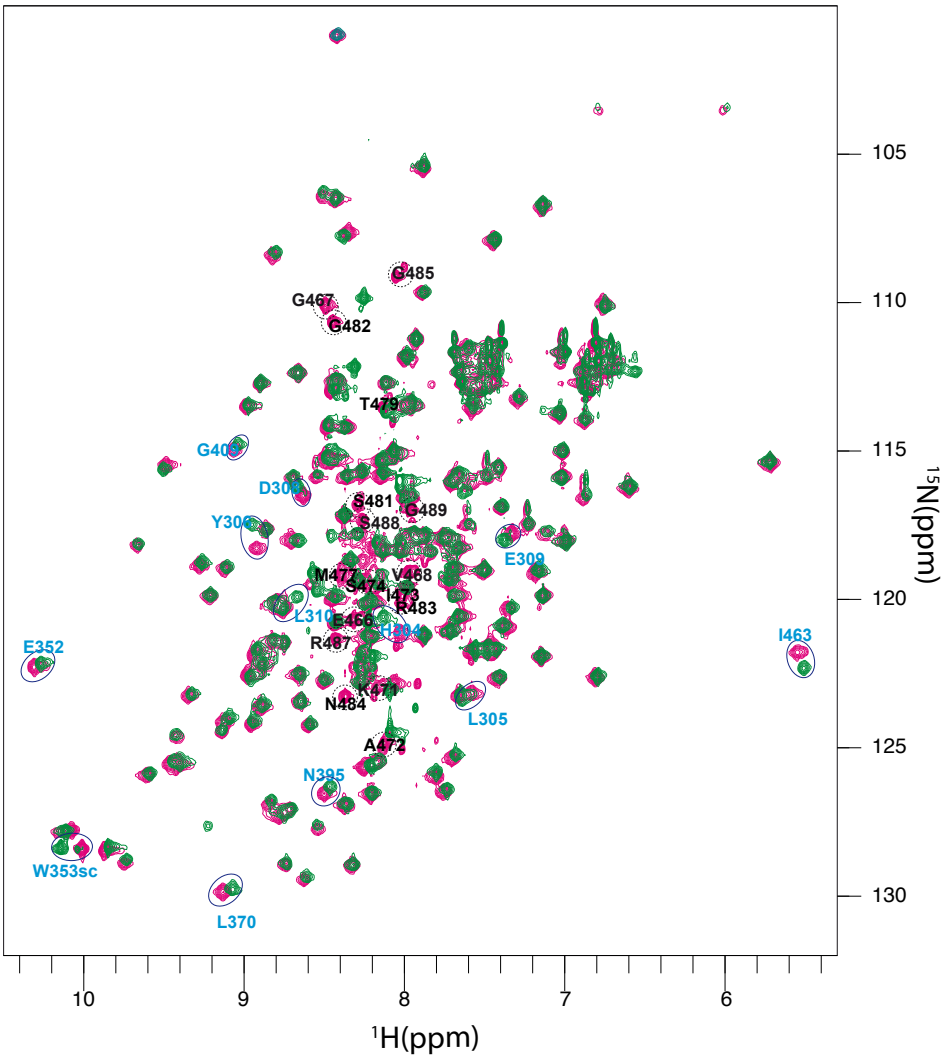
Supplementary Figure S1. Selectively unlabelled samples in $^{13}C/^{15}N$ background greatly facilitates NOEs assignments. Comparison between the standard 2D NOESY (**A**) and 2D $^{13}C,^{15}N$ double-filter NOESY of three samples obtained by selectively unlabelling Phe/Leu (**B**), Ile (**C**) and Arg (**C**). Samples were produce by growing bacterial cultures in $^{13}C/^{15}N$ medium and adding the aminoacid/metabolic precursor 30 min prior induction. Metabolic precursors used for each amino acid (Sigma-Aldrich) are explained in detail in (4,5) which the exception of L-citrulline and L-histidinol (Sigma-Aldrich) that have been used in this work for the first time for selective unlabelling Arg and His respectively.



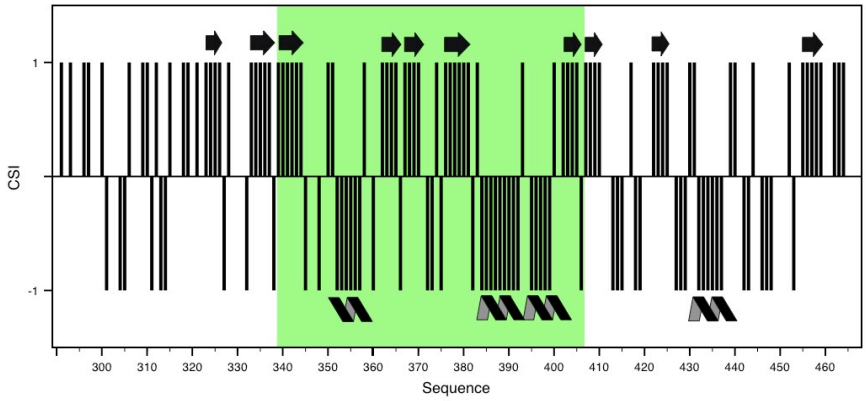
Supplementary Figure S2. Selected NMR data and derived distance restraints used in the calculation of 5O1T (full restraint set available at the PDB site). **A** Representation of all distance restrains (in red) over the NMR ensemble/one conformer (in black) r. **B** Representation of SD-RRM restrains subset (blue). **C** Zooms in of selected regions of various types of NOESY spectra. For simplicity, only the SD to RRM NOEs have been labelled.



