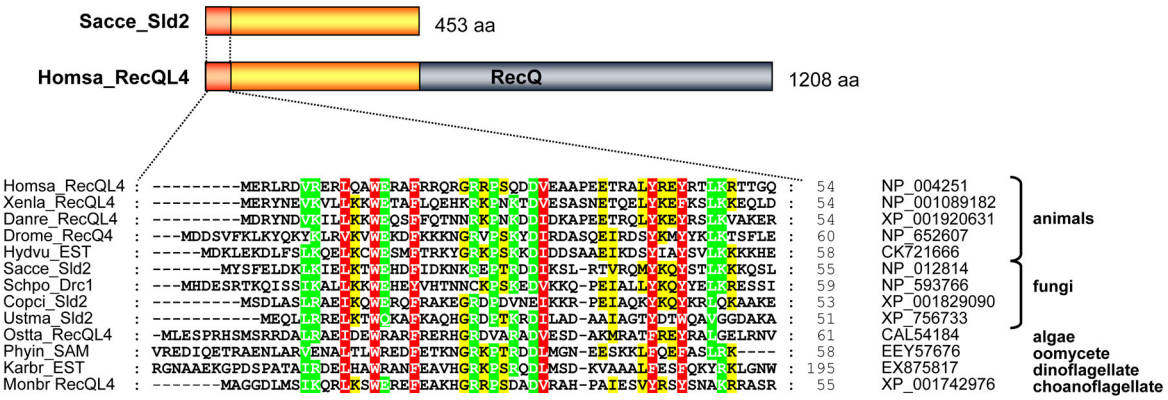
Supplemental Information

The N–terminus of the human RecQL4 helicase is a homeodomain-like DNA interaction motif

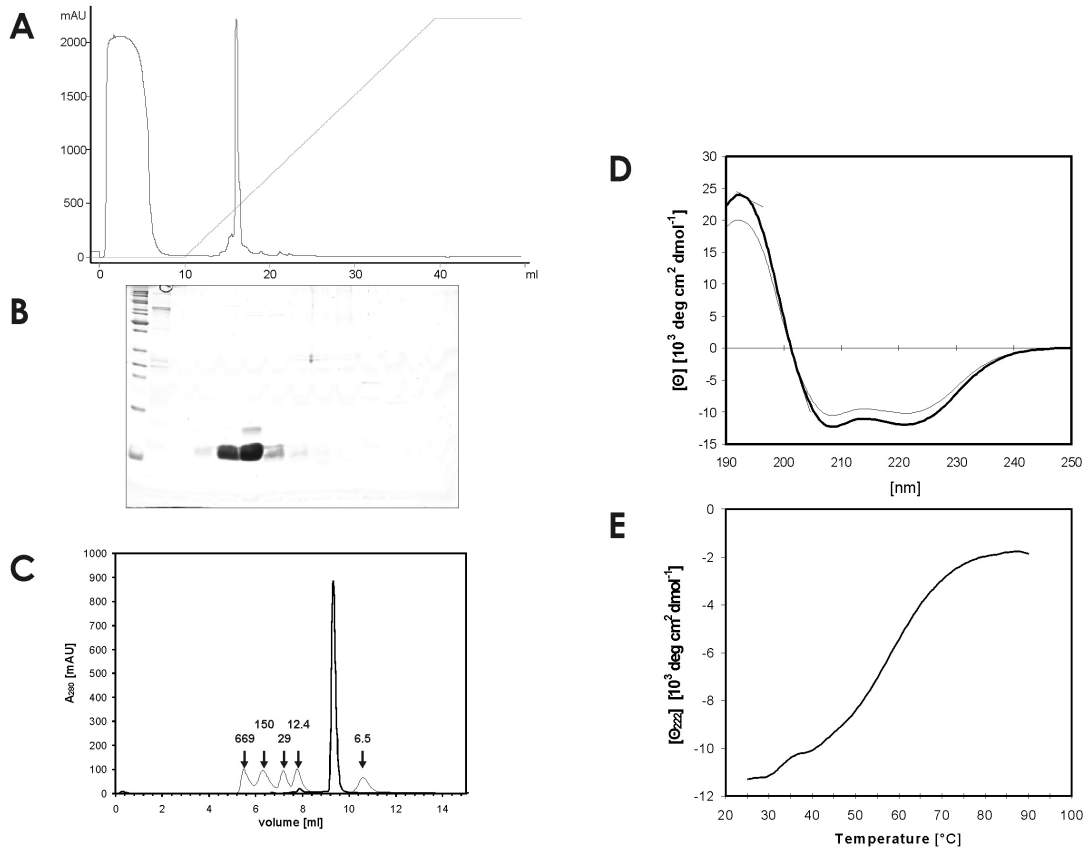
Oliver Ohlenschläger[1+], Anja Kuhnert[3+], Annerose Schneider[2], Sebastian Haumann[1], Peter Bellstedt[1], Heidi Keller[2], Hans-Peter Saluz[3], Peter Hortschansky[4], Frank Hänel[3], Frank Grosse[2], Matthias Görlach[1] and Helmut Pospiech[2,*]

