www.sciencemag.org/cgi/content/full/science.1247997/DC1

Supplementary Materials for

Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation

Martin Jinek,* Fuguo Jiang, David W. Taylor, Samuel H. Sternberg, Emine Kaya, Enbo Ma, Carolin Anders, Michael Hauer, Kaihong Zhou, Steven Lin, Matias Kaplan, Anthony T. Iavarone, Emmanuelle Charpentier, Eva Nogales,* Jennifer A. Doudna*

*Corresponding author. E-mail: jinek@bioc.uzh.ch (M.J.); enogales@lbl.gov (E.N.); doudna@berkeley.edu (J.A.D.)

Published 6 February 2014 on *Science* Express
DOI: 10.1126/science.1247997

This PDF file includes:

Materials and Methods

SpyCas9 expression and purification

Streptococcus pyogenes Cas9 (SpyCas9) was cloned into a custom pET-based expression vector encoding an N-terminal His_6-tag followed by Maltose-Binding Protein (MBP) and a TEV protease cleavage site (*8*). Point mutations were introduced into SpyCas9 using site-directed mutagenesis and verified by DNA sequencing.

For crystallization, wild-type (WT) and K848C mutant SpyCas9 proteins were expressed and purified essentially as described (*8*). The protein was purified by a combination of Ni-NTA affinity, cation exchange (SP sepharose) and gel filtration (Superdex 200) chromatography steps. The final gel filtration step was carried out in elution buffer containing 20 mM HEPES-KOH pH 7.5, 250 mM KCl and 1 mM TCEP. The protein was concentrated to 4-6 mg ml^{-1} and flash frozen in liquid N_2. Selenomethionine (SeMet)-substituted SpyCas9 was expressed as described (*62*) and purified as for native SpyCas9, except that all chromatographic solutions were supplemented with 5 mM TCEP.

For crosslinking and biochemical assays, WT and mutant SpyCas9 proteins were expressed as His_{10}-MBP-TEV fusions and purified as described (*8*), with the following modifications: All buffers contained 20 mM Tris-Cl pH 7.5, 5% glycerol, and 1 mM TCEP. The NaCl concentration was maintained at 500 mM during Ni-NTA chromatography and overnight dialysis with TEV protease. In order to remove TEV protease, His_{10}-MBP, and any uncleaved His_{10}-MBP-SpyCas9, the TEV-treated protein sample was run over Ni-NTA agarose resin again. SpyCas9 was dialyzed into Buffer A (20 mM Tris-Cl pH 7.5, 125 mM KCl, 5% glycerol, 1 mM TCEP) for 3 h at 4°C, and then applied onto a 5 ml HiTrap SP HP sepharose column (GE Healthcare). After washing with three column volumes of Buffer A, SpyCas9 was eluted using a linear gradient from 0-100% Buffer B (20 mM Tris-Cl pH 7.5, 1 M KCl, 5% glycerol, 1 mM TCEP) over 20 column volumes. The protein was further purified by gel filtration chromatography on a Superdex 200 16/60 column (GE Healthcare) in SpyCas9 Storage Buffer (20 mM Tris-Cl pH 7.5, 200 mM KCl, 5% glycerol, 1 mM TCEP).

SpyCas9 crystallization and structure determination

SpyCas9 crystals were grown using the hanging drop vapor diffusion method at 20 °C by mixing equal volumes (1.5 μl + 1.5 μl) of protein solution and crystallization buffer (0.1 M Tris-Cl pH 8.5, 0.2-0.3 M Li_2SO_4 and 14-15% (w/v) PEG 3350). Crystal nucleation and growth was gradually improved using iterative microseeding. For diffraction experiments, the crystals were cryoprotected in situ by stepwise exchange into a solution containing 0.1 M Tris-Cl pH 8.5, 0.1 M Li_2SO_4, 35% (w/v) PEG 3350, and 10% ethylene glycol in five steps executed at 5 min intervals. In each step, 0.5 μl of mother liquor was removed from the crystal drop and replaced with 0.5 μl cryoprotectant. After the final cryoprotectant addition, the crystals were incubated for an additional 5 min, transferred to a drop containing 100% cryoprotectant for 30 s, and then flash cooled in liquid N_2. Diffraction data were measured at beamlines 8.2.1 and 8.2.2 of the Advanced Light Source (Lawrence Berkeley National Laboratory), and beamlines PXI and PXIII of the Swiss Light Source (Paul Scherer Institute) and processed using XDS (*49*). Data collection statistics are shown in Table 1. The crystals belonged to space group $P2_12_12$ and contained two molecules of SpyCas9 in the asymmetric unit related by pseudotranslational, non-crystallographic symmetry. High-resolution native data to 2.62 Å resolution were measured from an unusually large crystal cryoprotected in the presence of 1 mM $MgCl_2$. A complete native data set was obtained by collecting four datasets (40° rotation per dataset) from different exposed parts of the crystal.

Phasing was performed as follows. A 4.2 Å resolution single-wavelength anomalous diffraction (SAD) dataset was measured at the selenium peak wavelength using a SeMet-substituted SpyCas9 crystal. However, due to small crystal size and low resolution, the anomalous signal in this dataset was too weak to locate the selenium sites. Additional phases were therefore obtained from SpyCas9 crystals soaked in sodium tungstate. The crystals were soaked by stepwise exchange of the lithium sulfate containing mother liquor with 0.1 M Tris-Cl pH 8.5, 0.1 M Na_2WO_4, 15% (w/v) PEG 3350, and then cryoprotected by stepwise exchange (as described above) of the soak solution with cryoprotectant solution supplemented with 10 mM Na_2WO_4. Using these crystals, a highly redundant SAD 3.9 Å dataset was measured at the tungsten L-III absorption edge (1.2149 Å), and 16 tungstate sites were located using SHELXD (*63*). Further phase information came from peak-wavelength SAD datasets obtained from a crystal of SpyCas9 K848C mutant soaked in 1 mM thimerosal for 6 hr prior to cryoprotection

(thimerosal soak), a WT SpyCas9 crystal soaked with 10 mM $CoCl_2$ during the cryoprotection procedure (Co soak), and a WT SpyCas9 crystal grown in the presence of 1 mM Er(III)-acetate. Refinement of the substructures and phase calculations were performed using the MIRAS procedure in AutoSHARP (*50*) by combining initial tungstate SAD phases with the additional SAD data sets (SeMet, Co, Er and thimerosal) and the high-resolution native data. Phases were improved by density modification and two-fold non-crystallographic symmetry averaging using the Resolve module of the Phenix suite (*51*, *64*). The resulting electron density maps were of excellent quality and allowed manual model building in COOT (*52*, *65*). Selenium positions aided in assigning the sequence register. The atomic model of SpyCas9 was completed by iterative model building in COOT and refinement using Phenix.refine (*53*). Refinement and model statistics are provided in Table 1.

The final atomic model has R_{work} and R_{free} values of 0.253 and 0.286, respectively, and good stereochemistry, as assessed with MolProbity (*66*), with 96.6% of the residues in the most favored regions of the Ramachandran plot and no outliers. The model contains two SpyCas9 molecules that superimpose with an overall rmsd of 1.1 Å over 1060 Cα atoms, the major difference being a ~5° hinge-like rotation of the HNH domain. In the atomic model, molecule A contains residues 4-102, 115-307, 314-447, 503-527, 540-567, 587-672, 677-714, 718-764, 775-791, 799-859, 862-902, 908-1027, 1036-1102, 1137-1146, 1159-1186, 1192-1242, and 1259-1363. Molecule B contains residues 4-103, 116-308, 310-447, 502-527, 539-570, 587-673, 676-713, 718-764, 773-791, 800-859, 862-902, 908-1025, 1036-1102, 1137-1148, 1160-1185, 1188-1241, and 1256-1363. The remaining residues do not appear ordered in electron density maps and could not be built. In the manuscript, the discussion of the SpyCas9 structure is based on molecule B, which is better ordered.

An additional dataset (at 3.1 Å resolution) was measured using a SpyCas9 crystal soaked in 20 mM $MnCl_2$ during the cryoprotection procedure. F_o-F_c difference maps calculated using the high-resolution model revealed two Mn^{2+} ions bound in the RuvC domain active site (fig. S3) and 4 additional Mn^{2+} ions bound to each of the two SpyCas9 molecules. The HNH domain active site remained poorly ordered in this structure, and no Mn^{2+} binding was observed. The model was refined to an R_{work} and R_{free} of 0.252 and 0.278, respectively.

Endonuclease cleavage assays with SpyCas9

A synthetic 42-nt crRNA targeting a protospacer from the bacteriophage λ genome was purchased from Integrated DNA Technologies (IDT) and purified via 10% denaturing PAGE. tracrRNA was *in vitro* transcribed from a synthetic DNA template (IDT) using T7 RNA polymerase and corresponds to nucleotides 15-87 as described previously (*8*). crRNA:tracrRNA duplexes (10 μM) were prepared by mixing equimolar amounts of crRNA and tracrRNA in Hybridization Buffer (20 mM Tris-Cl pH 7.5, 100 mM KCl, 5 mM $MgCl_2$), heating at 95 °C for 30 sec, and slow-cooling on the benchtop. SpyCas9:RNA complexes were reconstituted by mixing SpyCas9 with a 2X molar excess of the crRNA:tracrRNA duplex in Reconstitution Buffer (20 mM Tris-Cl pH 7.5, 100 mM KCl, 5 mM $MgCl_2$, 1 mM DTT) and incubating at 37°C for 10 minutes.

A 55 base-pair (bp) DNA target derived from the bacteriophage λ genome was prepared by mixing equimolar amounts of individual synthetic oligonucleotides (IDT) in Hybridization Buffer supplemented with 5% glycerol, heating for 1-2 minutes, and slow-cooling on the benchtop. Duplexes were separated from single-stranded DNA by 6% native PAGE conducted at 4°C, with 5 mM $MgCl_2$ added to the gel and the running buffer. The DNA was excised, eluted into 10 mM Tris-Cl, pH 8 at 4°C overnight, ethanol precipitated, and resuspended in Hybridization Buffer. Br-dU containing ssDNAs used in analytical crosslinking reactions were radiolabeled and hybridized with a 5X molar excess of the unlabeled complementary strand. Cleavage reactions were performed at room temperature in Reaction Buffer (20 mM Tris-Cl pH 7.5, 100 mM KCl, 5 mM $MgCl_2$, 5% glycerol, 1 mM DTT) using 1 nM radiolabeled dsDNA substrates and 1 nM or 10 nM Cas9:RNA. Aliquots (10 μl) were removed at various time points and quenched by mixing with an equal volume of formamide gel loading buffer supplemented with 50 mM EDTA. Cleavage products were resolved by 10% denaturing PAGE and visualized by phosphorimaging (GE Healthcare). The sequences of DNA and RNA oligonucleotides used in this study are listed in Supplementary Table S2.

Preparation of crosslinked peptide-DNA heteroconjugates for mass spectrometry

200 pmol of catalytically inactive (D10A/H840A) Cas9 was reconstituted with crRNA:tracrRNA and incubated with a 10X molar excess of Br-dU containing dsDNA substrate for 30 min at room

temperature in Reaction Buffer. Reactions were transferred into the lid of open PCR tubes and irradiated with UV-light (308 nm) for 30 min at room temperature. Crosslinked samples were denatured with 6 M urea for 1 h at 65°C, diluted to 0.5 M urea with 25 mM ammonium bicarbonate, and digested with 1 ng trypsin overnight at room temperature. Samples were concentrated to a final volume of 50 µL and desalted with Illustra MicroSpin G-25 Columns (GE Healthcare). Samples were then treated with 1,000 Units of Nuclease S1 (Sigma Aldrich) for 1 h at 37 °C in 30 mM ammonium acetate pH 5.7, 10 mM $CaCl_2$ and 0.1 mM $ZnCl_2$ in a total volume of 60 µL. In order to remove remaining phosphate groups at the crosslink site, 7 µL of 10X Antarctic Phosphatase buffer and 5 Units of Antarctic Phosphatase (New England BioLabs) were added to the reactions, and samples were incubated for an additional hour at 37 °C.

Liquid chromatography-tandem mass spectrometry (LS-MS/MS)

Tryptic digests of crosslinked proteins were analyzed using a Dionex UltiMate3000 RSLCnano liquid chromatograph that was connected in-line with an LTQ Orbitrap XL mass spectrometer equipped with a nanoelectrospray ionization source (nanoESI; Thermo Fisher Scientific). The LC was equipped with a C18 analytical column (Acclaim® PepMap RSLC, 150 mm length × 0.075 mm inner diameter, 2 µm particles, 100 Å pores, Thermo) and a 1 µL sample loop. Solvent A was 99.9% water/0.1% formic acid and solvent B was 99.9% acetonitrile/0.1% formic acid (v/v). Samples were placed in polypropylene autosampler vials with septa caps (Wheaton,) and loaded into the autosampler compartment (maintained at 4 °C) prior to analysis. The elution program consisted of isocratic flow at 5% B for 4 min, a linear gradient to 35% B over 98 min, isocratic flow at 95% B for 6 min, and isocratic flow at 5% B for 12 min, at a flow rate of 300 nL min^{-1}. The column exit was connected to the nanoESI emitter in the ion source of the mass spectrometer using polyimide-coated, fused-silica tubing (20 µm inner diameter × 280 µm outer diameter, Thermo).

Full-scan mass spectra were acquired in the positive ion mode over the range m/z = 350 to 1500 using the Orbitrap mass analyzer, in profile format, with a mass resolution setting of 60,000 (at m/z = 400, measured at full width at half-maximum peak height). Under these conditions, isotopic distributions of singly and multiply charged peptide ions were resolved in the full-scan mass spectra. Thus, a precursor ion's mass and charge were determined independently, i.e. the ion charge was determined from the reciprocal of the spacing between

adjacent isotope peaks in the *m*/*z* spectrum. In the data-dependent mode, the six most intense ions exceeding an intensity threshold of 30,000 counts were selected from each full-scan mass spectrum for tandem mass spectrometry (MS/MS) analysis using collision-induced dissociation (CID). MS/MS spectra were acquired using the linear ion trap, in centroid format, with the following parameters: isolation width 3 *m*/*z* units, normalized collision energy 28%, default charge state 2+, activation Q 0.25, and activation time 30 ms. Real-time charge state screening was enabled to exclude singly charged ions and unassigned charge states from MS/MS analysis. To avoid the occurrence of redundant MS/MS measurements, real-time dynamic exclusion was enabled to preclude re-selection of previously analyzed precursor ions, with the following parameters: repeat count 2, repeat duration 10 s, exclusion list size 500, exclusion duration 60 s, and exclusion mass width 20 ppm (relative to mass). Data were analyzed using Xcalibur (version 2.0.7 SP1, Thermo) and Proteome Discoverer (version 1.3, Thermo, SEQUEST algorithm) software. Validation of identified cross-linked peptides was by manual inspection of the MS/MS spectra, i.e. to verify the occurrence of b- and y-type fragment ions (*67*) that identify the peptide sequences.

DNA binding experiments

SpyCas9:crRNA:tracrRNA complexes (containing wild-type SpyCas9 or PAM loop mutants $PWN_{475-477}$→AAA, $DWD_{1125-1127}$→AAA , and $PWN_{475-477}/DWD_{1125-1127}$→AAA/AAA) were reconstituted for 10 min at 37 °C in Reaction Buffer before being incubated with ~1 nM radiolabeled DNA target for 60 minutes at 37 °C. Reactions were resolved by 5% native PAGE and visualized by phosphorimaging (GE Healthcare).

AnaCas9 expression and purification

Full-length *Actinomyces naeslundii* Cas9 (AnaCas9; residues 1-1101) was subcloned into a custom pET-based expression vector with an N-terminal His_{10}-tag followed by Maltose-Binding Protein (MBP) and a TEV protease cleavage site. The protein was overexpressed in *Escherichia coli* strain Rosetta (DE3) and was purified to homogeneity by immobilized metal ion affinity chromatography and heparin affinity chromatography. An additional gel filtration chromatography step (HiLoad 16/60 Superdex200, GE Healthcare) was added to further purify AnaCas9 and remove trace nucleic acid contaminants prior to crystallization. Purified AnaCas9

protein in gel filtration buffer (50 mM HEPES 7.5, 300 mM KCl, 2 mM TCEP, 5% glycerol) was snap frozen in liquid nitrogen and stored at -80°C. Selenomethionine–labeled AnaCas9 protein was expressed in Rosetta (DE3) cells grown in M9 minimal medium supplemented with 50 mg ml^{-1} L-SeMet (Sigma) and specific amino acids to inhibit endogenous methionine synthesis. The SeMet-substituted protein was then purified using the same procedure as for the native AnaCas9 protein.

AnaCas9 crystallization and structure determination

Crystals of native and SeMet-substituted AnaCas9 were grown by the hanging drop vapor diffusion method at 20 °C. Aliquots (2.5 µl) of 4.5 mg ml^{-1} native AnaCas9 protein in 50 mM HEPES 7.5, 300 mM KCl, 2mM TCEP, 5% glycerol were mixed with 2.5 µl of reservoir solution containing 10% (w/v) PEG 8000, 0.25 M calcium acetate, 50 mM magnesium acetate and 5 mM spermidine. Crystals appeared after 1–2 days, and they grew to a maximum size of 0.15 × 0.20 × 0.35 mm over the course of 6 days. SeMet-substituted AnaCas9 crystals were grown and optimized under the same conditions. For cryogenic data collection, crystals were transferred into crystallization solutions containing 30% (v/v) glycerol as the cryoprotectant and then flash-cooled at 100 K. Native and SeMet single-wavelength anomalous diffraction (SAD) datasets were collected at beamline 8.3.1 of the Advanced Light Source, Lawrence Berkeley National Laboratory. Data from manganese-soaked AnaCas9 crystals were collected at the 8.2.2 beamline of the Advanced Light Source, Lawrence Berkeley National Laboratory. All diffraction data were integrated using Mosflm and scaled in SCALA (*54*, *55*).

