Supporting Information

Khazina and Weichenrieder 10.1073/pnas.0809964106

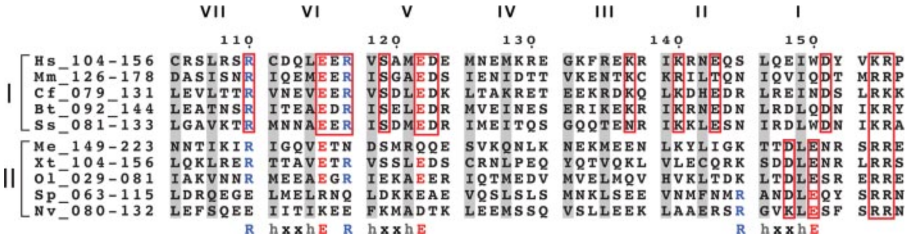

Fig. S1. Conservation of trimerization motifs in the N-terminal coiled-coil (cc) domain of type II NLR ORF1p proteins. The seven C-terminal heptad repeats of the cc domain are numbered I-VII starting from the C-terminal side. The hydrophobic positions one and four of each repeat are shaded in gray. Highly conserved residues are boxed in red. Surface residues only conserved in placental mammals (group I) or only outside of placental mammals (group II) are boxed separately. The highly conserved RhxxxhE sequence motifs promoting a trimeric parallel coiled-coil are indicated below [see: Kammerer RA *et al.* (2005). A conserved trimerization motif controls the topology of short coiled coils. *Proc Natl Acad Sci USA* 102:13891–13896]. h, hydrophobic; x any amino acid. Gene identifiers: Hs, *Homo sapiens* (gi:307098); Mm, *Mus musculus* (gi:198644); Cf, *Canis familiaris* (gi:116175029); Bt, *Bos taurus* (gi:66734172); Ss, *Sus scrofa* (gi:148645275); Me, *Macropus eugenii* (gi:151302550); Xt, *Xenopus tropicalis* (gi:85740540); Ol, *Oryzias latipes* (gi:3746501), Sp, *Strongylocentrotus purpuratus* (gi:111740418); Nv, *Nematostella vectensis* (gi:149338150).

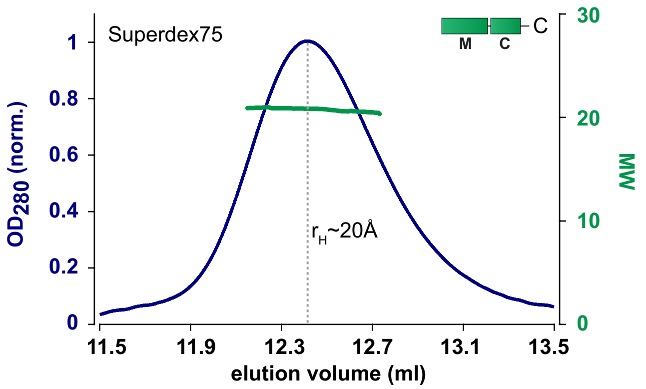

Fig. S2. Monomeric state of the human L1 ORF1p RRM-CTD double domain. Size-exclusion chromatography and MALLS support a rather globular ($r_H \approx 20$ Å), monomeric state of this 21-kDa fragment. (*Inset*) Schematic representation of the construct.