Research groups [1]Biomolecular NMR Spectroscopy and [2]Biochemistry, Leibniz Institute for Age Research – Fritz Lipmann Institute, Beutenbergstr. 11, D–07745 Jena, Germany, [3]Dept. of Cell and Molecular Biology, and [4]Dept. Molecular and Applied Microbiology, Leibniz Institute for Natural Product Research and Infection Biology – Hans Knöll Institute, Beutenbergstr. 11a, D–07745 Jena, Germany

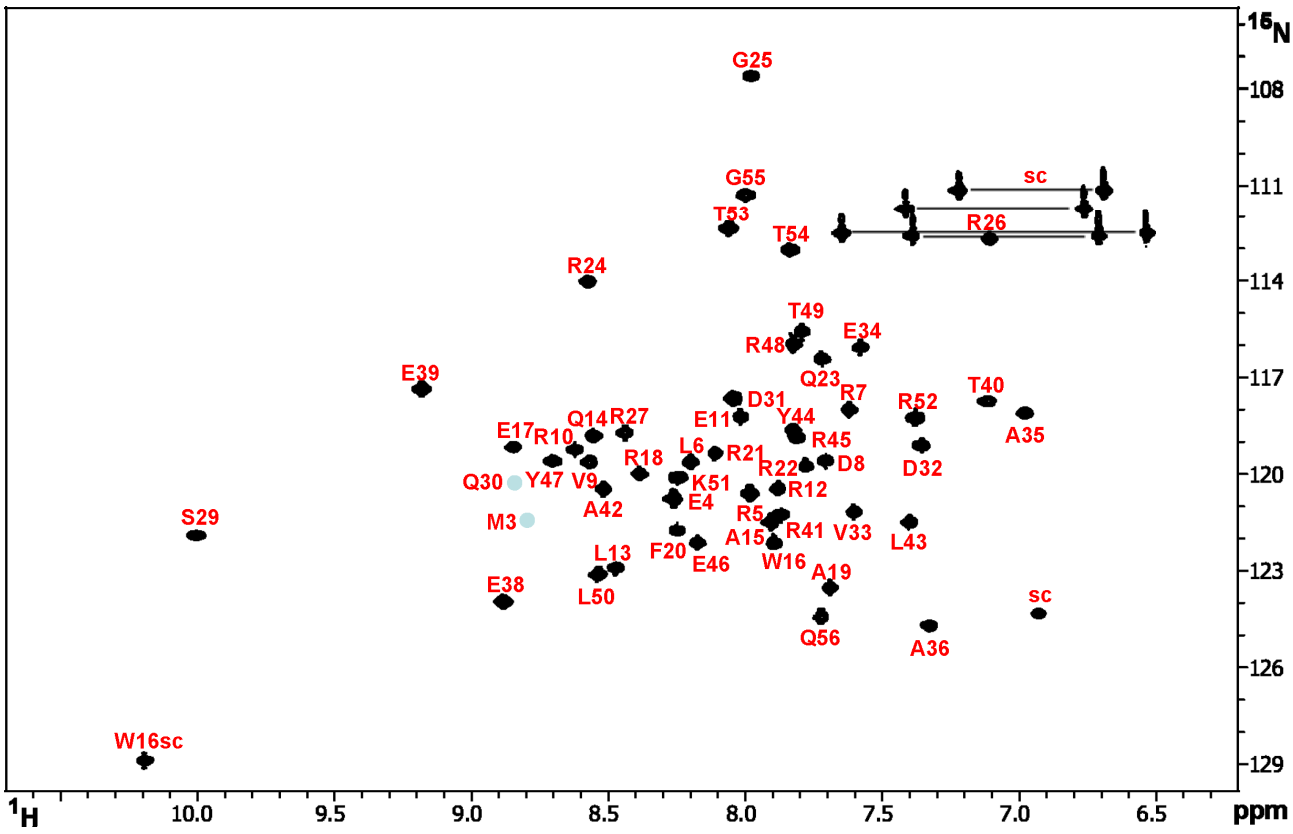
Supplemental Figures



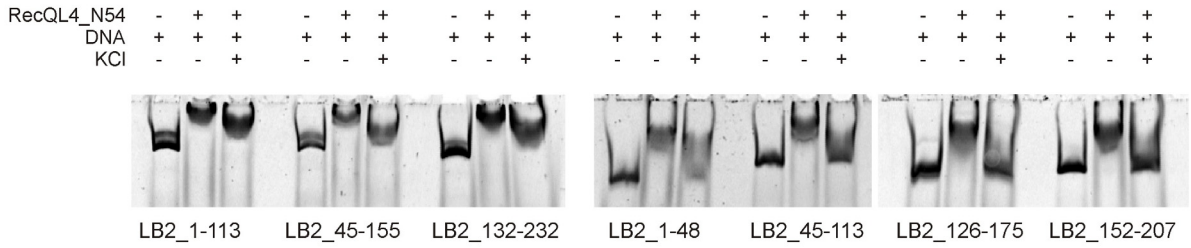
Supplemental Figure S1: Domain organisation of yeast Sld2 and human RecQL4, and conservation of the amino acid sequence in the N-terminal domain of the RecQL4/Sld2 protein family. Genebank accession numbers are given on the right. Abbreviations are as follows: Homsa, *Homo sapiens*; Xenla, *Xenopus laevis*; Danre, *Danio rerio*; Drome, *Drosophila melanogaster*; Hydvu, *Hydra vulgaris*; Sacce, *Saccharomyces cerevisiae*; Copci, *Coprinopsis cinerea*; Ustma, *Ustilago maydis*; Ostta, *Ostreococcus tauri*; Phyin, *Phytophthora infestans*; Karbr, *Karenia brevis*; Monbr, *Monosiga brevicollis*.



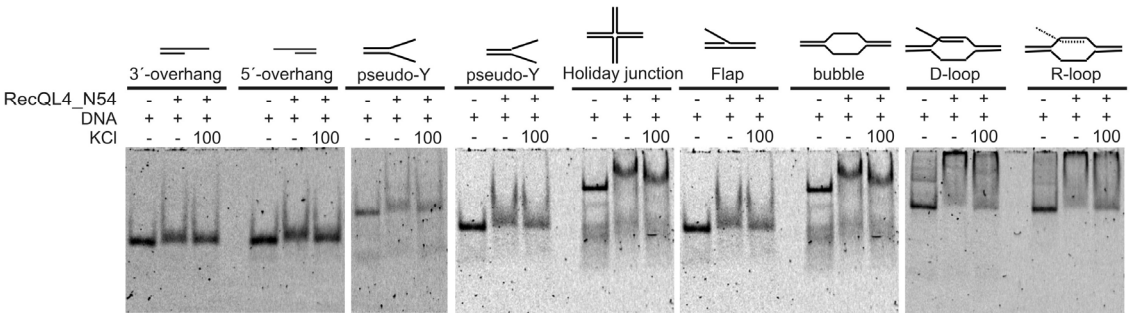