The AnaCas9 structure was solved using the single anomalous dispersion phasing method. Using SeMet data between 79.0 and 3.2 Å resolution, both SHELXD/HKL2MAP (*63*) and HySS in Phenix (*68*) detected a total of 13 out of 18 possible selenium sites in the asymmetric unit. Initial phases were calculated using SOLVE followed by solvent flattening with RESOLVE to produce an electron-density map into which most of the protein residues could be unambiguously built (*51*). The initial model automatically generated from Phenix AutoBuild module was subjected to subsequent iterative rounds of manual building with COOT (*52*) and refinement against the 2.2 Å native data in Refmac (*56*) and Phenix (*53*). The final model contains one zinc ion, two magnesium ions, AnaCas9 residues 8-49, 65-98, 134-170, and 225-1101, and has R_{work} and R_{free} values of 0.19 and 0.23, respectively. The N terminus (residues 1–

7), loop regions (residues 50-64), and a portion of the alpha-helical lobe (residues 99-133, 171-224) are completely disordered. Model validation showed 94% of the residues in the most favored and 5.8% in the allowed regions of the Ramachandran plot. The structure of Mn^{2+}-bound AnaCas9 was obtained by molecular replacement using the program Phaser (*69*), which revealed two unambiguously refined Mn^{2+} ions present in the RuvC active site. All statistics of the data processing and structure refinement of AnaCas9 are summarized in Table 2.

Complex reconstitution for negative-stain EM

All samples for EM (10 µl volumes) were prepared in Reaction Buffer at a final Cas9 concentration of 1 µM. Cas9:RNA complexes contained 2 µM crRNA:tracrRNA duplex and were incubated at 37 °C for 10 minutes before storing on ice until grid preparation. Cas9:RNA:DNA complexes were prepared by first generating Cas9:RNA as before and then adding the DNA duplex at 5 µM (unlabeled) or 2 µM (biotin labeled) and incubating an additional 10 minutes at 37 °C. When present, streptavidin (New England Biolabs) was added after formation of Cas9:RNA or Cas9:RNA:DNA complexes at a 2X unit excess over the biotinylated species, according to the manufacturer's unit definition (~65 ng/µL in the final reaction volume), followed by an additional 10 minute incubation at 37 °C before storing on ice. Catalytically inactive Cas9 (D10A/H840A) was used to generate the following samples: unlabeled Cas9:RNA:DNA, Cas9:RNA:DNA containing biotin modifications on one or both ends of the duplex, and Cas9:RNA:DNA containing an N-terminal MBP. Wild-type Cas9 was used to generate apo-Cas9 and all Cas9:RNA complexes.

Negative-stain electron microscopy

We diluted Cas9 complexes for negative-stain EM to a concentration of ~25-60 nM in 20 mM Tris-HCl pH 7.5, 200 mM KCl, 1 mM DTT, and 5% glycerol immediately before applying the sample to glow-discharged 400 mesh continuous carbon grids. After adsorption for 1 min, we stained the samples consecutively with six droplets of 2% (w/v) uranyl acetate solution, gently blotted off the residual stain, and air-dried the sample in a fume hood. Data were acquired using a Tecnai F20 Twin transmission electron microscope operated at 120 keV at a nominal magnification of either 80,000X (1.45 Å at the specimen level) or 100,000X (1.08 Å at the

specimen level) using low-dose exposures (~20 e^{-}Å^{-2}) with a randomly set defocus ranging from −0.5 to −1.3 μm. A total of 300–400 images of each Cas9 sample were automatically recorded on a Gatan 4k x 4k CCD camera using the MSI-Raster application within the automated macromolecular microscopy software LEGINON (*57*).

Single-particle pre-processing

All image pre-processing and two-dimensional classification was performed in Appion as described previously (*44*). The contrast transfer function (CTF) of each micrograph was estimated, and particles were selected concurrently with data collection using ACE2 (*70*) and a template-based particle picker (*71*), respectively. Micrograph phases were corrected using ACE2 (*70*), and the negatively-stained Cas9 particles were extracted using a 288 × 288-pixel box size. The particle stacks were binned by a factor of 2 for processing, and particles were normalized to remove pixels whose values were above or below 4.5-σ of the mean pixel value using XMIPP (*72*).

Random conical tilt reconstruction

Initial models for reconstructions of both apo-Cas9 and Cas9:RNA:DNA samples were determined using random conical tilt (RCT) methodology (*34*). Briefly, tilt-pairs of micrographs were recorded manually at 0° and 55°, and *ab initio* models were generated using the RCT module (*73*) in Appion (*74*). Particles were correlated between tilt-pairs using TiltPicker (*75*), binned by 2, and extracted from raw micrographs. Reference-free class averages were produced from untilted particle images by iterative 2D alignment and classification using MSA-MRA in IMAGIC (*76*). These class averages served as references for SPIDER (*77*) reference-based alignment and classification, and RCT volumes were calculated for each class average using back-projection in SPIDER based on these angles and shifts. The RCT model from the most representative class (largest number of particles) was low-pass filtered to 60-Å resolution and used to assign Euler angles to the entire data set of reference-free class averages. The resulting low-resolution model was again low-pass filtered to 60-Å resolution and used as the initial model for refinement of the three-dimensional structure by iterative projection matching using the

untilted particle images as previously described (*78*), with libraries from EMAN2 and SPARX software packages (*58*, *59*).

Domain mapping and localization of RNA- and DNA-ends

Particle stacks were binned by a factor of 2 and subjected to five rounds of iterative multivariate statistical analysis (MSA) and multi-reference alignment (MRA) using the IMAGIC (*76*) software package, to generate two-dimensional class averages of each complex. The resulting set of class averages for each species was normalized using 'proc2d' in EMAN (*79*). The EMAN classification program 'classesbymra' was used to match the labeled class average to the best-matching unlabeled class average based on cross-correlation coefficients. The difference maps were calculating by subtracting the unlabeled class average from the labeled class averages using 'proc2d' in EMAN. This same strategy was used to match the unlabeled class average to the best-matching reprojection of the corresponding structure. The Euler angles used for creating the reprojection were applied to the 3D electron density using 'proc3d,' and the surface representation visualized in Chimera (*80*) is shown along with its corresponding reprojection.

3D reconstruction and analysis

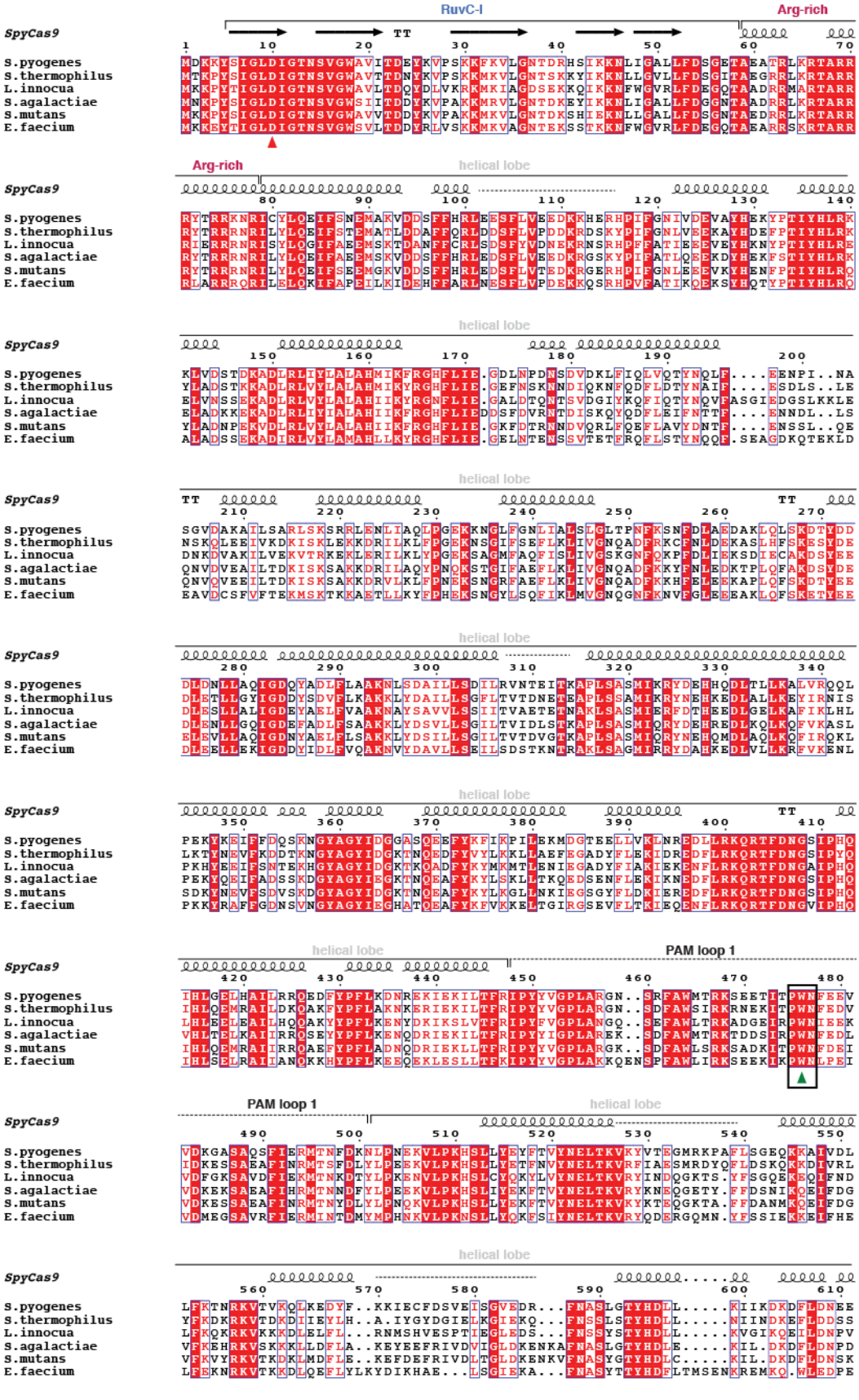
Three-dimensional reconstructions were all performed using an iterative projection-matching refinement with libraries from the EMAN2 and SPARX software packages (*58*, *59*). Refinement of the RCT starting models began using an angular increment of 25°, progressing down to 4° for all reconstructions. The resulting model was again low-pass filtered to 60-Å resolution and subjected to iterative projection-matching refinement to obtain the final structure. In an alternative approach for apo-Cas9 and Cas9:RNA:DNA, we used a low-pass filtered model of the other structure after initial refinement with untilted particles as an initial model for the above-mentioned projection matching refinement. This led to EM densities with similar structural features as the RCT models (Fig. S1H, S3I), and the structures converged to the final models presented in Fig. 1B,C. The resolution was estimated by splitting the particle stack into two equally sized data sets and calculating the Fourier shell correlation (FSC) between each of the back-projected volumes. The final reconstructions of Cas9, Cas9:RNA, and Cas9:RNA:DNA showed structural features to ~19-Å, ~21-Å, and ~19-Å resolution, respectively, based on the 0.5

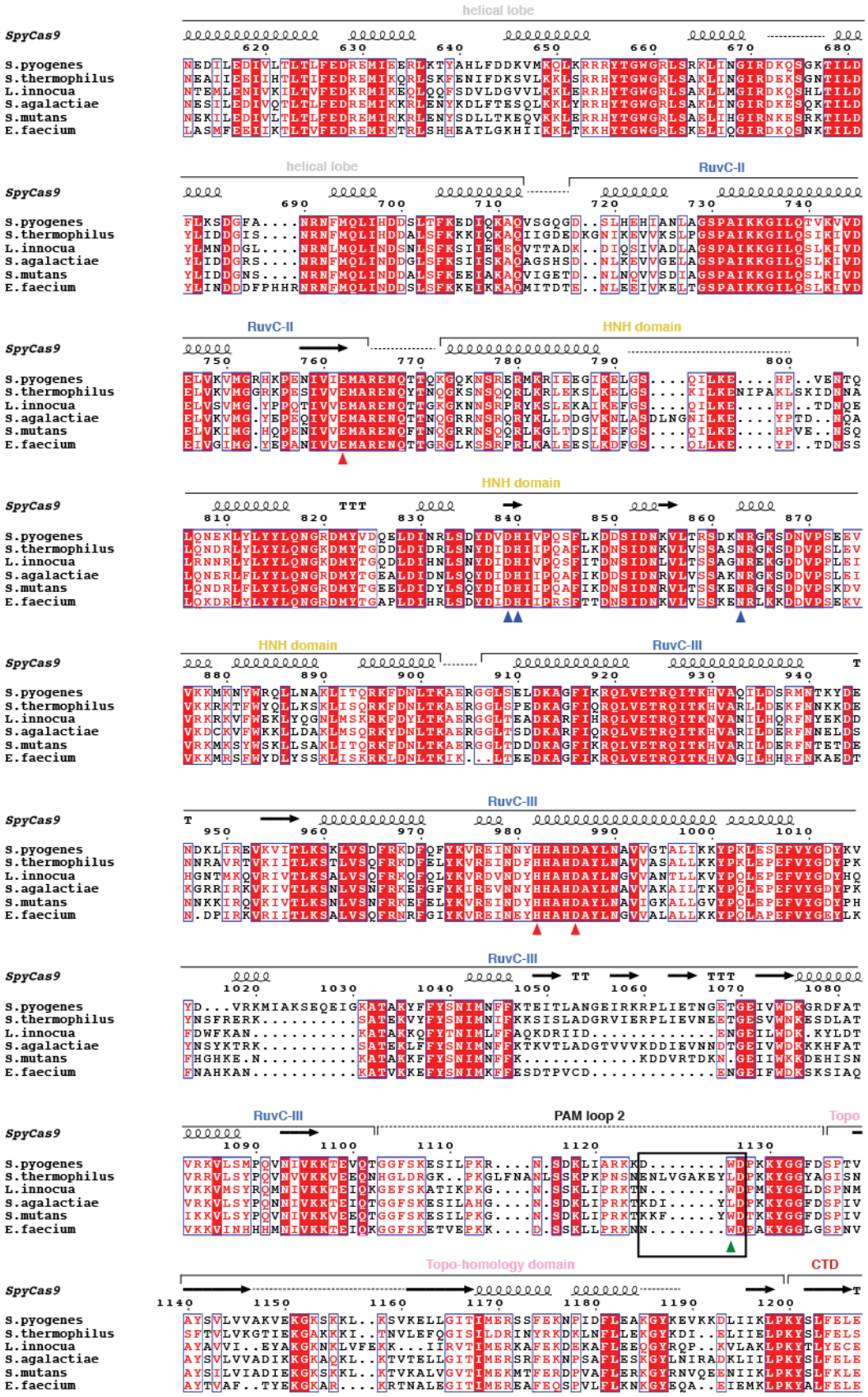
Fourier shell correlation criterion. Reprojections of the final three-dimensional reconstruction showed excellent agreement with the reference-free class averages (Fig. S1G, S3H, S5D) and displayed a large distribution of Euler angles, despite some preferential orientations of the particles on the carbon film (Fig. S1E, S3F, S5E).

The final reconstruction was segmented using Segger (*81*) in Chimera (*80*) based on inspection of the similarities between lobes in the apo-Cas9 and Cas9:RNA:DNA reconstructions. A modeled A-form duplex was manually docked into the map with Chimera, using information from the labeling experiments and map segmentation, and by accommodating the substrate within the channel in the EM reconstruction. While the absolute handedness of our apo-Cas9 reconstruction could be confirmed using the X-ray crystal structure, the relative handedness of our Cas9:RNA:DNA reconstruction is uncertain. Free hand tests performed on this sample failed, likely due to the small and/or dynamic nature of the enzyme. The model we present (Fig. 1C) is based on the alpha-helical domain from the crystal structure having a more optimal CCC with the larger lobe of this reconstruction (0.83) than this lobe using the reconstruction of opposite handedness (0.74) (Fig. S4).

Enzymatic footprinting experiments

DNA targets (55 bp) were prepared by 5'-radiolabeling either the target or displaced non-target strand and then hybridizing it to a 5X molar excess of unlabeled complementary strand. After incubating catalytically inactive (D10A/H840A) SpyCas9:crRNA:tracrRNA complexes (100 nM) with ~1 nM DNA substrate for 30 minutes at 37 °C in Reaction Buffer, 100 units of exonuclease III (NEB) or 1.2 μg nuclease P1 (Sigma) was added and reactions were incubated an additional 10 minutes at 37 °C before quenching with formamide gel loading buffer supplemented with 50 mM EDTA. Reaction products were resolved by 15% denaturing (7M urea) PAGE and visualized by phosphorimaging (GE Healthcare). Control reactions contained a non-targeting crRNA that is not complementary to the 55-bp DNA substrate. To define the sequence register of enzymatic reaction products, a DNA ladder was generated by 5'-radiolabeling the synthetic target or non-target strand without prior gel purification and compared to DNA cleavage products using active SpyCas9:RNA or FokI and BglI restriction enzymes (NEB). Note that we observed SpyCas9:RNA cleaving the non-target strand between nucleotides 4 and 5 from the PAM end, in contrast to the cleavage site observed previously (*8*).