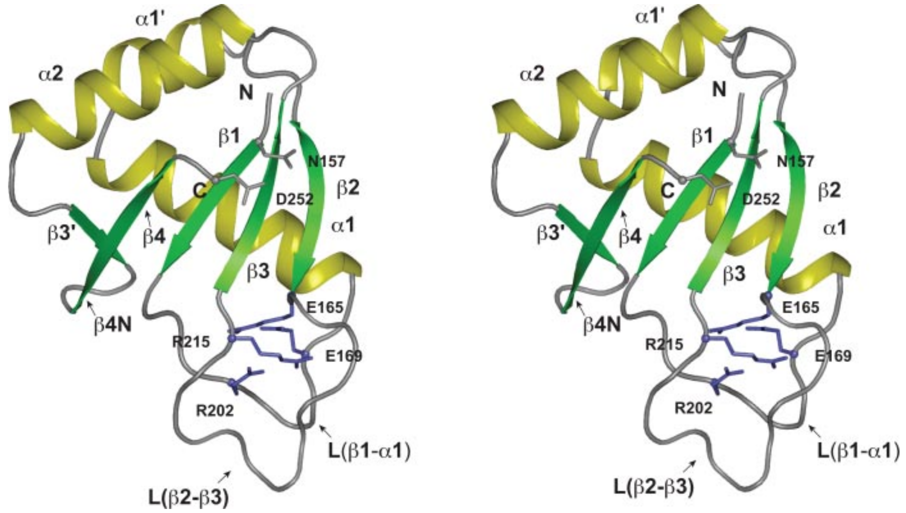

Fig. S3. Crystal structure of the RRM domain of human L1 ORF1p (stereo). Ribbons representation with α-helices in yellow and β-strands in green. Blue sticks, side chains forming the conserved salt-bridges; gray sticks, side chains of the N-terminal N157 and of the C-terminal D252.

