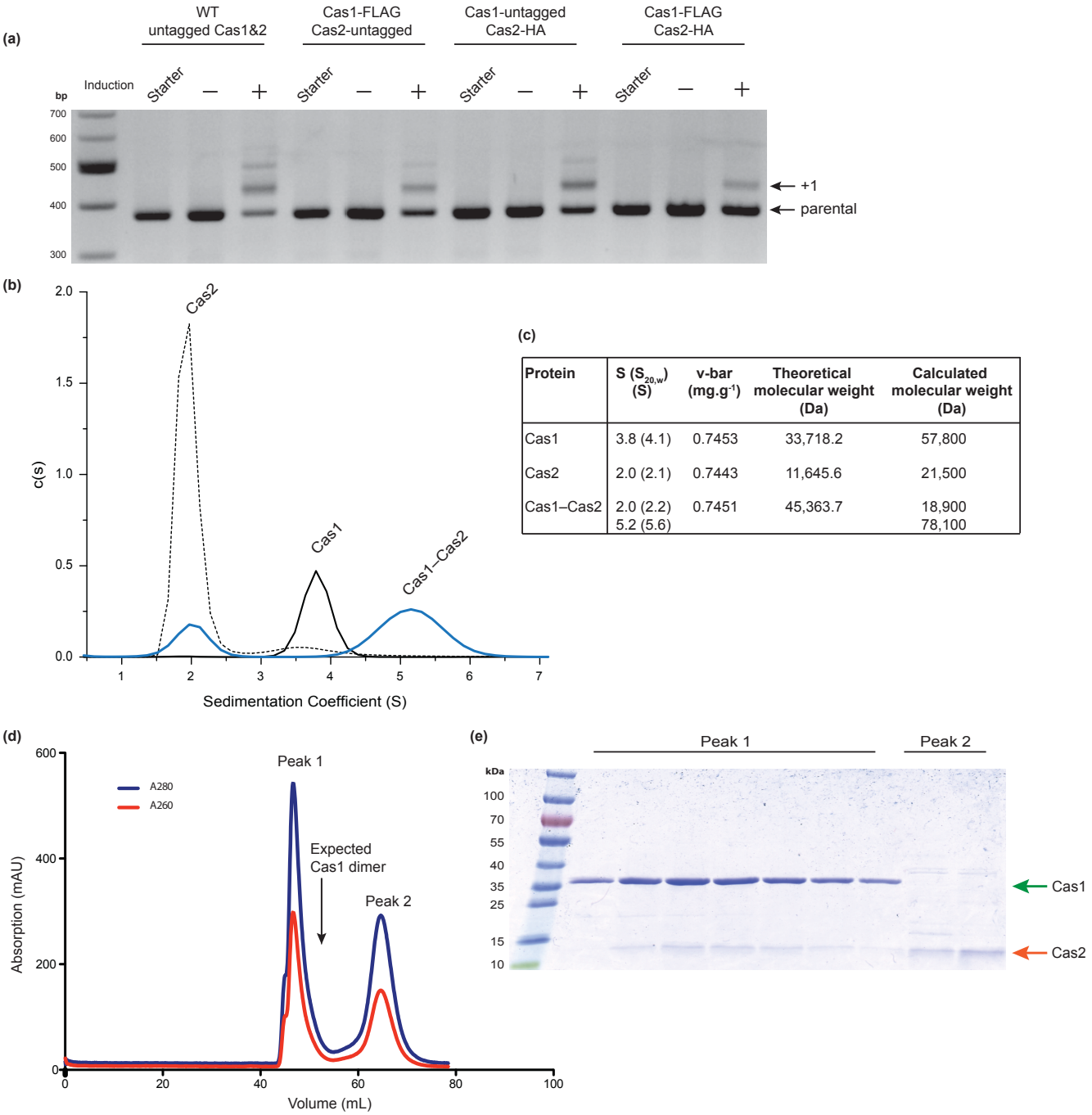
Supplementary Information

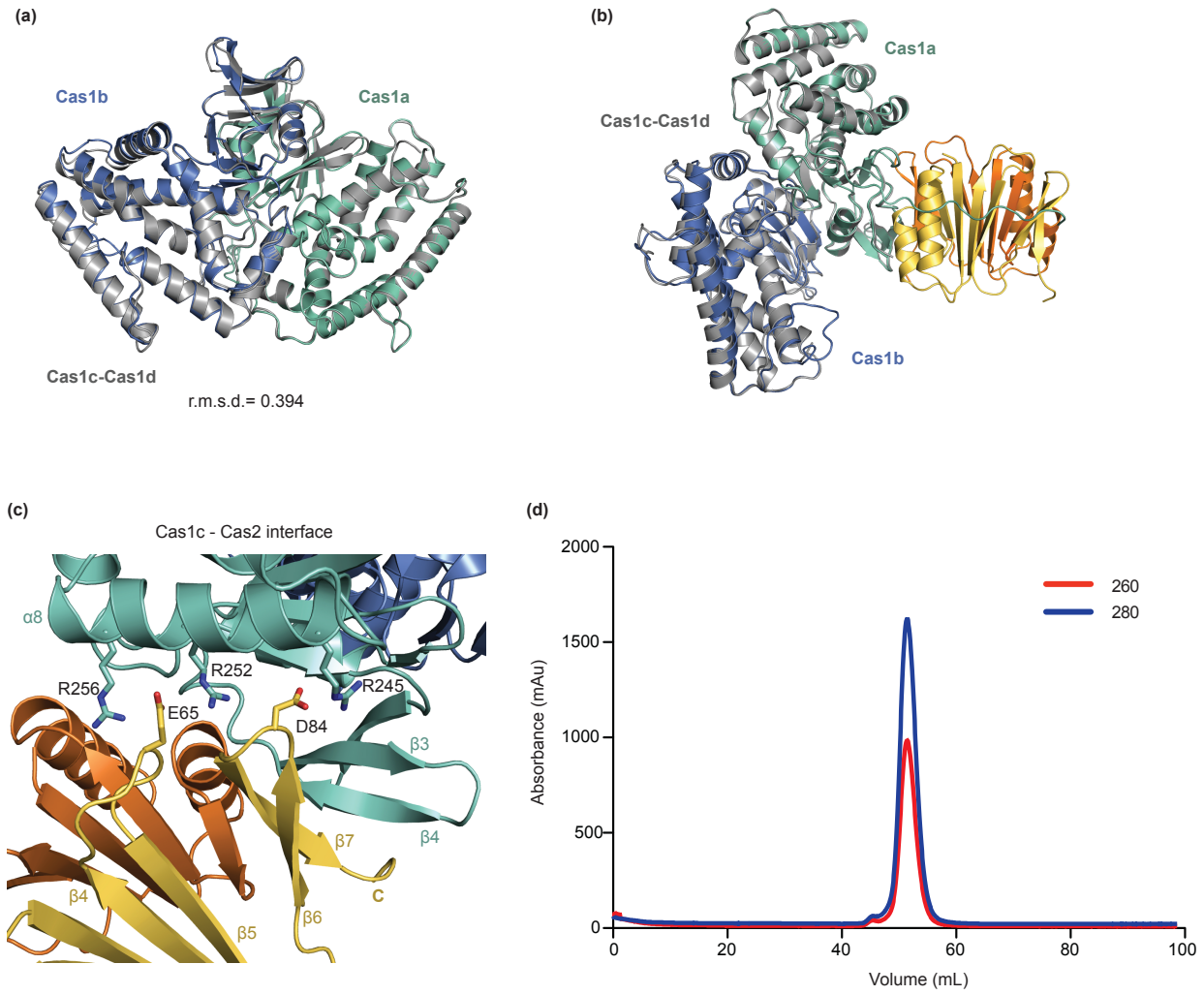
Cas1–Cas2 complex formation mediates spacer acquisition during CRISPR–Cas adaptive immunity

James K. Nuñez, Philip J. Kranzusch, Jonas Noeske, Addison V. Wright, Christopher W. Davies & Jennifer A. Doudna

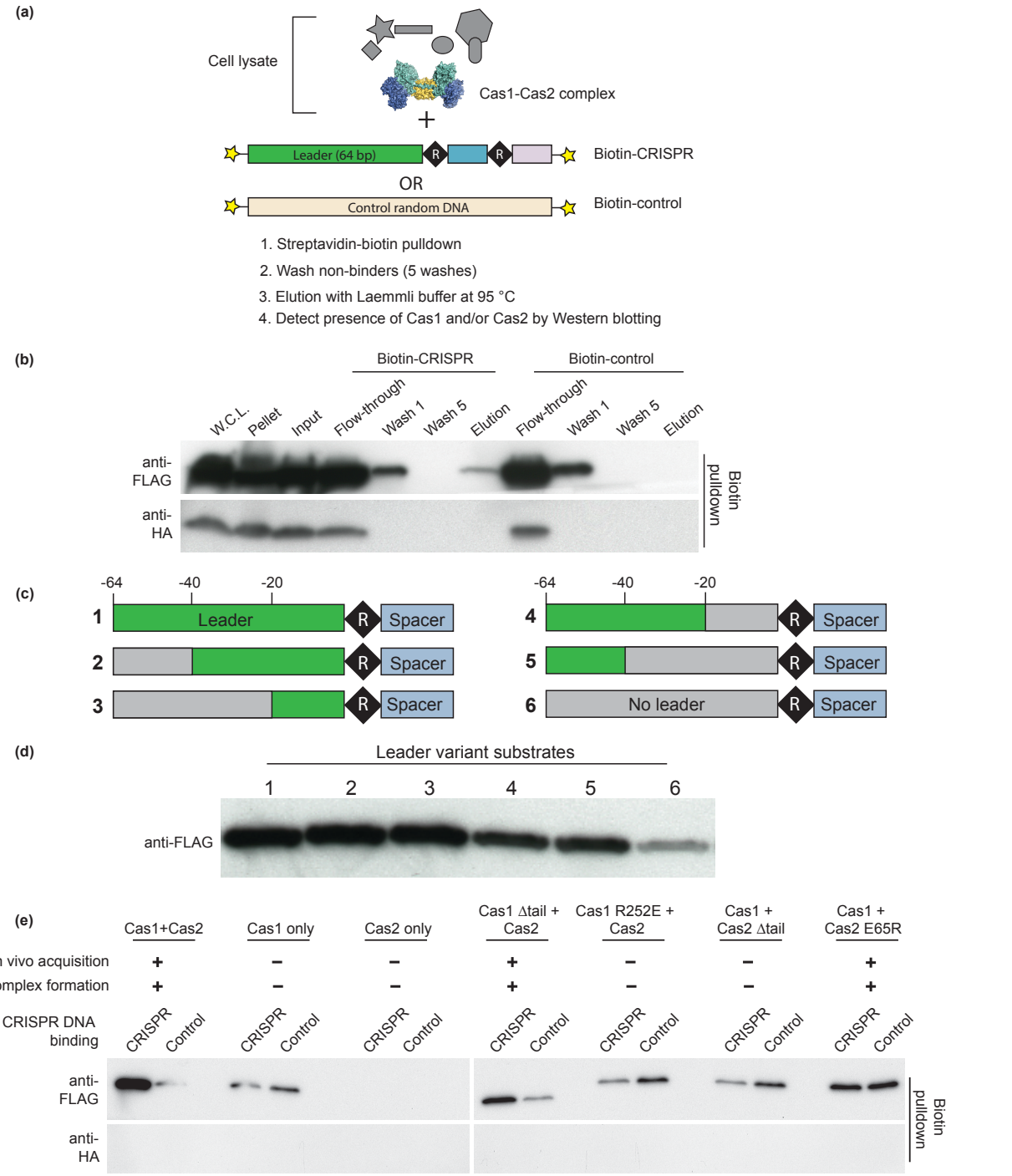


Protein	S ($S_{20,w}$) (S)	v-bar ($mg.g^{-1}$)	Theoretical molecular weight (Da)	Calculated molecular weight (Da)