RuvC-I
Arg-rich
SpyCas9
S.pyogenes
S.thermophilus
L.innocua
S.agalactiae
S.mutans
E.faecium
helical lobe
PAM loop 1


helical lobe
SpyCas9
S.pyogenes
S.thermophilus
L.innocua
S.agalactiae
S.mutans
E.faecium
helical lobe
RuvC-II
RuvC-II
HNH domain
HNH domain
RuvC-III
RuvC-III
RuvC-III
PAM loop 2
Topo
Topo-homology domain
CTD

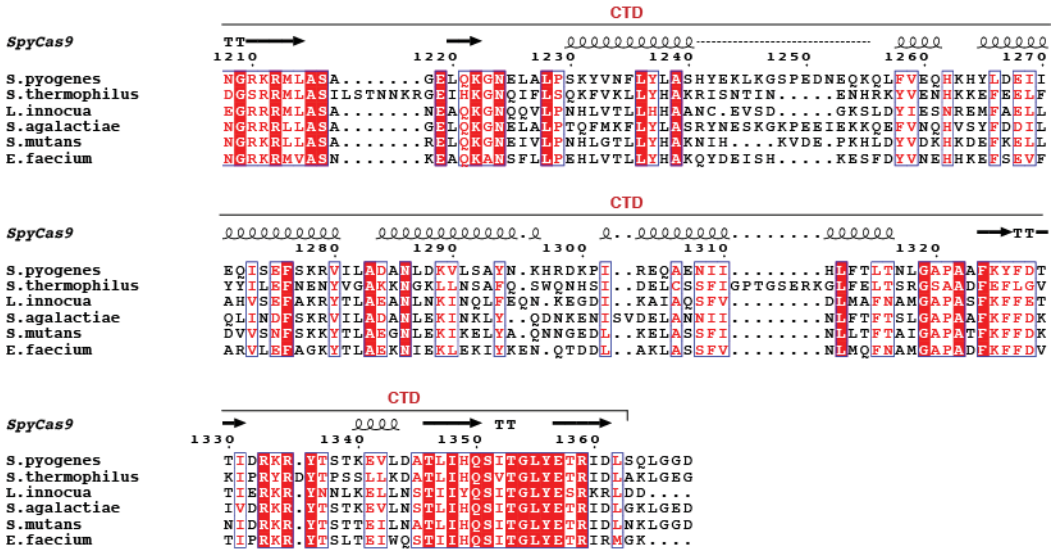


Figure S1. Multiple sequence alignment of Cas9 proteins associated with Type II-A CRISPR loci. Primary sequences of Cas9 proteins from *Streptococcus pyogenes* (GI 15675041), *Streptococcus thermophilus* LMD-9 (GI 11662823), *Listeria innocua* Clip 11262 (GI 16801805), *Streptococcus agalactiae* A909 (GI 76788458), *Streptococcus mutans* UA159 (GI 24379809), and *Enterococcus faecium* 1,231,408 (GI 257893735) were aligned using MAFFT (*82*). The alignment was generated in ESPript (*83*) using default settings. Strictly conserved residues are shown with white letters on red background. Residues with >70% similarity are shown in red and boxed in blue. The domain organization of SpyCas9 (as in Fig. 1A) and secondary structure are shown above the sequences. Disordered segments of the polypeptide chain are indicated with dashed lines. RuvC domain catalytic residues are denoted with red arrowheads. HNH domain active site residues are denoted with blue arrowheads. Tryptophan residues that crosslinked to nucleotides flanking the PAM are denoted with green arrowheads, and tryptophan-containing motifs mutated in Fig. 3D are boxed in black.

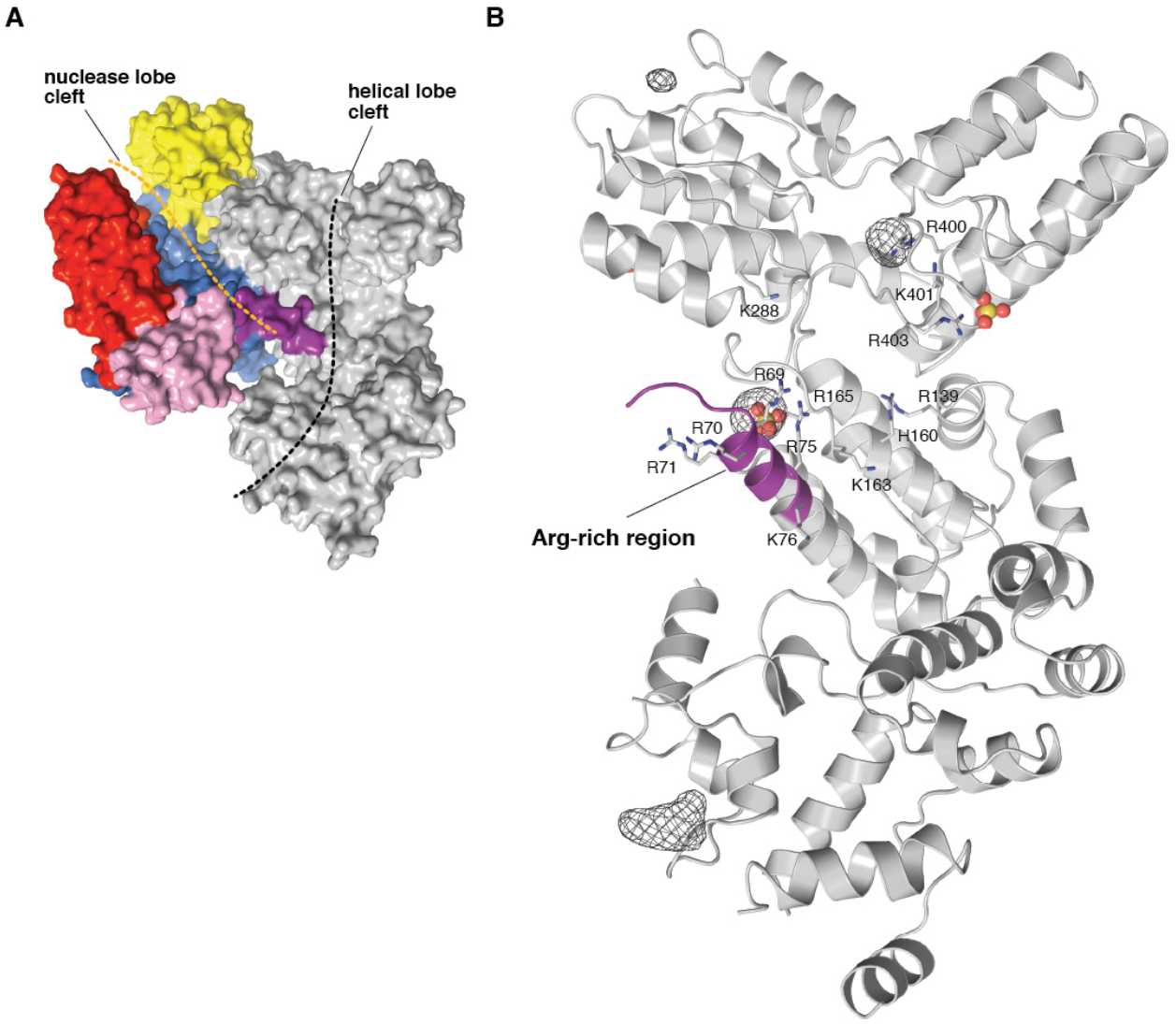


Figure S2. The helical lobe of SpyCas9 features a putative nucleic acid binding cleft. (A) Surface representation of SpyCas9, colored according to the scheme in Fig. 1A. The surface clefts located on the nuclease and alpha-helical lobes of the protein are indicated with orange and black dashed lines, respectively. (**B**) Close-up view of the helical lobe of SpyCas9. Arg-rich region is depicted in purple. Conserved basic (Arg, Lys) residues lining the cleft are shown in stick format. Sulfate ions bound to the cleft are shown in ball-and-stick format. Anomalous difference electron density map (black mesh, contoured at 5.0) indicates positions of tungstate ions bound to SpyCas9 in crystals soaked with 10 mM Na_2WO_4.

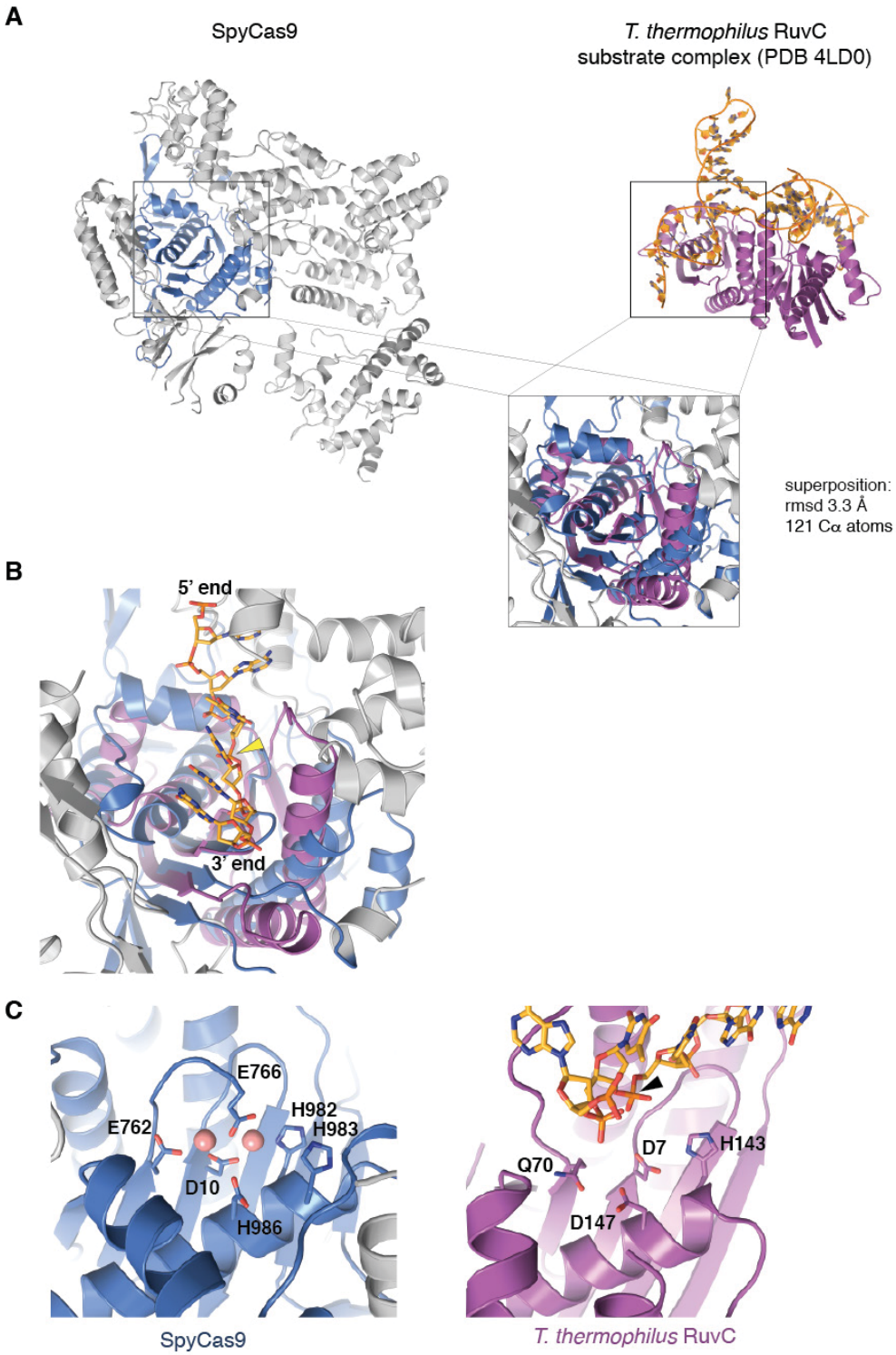


Figure S3. Structural superposition of SpyCas9 with RuvC resolvase defines the directionality of non-target DNA strand in DNA-bound SpyCas9 holoenzyme. (A) Structural superposition of SpyCas9 with *Thermus thermophilus* RuvC resolvase bound to a Holliday junction substrate (PDB entry 4LD0) (*28*). The structures were superimposed using DALI (*84*) and are shown in the same orientation. The SpyCas9 RuvC domain is depicted in blue, and the RuvC resolvase is colored purple. Inset shows the superposition of the two structures. The proteins superimpose with an rmsd of 3.3 Å over 121 Cα atoms. (**B**) Close-up view of the SpyCas9 nuclease lobe cleft harbouring the RuvC active site. Six nucleotides of single stranded DNA are modeled in the cleft (stick format, colored orange) based on the superposition in (A). The position of the scissile phosphate is indicated with a yellow

arrowhead. (**C**) Close-up views of the catalytic sites in SpyCas9 (left) and *T. thermophilus* RuvC (right). Active site residues are shown in stick format. Pink spheres represent two Mn^{2+} ions bound to the SpyCas9 RuvC domain in crystals soaked with 20 mM $MnCl_2$. The DNA substrate is show in stick format, and the position of the scissile phosphate is indicated with a black arrowhead.

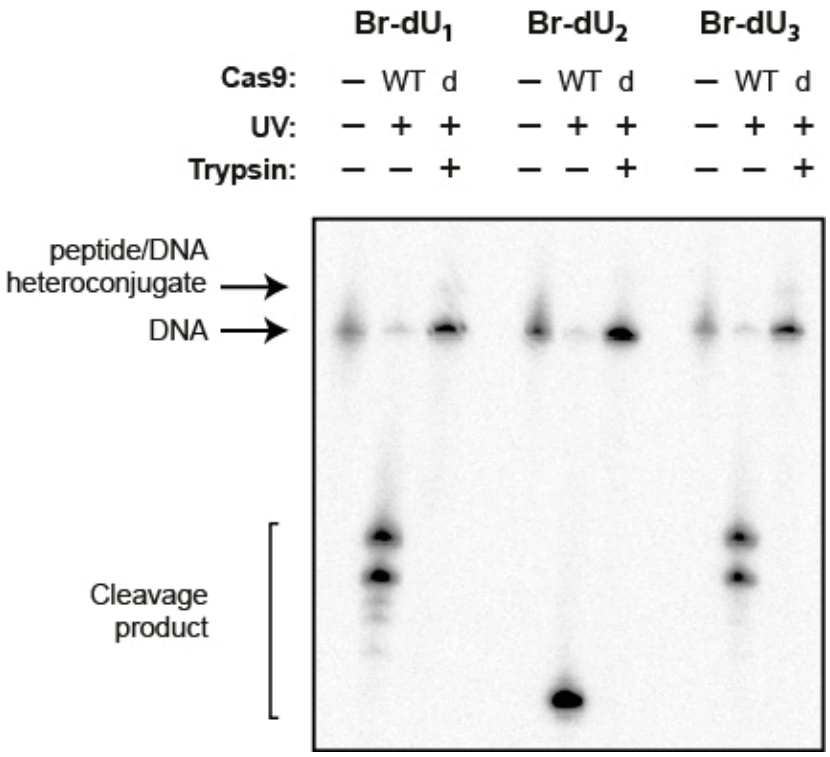
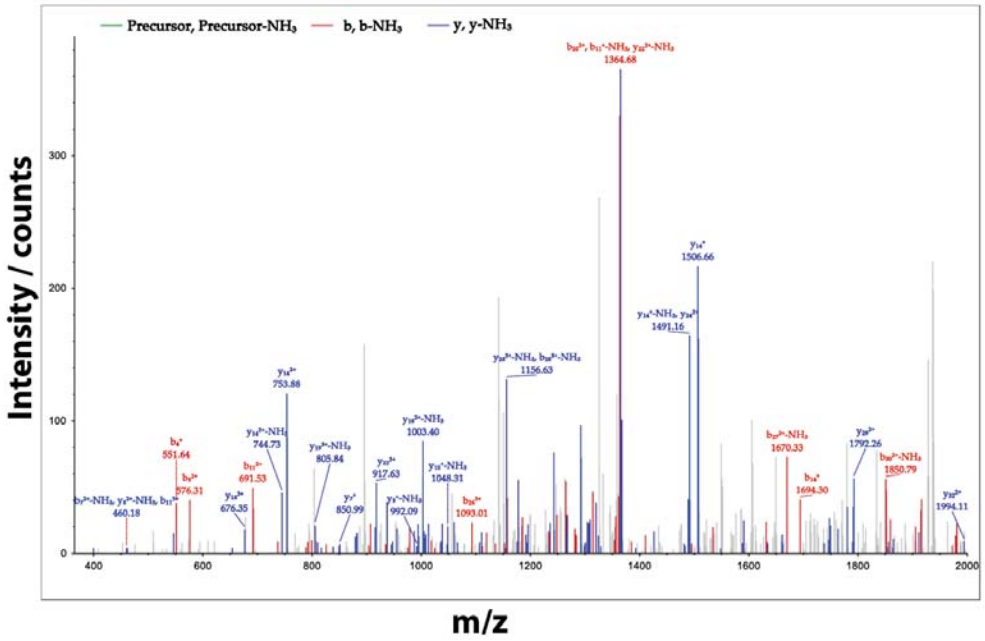
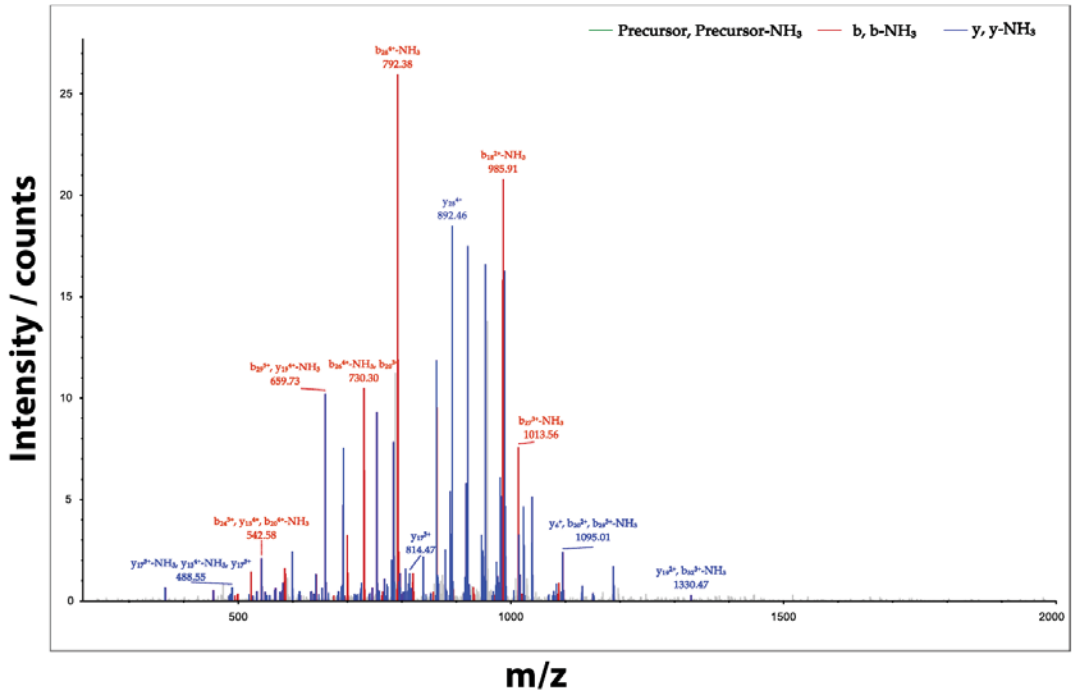


Figure S4. Br-dU containing dsDNA substrates are cleaved by WT SpyCas9 and crosslink to catalytically inactive dCas9. DNA cleavage assays were performed and analysed by denaturing PAGE to verify that modified dsDNA substrates do not impair cleavage by WT SpyCas9. Sequences for each substrate (Br-dU_1, Br-dU_2, and Br-dU_3) can be found in Supplementary Table S2. Reactions with catalytically inactive (D10A/H840A) dCas9 that can bind but not cleave DNA showed an additional band of higher molecular weight following UV irradiation and trypsin digestion, providing evidence for the generation of a peptide-DNA heteroconjugate. Crosslinking reactions with Br-dU_1, Br-dU_2, and Br-dU_3 were analyzed by LC-MS/MS, but only reactions with Br-dU_1 and Br-dU_3 dsDNA substrates resulted in the identification of crosslinked peptides.