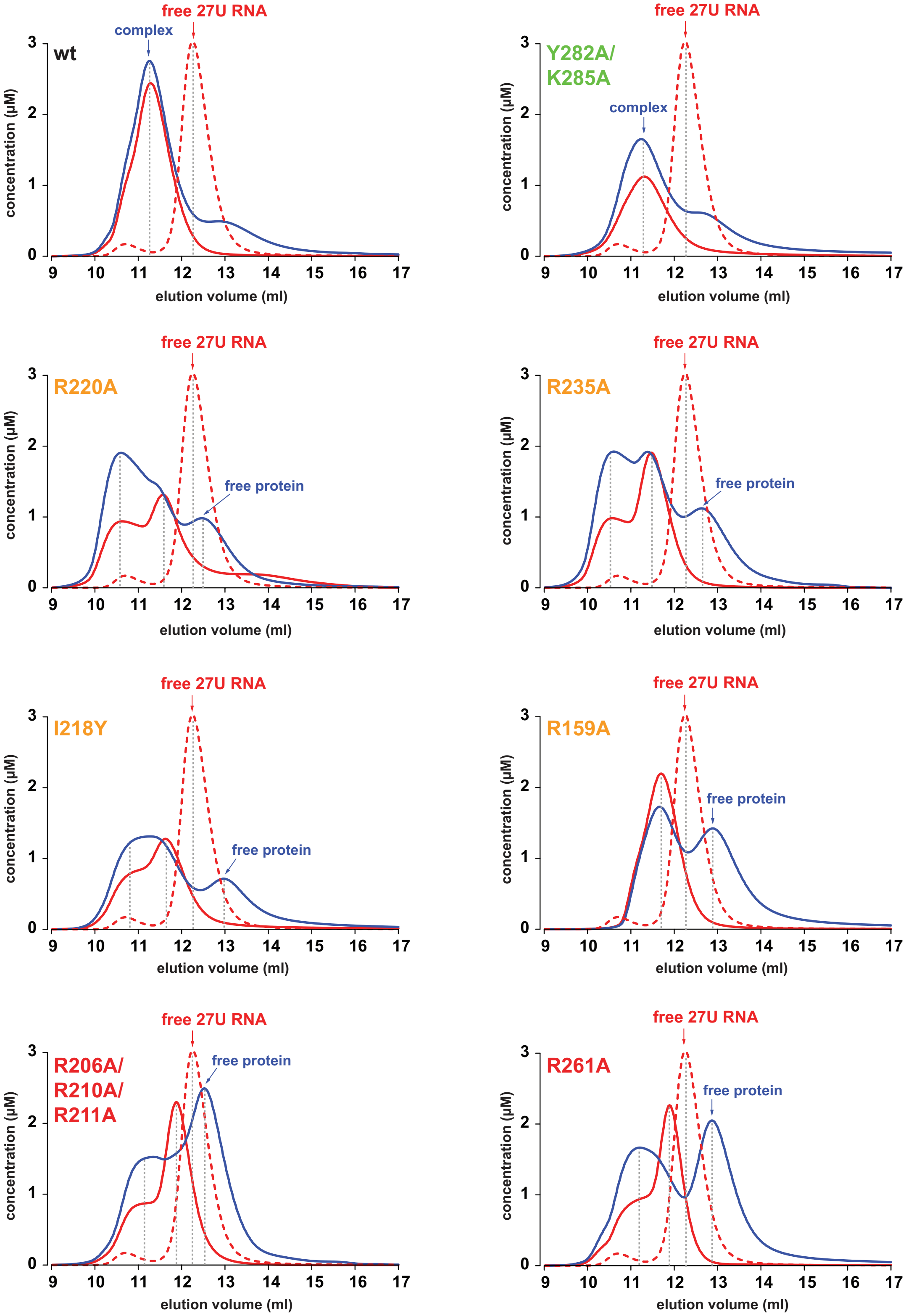

Fig. S4. Mutational analysis of the RNA-binding properties of the human L1 ORF1p RRM-CTD double domain. Size-exclusion chromatography was done with 27-mer poly(U) RNA (27U RNA, 40 μM at start, red lines) in the absence (dashed lines) or in the presence (solid lines) of RRM-CTD (hL1ORF1p-MC^{H6}) protein variants (50 μM at start, blue solid lines). Elution volumes of the free components and of the complexes are indicated by arrows and dashed gray lines, while apparent concentrations are calculated from the relative absorption properties of the components. The respective mutations are indicated. Colors signal strong (red), moderate (orange), and negligible (green) effects on RNA-binding.

Table S1. Identification of RRM domains in selected NLR-ORF1p proteins

Name	Species	gi-number	Query residues*	Probability (%)	E value	*P* value	PDB-ID: Residues	No. of RRMs
Type I NLR ORF1p								
I clade								
I	Dr	20146016	081–233	96.1	0.13	3.3E-06	2adc_A: 102–211	2
Jockey clade								
Jockey	Dm	157823	233–404	97.4	0.0045	1.1E-07	1cvj_A: 235–401	2
TART-A	Dm	48596445	564–717	97.0	0.027	6.7E-07	2qfj_A: 031–210	2
Het_A	Dm	14030851	471–623	97.1	0.022	5.4E-07	2qfj_A: 031–208	2
R1 clade								
pilger	Dm	9369277	235–408	97.1	0.0023	5.8E-08	2dgx_A: 326–396	1
Tad1 clade								
Tad1_1	Nc	409759	176–349	97.5	0.00094	2.4E-08	2dny_A: 250–343	1
L1(plant) clade								
ATLINE1	At	12321249	082–258	97.0	0.021	5.4E-07	1fje_B: 082–219	2
L1(Tx) clade								
Tx1L	Xl	214844	141–241	96.1	0.0088	2.2E-07	2o3d_A: 149–240	1
Type II NLR ORF1p								
L1(vertebrate) clade								
L1.3	Hs	307098	157–252	77.5	8.3	0.00021	3bs9_A: 164–237	1
L1(Tx) clade								
L1(Tx)	Nv	149795606	133–221	73.9	9.9	0.00025	2nlw_A: 134–216	1
CR1 clade								
CR1	Nv	149844706	076–174	48.8	14	0.00035	2ghp_A: 076–158	1
CR1	Sp	111740418	116–200	23.1	88	0.0022	2dgp_A: 117–212	1
Type III NLR ORF1p								
CR1 clade								
Q	Ag	432429	193–294	93.9	0.1	2.6E-06	2dng_A: 193–287	1
T	Ag	159642	215–312	93.0	0.16	3.9E-06	2dng_A: 220–305	1

Identification was done with HHpred. See Söding J, Biegert A, Lupas AN (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* 33:W244–W248. Hs, *Homo sapiens*; Dm, *Drosophila melanogaster*; Dr, *Danio rerio*; Nc, *Neurospora crassa*; At, *Arabidopsis thaliana*; Ag, *Anopheles gambiae*; Xl, *Xenopus laevis*; Sp, *Strongylocentrotus purpuratus*; Nv, *Nematostella vectensis*.