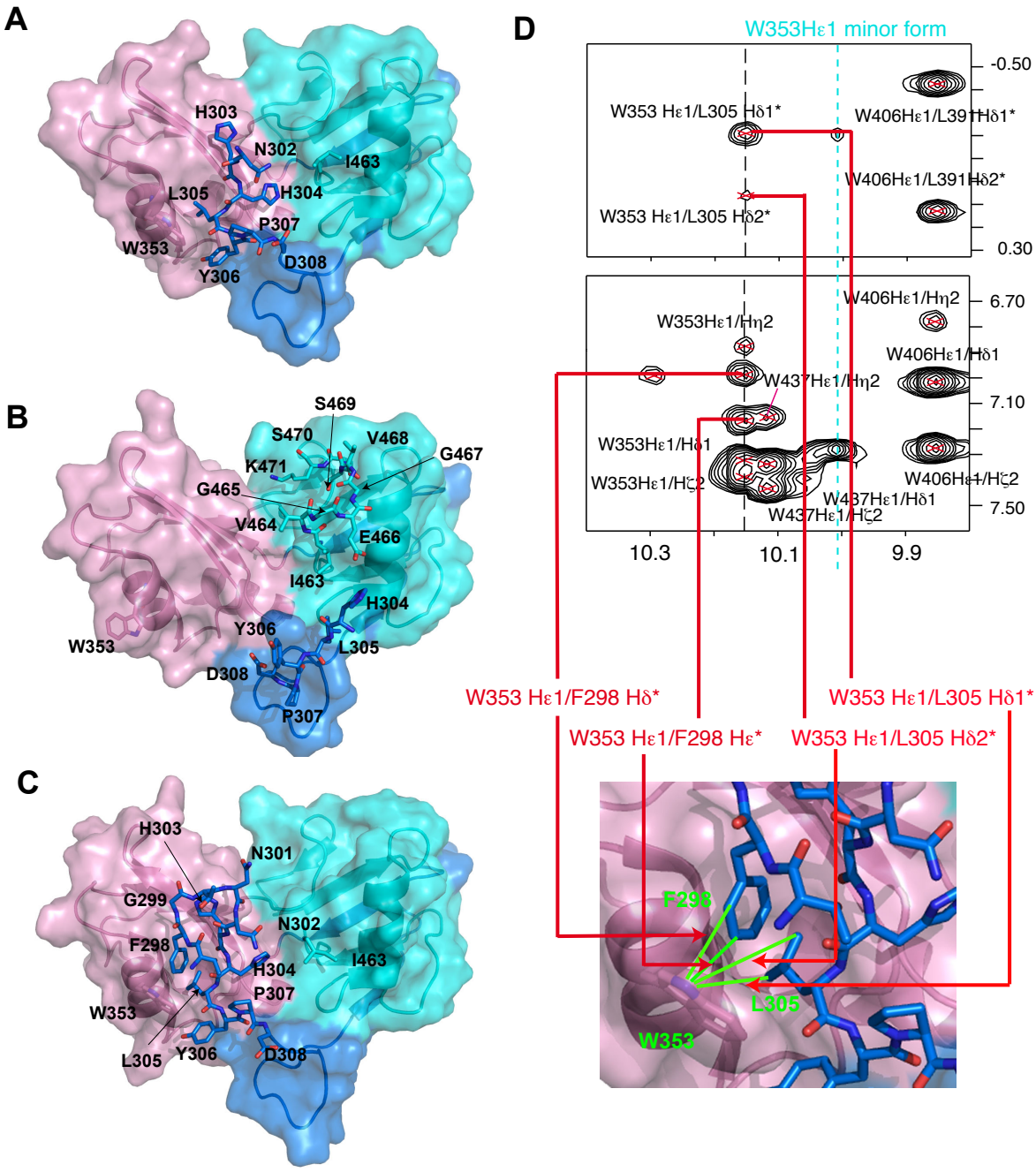
Supplementary Figure S3. Sequence alignment of RNA binding domains of 39 Nrd1 orthologues in the OMA group 108412 (fingerprint: GTHKLGV. http://omabrowser.org/). Sequences are shorted according sequence similarity to *S. cerevisiae* Nrd1 (on top). Numbering corresponds to *S. cerevisiae* Nrd1. The percentage of sequence conservation within the group is plotted on the bar chart below each block and conserved residues are coloured by amino acid type according to the key code below. Highly conserved positions of yeast Nrd1 are marked (: > 75% and * 100%). The predicted location of the RRM is framed by a rectangle.