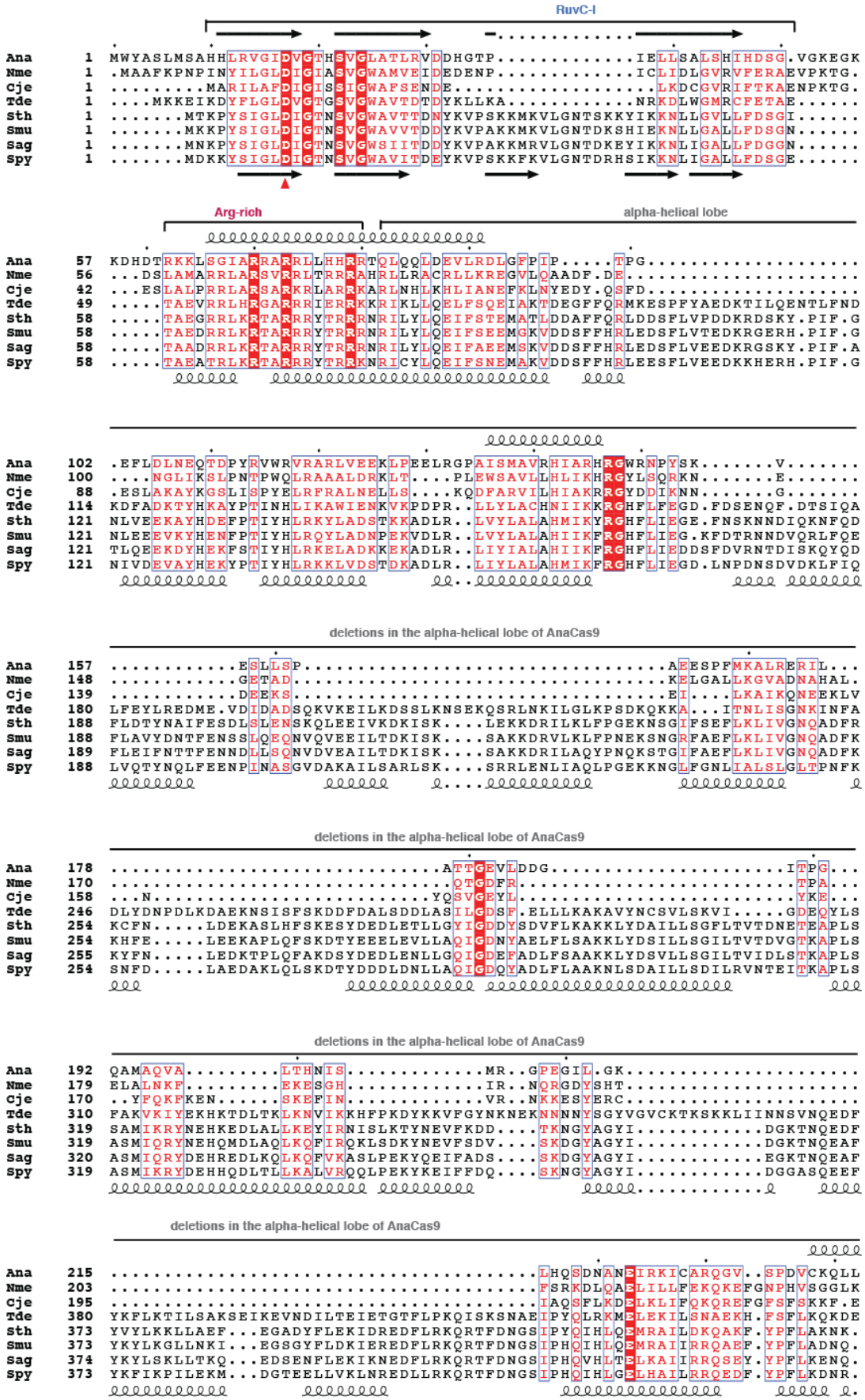
#1	b+1	b+2	b+3	Seq.	y+1	y+2	y+3	#2
1	148.07570	74.54149	50.03008	F				33
2	219.11282	110.06005	73.70912	A	3984.87105	1992.93916	1328.96187	32
3	405.19214	203.09971	135.73556	W	3913.83393	1957.42060	1305.28283	31
4	552.22755	276.61741	184.74737	M-Oxidation	3727.75461	1864.38094	1243.25639	30
5	653.27523	327.14125	218.42993	T	3580.71919	1790.86323	1194.24458	29
6	809.37635	405.19181	270.46363	R	3479.67151	1740.33939	1160.56202	28
7	937.47132	469.23930	313.16196	K	3323.57039	1662.28883	1108.52831	27
8	1024.50335	512.75531	342.17263	S	3195.47542	1598.24135	1065.82999	26
9	1153.54595	577.27661	385.18683	E	3108.44339	1554.72533	1036.81931	25
10	1282.58855	641.79791	428.20103	E	2979.40079	1490.20403	993.80511	24
11	1383.63623	692.32175	461.88359	T	2850.35819	1425.68273	950.79091	23
12	1496.72030	748.86379	499.57828	I	2749.31051	1375.15889	917.10835	22
13	1597.76798	799.38763	533.26084	T	2636.22644	1318.61686	879.41366	21
14	1694.82075	847.91401	565.61177	P	2535.17876	1268.09302	845.73110	20
15	2106.95903	1053.98315	702.99119	W-dU	2438.12599	1219.56663	813.38018	19
16	2221.00196	1111.00462	741.00550	N	2025.98771	1013.49749	676.00075	18
17	2368.07038	1184.53883	790.02831	F	1911.94478	956.47603	637.98644	17
18	2497.11298	1249.06013	833.04251	E	1764.87636	882.94182	588.96364	16
19	2626.15558	1313.58143	876.05671	E	1635.83376	818.42052	545.94944	15
20	2725.22400	1363.11564	909.07952	V	1506.79116	753.89922	502.93524	14
21	2824.29242	1412.64985	942.10232	V	1407.72274	704.36501	469.91243	13
22	2939.31937	1470.16332	980.44464	D	1308.65432	654.83080	436.88962	12
23	3067.41434	1534.21081	1023.14296	K	1193.62737	597.31732	398.54731	11
24	3124.43581	1562.72154	1042.15012	G	1065.53240	533.26984	355.84898	10
25	3195.47293	1598.24010	1065.82916	A	1008.51093	504.75910	336.84183	9
26	3282.50496	1641.75612	1094.83984	S	937.47381	469.24054	313.16279	8
27	3353.54208	1677.27468	1118.51888	A	850.44178	425.72453	284.15211	7
28	3481.60066	1741.30397	1161.20507	Q	779.40466	390.20597	260.47307	6
29	3568.63269	1784.81998	1190.21575	S	651.34608	326.17668	217.78688	5
30	3715.70111	1858.35419	1239.23855	F	564.31405	282.66066	188.77620	4
31	3828.78518	1914.89623	1276.93324	I	417.24563	209.12645	139.75339	3
32	3957.82778	1979.41753	1319.94744	E	304.16156	152.58442	102.05870	2
33				R	175.11896	88.06312	59.04450	1

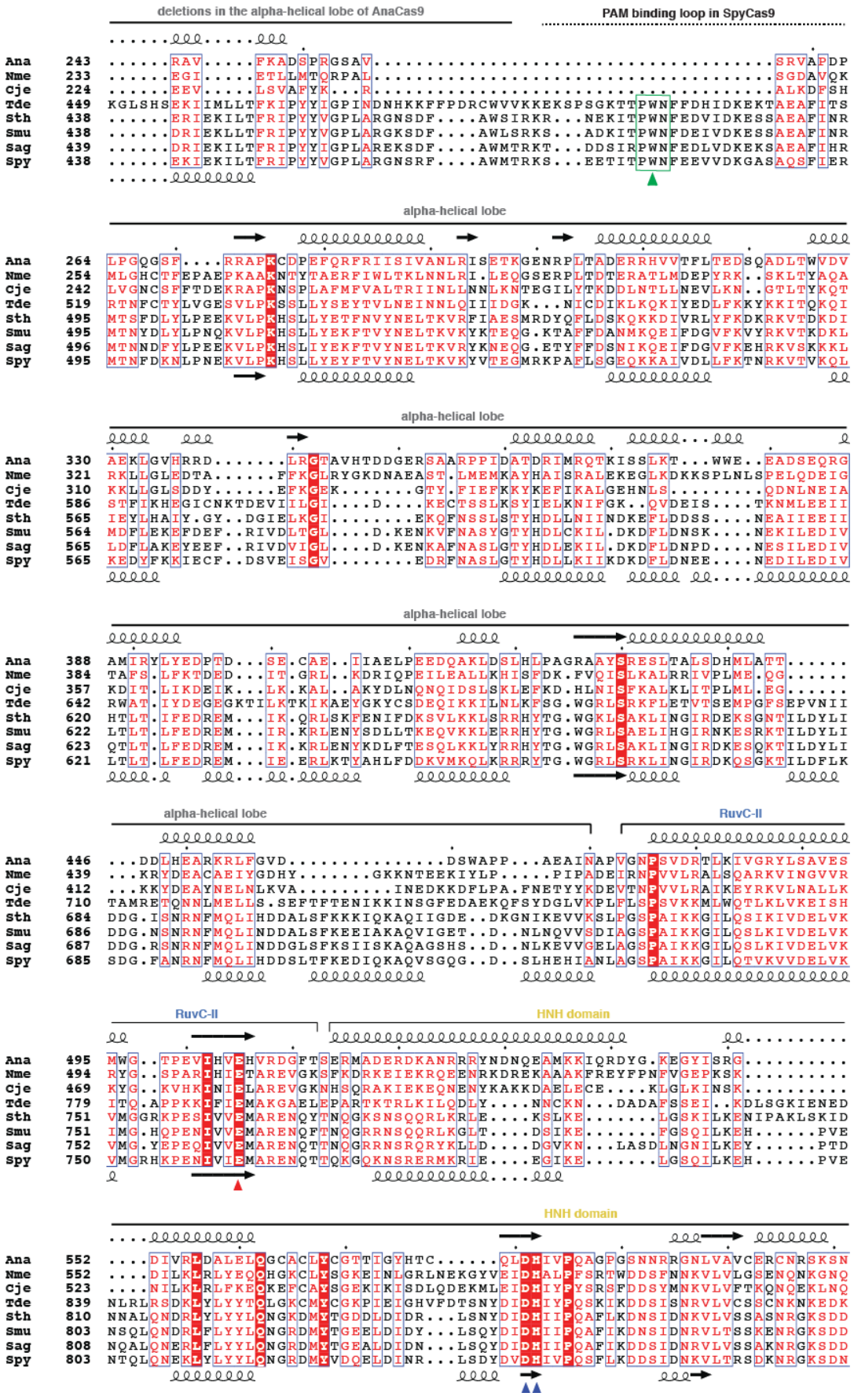
Figure S5. Trp476^{Spy} crosslinks to Br-dU$_1$ dsDNA target. Tandem mass spectrum (MS/MS) and fragment ion list resulting from collision-induced dissociation (CID) of the 3+ ion occurring at mass-to-charge ratio *m*/z = 1377.9835. This corresponds to the $[M + 3H]^{3+}$ ion of the peptide, FAWMTRKSEETITP(**W-dU**)NFEEVVDKGASAQSFIER, which corresponds to residues 462-494 of SpyCas9, in which Trp476^{Spy} is crosslinked to deoxyuridine (dU) and Met468^{Spy} is oxidized (i.e. methionine sulfoxide). (Crosslinking to deoxyuridine and oxidation result in exact, monoisotopic mass additions of 226.05896 Da and 15.994915 Da, respectively.) Fragment ions b15 through b32 and y19 through y32 exhibit the deoxyuridine mass addition. Detected b-ions are shown in red and y-ions are shown in blue.