Cas1	3.8 (4.1)	0.7453	33,718.2	57,800
Cas2	2.0 (2.1)	0.7443	11,645.6	21,500
Cas1–Cas2	2.0 (2.2) 5.2 (5.6)	0.7451	45,363.7	18,900 78,100

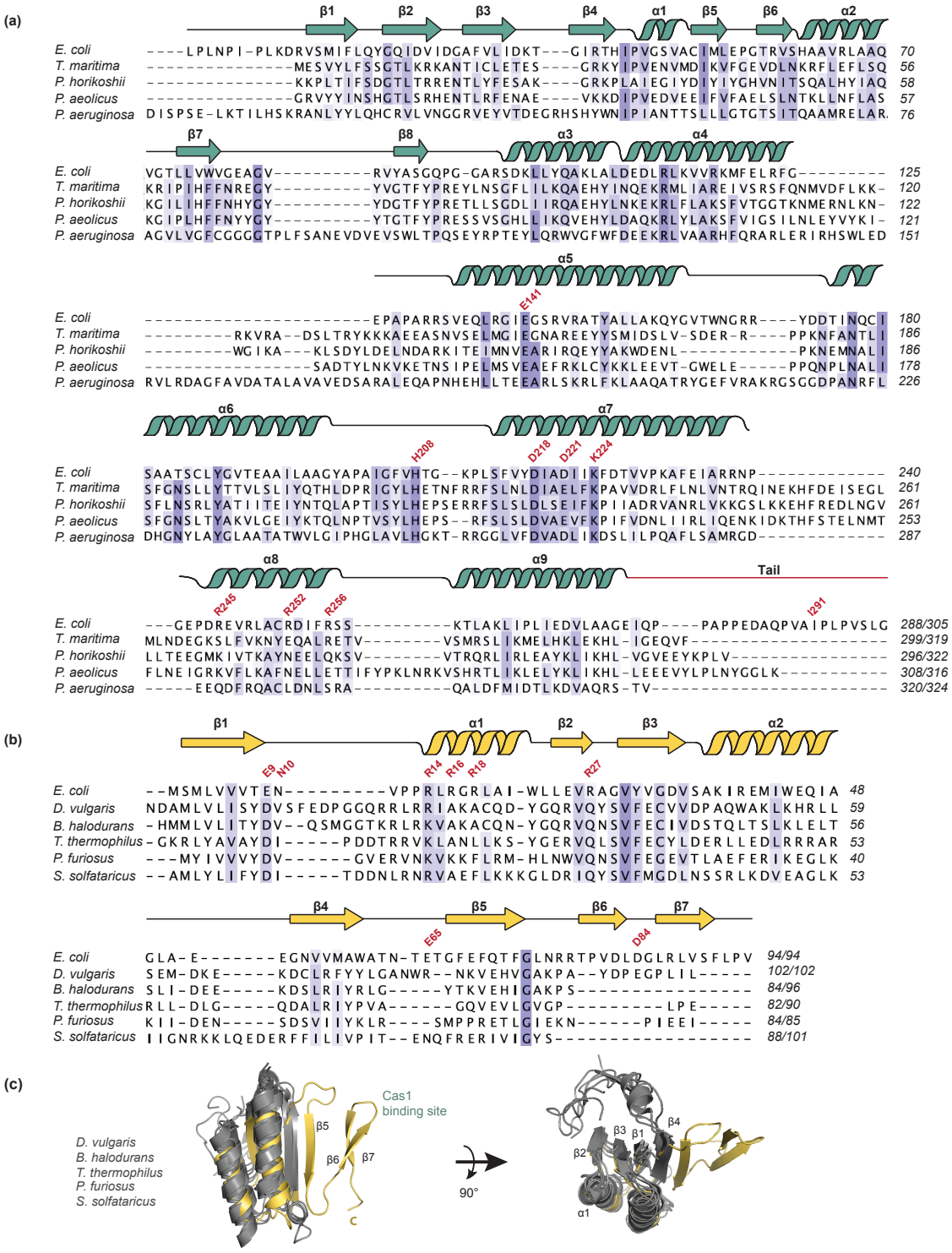
Supplementary Figure 1. *In vivo* acquisition with epitope-tagged Cas1 and Cas2 and *in vitro* reconstitution of the complex. (**a**) Agarose gel of acquisition assays in BL21-AI cells overexpressing Cas1 and Cas2 with or without epitope tags. Lanes labeled 'starter' indicate the starter culture before inoculation into cultures in inducing (+) or non-inducing (-) conditions. The Cas1-FLAG and Cas2-HA constructs were used for the immunoprecipitation and DNA affinity precipitation experiments in this study. (**b**) An overlay of the c(s) distributions of Cas1 only (solid black), Cas2 only (dotted) and Cas1–Cas2 complex (solid blue). (**c**) AUC data table highlighting the s-values and the calculated apparent molecular weights. (**d**) Gel filtration chromatogram of pre-incubated, separately purified Cas1 and Cas2, as described in the Methods section. The arrow points to the expected elution peak of Cas1 dimer, based on our protein purification. (**e**) Coomassie-stained SDS-PAGE of fractions corresponding to the two peaks in (**b**). The green arrow points to Cas1 (33.7 kDa) and the orange arrow points to Cas2 (11.6 kDa).