*Amino acids are counted starting with the first methionine.

Table S2. Query sequences used for Table S1

Type I NLR ORF1p	
I_I-DR_Dr_gi_20146016_(081–233)	VFVRLVQEGATFEDWSPIQLTKALYKEIGEVRCAKKLRNGCLLVSCKDEAQQKKAIKVNKINGKKVKCSEVYDRKLIRGVITGIPVSESLNNVIEGITNAKIKEAKRLKTRWNGAICDSLSIMLTFDETKLPDKVFIGYMSYEVKMYIPPPVR
Jockey_Jockey_Dm_gi_157823_(233–404)	KPPAICVPSVSDPVTLERALNLSTGSSNYYIRISRFGVSRIYTANPDAFRTAVKELNKLNCQFWHHQLKEEKPYRVVLKGIHANVPSSQIEQAFSDHGYEVLNIYCPRKSDWKNIQVNEDDNEATKNFKTRQNLFYINLKQGPNVKESLKITRLGRYRVTVERATRRKELLQ
Jockey_TART-A_Dm_gi_48596445_(564_717)	IFLSNIQQIIPLIEKLNYKAGVNSFTTKSELGNNIRIQAKTMDAYKAIQNVLLGANIPLHSHQPKSAKGFQIVIRHLHQSTPTKWIESQLQDIGIATKFIRAMQFRDTRNPMRIHEVEVVPKADGSHLKVLLLKSLGGQTVKVERKRVSKDPTQ
Jockey_HeT-A_Dm_gi_14030851_(471_623)	ILVNDVKEIVPLLEKLNYTAGVSSYTTRAIEGNGVRIQAKDMTAYNKIKEVLVANGLPLFTNQPKSERGFRVIIRHLHHSTPCSWIVEELLKLGFQARFVRNMTNPATGGPMRMFEVEIVMAKDGSHDKILSLKQIGGQRVDIERKNRTREP
R1_pilger_(waldo)_Dm_gi_9369277_(235–408)	AKVKPKRLRKKPEALILKKTGEVTYSDMLRKMKAEPSLTEFGKHVRKIRRTQQGELLLELEGKASEVIPSFKNELEATLKEIASVRTGAHRTALICSGLDETTTAQDLHNSLVSQFQGIRLEPEDVRGLRRRRDGTQIASVLMCANDAIAVINRGVVTVGWSRCRIAQDVRPIR
R1_TRAS1_Bm_gi_940388_(202–378)	RQPPKCTTLHSIMVSSKDENETGDGILTELRKTASEDEGWVRVERVRKIKDRKIIMSYRTEEERTKATQRLKKSEGELVVEEIKNKDPLLILYNVLKMHSDEDLQKALRSKNKDLFRNLNKEDDRIEVKYKKSARNPHTHHVVLKVSPTIWNRALSMGSLHIDIQPVRVADQTPLVQ
Tad1_Tad11_Nc_gi_409759_(176–349)	RQLTIKGATIAAEFVNRSNEDTKTTLATCLGKKKPGLIVRAATRMPTTGDYVIVFDEPTRTWCWRNQAWAKEVFGPDAFITMSTVGVLVRGVPWDSVDNYTTAEAISNVAKERNPEASIIRVRPWKRRDGESRGLLLVEVATASAACFLQDNLFLWDGGAYPCEPFQASSNVQQ
L1_ATLINE1_1_At_gi_12321249_(082–258)	GLEVFEAMNSLWKNCMLVKVLGRSVPIAVLSKKLRELWKPIGAMHVVDLPRQYFMVRFESEEEYLTALTGGPWRVFGSYLLVQAWSPDFDPMKDEIVTTPVWVRLSNIPLNLYHPSILMGITGGLGNLIKVDMTTLTCERARFARVCVEVNLRKPLKGTVMINEDRYFVAYEGLTNI
L1Tx_Tx1L_Xl_gi_214844_(141–241)	GGSYVPVEPLEGLGTRVVLSNVPPFLQDHLLYPHLQALGELKSNMSRIPLGCKESRLRHVLSFKRQVQLLLPRGQDTIEGSFGVPFEGVLYKIFYSTEEVR
Type II NLR ORF1p	
L1_L13_Hs_gi_307098_(157–252)	NLRLIGVPESDVENGTKLENTLQDIIQENFPNLARQANVQIQEIQRTPQRYSSRRATPRHIIVRFTKVEMKEKMLRAAREKGRVTLKGKPIRLTVD
L1(Tx)_L1(Tx)_Nv_gi_149795606_(133–221)	NLRFFGIPEGTNESWNGTEEAVRDFIHKNLKAGPKQAGDVSFERVHRTGTEDKSSPRPIIAKFSFFKDKEEVRSLAKNLAGTSFGIAED
Cr1_Cr1_Nv_gi_149844706_(076–174)	CLEFKGIPSLEDENTNDLVIQVAQLAGVELDEDDISISHRLPAANNREWSDYEGNVHPPSPPTIIAKFVRRDIKDEIYKARFSLKDKTTQDLEHFNCTD
Cr1_Cr1_Sp_gi_111740418_(116–200)	SVRIFGVPESKGEVTDQLVIKAVSDHLPCEISPSDIDRSHRSGKPRPDAKKPRPILVKFTQYKKKAAMMKDRRRLKGSGISIQED
Type III NLR ORF1p	
CR1_Q_Ag_gi_432429_(193–294)	PFTDRIWIRLSAYQRPSLWNKWSLSVKRRLATDDVIAYCLLRRGVSVDSMNWLSFKVRVPAILRDAALTPSTWPVGIGVREFFQSRQHDHQTSSPIATRNRF
CR1_T_Ag_gi_159642_(215–312)	GIAEKVWLYFTNIKSHVSADDMRVWLKAVLPTDNIDVYRLTKKGANLDLMSFISFKVSIPKSLKDLALQSTIWPVSLTVREFVDRGLPKQRIHERARF