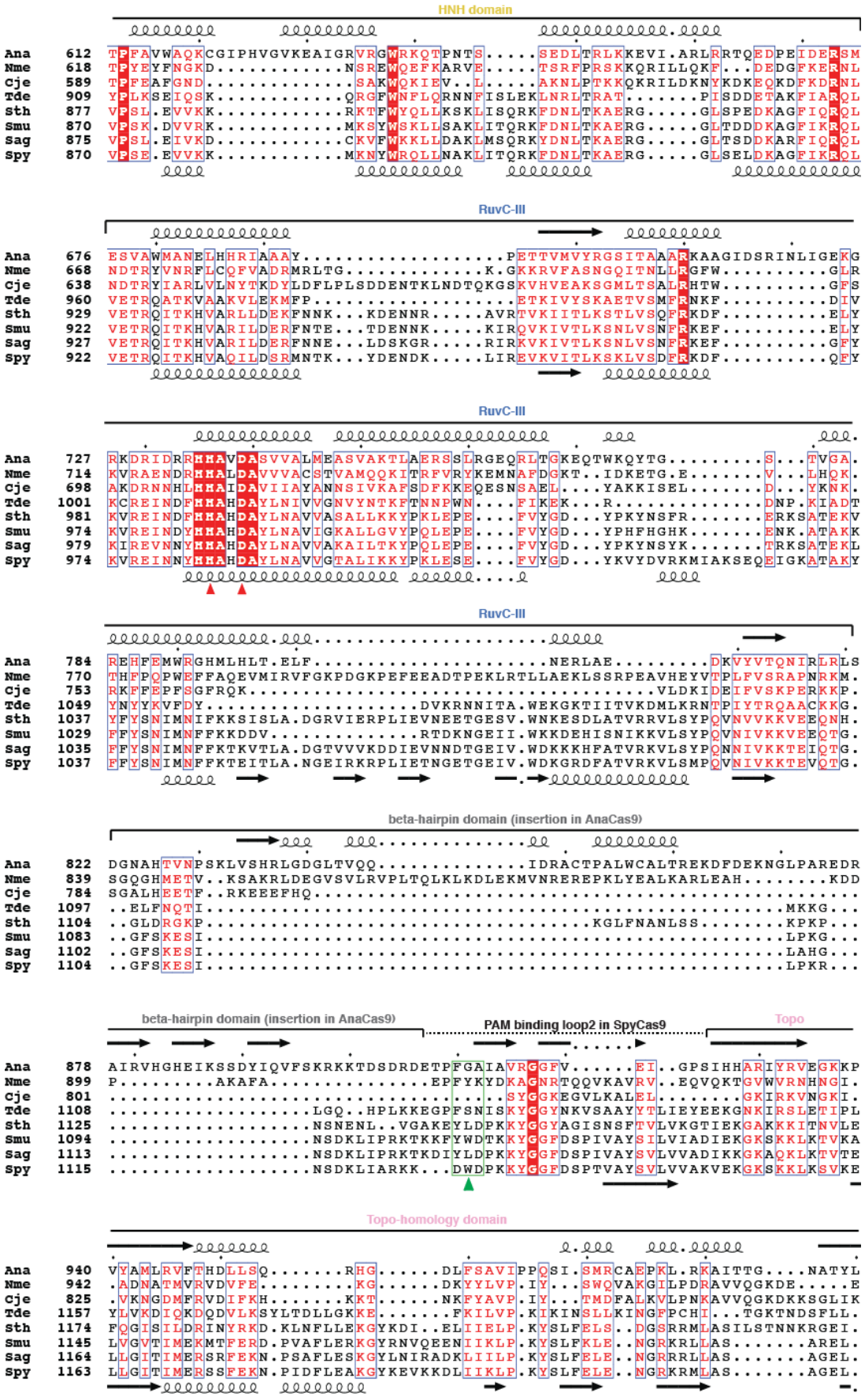


#1	b+1	b+2	b+3	b+4	b+5	Seq.	y+1	y+2	y+3	y+4	y+5	#2
1	129.10225	65.05476	43.70560	33.03102	26.62627	K						33
2	230.14993	115.57860	77.38816	58.29294	46.83581	T	4021.01806	2011.01267	1341.01087	1006.00997	805.00943	32
3	359.19253	180.09990	120.40236	90.55359	72.64433	E	3919.97038	1960.48883	1307.32831	980.74805	784.79990	31
4	458.26095	229.63411	153.42517	115.32069	92.45801	V	3790.92778	1895.96753	1264.31411	948.48740	758.99138	30
5	586.31953	293.66340	196.11136	147.33534	118.06973	Q	3691.85936	1846.43332	1231.29130	923.72030	739.17769	29
6	687.36721	344.18724	229.79392	172.59726	138.27926	T	3563.80078	1782.40403	1188.60511	891.70565	713.56598	28
7	744.38868	372.69798	248.80108	186.85263	149.68356	G	3462.75310	1731.88019	1154.92255	866.44373	693.35644	27
8	801.41015	401.20871	267.80823	201.10799	161.08785	G	3405.73163	1703.36945	1135.91539	852.18836	681.95215	26
9	948.47857	474.74292	316.83104	237.87510	190.50153	F	3348.71016	1674.85872	1116.90824	837.93300	670.54785	25
10	1035.51060	518.25894	345.84172	259.63311	207.90794	S	3201.64174	1601.32451	1067.88543	801.16589	641.13417	24
11	1163.60557	582.30642	388.54004	291.65685	233.52693	K	3114.60971	1557.80849	1038.87475	779.40788	623.72776	23
12	1292.64817	646.82772	431.55424	323.91750	259.33545	E	2986.51474	1493.76101	996.17643	747.38414	598.10877	22
13	1379.68020	690.34374	460.56492	345.67551	276.74186	S	2857.47214	1429.23971	953.16223	715.12349	572.30025	21
14	1492.76427	746.88577	498.25961	373.94652	299.35867	I	2770.44011	1385.72369	924.15155	693.36548	554.89384	20
15	1605.84834	803.42781	535.95430	402.21754	321.97549	L	2657.35604	1329.18166	886.45686	665.09447	532.27703	19
16	1702.90111	851.95419	568.30522	426.48073	341.38804	P	2544.27197	1272.63962	848.76217	636.82345	509.66021	18
17	1830.99608	916.00168	611.00354	458.50448	367.00504	K	2447.21920	1224.11324	816.41125	612.56026	490.24966	17
18	1987.09720	994.05224	663.03725	497.52976	398.22526	R	2319.12423	1160.06575	773.71293	580.53651	464.63067	16
19	2101.14013	1051.07370	701.05158	526.04049	421.03385	N	2163.02311	1082.01519	721.67922	541.51123	433.41044	15
20	2188.17216	1094.58972	730.06224	547.79850	438.44025	S	2048.98018	1024.99373	683.66491	513.00050	410.60185	14
21	2303.19911	1152.10319	768.40455	576.55523	461.44564	D	1961.94815	981.47771	654.65423	491.24249	393.19545	13
22	2474.29989	1237.65358	825.43815	619.33043	495.66580	K-Carbamyl	1846.92120	923.96424	616.31192	462.48576	370.19006	12
23	2587.38396	1294.19562	863.13284	647.60145	518.28261	L	1675.82041	838.41384	559.27832	419.71056	335.96990	11
24	2700.46803	1350.73765	900.82753	675.87247	540.89943	I	1562.73634	781.87181	521.58363	391.43954	313.35309	10
25	2771.50515	1386.25621	924.50657	693.63175	555.10685	A	1449.65227	725.32977	483.88894	363.16853	290.73628	9
26	2927.60627	1464.30677	976.54027	732.65703	586.32708	R	1378.61515	689.81121	460.20990	345.40925	276.52885	8
27	3055.70124	1528.35426	1019.23860	764.68077	611.94607	K	1222.51403	611.76065	408.17619	306.38397	245.30863	7
28	3183.79621	1592.40174	1061.93692	796.70451	637.56508	K	1094.41906	547.71317	365.47787	274.36022	219.68963	6
29	3298.82316	1649.91522	1100.27924	825.46125	660.57045	D	966.32409	483.66568	322.77955	242.33648	194.07064	5
30	3790.92777	1895.96752	1264.31411	948.48740	758.99138	W-pU	851.29714	426.15221	284.43723	213.57974	171.06525	4
31	3905.95472	1953.48100	1302.65642	977.24414	781.99677	D	359.19253	180.09990	120.40236	90.55359	72.64433	3
32	4003.00749	2002.00738	1335.00735	1001.50733	801.40732	P	244.16558	122.58643	82.06004	61.79685	49.63894	2
33						K	147.11281	74.06004	49.70912	37.53366	30.22838	1

Figure S6. Trp1126^{Spy} crosslinks to Br-dU$_3$ dsDNA target. MS/MS spectrum and fragment ion list resulting from CID of the 5+ ion occurring at *m*/*z* = 830.6273. This orresponds to the $[M + 5H]^{5+}$ ion of the peptide, KTEVQTGGFSKESILPKRNSDKLIARKKD(**W-pdU**)DPK, which corresponds to residues 1097-1129 of SpyCas9, in which Trp1126^{Spy} is crosslinked to deoxyuridine monophosphate (pdU) and Lys1121^{Spy} is carbamylated. (Cross-linking to deoxyuridine monophosphate and carbamylation result in exact, monoisotopic mass additions of 306.02529 Da and 43.005814 Da, respectively.) Fragment ions b30 through b32 and y4 through y32 exhibit the deoxyuridine monophosphate mass addition. Detected b-ions are shown in red and y-ions are shown in blue.


RuvC-I
Arg-rich
alpha-helical lobe
deletions in the alpha-helical lobe of AnaCas9
deletions in the alpha-helical lobe of AnaCas9
deletions in the alpha-helical lobe of AnaCas9
deletions in the alpha-helical lobe of AnaCas9


deletions in the alpha-helical lobe of AnaCas9
PAM binding loop in SpyCas9
alpha-helical lobe
alpha-helical lobe
alpha-helical lobe
alpha-helical lobe
RuvC-II
RuvC-II
HNH domain
HNH domain
Ana 243
Nme 233
Cje 224
Tde 449
Sth 438
Smu 438
Sag 439
Spy 438
Ana 264
Nme 254
Cje 242
Tde 519
Sth 495
Smu 495
Sag 496
Spy 495
Ana 330
Nme 321
Cje 310
Tde 586
Sth 565
Smu 564
Sag 565
Spy 565
Ana 388
Nme 384
Cje 357
Tde 642
Sth 620
Smu 622
Sag 623
Spy 621
Ana 446
Nme 439
Cje 412
Tde 710
Sth 684
Smu 686
Sag 687
Spy 685
Ana 495
Nme 494
Cje 469
Tde 779
Sth 751
Smu 751
Sag 752
Spy 750
Ana 552
Nme 552
Cje 523
Tde 839
Sth 810
Smu 803
Sag 808
Spy 803


HNH domain
RuvC-III
RuvC-III
RuvC-III
beta-hairpin domain (insertion in AnaCas9)
beta-hairpin domain (insertion in AnaCas9)
PAM binding loop2 in SpyCas9
Topo
Topo-homology domain
Ana 612 TPFAVWAQKCGIPHVGVKEAIGRVRGWRKQTPNTS.....SEDLTRLKKEVI.ARLRRTQEDPEIDERSM
Nme 618 TPYEYFNGKD.............NSREWQEFKARVE.....TSRFPRSKKQRILLQKF...DEDGFKERNL
Cje 589 TPFEAFGND..............SAKWQKIEV..L.....AKNLPTKKQKRILDKNYKDKEQKDFKDRNL
Tde 909 YPLKSEIQSK.............QRGFWNFLQRNNFISLEKLNRLTRAT.......PISDDETAKFIARQL
Sth 877 VPSL.EVVKK.............RKTFWYQLLKSKLISQRKFDNLTKAERG.....GLSPEDKAGFIQRQL
Smu 870 VPSK.DVVRK.............MKSYWSKLLSAKLITQRKFDNLTKAERG.....GLTDDDKAGFIKRQL
Sag 875 VPSL.EIVKD.............CKVFWKKLLDAKLMSQRKYDNLTKAERG.....GLTSDDKARFIQRQL
Spy 870 VPSE.EVVKK.............MKNYWRQLLNAKLITQRKFDNLTKAERG.....GLSELDKAGFIKRQL
Ana 676 ESVAWMANELHHRIAAAY...................PETTVMVYRGSITAAARKAAGIDSRINLIGEKG
Nme 668 NDTRYVNRFLCQFVADRMRLTG..............K.GKKRVFASNGQITNLLRGFW..........GLR
Cje 638 NDTRYIARLVLNYTKDYLDFLPLSDDENTKLNDTQKGSKVHVEAKSGMLTSALRHTW..........GFS
Tde 960 VETRQATKVAAKVLEKMFP..................ETKIVYSKAETVSMFRNKF..........DIV
Sth 929 VETRQITKHVARLLDEKFNNK...KDENNR.....AVRTVKIITLKSTLVSQFRKDF..........ELY
Smu 922 VETRQITKHVARILDERFNTE...TDENNK.....KIRQVKIVTLKSNLVSNFRKEF..........ELY
Sag 927 VETRQITKHVARILDERFNNE...LDSKGR.....RIRKVKIVTLKSNLVSNFRKEF..........GFY
Spy 922 VETRQITKHVAQILDSRMNTK...YDENDK.....LIREVKVITLKSKLVSDFRKDF..........QFY
Ana 727 RKDRIDRRMHAVDASVVALMEASVAKTLAERSSLRGEQRLTGKEQTWKQYTG.........S...TVGA.
Nme 714 KVRAENDRMHALDAVVVACSTVAMQQKITRFVRYKEMNAFDGKT..IDKETG.E.......V...LHQK.
Cje 698 AKDRNNHLMHAIDAVIIAYANNSIVKAFSDFKKEQESNSAEL....YAKKISEL.......D...YKNK.
Tde 1001 KCREINDFMHAHDAYLNIVVGNVYNTKFTNNPWN.....FIKEK...R..............DNP.KIADT
Sth 981 KVREINDFMHAHDAYLNAVVASALLKKYPKLEPE....FVYGD...YPKYNSFR.......ERKSATEKV
Smu 974 KVREINDYMHAHDAYLNAVIGKALLGVYPQLEPE....FVYGD...YPHFHGHK.......ENK.ATAKK
Sag 979 KIREVNNYMHAHDAYLNAVVAKAILTKYPQLEPE....FVYGD...YPKYNSYK.......TRKSATEKL
Spy 974 KVREINNYMHAHDAYLNAVVGTALIKKYPKLESE....FVYGD...YKVYDVRKMIAKSEQEIGKATAKY
Ana 784 REHFEMWRGHMLHLT.ELF.....................NERLAE.........DKVYVTQNIRLRLS
Nme 770 THFPQPWEFFAQEVMIRVFGKPDGKPEFEEADTPEKLRTLLAEKLSSRPEAVHEYVTPLFVSRAPNRKM.
Cje 753 RKFFEPFSGFRQK.....................................VLDKIDEIFVSKPERKKP.
Tde 1049 YNYYKVFDY.....................DVKRNNITA.WEKGKTIITVKDMLKRNTPIYTRQAACKKG.
Sth 1037 YFYSNIMNIFKKSISLA.DGRVIERPLIEVNEETGESV.WNKESDLATVRRVLSYPQVNVVKKVEEQNH.
Smu 1029 FFYSNIMNFFKKDDV.............RTDKNGEII.WKKDEHISNIKKVLSYPQVNIVKKVEEQTG.
Sag 1035 FFYSNIMNFFKTKVTLA.DGTVVVKDDIEVNNDTGEIV.WDKKKHFATVRKVLSYPQNNIVKKTEIQTG.
Spy 1037 FFYSNIMNFFKTEITLA.NGEIRKRPLIETNGETGEIV.WDKGRDFATVRKVLSMPQVNIVKKTEVQTG.
Ana 822 DGNAHTVNPSKLVSHRLGDGLTVQQ.............IDRACTPALWCALTREKDFDEKNGLPAREDR
Nme 839 SGQGHMETV..KSAKRLDEGVSVLRVPLTQLKLKDLEKMVNRREREPKLYEALKARLEAH........KDD
Cje 784 SGALHEETF..RKEEEFHQ...
Tde 1097 ..ELFNQTI...MKKG...
Sth 1104 ..GLDRGKP.....................................KGLFNANLSS.....KPKP...
Smu 1083 ..GFSKESI...LPKG...
Sag 1102 ..GFSKESI...LAHG...
Spy 1104 ..GFSKESI...LPKR...
Ana 878 AIRVHGHEIKSSDYIQVFSKRKKTDSDRDETPFGAIAVRGGFV......EI..GPSIHHARIYRVEGKKP
Nme 899 P.........AKAFA..............EPFYKYDKAGNRTQQVKAVRV..EQVQKTGVWVRNHNGI.
Cje 801 ..........................SYGGKEGVLKALEL.........GKIRKVNGKI.
Tde 1108 ....LGQ..HPLKKEGPFSNISKYGGYNKVSAAYYTLIEYEEKGNKIRSLETIPL
Sth 1125 ....NSNENL..VGAKEYLDPKKYGGYAGISNSFTVLVKGTIEKGAKKITNVLE
Smu 1094 ..................NSDKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKA
Sag 1113 ..................NSDKLIPRKTKDIYLDPKKYGGFDSPIVAYSVLVVADIKKGKAQKLKTVTE
Spy 1115 ....NSDKLIARKK...DWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKE
Ana 940 VYAMLRVFTHDLLSQ.......RHG......DLFSAVIPPQSI.SMRCAEPKL.RKAITTG....NATYL
Nme 942 .ADNATMVRVDVFE.........KG......DKYYLVP.IY...SWQVAKGILPDRAVVQGKDE.....E
Cje 825 .VKNGDMFRVDIFKH.......KT......NKFYAVP.IY...TMDFALKVLPNKAVVQGKDKKSGLIK
Tde 1157 YLVKDIQKDQDVLKSYLTDLLGKKE........FKILVP.KIKINSLLKINGFPCHI...TGKTNDSFLL.
Sth 1174 FQGISILDRINYRKD.KLNFLLEKGYKDI..ELIIELP.KYSLFELS..DGSRRMLASILSTNNKRGEI.
Smu 1145 LVGVTIMEKMTFERD.PVAFLERKGYRNVQEENIIKLP.KYSLFKLE..NGRKRLLAS.......AREL.
Sag 1164 LLGITIMERSRFEKN.PSAFLESKGYLNIRADKLIILP.KYSLFELE..NGRRLLLAS.......AGEL.
Spy 1163 LLGITIMERSSFEKN.PIDFLEAKGYKEVKKDLIIKLP.KYSLFELE..NGRKRMLAS.......AGEL.

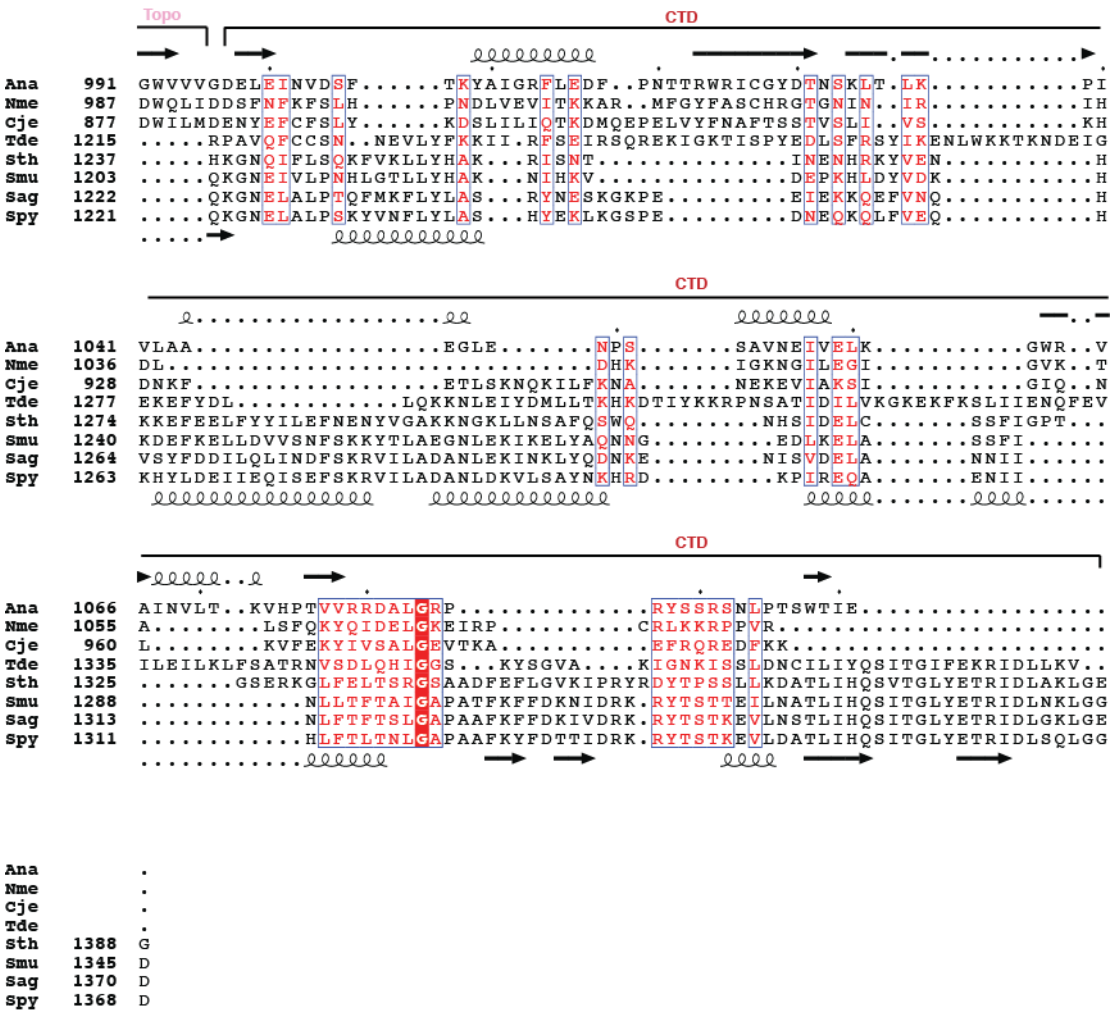


Figure S7. Multiple sequence alignment of Type II-A and II-C Cas9 orthologs. The primary sequences of Type II-C Cas9 orthologs from *Actinomyces naeslundii* (Ana), *Neisseria meningitidis* (Nme) and *Campylobacter jejuni* (Cje), together with type II-A Cas9 orthologs from *Treponema denticola* (Tde), *Streptococcus thermophilus* (Sth), *Streptococcus mutans* (Smu), *Streptococcus agalactiae* (Sag) and *Streptococcus pyogenes* (Spy) were aligned using CLUSTALW (*85*). The alignment was generated in ESPript (*83*) using default settings**.** Absolutely conserved residues are shown as white text on a red background, while similar residues are shown as red text with a white background. Red triangles indicate conserved residues in the RuvC active site, whereas conserved residues located in the HNH active site are denoted with a blue triangle. Green triangles indicates the tryptophan residues involved in PAM binding based on SpyCas9 crosslinking assay. The secondary structure of AnaCas9 derived from the crystal structure is marked on the top of the sequence alignment, whereas the secondary structure of SpyCas9 is shown at the bottom. Accession numbers for each Cas9 ortholog are as follows: Ana (*Actinomyces naeslundii* str. Howell 279, EJN84392.1), Nme (*Neisseria meningitidis*, WP_019742773.1), Cje (*Campylobacter jejuni*, WP_002876341.1), Tde (*Treponema denticola*, WP_002676671.1), Sth (*Streptococcus thermophilus* LMD-9, YP_820832.1), Smu (*Streptococcus mutans*, WP_019803776.1), Sag (*Streptococcus agalactiae*, WP_001040088.1), and Spy (*Streptococcus pyogenes*, YP_282132.1).