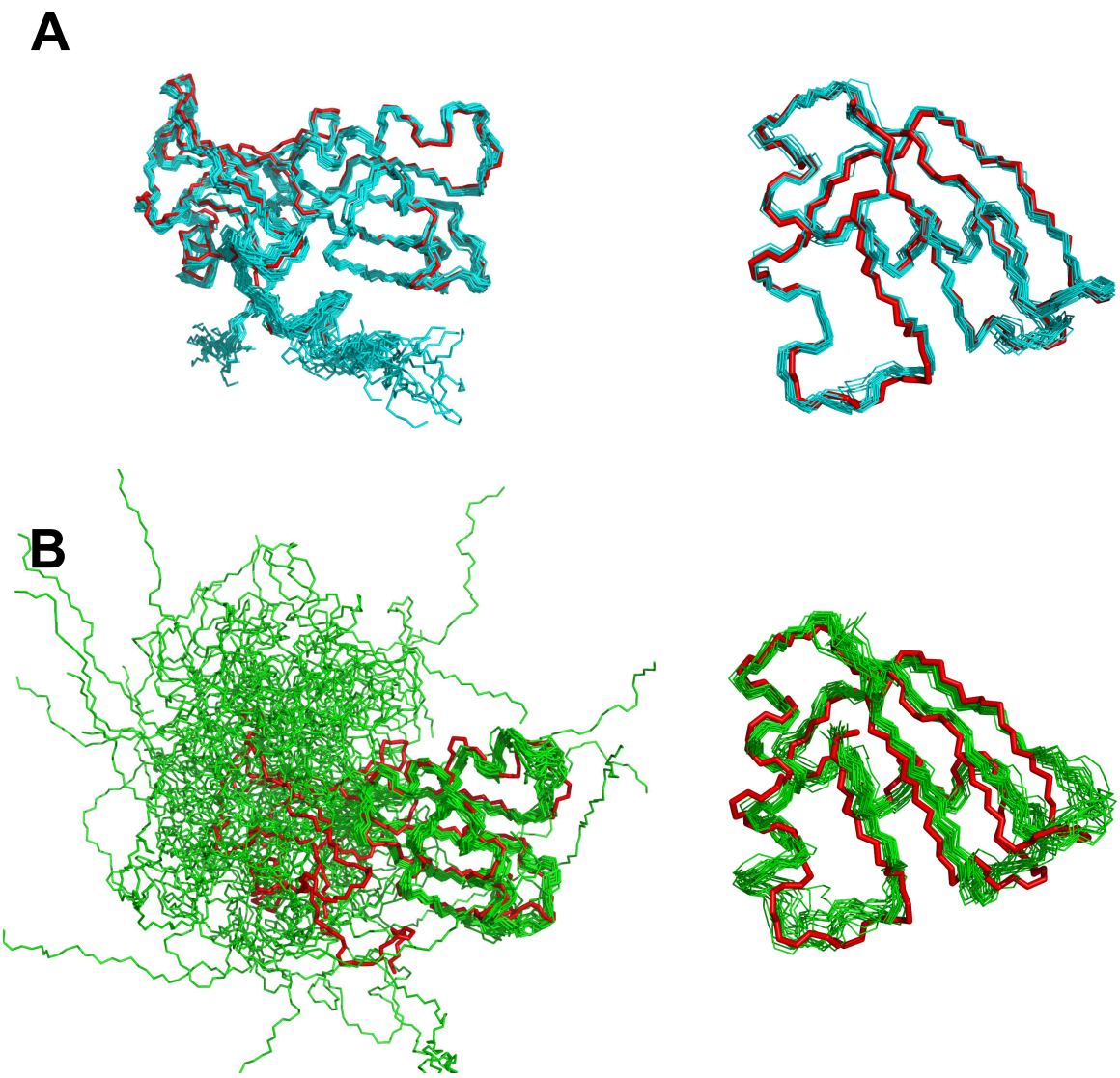
Supplementary Figure S4. Comparison of ^{1}H-^{15}N HSQC NMR spectra of $Nrd1_{301-489}$ (magenta) and $Nrd1_{290-468}$ (green). Signals for residues in the unfolded $Nrd1_{301-489}$ C-terminus (residues 466-489) have been marked (dashed cycled and black labels). Crosspeaks exhibiting larger differences between the two constructs have been marked (solid ovals and cyan labels). Both spectra have been recorded at identical experimental conditions (see materials and methods) in a Bruker AV 800 MHz spectrometer.