Supplemental Figure S2: Purification and biophysical analysis of RecQL4_N54. **A,** Elution profile of RecQL4_N54 from a MonoS column (top). RecQL4_N54 was purified as described in Materials and Methods. Elution was performed by a linear gradient of 0 – 2 M NaCl in 20 mM potassium phosphate, pH 7.4, 1 mM EDTA, 1 mM DTT with an ÄKTA system. **B**, Equal volumes of the MonoS peak fractions were analysed by SDS-PAGE following Coomassie brilliant blue staining; **C,** Analytical gel filtration. A sample of the pooled RecQL4_N54 MonoS peak was applied to an analytical G3000PWXL gel filtration column developed with 10 mM potassium phosphate, pH 7.4. The position of the molecular weight markers (thyroglobulin, 669 kDa; alcohol dehydrogenase, 150 kDa; carbonic anhydrase, 29 kDa; cytochrom c, 12.4 kDa; aprotinin, 6.5 kDa) analysed in a separate experiment are indicated (grey line); **D,** Far-UV CD spectrum of RecQL4_N54 in 10 mM potassium phosphate, pH 7.4. The grey line represents the corresponding CD spectrum after the temperature scan, indicating the thermal denaturation is largely reversible. **E,** CD spectroscopic melting profile of RecQL4_N54 in 10 mM potassium phosphate. The protein was subjected to increasing temperature in order to evaluate its thermal stability. Amount of secondary structure was monitored as a function of the reduction in ellipticity. The molar ellipticity at 222 nm as function of temperature is shown.