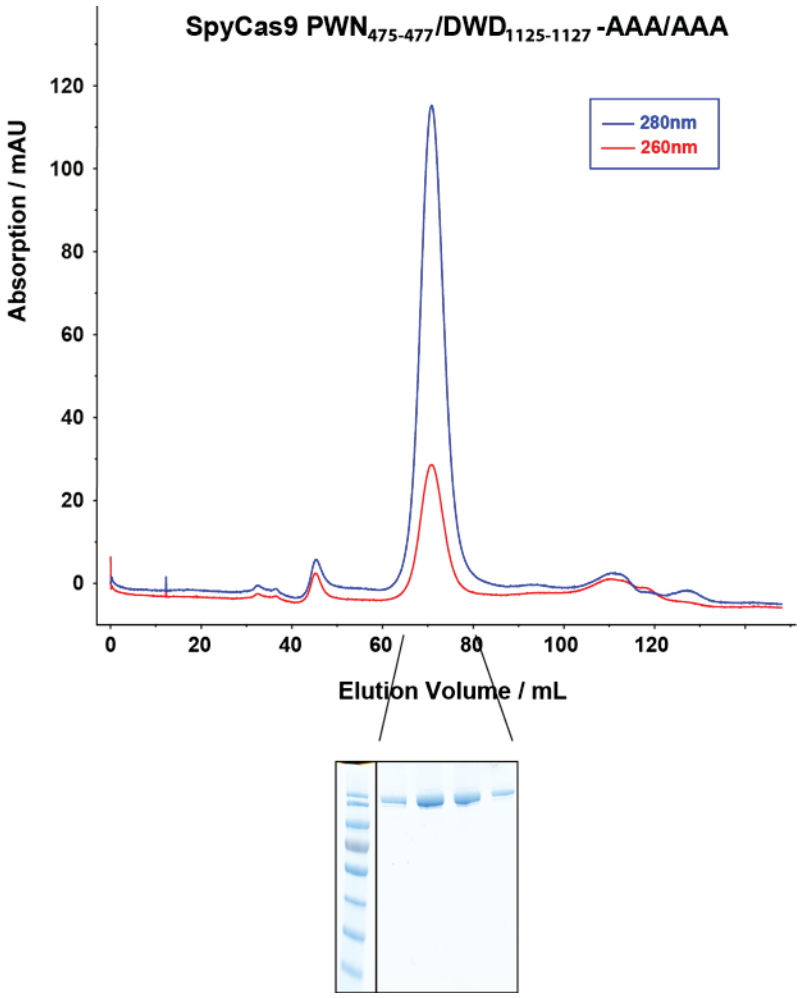


Figure S8. Size exclusion chromatogram of SpyCas9 PWN$_{475-477}$/DWD$_{1125-1127}$→AAA/AAA mutant. All SpyCas9 mutants in this study showed the same properties during purification as observed for the wild-type SpyCas9. The retention time during gel filtration chromatography on a Superdex 200 16/60 column (GE Healthcare) is comparable to WT SpyCas9 (*8*).

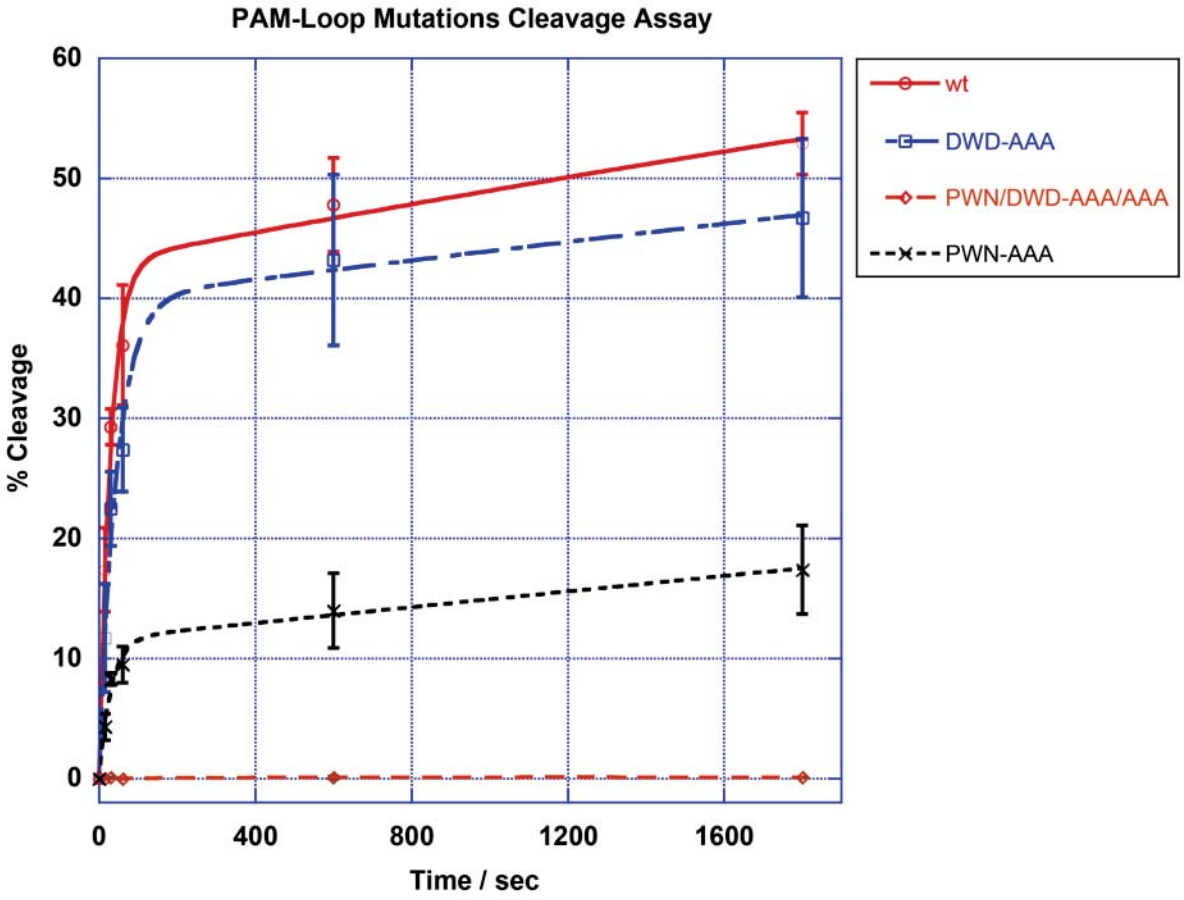


Figure S9. Quantification of DNA cleavage experiments with PAM-binding mutants. For cleavage experiments, 1 nM radiolabeled 55-bp dsDNA substrate was incubated with equimolar Cas9:RNA variants (wildtype, $PWN_{475\text{-}477}$→AAA and/or $DWD_{1125\text{-}1127}$→AAA/AAA) at room temperature. The reactions were quenched at various time points and resolved by 10% denaturing PAGE. DNA was visualized by phosphorimaging, quantified with ImageQuant (GE Healthcare), and analyzed with Kaleidagraph (Synergy Software). The results presented here show a decreased cleavage activity for the $PWN_{475\text{-}477}$→AAA mutant, whereas SpyCas9 mutated in both regions leads to a severe defect in dsDNA cleavage.

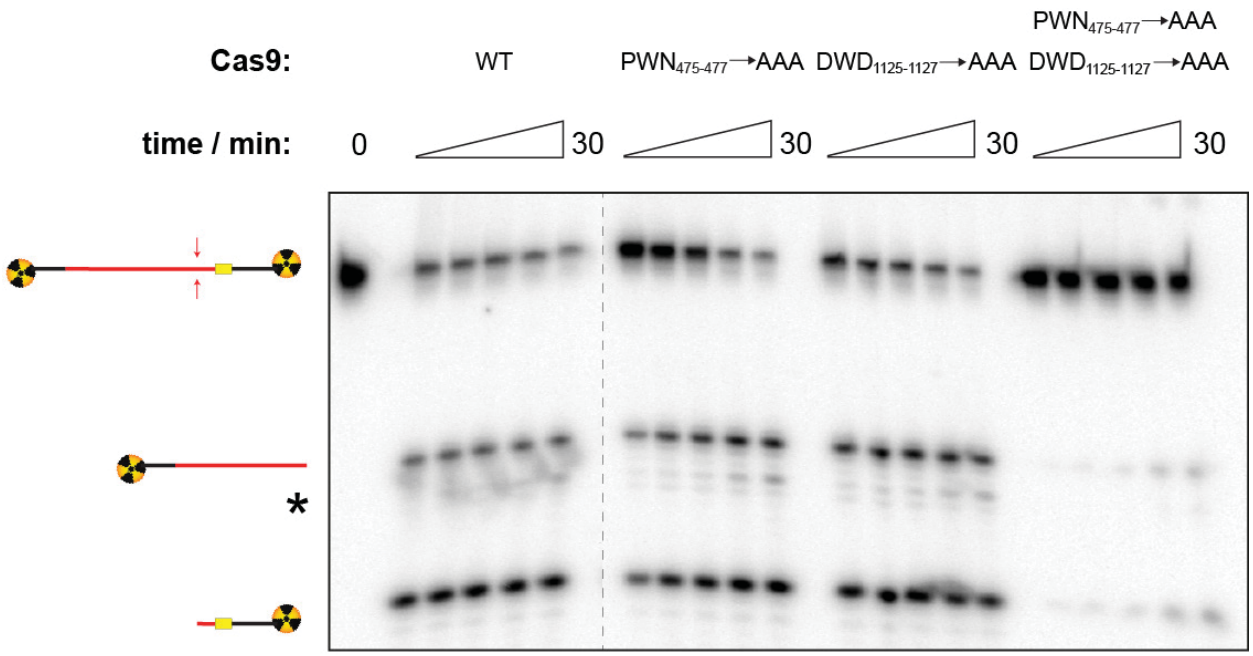


Figure S10. SpyCas9 $PWN_{475-477}/DWD_{1125-1127}$→AAA/AAA mutant is impaired in dsDNA substrate cleavage. In addition to equimolar cleavage conditions (Fig. 3D), reconstituted SpyCas9 variants were also tested at a 10-fold molar excess over dsDNA substrate concentration. Reactions contained 1 nM radiolabeled DNA substrate and 10 nM Cas9:RNA complex, and were conducted at room temperature. Aliquots were removed at 0.25, 0.5, 1, 10, and 30 minutes, quenched by mixing with formamide gel loading buffer containing 50 mM EDTA, and resolved by 10% denaturing PAGE. Reaction products were visualized by phosphorimaging.

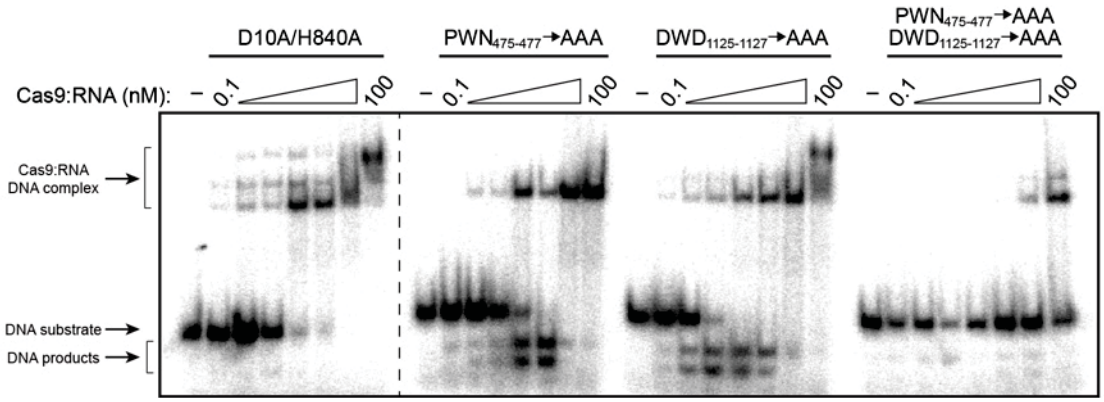


Figure S11. SpyCas9 $PWN_{475-477}$/$DWD_{1125-1127}$→AAA/AAA mutant is impaired in dsDNA binding. Target 55-bp dsDNA was incubated with increasing concentrations of the indicated Cas9:RNA mutants for 60 min before being resolved by 5% native PAGE. SpyCas9 mutated individually at $PWN_{475-477}$→AAA or $DWD_{1125-1127}$→AAA/AAA binds dsDNA with an affinity similar to catalytically inactive dCas9 (D10A/H840A), whereas SpyCas9 mutated in both regions is defective in dsDNA binding. Note that unbound DNA cleavage products exhibit a distinct mobility from intact substrate DNA.

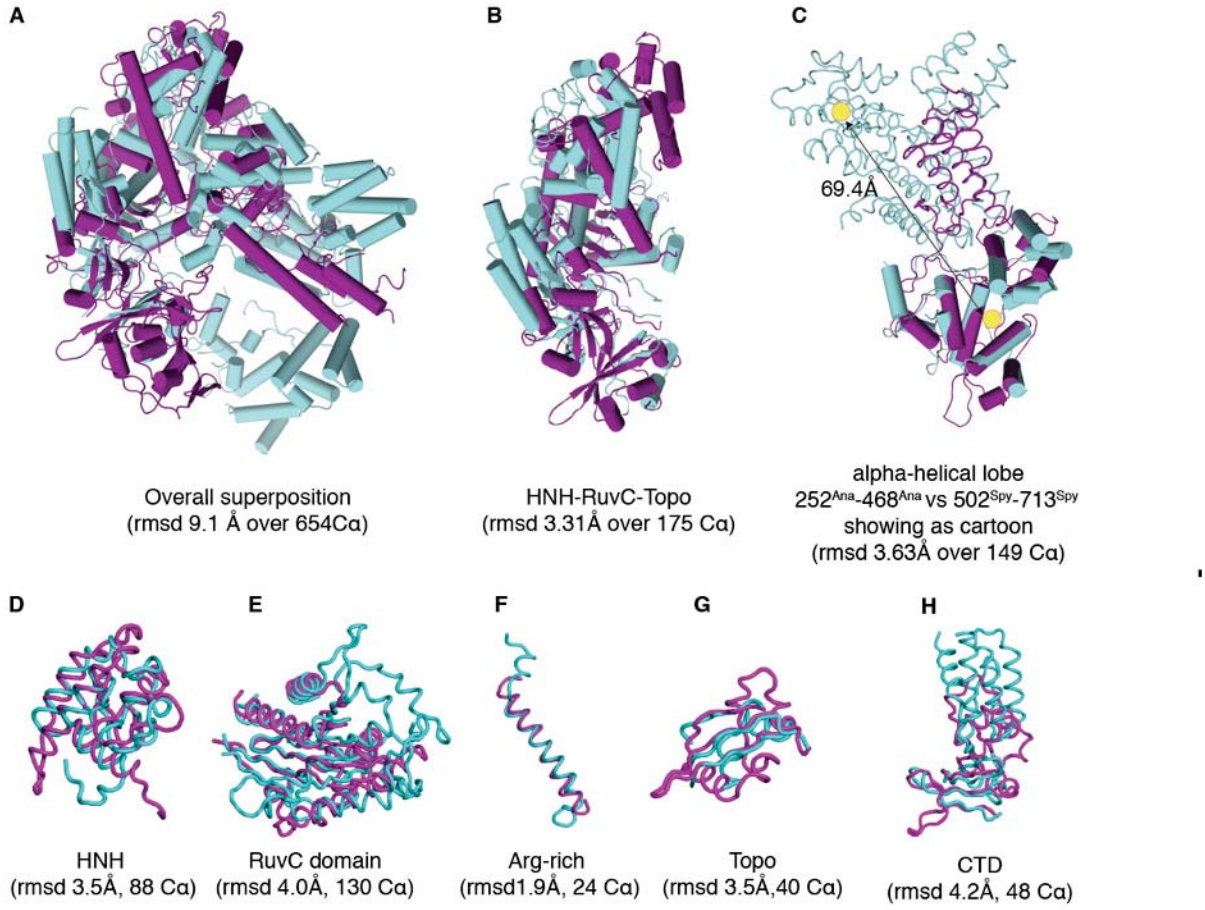


Figure S12. Pairwise structural comparisons of SpyCas9 and AnaCas9. (**A**) Overall structural alignment of AnaCas9 (purple) and SpyCas9 (cyan) showing a good alignment of the nuclease lobe but distinct structural features in the alpha-helical lobe. The superpositions were generated using the jCE algorithm (http://source.rcsb.org/jfatcatserver/). (**B**) Superposition of the catalytic core. For clarity, the alpha-helical lobe is not shown. (**C**) Superposition of the alpha-helical lobe, revealing structural similarity between 252^{Ana}-468^{Ana} and 502^{Spy}-713^{Spy}, with a large displacement of 69.4 Å towards the RuvC domain and an approximately 35° rotation about the junction between two domains in AnaCas9. The putative domain centers are labeled with yellow circles. (**D-H**) Individual domains of AnaCas9 superimposed onto the corresponding domains in SpyCas9 with root mean square deviation (rmsd) values for the equivalent alpha-carbons indicated.