Query sequences are labeled according to clade_name_species_gi_number_(residue range).

Table S3. Data collection, phasing, and refinement statistics

	SeMet data (SAD)
Data collection	
Wavelength, Å	0.97154
Resolution range, Å	56 - 1.4
Space group	$P2_1$
Unit cell	
dimensions (a / b / c), Å	32.4 / 54.7 / 57.8
angles (α / β / γ), °	90 / 103.0 / 90
R_{merge}, %	5.4 (53.3)*
Completeness, %	99.4 (96.6)*
Completeness (anomalous), %	97.9 (94.7)*
Mean I/σ(I)	13.9 (2.8)*
Number of unique reflections	38608 (2754)*
Multiplicity	3.7 (3.5)*
Multiplicity (anomalous)	1.9 (1.8)*
Phasing	
R_{cullis}	0.922
Phasing power	0.60
Mean figure of merit	0.20
Refinement	
R_{cryst}, %	13.8
R_{free}, %	18.5
Number of molecules per asymmetric unit	
protein molecules	2
malonate ions	1
atoms (excluding water)	1710
water molecules	298
Average B-factor (anisotropic), Å^2	14.1
Ramachandran plot	
most favored regions, %	98.0
disallowed regions, %	0.0
R.m.s.d. from ideal geometry	
bond lengths, Å	0.018
bond angles, °	1.69

*Values in parentheses correspond to those in the outer resolution shell (1.40–1.44 Å).