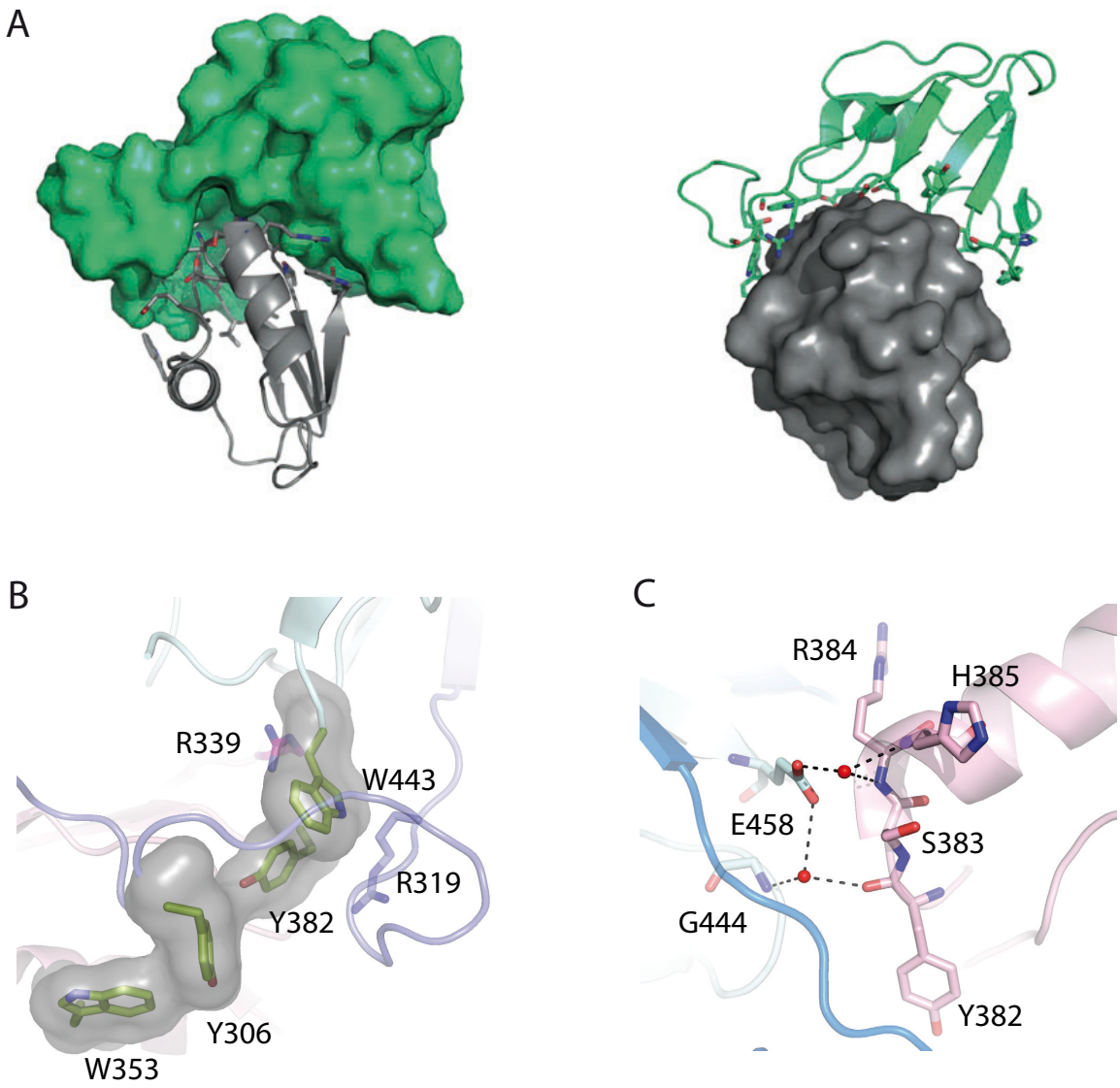
Supplementary Figure S5. Chemical Shift Index by residue for $Nrd1_{290-468}$ RNA binding domain and predicted secondary structure elements (1: β-sheet, -1: α-helix). CSI defined from $^{13}C\alpha$, $^{13}C'$ and $^{13}C\beta$ chemical shifts. RRM predicted limits are shadowed in green.



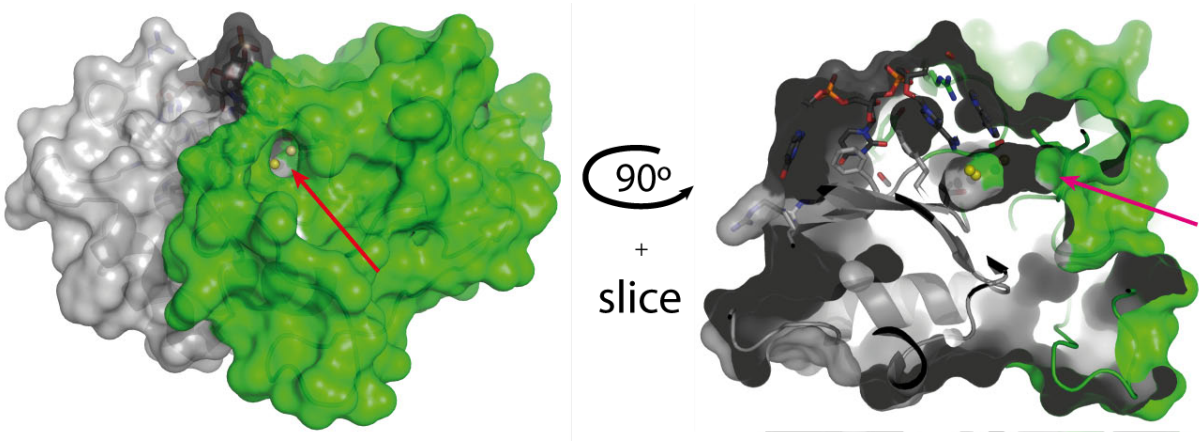
Supplementary Figure S6. Nrd1 RBD structures show conformational differences at their termini. **A** $Nrd1_{290\text{-}468}$ crystal structure. **B** $Nrd1_{301\text{-}489}$ crystal structure. **C** $Nrd1_{290\text{-}468}$ NMR structure (a representative conformer). The unstructured 290-297 segment omitted for clarity. Domains coloured with the same code as in Figure 1. The Nrd1 RBD structural core (residues 308-462) has been represented as ribbon and translucent surface. The C- and N-termini of different length in the two Nrd1 RBD constructs have been represented in sticks and each residue labelled. Trp 353 of the RRM has been also labelled represented in sticks to show the differences on its conformation among the structures. **D.** Detail of 2D NOESY showing the NOEs involving various Nrd1 tryptophan Hε1 protons. Long-range W335 NOEs are labelled in red and correlated with their corresponding distance restrain in the structure (zoom in of C). The position of W353 Hε1 minor form is represented by a dashed cyan line.