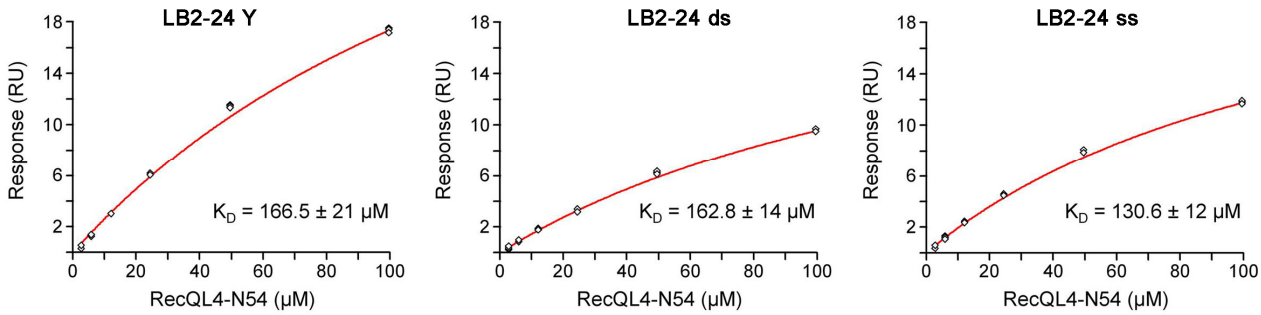
Supplemental Figure S3: [^{1}H,^{15}N]–HSQC spectrum of RecQL4_N54. Backbone amide resonance assignments are indicated as one–letter code. Shaded circles indicate cross peaks observable only at lower thresholds. *Of note*: the first native RecQL4 residue is M3. The numbering of the a.a. positions in RecQL4_N54 starts with two non-native residues derived from the expression construct to maintain consistency with the related PDB (2KMU) and BMRB (16544) data base entries.



Supplemental Figure S4. RecQL4_N54 binds DNA without sequence specificity. dsDNA fragments covering different parts of 232 bp genomic DNA derived from the human replication origin at the LB2 locus were analysed by EMSA. All fragments assessed were shifted with similar efficiency by RecQL4_N54. Binding reactions were performed with RecQL4_N54 (2 µg / ~300 nmol) and 200 ng of each DNA fragment. Details for the DNA fragments are provided in Supplemental Table S1. The DNA binding buffer was supplemented with 100 mM KCl where indicated.



Supplemental Figure S5. RecQL4_N54 binds branched DNA substrates. Binding of RecQL4_N54 to various DNA substrates as schematically depicted on top was assessed by EMSA. Binding reactions were performed with RecQL4_N54 (2 µg / ~300 nmol) and 30 nmol of each DNA fragment. Details for the DNA fragments are provided in Supplemental Table S3. The DNA binding buffer was supplemented with 100 mM KCl where indicated.



Supplemental Figure S6. Real-time binding analysis of RecQL4-N54 to Y-shaped (Y), double stranded (ds) and single stranded (ss) DNA, respectively. Fits of the surface plasmon resonance equilibrium responses for DNA binding of RecQL4_N54 using HEPES buffer containing 150 mM NaCl.

Supplemental Tables

Supplemental Table S1. Primers used to amplify defined human origins of DNA replication.