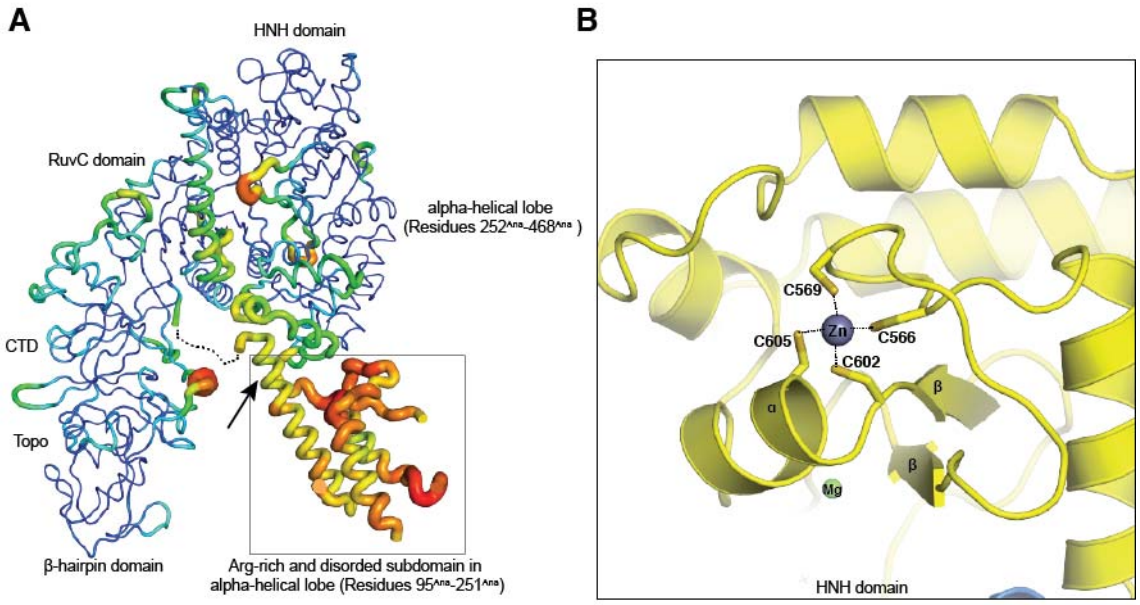


Figure S13. (**A**) **AnaCas9 displayed by B-factor putty.** Thin blue loops represent low B-values, while broad red tubes represent high B-values. The Arg-rich region and the neighboring alpha-helical part (box) have the highest B-factors in the structure, suggesting high flexibility in these regions. The hinge connecting the RuvC domain and the Arg-rich region is drawn as a dotted line. (**B**) **Close-up view of the zinc-binding site in the HNH domain of AnaCas9.** The zinc site is coordinated by residues C566^{Ana}, C569 ^{Ana}, C602^{Ana} and C605^{Ana}, and may serve to stabilize the AnaCas9 HNH domain architecture (ββα-Me fold).

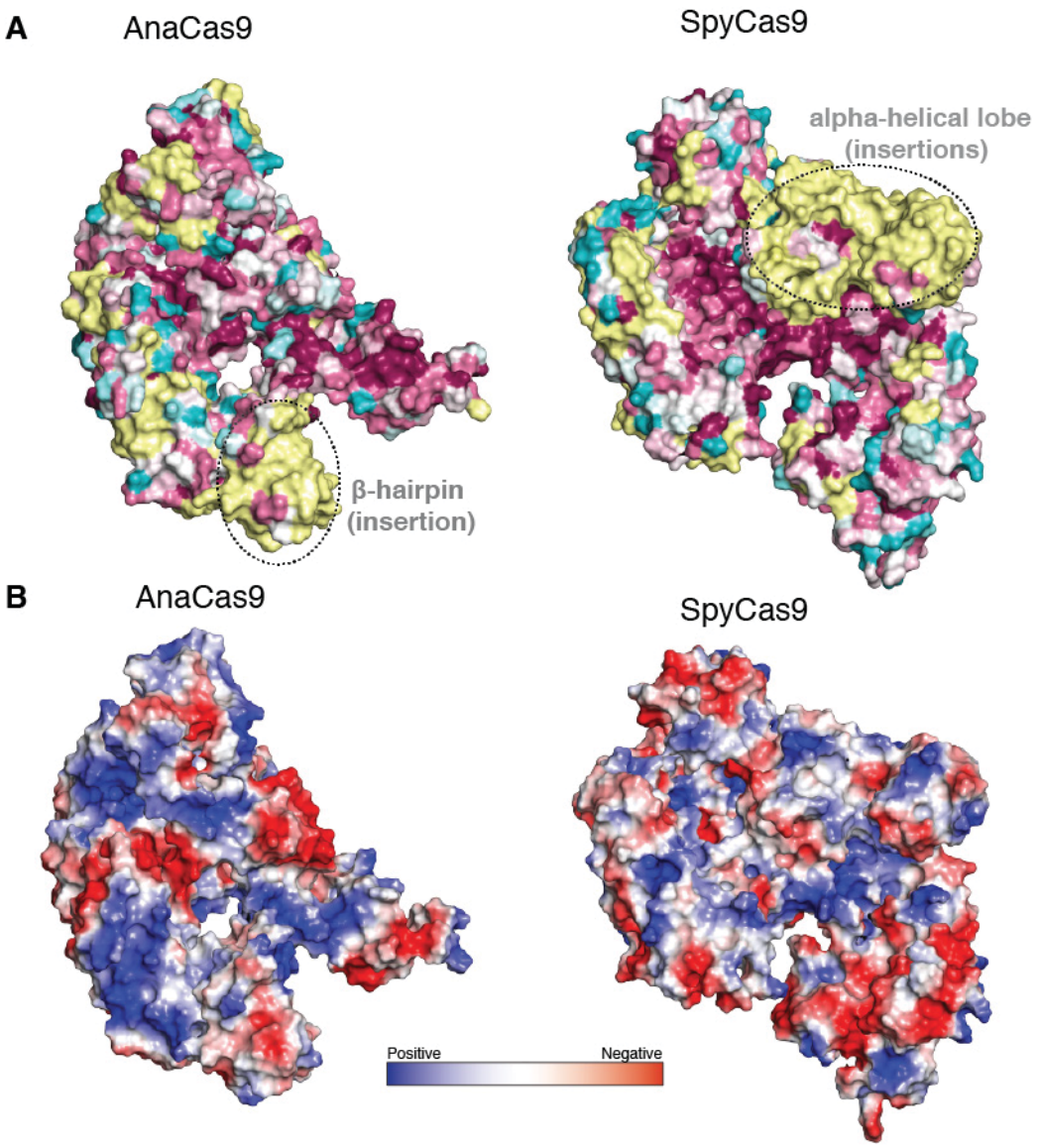


Figure S14. Surface features of SpyCas9 and AnaCas9 based on sequence conservation and electrostatic potential. (**A**) Surface conservation of AnaCas9 (left) and SpyCas9 (right), with the same orientation as in Fig 4. The surface is colored according to amino acid conservation among the Type-II Cas9 proteins shown in Fig. S7 by the Consurf Server (*61*), where purple/red represents highly conserved residues, while yellow/light green denotes the most variant residues in Type-II Cas9 orthologs. Notably, AnaCas9 harbors a β-hairpin domain insertion, whereas SpyCas9 has a large insertion in the alpha-helical lobe. (**B**) The same molecular surface representations of AnaCas9 (left) and SpyCas9 (right) are color-coded by electrostatic potential, as calculated by APBS (*60*) electrostatics in PyMOL (The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC).

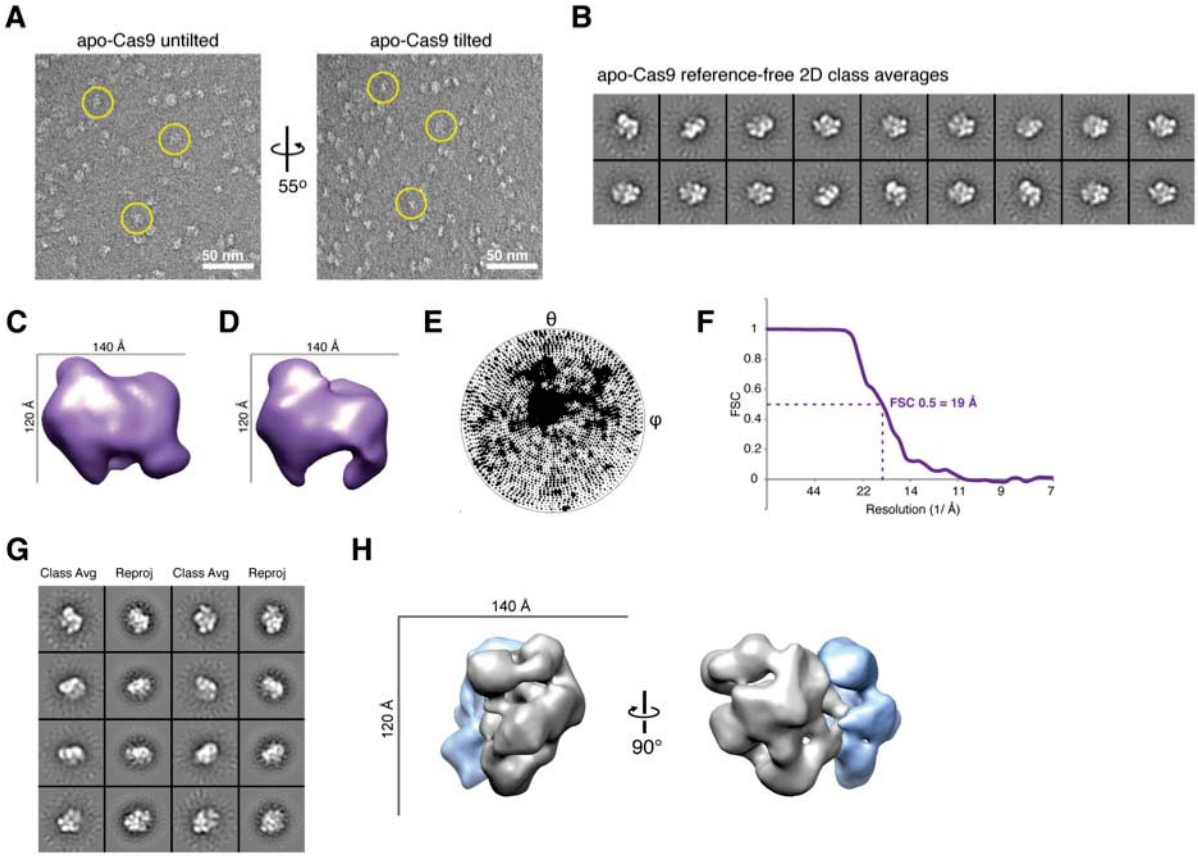


Figure S15. Molecular architecture of apo-SpyCas9. (**A**) Representative untilted (left) and tilted (right) micrographs of negatively stained apo-SpyCas9. Scale bar indicates 50 nm. (**B**) Reference-free 2D class averages of apo-SpyCas9. The width of the boxes is ~316 Å. (**C**) Random conical tilt (RCT) class volume showing the *ab initio* structure of apo-SpyCas9. (**D**) Initial model generated by assigning Euler angles of the reference-free class averages with respect to the RCT volume. This initial model was used for refinement of the raw particle images of apo-SpyCas9. (**E**) Euler angle distribution for the final reconstruction. (**F**) Fourier shell correlation (FSC) curve for the final reconstruction, showing the resolution to be ~19 Å using the 0.5 FSC criterion. (**G**) Reference-free 2D class averages of apo-SpyCas9 (first and third columns) matched to reprojections of the final reconstruction (second and fourth columns). The width of the boxes is ~316 Å. (**H**) Final reconstruction of apo-SpyCas9 using the map in (D) as the initial model for refinement. The final map is segmented and colored as in Fig. 5A.

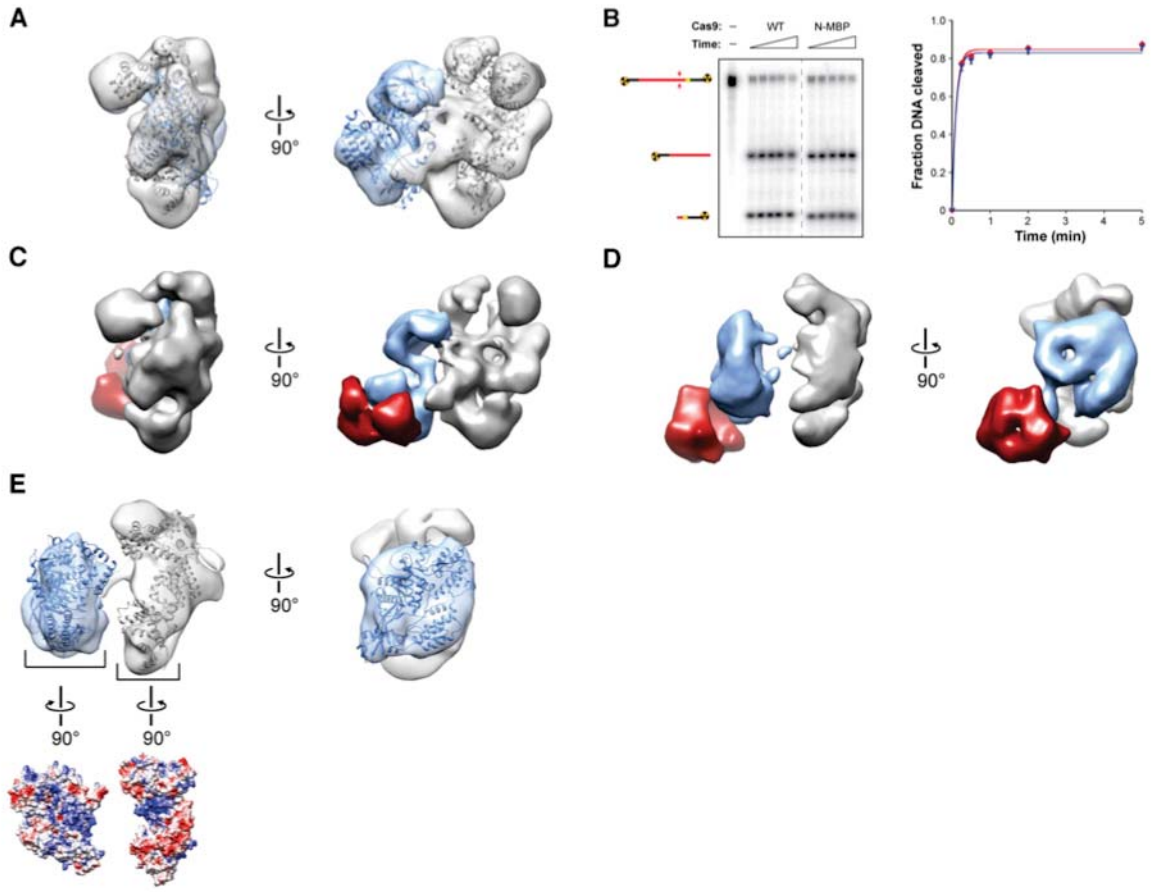


Figure S16. Structural similarities between the apo-SpyCas9 EM structure and X-ray crystal structure. (**A**) The X-ray crystal structure of SpyCas9 was split into the alpha-helical lobe (residues 66-713) and the RuvC nuclease-containing lobe (residues 1-65 and 744-1363). Both lobes were computationally docked into the apo-SpyCas9 EM density as separate rigid bodies using SITUS (*86*), due to flexibility in the RuvC nuclease-containing lobe (blue) in the absence of bound nucleic acids (see Fig. S1B, blurry, smaller lobe in class averages). The HNH domain was excluded from docking for the same reason. (**B**) Activity assay with WT and N-MBP SpyCas9. DNA cleavage experiments were performed and resolved by 10% denaturing polyacrylamide gel electrophoresis (left). The data were plotted (right) and fit with single-exponentials (solid lines); error bars represent the standard deviation from three independent experiments and are not always visible. (**C, D**) 3D difference maps (> 7-σ) (red density) between the N-terminal MBP-labeled and unlabeled reconstructions of apo-SpyCas9 (**C**) and SpyCas9:RNA:DNA (**D**) were mapped onto the corresponding unlabeled reconstructions. (**E**) The X-ray crystal structure of SpyCas9 was again split into the alpha-helical lobe and nuclease-containing lobe and both lobes were computationally docked into the SpyCas9:RNA:DNA EM density as separate rigid bodies using SITUS (*86*) (top). This docking result is consistent with a 100° rigid body rotation of the nuclease lobe toward the alpha-helical lobe and places the two nucleic acid binding clefts across from one another (electrostatic surface potential, below). Further experiments and/or higher-resolution structures will be required to verify this working model.