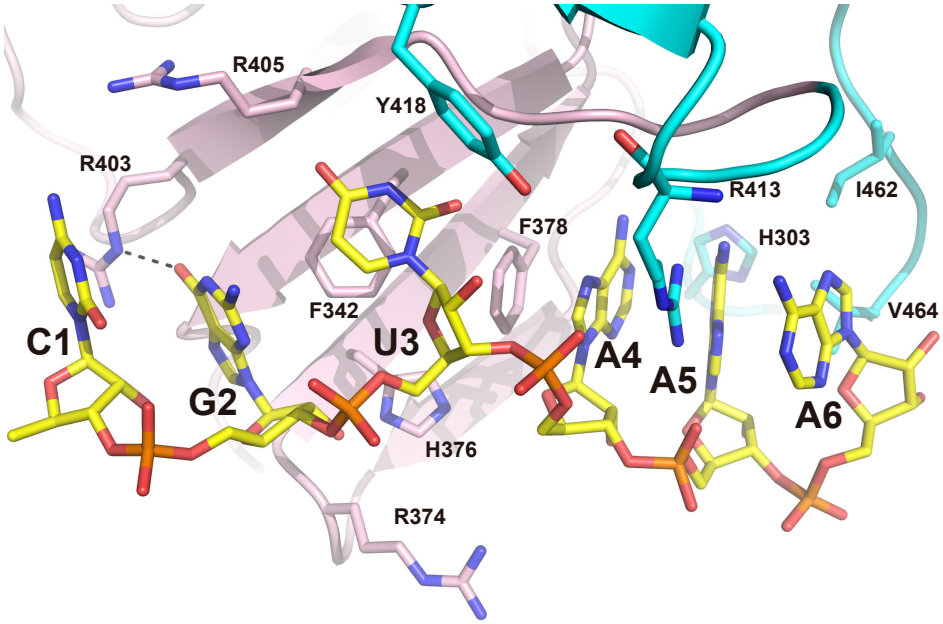
Supplementary Figure S7. Comparison between X-ray and NMR structures of Nrd1 RBD. **A.** Superposition between the backbone of X-ray (red) and NMR (cyan) structures of $Nrd1_{290-468}$ (this work). Detail of the RRM alignment is shown on the right. **B.** Superposition between the backbone of $Nrd1_{290-468}$ X-ray structure (in red) (this work) and previously determined NMR of $Nrd1_{307-490}$ (PDB:2m88) (in green) (6), including a comparison of the RRMs domains on the right. The lack of definition of the SD domain in the $Nrd1_{307-490}$ might be due to partial truncation of the N-terminus of this subdomain or, most likely to protein misfolding, caused by overexpression at 30ºC. Our earlier studies showed evidences that our construct $Nrd1_{301-489}$ is misfolded when expressed at temperatures of >12ºC: inhomogeneous signals in the ^{1}H-^{15}N-HSQC, faster degradation and unsuccessful crystallization



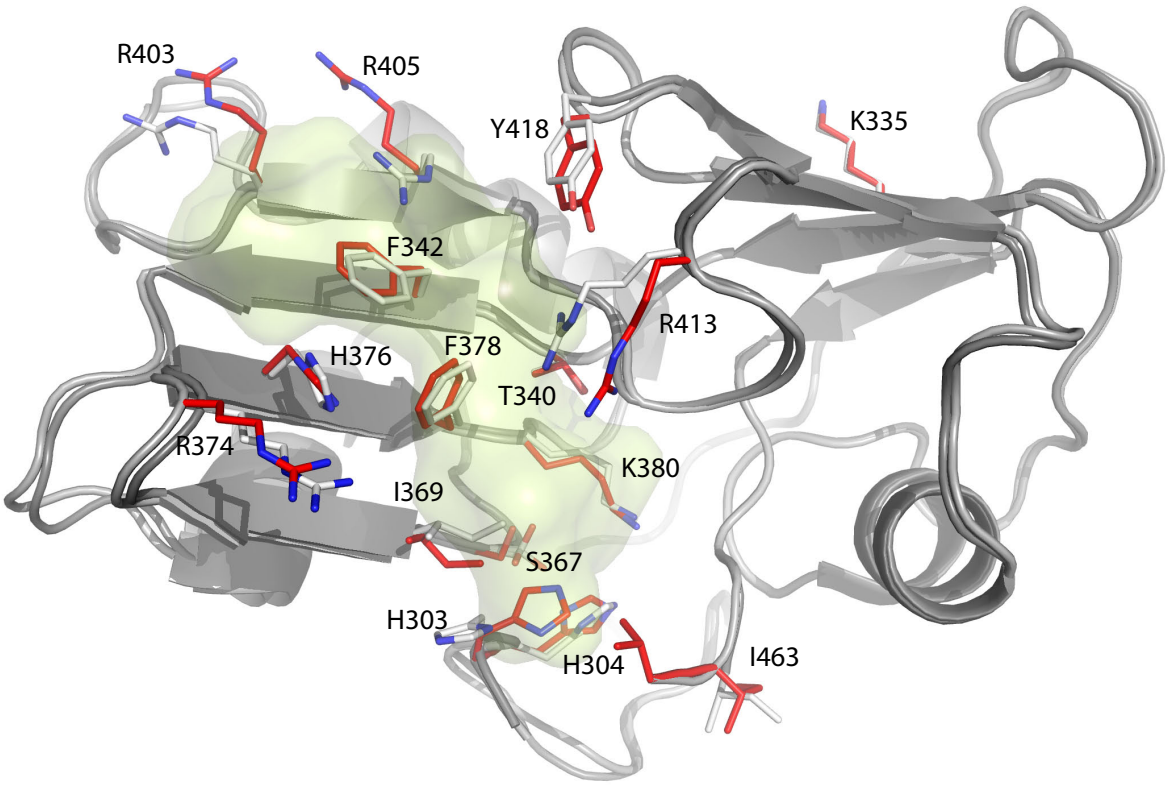
Supplementary Figure S8. Nrd1 RRM and SD subdomains interface details. **A**. RRM (grey) and SD (green) interact through a curved interface. **B**. Aromatic layer separating the Arg319 and Arg339 polar clusters. The residues forming the layer are shown as green sticks and surrounded by a grey surface representation. **C**. Only two structural waters are found at the SD-RRM interface despite its polar nature. The two water molecules (shown as red spheres) make hydrogen bonds with residues from both domains and stabilise the buried Glu458 residue. Equivalent waters are built in crystal structures with best resolution from constructs $Nrd1_{290\text{-}468}$ and $Nrd1_{301\text{-}489}$.