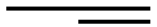
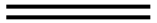
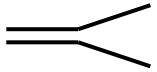
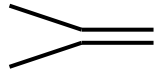
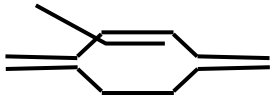
GenBank accession no. / primer	Sequence (5' to 3')	Primer length (bp)	Tm (°C)	Map position (bp)	Product length (bp)
U63630*a*	**Connelly, M. A., H. Zhang, J. Kieleczawa, and C. W. Anderson.** 1998. The promoters for human DNA-PKcs (PRKDC) and MCM4: divergently transcribed genes located at chromosome 8 band q11. Genomics **47:**71–83.				
MCM4upr-F	AATTAAGGACGGGAGGTGGT	20	59.69	13167 - 13408	242
MCM4upr-R	TTGGGTGGCTACTTGGTGTT	20	60.41		
	AATTAAGGAC GGGAGGTGGT GGGGATGGTC TCATTATGTT GCCCAGACTG GTCTCGAACT CCTGCGATCC CCCCGCCTCA GCTTCCTAAA GTGTTGGGAC TGCAGGCGGG CATCACCGCC CCCGACCGCG TTTGTCGTTT TTTATCGAGG AACAAACTTG GAACTCTTGA CCTAGGCCCC TCGCTTGTTT TATCTGCCTC TGGTATTTAT TTAGCCAAGT CCAACACCAA GTAGCCACCC AA				
MCM4in1-F	ATCTCGCCTAATCCCACCAGTACC	24	65.57	14364 – 14656	293
MCM4in2-R	CATATTCACTACTAGACCCTCCGG	24	60.26		
	ATCTCGCCTA ATCCCACCAG TACCTTTTCC AAACCCCTCT ATTTGCCGTT CCTACTTTGA GAAAAATTGA AGAAATGTAA TTGAAGTCAT GTTCTTTTAA CTCTTTAGCA ATGCCTGGGT GGGAACTTTT TCCCCTTTGT AGTAAAAGTG CCTCTTTTTA GCATTTTAAA TATTCAAATA CTTAATTATG TGGATCTCCT GTATTCCTTC TTTAAAACAC GAGTATGCAG ACCTTGATGT TGTGCCCTAT TAAAAGGTGA GATTTTCTTC CGGAGGGTCT AGTAGTGAAT ATG				
M94363*b*	**Biamonti, G., M. Giacca, G. Perini, G. Contreas, L. Zentilin, F. Weighardt, M. Guerra, G. Della Valle, S. Saccone, S. Riva, et al.** 1992. The gene for a novel human lamin maps at a highly transcribed locus of chromosome 19 which replicates at the onset of S phase. Mol. Cell. Biol. **12:**3499–3506.				
LB2-F	GGCTGGCATGGACTTTCATTTCAG	24	67.97	3839 – 4070	232
LB2-R	GTGGAGGGATCTTTCTTAGACATC	24	59.51		
	GGCTGGCATG GACTTTCATT TCAGAGATTC GGTTTTTAAG AAGATGCATG CCTAGCGTGT TCTTTTTTTT TTCCAATGAT TTGTAATATA CATTTTATGA CTGGAAACTT TTTTGTACAA CACTCCAATA AACATTTTGA TTTTAGGTTC TGCCTCTGAG TTTATTCCTG AGGGGAAGCT CGAGCCGGGC CTCTGCCCTA ATGAAGCGGA TGTCTAAGAA AGATCCCTCC AC				
LB2C1-F	GTTAACAGTCAGGCGCATGGGCC	23	70.72	1 – 240	240
LB2C1-R	CCATCAGGGTCACCTCTGGTTCC	23	68.27		
	GTTAACAGTC AGGCGCATGG GCCATGGAGC CTCCTGGGTG TCAGATCCCA GTTCAGCCCC TGCTTGCCAA GAGCCCTTGG CCTCTGTGAA CCCTGGTCTC TTCTTGTGTA TCCTGGGGAC CATCACCCTG GATAAGCCTG CGTCGGGCTG GAAGGACTGA ACGCCCAGGG CATTGAAGGG GCTTTGTGGT AGGGGCCACT GGCTGGCCCA CGTCCCTGGA ACCAGAGGTG ACCCTGATGG				
AL035652	**Keller C, Ladenburger EM, Kremer M, Knippers R** (2002)**.** The origin recognition complex marks a replication origin in the human TOP1 gene promoter. J. Biol. Chem. **277**:31430-40.				
TopProm-F	CACTGCCTAGCAGAGGGGCTG	21	66.74	12461 - 12749	295
TopProm-R	GCAGTTGTGTAACAGCCTAAGTTCG	25	63.39		
	CACTGCCTAG CAGAGGGGCT GGGGTTCCGA GAAAAAGCGT CTGGAGAGGA AGGGGAGGCT GCTTGGGGTG GGACGCCGTG GGAGGAGTCG GCTCCTTATG CAAATCACAG CGGAGCGCGC ACGGTCCGGA GGCGGGGCTT GCGATGCAAA GACAGGTCCG TCTGGCGAAC AGCGAGGGGG CGGGCCGCAA CCCTCTGCCT CTTTCCGCGA GCGCTGACGT CGCCGACGTG TTGTTTAAAA GCGGCCGCGC AGGCGCAGTG AGCCCAAATG CGAACTTAGG CTGTTACACA ACTGC				
TopM1-F	CATGTCTTAAAGTTTATATTCCAAA	25	54.03	17370 - 17550	181
TopM1-R	CAGAATCATTTTCCATACCATTAG	24	56.86		
	CATGTCTTAA AGTTTATATT CCAAAATTGG GCTGTGAGGT TTTTTTGTTT TGCTTGTTTT CAGTGCATGT GTGGGGTGGG GAGGTGCAAA AAAATGTTTG CCTTCTAATA TACAGCTTTG TTGTATATTA ATTATATAGT AACAGTTGTC CCTGCTGCTA ATGGTATGGA AAATGATTCT G				