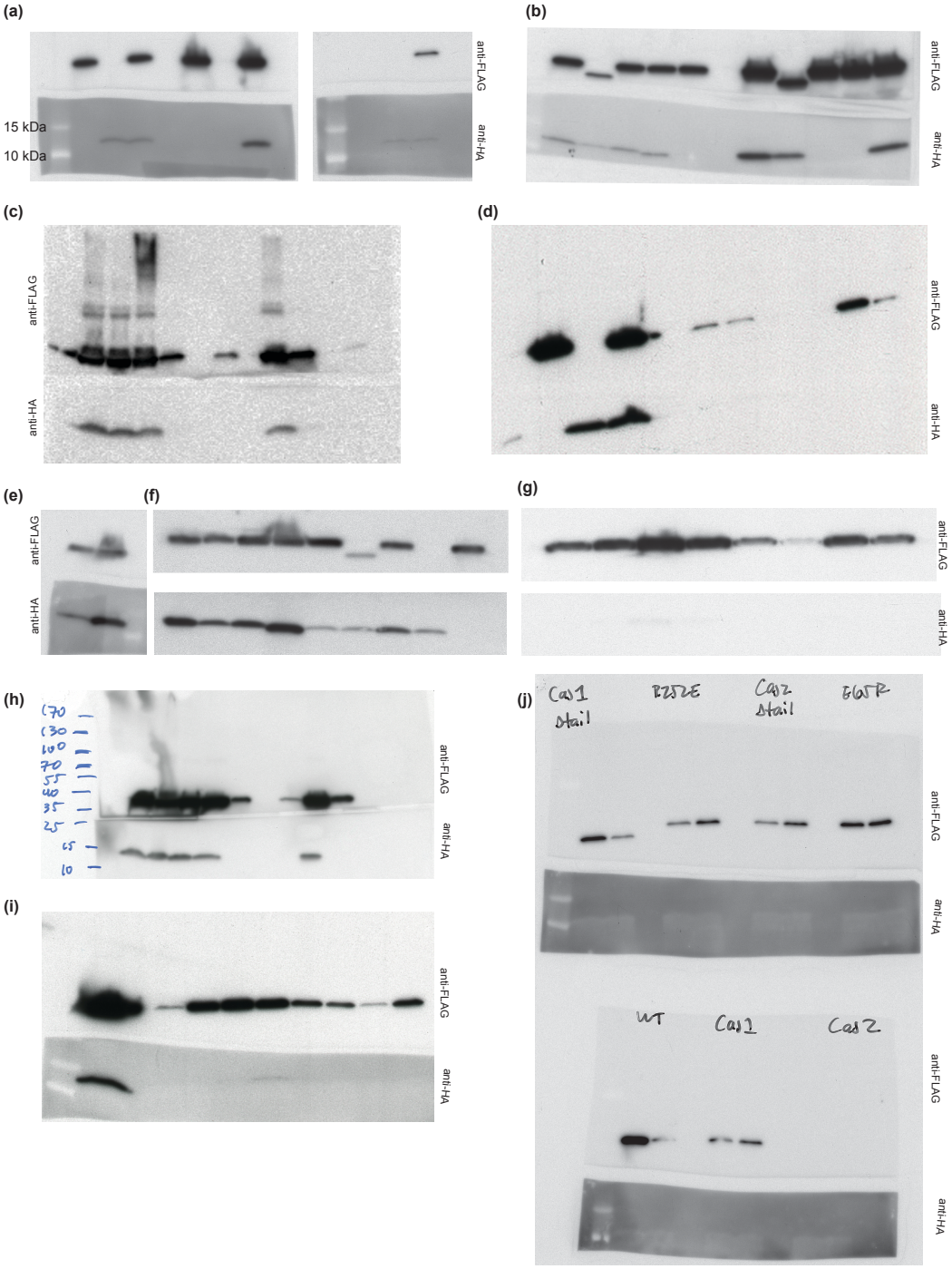
Supplementary Figure 2. The two Cas1 dimers are structurally similar. (**a,b**) Two views of the superposition of the Cas1a-b dimer (teal and blue) with the Cas1c-d dimer (gray). The root-mean-square deviation for the C-alpha backbone is 0.394. (**c**) The Cas1c–Cas2 interface is similar to the Cas1a–Cas2 interface, shown in Figure 3b. (**d**) Gel filtration chromatogram of purified Cas1 R252E using a Superdex 75 (16/60) size exclusion column.



Supplementary Figure 3. Cas1–Cas2 complex formation is required for CRISPR DNA binding. (**a**) A schematic of the biotinylated DNA affinity precipitation experiments conducted in this study, as further described in the Methods section. (**b**) Western blot of the fractions throughout the experiment using magnetic Streptavidin beads, as opposed to streptavidin-agarose resin shown in Fig. 4e. (**c**) A schematic of the six 125-bp, biotinylated DNA substrates with scrambled regions within the leader sequence, shown in gray. Substrate 6 has a random DNA sequence upstream of the repeat with no similarity to the leader sequence (**d**) Western blot to detect Cas1 levels in the elution fractions of DNA affinity precipitations using the six biotinylated DNA substrates in BL21-AI cells overexpressiong Cas1 and Cas2. (**e**) Elution samples of DNA affinity precipitations in lysates from BL21-AI cells overexpressing the indicated Cas1 and Cas2 mutants. The +/- annotations are results from Fig. 3.



Supplementary Figure 4. Structure-based Cas1 and Cas2 alignments. (**a,b**) The *E. coli* Cas1 (**a**) or Cas2 (**b**) protein was aligned to its respective homologs for which crystal structures are available, as described in the Methods section. The number annotations at the C-termini refer to the last amino acid residue resolved in the structure, followed by the last residue of the full-length protein. The secondary structure cartoon is for the *E. coli* protein and the annotated residues (red) refer to the ones that were mutated in this study. The BLOSUM62 score conservation threshold is set to 50%, as reflected by the blue colors of the alignment. (**c**) Two views of a structure alignment of the Cas2 in the Cas1–Cas2 complex with all the available crystal structures of Cas2 homologs (all shown in monomers).



Supplementary Figure 5. Uncropped western blots of immunoprecipitation and DNA affinity precipitation assays. (**a**) Related to Fig. 1c. (**b**) Related to Fig. 3f. (**c**) Related to Fig. 3e. (**d**) Related to Fig. 3f. (**e**) Related to Fig. 3g. (**f**) Related to Fig. 3h. (**g**) Related to Fig. 3h. (**h-j**) Related to Supplementary Fig. 3.