Supplementary Figure S9. The structure of the Nrd1-RNA complex contains an internal water pocket (4 molecules; yellow). Magenta arrows show the entry access to this pocket.



Supplementary Figure S10. Structure of $Nrd1_{290\text{-}468}$ in complex with CGUAAA obtained in phosphate buffer. Colour code is as in Figure 3. The structure reveals the hydrogen bond Arg403 Nε-Gua2 O6 that was not detected in the complex with GUAA (Figure 3) due to the interference of a tris molecule (present in the protein buffer in that case).

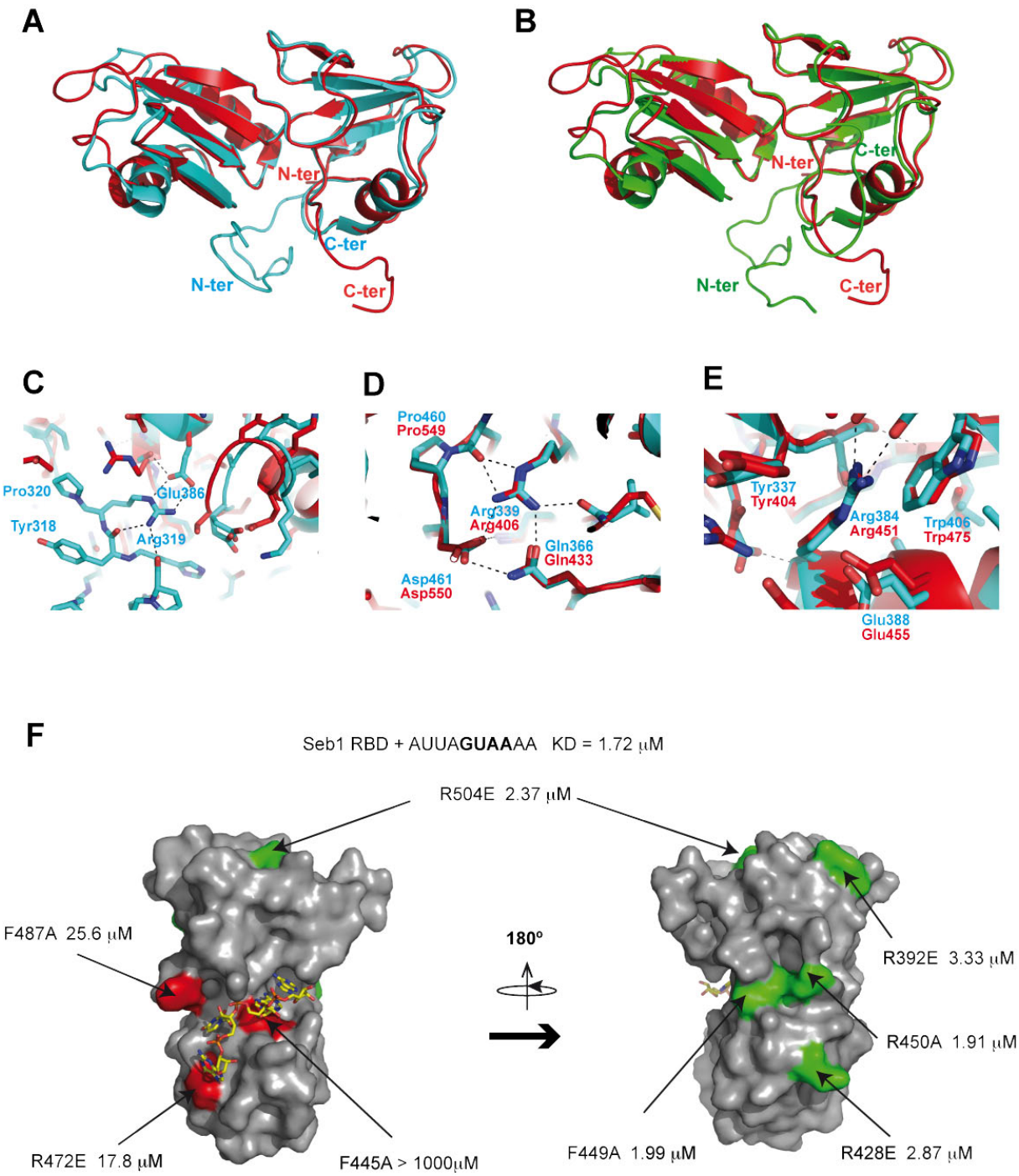


Supplementary Figure S11. The Nrd1 RBD experiences minor changes upon GUAA recognition. Superposition of the crystal structures of $Nrd1_{290-468}$ in its free (sidechain carbons coloured in white), and GUAA-bound (sidechains in red) states. The location of the GUAA sequence in the complex is shown as green transparent surface. The residues in contact with the RNA in the complex have been represented. The rest of the residues do not suffer significant changes upon binding. Arg403 and Arg405 are poorly defined in the structure of the free form and interact with a tris molecule in the complex (not shown here), thus their changes are probably not attributed to RNA recognition. The side chain of Lys335, mentioned several times in the discussion, has been included to emphasize its opposite position to the RNA interface.