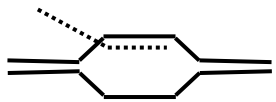
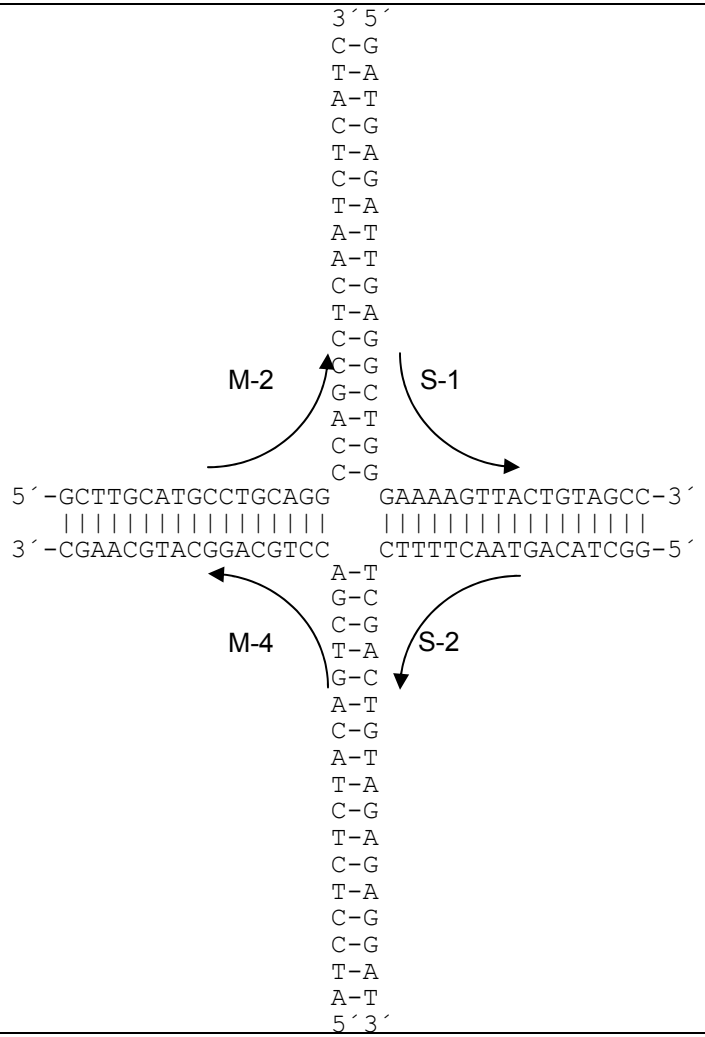
Supplemental Table S2. Primers/oligonucleotides used to produce fragments covering portions of the *LB2* origin of DNA replication region.