Source	Strand	PAM	Length	Sequence	Location
plasmid	+	ATG	33	GTTAGTCATGCCCCGCGCCCACCGGAAGGAGCT	
plasmid	+	TCG	33	GCTCGAGCTGAAGGAGATATACCATGAGTATGT	*cas1–rbs–cas2*
plasmid	+	AAG	33	GAATGTCATTGCGCTGCCATTCTCCAAATTGCA	
plasmid	-	TTG	33	GTATTTCTCCAGCGGCAAGCACGTCCTCTATAA	*cas1*
plasmid	-	CCG	33	GGAAGCAGTGTGACCGTGTGCTTCTCAAATGCC	
plasmid	-	TGT	33	TTATATCCCGCCGTTAACCACCATCAAACAGGA	*lacI*
plasmid	-	AAG	33	GCATAAACACGAACGCCCGCTTCCCCCACCCAT	*cas1*
plasmid	-	GTG	33	GGAAGCGGCGATGGCGGAGCTGAATTACATTCC	*lacI*
plasmid	+	CAG	33	GCAATGGCATCCTGGTCATCCAGCGGATAGTTA	*lacI*
plasmid	-	ACT	33	TGGAAAGCGGGCAGTGAGCGCAACGCAATTAAT	*lacI*
plasmid	+	TGC	33	CTGAAACCTCAGGCATTTGAGAAGCACACGGTC	
plasmid	-	TTA	33	AATTAAGCTGCGCTAGTAGACGAGTCCATGTGC	
plasmid	-	AAG	33	GATCATGGAGACGCGATCTTTGAGTGGAATGGG	*cas1*
plasmid	+	AAT	33	TGAATCGGCCAACGCGCGGGGAGAGGCGGTTTG	*lacI*
plasmid	-	AAT	33	TGGCAACAGGCTGTGCATCTTCAGGTGGGGCCG	*cas1*
plasmid	+	GCG	33	GGGAAACGGTCTGATAAGAGACACCGGCATACT	*lacI*
plasmid	-	CAC	33	CCCTGGCGCCCAATACGCAAACCGCCTCTCCCC	*lacI*
plasmid	-	GAT	33	TGGCTAATCTGCCTCGTAAGCGCGGAGGTACAT	*cas2*
plasmid	-	CCC	33	CAATACGCAAACCGCCTCTCCCCGCGCGTTGGC	*lacI*
plasmid	-	AAG	34	GTCCTGTCGGGTTTCGCCCGTACTGTCAGATTCA	
plasmid	+	AAG	33	GTCAGCCCCATACGATATAAGTTGTAATTCTCA	
plasmid	+	GGA	33	AGAAATACAACCGCCGGCCCCACCTGAAGATGC	*cas1*
plasmid	+	TAA	33	AACAAATAGCTAGCTCACTCGGTCGCTACGCTC	
plasmid	-	AAG	34	GCGGCGATGGCGGAGCTGAATTACATTCCCAACC	*lacI*
plasmid	+	TGG	33	GCGATGACCTGGCTTCCCCTTAATCCCATTCCA	*cas1*
plasmid	+	GCG	33	GGGAAACGGTCTGATAAGAGACACCGGCATACT	*lacI*
genomic	+	AAG	33	GTCCAGTTGTGCGGGAAGACTTCGAGCATTTGT	*mannonate/altronate dehydratase*
genomic	+	GTT	33	TAATACCGTTGAAATGATGGTCCATATCCATTG	*serine acetyltransferase*
genomic	+	AAT	33	TCTTGATTCCTGAACTGATAGGCTACCTGGCGA	*methyl-directed mismatch repair protein*
genomic	-	GTA	33	ATGCAGGCATGATAGCAAAAATGGCGAGGATGG	*Yick MFS transporter & inhibitor of heme biosynthesis*
genomic	-	AAG	33	GCCATTAATGGCGCAGGCGCTGCATATGGGGGA	*protein of unknown function*
genomic	+	GAT	33	TGGCAGTTTCAATCAGTTCTTGCCGGGTCTTCA	*EnvR DNA binding transcriptional repressor*

Supplementary Table 1. Sequences of newly acquired spacers. A list of a sample of new spacers acquired into the CRISPR locus of BL21-AI. The PAM is defined as the third base being the first nucleotide of the spacer and the first two nucleotides are the −1 and −2 position in the DNA source[5,28].