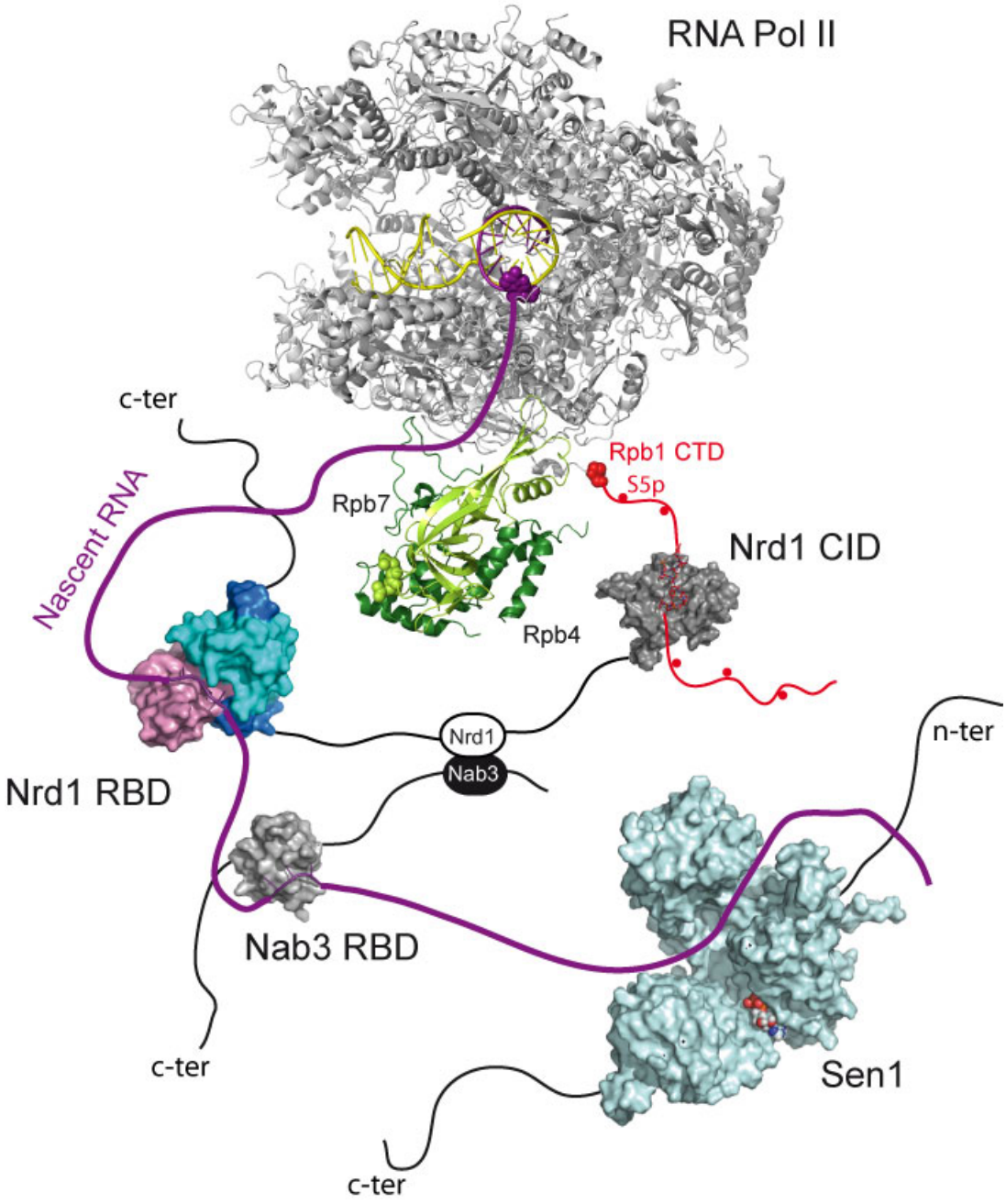
```
S. cere.        DSGRNGPRSGKPNKSGSISS----------ISPVP-YGNAPLASPPP------QQYVQPM 522
T. pha.         DSIRSQQNYQNVMHHHP------QPPQQQQQQ--MMYGGYQ-----D------TKYNNNN 544
K. pol.         DSNRNGPRSGKGEMVPTQ--------------NVPQYGAYP----PE------Q------ 538
S. cas.         DS-RHGPKSAKAFNSPPAAA----------PIPMQQTQGMPYGST-N------QMYPQ-- 523
S. dai.         DGTRNGPRSLKRNPMPLQSGSYATPPQQQQPP--QMVQPQPYGSPN-------QLYGQLQ 566
C. gla.         DSGRNGPRSNKAKNMAPQYGP----PQAAVISPSQGYAGMQQGYPQQQQQQQMGMYGQPM 550
C. col          DSGRNGPRSGKSTKTSV--S----------SPSMPGYATNQIASPG-------QMYSQ-- 529
Z. rou.         DSGRNGPRSGKGNSRSSLSV----------SPPVTKYSMAPVPSPG-------HVYGQ-- 517
                *. *   .  :

MQQPY--------GYAPNQ---------------PLPSQ--GPAAAAPPVPQQQFDPTAQLNSLMNMLNQQQQQQQQS 575
FMPPPPPQQQYNQHMPPQPHMP------------HMP----PQPHMPPQQQPQGFDPTAQLNSLMSMLNQQPPR---- 602
RYSGYPPQQNMGYNQPMPQQGVDE----YQQPPPPLPPQ------QQQQQPQQSFDPTAHLNSLMSMLNQQQR----- 601
MQGPPPPQQQMGYGQ--P-----------------YPQPQQAHMAPPPPQQQGNIDPTAQLNSLMNMLSQQQQRN--- 579
-PNMLPPQQQQQQQSYYAMN--------GQQQVPPPPP--------PQQQVQQNVADPTAQLNSLMNMLNQQQQRS--- 624
MQQPINAYAS------PAQQQ-----PIMQAPPPPAP--------PSQQQQSSNYDPTAQLNSLMNILNQQQQ----- 604
MQQPYPPQQAAPQQAAPPYGQP------------PVAAPM-A-GYTQPPQQQQPFDATAQLNSLMSMLNQQQN----- 588
IPPAYPPQPVAAYNGTPIAAAPQMAPAAATPSYPPAPAPQ-QQAPAPPVPQQQSFDPAAQLNSLMSMLNQQQTQ---- 590
                                                  * :*:*****.:*.**
```