Fragment	Primer	Sequence (5' to 3')	Primer length (bp)	Tm (°C)	Map position (bp)[a]	Product length (bp)
LB2	LB2-F	GGCTGGCATGGACTTTCATTTCAG	24	68	3839 – 4070	232
	LB2-R	GTGGAGGGATCTTTCTTAGACATC	24	60		
LB2_ 1-113	LB2-F	GGCTGGCATGGACTTTCATTTCAG	24	68	3839 – 3951	113
	LB2-R113	AAAAAGTTTCCAGTCATAAAATGTATATTACAA	33	59		
LB2_ 45-155	LB2-F45	TGCATGCCTAGCGTGTTCTT	20	58	3884 – 3995	111
	LB2-R155	AGGCAGAACCTAAAATCAAAATGTTTA	27	59		
LB2_ 132-232	LB2-F132	ACATTTTGATTTTAGGTTCTGCCTCT	25	59	3970 – 4070	101
	LB2-R	GTGGAGGGATCTTTCTTAGACATC	24	60		
LB2_ 1-48	LB2-F	GGCTGGCATGGACTTTCATTTCAG	24	68	3839 – 3886	48
	LB2-R48	TGCATCTTCTTAAAAACCGAATCTCT	26	60		
LB2_ 45-113	LB2-F45	TGCATGCCTAGCGTGTTCTT	20	58	3884 – 3951	69
	LB2-R113	AAAAAGTTTCCAGTCATAAAATGTATATTACAA	33	59		
LB2_ 126-175	LB2-F126	CAATAAACATTTTGATTTTAGGTTCTGC	28	59	3965 - 4053	50
	LB2-R175	CCCCTCAGGAATAAACTCAGAGG	23	59		
LB2_ 152-207	LB2-F152	GCCTCTGAGTTTATTCCTGAGGG	23	59	3991- 4045	56
	LB2-R207	GCTTCATTAGGGCAGAGGCC	20	60		
LB2_ 1-24	LB2-F	GGCTGGCATGGACTTTCATTTCAG	24	68	3839 – 3862	24
	LB2-R24	CTGAAATGAAAGTCCATGCCAGCC	24	63		
LB2_ 1-24Y	LB2-F	GGCTGGCATGGACTTTCATTTCAG	24	68	n.a.	n.a.
	LB2-R24Y	GACTTTACAAAGTCCATGCCAGCC	24	63		
LB2_ 13-36	LB2-F13	CTTTCATTTCAGAGATTCGGTTTT	24	56	3851 – 3874	24
	LB2-R36	AAAACCGAATCTCTGAAATGAAAG	24	56		
LB2_ 23-48	LB2-F23	AGAGATTCGGTTTTTAAGAAGATGCA	26	68	3861 – 3886	26
	LB2-R48	TGCATCTTCTTAAAAACCGAATCTCT	26	59		

[a]Map positions are indicated according to: **Biamonti, G., M. Giacca, G. Perini, G. Contreas, L. Zentilin, F. Weighardt, M. Guerra, G. Della Valle, S. Saccone, S. Riva, et al.** 1992. The gene for a novel human lamin maps at a highly transcribed locus of chromosome 19 which replicates at the onset of S phase. Mol. Cell. Biol. **12:**3499–3506.

Supplemental Table S3. Oligonucleotides used to assemble the DNA substrates utilized in Supplemental Fig. S5.

Construct	Sequence	Structure
M2-M1	5′-GCTTGCATGCCTGCAGGCCAGCCTCAATCTCATC |||||||||||||||| 3′-CGAACGTACGGACGTCC	3´-overhang 
M3-M1	5′-CTACTCTAACTCCGACCGCTTGCATGCCTGCAGG |||||||||||||||| 3′- CGAACGTACGGACGTCC	5´-overhang 
M3-S3	5′-CTACTCTAACTCCGACCGCTTGCATGCCTGCAGG |||||||||||||||||||||||||||||||||| 3′-GATGAGATTGAGGCTGGCGAACGTACGGACGTCC	dsDNA 
M2-M4	CCAGCCTCAATCTCATC 5′-GCTTGCATGCCTGCAGG |||||||||||||||| 3′-CGAACGTACGGACGTCC AGCTGAGATCTCTCCTA	Y-shaped DNA 
S4-M3	5′-ATCCTCTCTAGAGTCAA GGTCGGAGTTAGAGTAG |||||||||||||||| CCAGCCTCAATCTCATC 3′-GGACGTCCGTACGTTCG	Y--shaped DNA 
M1-S4-M3	5′-ATCCTCTCTAGAGTCAA \ 5′-CCTGCAGGCATGCAAGC GGTCGGAGTTAGAGTAG |||||||||||||||| |||||||||||||||| 3′-GGACGTCCGTACGTTCG-CCAGCCTCAATCTCATC	FLAP 
M5-M6	CGACCTGCAGGCATGCAAGCTT 5′-GTACCCGGGGATCCTCTAGAGT GGCACTGGCCGTCGTTTTACAAC |||||||||||||||||||| |||||||||||||||| 3′-CATGGGCCCCTAGGAGATCTCA CCGTGACCGGCAGCAAAATGTTG CTACTCTAACTCCGACCCTTTT	bubble 
M5-M6-M2	3′- CTACTCTAACTCCGACC CGACCTGCAGGCATGCAAGCTT 5′-GTACCCGGGGATCCTCTAGAGT |||||||||||||||| GGCACTGGCCGTCGTTTTACAAC |||||||||||||||||||| GGACGTCCGTACGTTCG |||||||||||||||||||| 3′-CATGGGCCCCTAGGAGATCTCA CCGTGACCGGCAGCAAAATGTTG CTACTCTAACTCCGACCCTTTT	D-loop 

M5-M6-R2	3′- CUACUCTAACUCCGACC CGACCTGCAGGCATGCAAGCTT 5′-GTACCCGGGGATCCTCTAGAGT |||||||||||||||| GGCACTGGCCGTCGTTTTACAAC |||||||||||||||||||| GGACGUCCGUACGUUCG |||||||||||||||||||||||| 3′-CATGGGCCCCTAGGAGATCTCA CCGTGACCGGCAGCAAAATGTTG CTACTCTAACTCCGACCCTTTT	R-loop 
M2-S1-M4-S2		Hollyday junction

Supplemental Table S4. SPR equilibrium data for DNA binding of RecQL4_N54 (HBS-EP150)

DNA	MW DNA (Da)	DNA bound (RU)	R_{max} calc.* (RU)	R_{max} (RU)	% of R_{max}	K_D (μM)
LB2-24 Y	15061	45.0	20.4	46.3	227.0	166.5 ± 21
LB2-24 ds	15110	43.7	19.8	27.1	136.9	162.8 ± 14
LB2-24 ss	7780	23.4	20.6	25.1	121.8	130.6 ± 12

* R_{max} calc. = 6836 Da (MW RecQL4-N54) / Mw DNA x DNA bound

Supplemental Methods

CD spectroscopy. CD spectroscopy was carried out at 20°C using a Jasco J–715 spectropolarimeter with a Jasco PFD–3505 temperature controller. Far–UV spectra of the protein (135 μg/ml – 20 μM) were recorded from 190 to 250 nm in 10 mM potassium phosphate, pH 7.5, with the following settings: response time 1 s; speed 50 nm/min; path length 1 mm; band width 1 nm; average of 5 scans. The mean molar ellipticity was calculated with the Jasco software. Protein melting was monitored simultaneously at 192, 208 and 222 nm in steps of 1°C from 20 to 98°C with the following instrument settings: temperature increase 30°C/h; response time 16 s; and path length 1 mm. Reversibility of the unfolding was assessed from a far-UV spectrum as described above recorded after the temperature scan.

In silico analyses. Homologues of RecQL4_N54 were retrieved from the NCBI gene bank using PSI-BLAST (65) and the amino acid of the amino–terminal region of human RecQL4 as a bait. Additional sequences were retrieved from the Genebank EST collection using the TBLASTN program and human RecQL4_N54 or its homologues as baits. Multiple sequence alignment was generated from the PSI-BLAST output and managed with the GENEDOC sequence editor (66). Secondary structure prediction data were generated at the PredictProtein server (67). Regions of disorder were predicted using the DISMETA server (www-nmr.cabm.rutgers.edu/bioinformatics/disorder). Structural homologues of RecQL4_N54 were identified with the DALI server (www.ebi.ac.uk/dali) (32) using the structural model 1 as a bait.

Supplemental References

65. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, **25**, 3389-3402.

66. Nicholas, K.B., Nicholas, H.B., Jr. and Deerfield, D.W., 2nd. (1997) GeneDoc: Analysis and visualization of genetic variation. *EMBNEW News*, **4**.

67. Rost, B., Yachdav, G. and Liu, J. (2004) The PredictProtein server. *Nucleic Acids Res*, **32**, W321-326.