Supplementary Figure S12. Conservation of C-terminal elements of Nrd1 in different yeasts: *Tetrapisispora phaffii, Kluyveromyces polysporus, Saccharomyces castellii, Saccharomyces dairenensis, Candida glabrata, Candida colliculosa* and *Zygosaccharomyces rouxii*



Supplementary figure S13. Structural comparison between Nrd1 and Seb1 (PDB: 5MD4) RNA binding domains. **A.** Superposition between $Nrd1_{290\text{-}468}$ (cyan) and Seb1 (red) crystal. **B.** Superposition between $Nrd1_{301\text{-}489}$ (green) and Seb1 (red) crystal. **C-E**. Detailed view of the three arginine-centred clusters at the $Nrd1_{290\text{-}486}$ RRM-SD interface (Figure 1E-G) superimposed with the equivalent regions of Seb1. Key residues in Nrd1 and Seb1 have been labelled in cyan and red respectively. **F.** Structure model of GUAA recognition by Seb1. The RNA coordinates have been taken from the $Nrd1_{290\text{-}468}$:GUAA complex after structural alignment between Nrd1 and Seb1 RBDs. Seb1 has been represented in grey surface and mutation sites (7) has been coloured according to their effect on RNA affinity (in red K_D > 10 –fold of wt; in green K_D < 3-fold of *wt*). Mutants have been labelled together with their corresponding K_D. The K_D of wild-type Seb1 and the RNA probe used in the mutagenesis experiments are presented above for reference.



Supplementary figure S14: Structural model of the NNS complex and elongating RNA polymerase. RNA polymerase II is shown in cartoons with DNA coloured in yellow, nascent RNA in purple and the stalk subunits: Rpb7 and Rbp4 in two shades of green. Nascent RNA has been represented as a purple line showing the interactions with Rpb7 and Sen1, with the exception of the regions of GUAA and UCUU terminators bound to Nrd1 (this work) and Nab3 (PDB: 2XNR) RBD. The Rpb7 region that interacts with Nrd1 is shown as green spheres (8). The Rpb1 CTD has been represented as a red line connecting the last Rpb1 residue seen in the RNA Pol II structure (1Y1W) and the CTD Ser2p peptide on the Nrd1 CID-CTD complex (PDB: 1LO6). The structure of Sen1 (PDB: 5MZN) is on its RNA-free state; the expected location of the RNA has been drawn in an equivalent position to that reported for Upf1 RNA helicase complex (PDB: 2XAO not shown in this picture). All the atomic models have been represented at the same scale. The loading order of the Nrd1, Nab3 and Sen1 on the nascent RNA does not reflect the biochemical knowledge about the real order in the NNS complex.

REFERENCES

1. Steinmetz, E.J. and Brow, D.A. (1996) Repression of gene expression by an exogenous sequence element acting in concert with a heterogeneous nuclear ribonucleoprotein-like protein, Nrd1, and the putative helicase Sen1. *Mol. Cell. Biol.*, **16**, 6993-7003.
2. Steinmetz, E.J., Ng, S.B.H., Cloute, J.P. and Brow, D.A. (2006) cis- and trans-Acting determinants of transcription termination by yeast RNA polymerase II. *Mol. Cell. Biol.*, **26**, 2688-2696.
3. Vasiljeva, L. and Buratowski, S. (2006) Nrd1 interacts with the nuclear exosome for 3' processing of RNA polymerase II transcripts. *Mol. Cell*, **21**, 239-248.
4. Prasanna, C., Dubey, A. and Atreya, H.S. (2015) Amino acid selective unlabeling in protein NMR spectroscopy. *Methods Enzymol.*, **565**, 167-189.
5. Rasia, R.M., Brutscher, B. and Plevin, M.J. (2012) Selective isotopic unlabeling of proteins using metabolic precursors: application to NMR assignment of intrinsically disordered proteins. *ChemBioChem*, **13**, 732-739.
6. Bacikova, V., Pasulka, J., Kubicek, K. and Stefl, R. (2014) Structure and semi-sequence-specific RNA binding of Nrd1. *Nucleic Acids Res.*, **42**, 8024-8038.
7. Wittmann, S., Renner, M., Watts, B.R., Adams, O., Huseyin, M., Baejen, C., El Omari, K., Kilchert, C., Heo, D.H., Kecman, T. *et al.* (2017) The conserved protein Seb1 drives transcription termination by binding RNA polymerase II and nascent RNA. *Nat. Commun.*, **8**, 14861.
8. Mitsuzawa, H., Kanda, E. and Ishihama, A. (2003) Rpb7 subunit of RNA polymerase II interacts with an RNA-binding protein involved in processing of transcripts. *Nucleic Acids Res.*, **31**, 4696-4701.