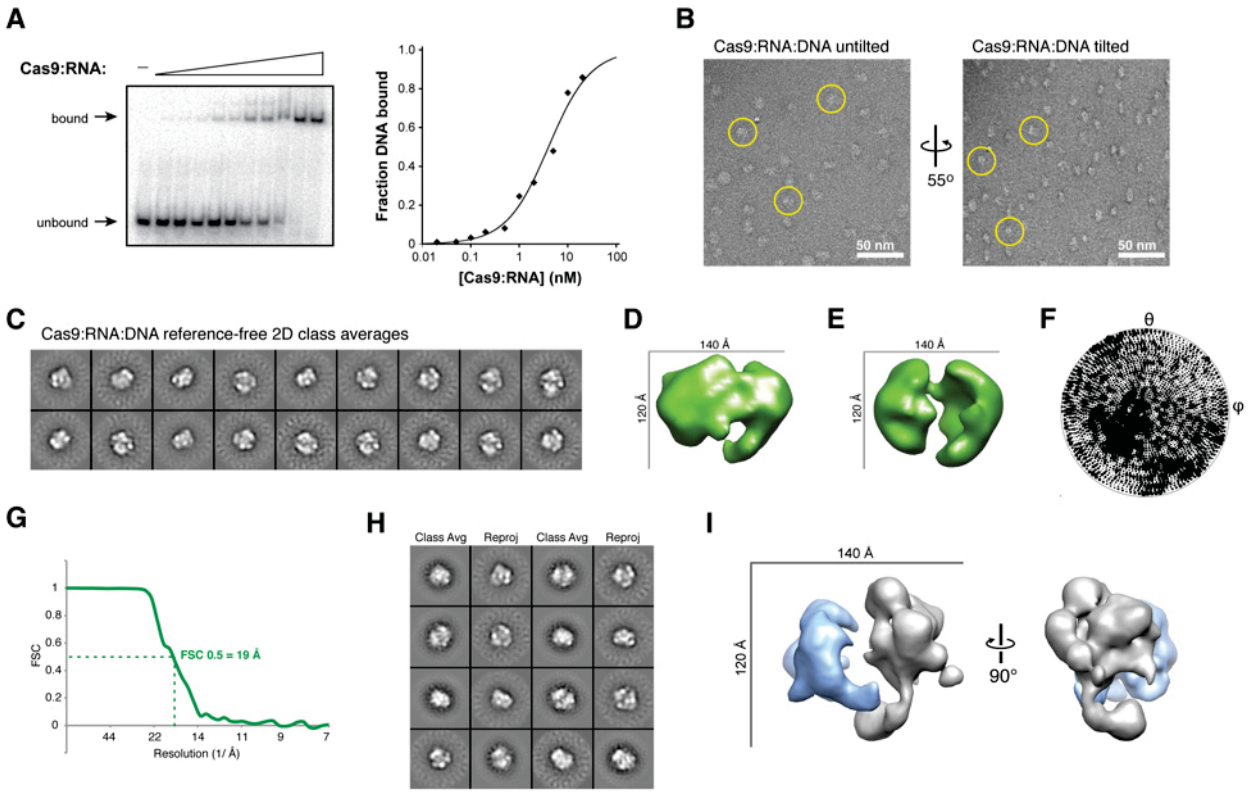


Figure S17. Molecular architecture of SpyCas9:RNA:DNA. (A) Electrophoretic mobility gel shift assay (left) with radiolabeled target DNA and increasing concentrations of catalytically inactive (D10A/H840A) SpyCas9:RNA complex. Fitting of the quantified data with a standard binding isotherm (solid line, right) yields an equilibrium dissociation constant (K_d) of 4.0 ± 0.4 nM. (**B**) Representative untilted (left) and tilted (right) micrographs of negatively stained SpyCas9:RNA:DNA. Scale bar indicates 50 nm. (**C**) Reference-free 2D class averages of SpyCas9:RNA:DNA. The width of the boxes is ~316 Å. (**D**) Random conical tilt (RCT) class volume showing the *ab initio* structure of SpyCas9:RNA:DNA. (**E**) Initial model generated by assigning Euler angles of the reference-free class averages with respect to the RCT volume. This initial model was used for refinement of the raw particle images of SpyCas9:RNA:DNA. (**F**) Euler angle distribution for the final reconstruction. (**G**) Fourier shell correlation (FSC) curve for the final reconstruction, showing the resolution to be ~19 Å using the 0.5 FSC criterion. (**H**) Reference-free 2D class averages of SpyCas9:RNA:DNA (first and third columns) matched to reprojections of the final reconstruction (second and fourth columns). The width of the boxes is ~316 Å. (**I**) Final reconstruction of SpyCas9:RNA:DNA using the map in (E) as the initial model for refinement. The final map is segmented and colored as in Fig. 5B.

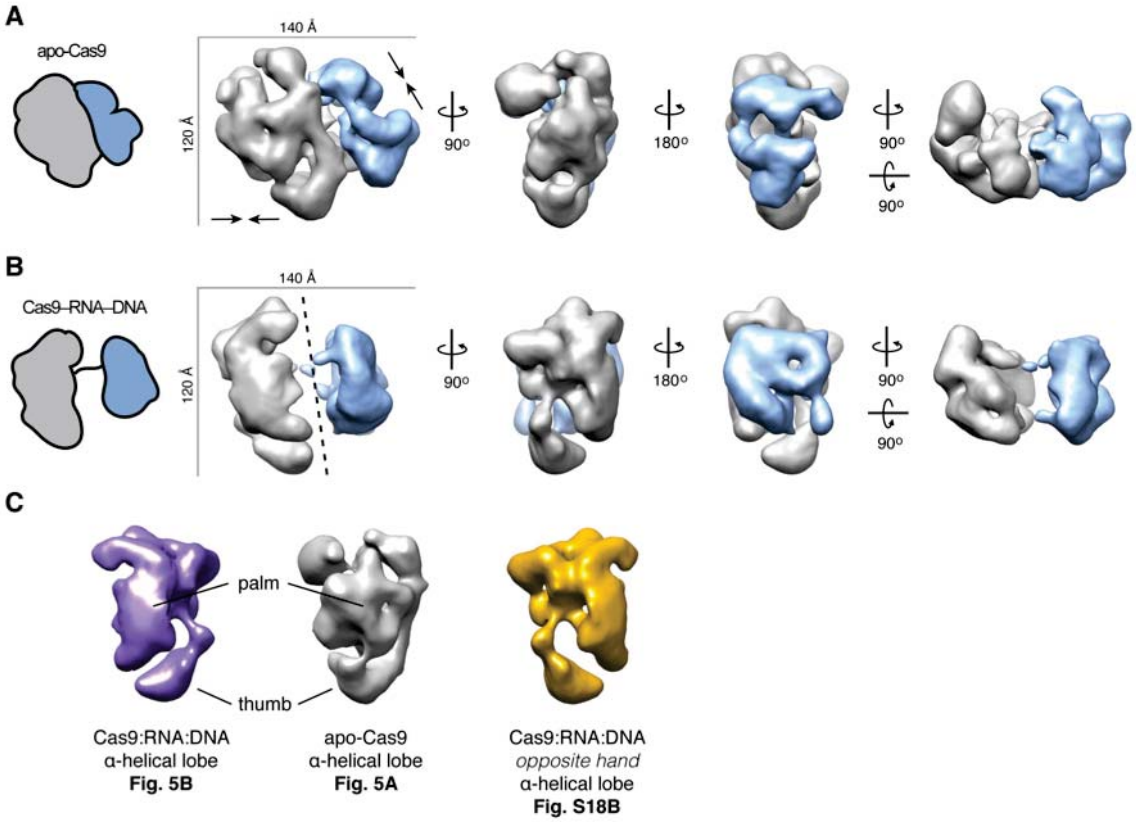


Figure S18. Alternative model for the conformational change in SpyCas9:RNA:DNA complex considering the opposite handedness. (**A, B**) Single particle EM reconstructions of negatively stained apo-SpyCas9 (**A**) (as in the main text) and SpyCas9:RNA:DNA (**B**) with opposite handedness as the structure presented in Fig. 5B. Cartoon representations are shown on the left. In this alternative model, the movement of the smaller lobe with respect to the larger one is subtler. The blue lobe rotates in towards the larger lobe and reorganizes to form the central channel spanning the length of the enzyme (black dashed line). Note that the grey lobes of the two structures are aligned differently than the alignment in the main text, to maintain the blue lobe in a similar relative position for the two structures. (**C**) From left to right: the α-helical lobe of SpyCas9:RNA:DNA from Fig. 5B (purple), apo-Cas9 (grey), and Cas9:RNA:DNA with opposite handedness from (**B**) (gold), aligned to one another based on optimal cross correlation coefficient (CCC). The favored model presented in Fig. 5B of the main text is based on the more obvious, direct correspondence between the features of the apo-SpyCas9 α-helical lobe (grey) and the SpyCas9:RNA:DNA α-helical lobe (purple). Additionally, the α-helical domain from the crystal structure exhibits a higher CCC with the α-helical lobe from the model presented in Fig. 5B of the main text (purple) than the α-helical lobe of opposite handedness (gold) (0.83 versus 0.74).

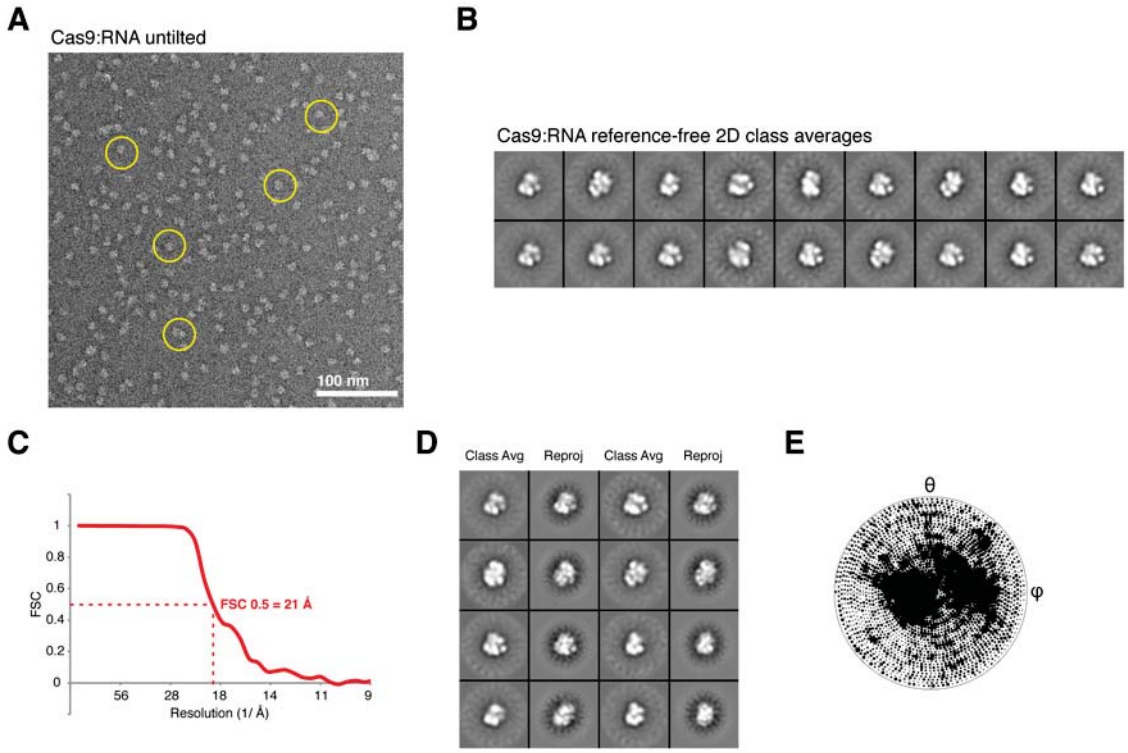


Figure S19. Molecular architecture of SpyCas9:RNA. (**A**) Representative untilted micrograph of negatively stained SpyCas9:RNA. Scale bar indicates 100 nm. (**B**) Reference-free 2D class averages of SpyCas9:RNA. The width of the boxes is ~316 Å. (**C**) Fourier shell correlation (FSC) curve for the final reconstruction, showing the resolution to be ~21 Å using the 0.5 FSC criterion. (**D**) Reference-free 2D class averages of SpyCas9:RNA (first and third columns) matched to reprojections of the final reconstruction (second and fourth columns). The width of the boxes is ~316 Å. (**E**) Euler angle distribution for the final reconstruction.

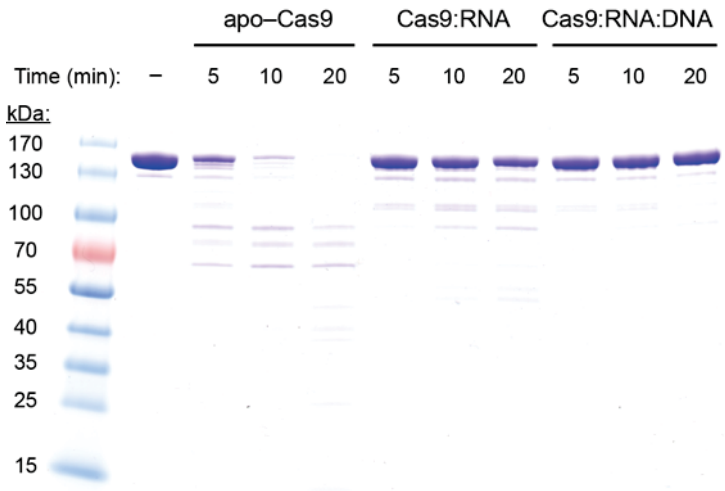


Figure S20. Limited proteolysis of SpyCas9 with and without nucleic acid substrates suggests that nucleic acid-bound complexes adopt similar structural states. Apo-SpyCas9, SpyCas9 bound to full-length crRNA and tracrRNA (SpyCas9:RNA), or RNA-programmed SpyCas9 in complex with target DNA (SpyCas9:RNA:DNA) were prepared at a concentration of 2.5 μM and incubated with 2 ng/μl trypsin at 37 °C for the indicated time before quenching with 2X SDS gel-loading buffer. Samples were resolved by SDS-PAGE on a 4-20% gradient polyacrylamide gel (Bio-Rad). Apo-SpyCas9 is rapidly proteolyzed, whereas both SpyCas9:RNA and SpyCas9:RNA:DNA complexes are resistant to digestion by trypsin, suggesting that SpyCas9 undergoes similar structural rearrangements in both cases that mitigate proteolysis. Complexes were prepared with catalytically inactive D10A/H840A-SpyCas9 under the same conditions used to prepare samples for electron microscopy imaging.

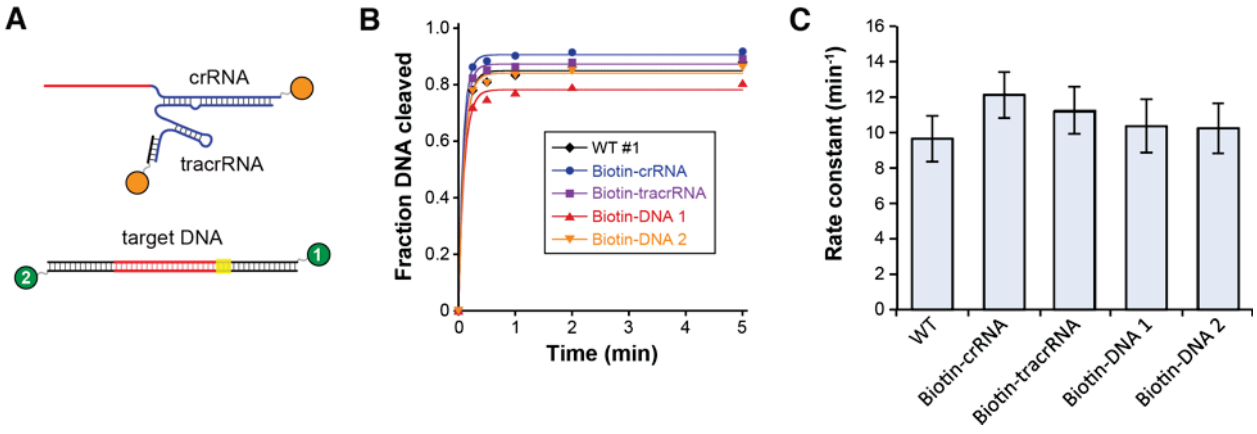


Figure S21. Activity assays with biotin-RNA and biotin-DNA substrates used in streptavidin labeling experiments. (A) Schematic depicting the attachment of biotin (orange and green circles) to each nucleic acid substrate. Note that the crRNA and each strand of the DNA target are covalently linked to biotin at their 3' ends, whereas tracrRNA is hybridized to a short biotinylated DNA oligonucleotide at its 3' end. **(B)** DNA cleavage assays were conducted with biotin-labeled nucleic acids to verify that the modification does not perturb DNA recognition and cleavage. Data from representative time courses were plotted and fit with single-exponentials (solid line) to yield first-order rate constants for the DNA cleavage reaction. Note that the steep part of the curve (<15 seconds) could not be well defined due to the rapid reaction rate, limiting the accuracy of these measurements. **(C)** Three independent DNA cleavage time courses were conducted for each SpyCas9 construct, and the averaged rate constants are shown in the bar graph. The fitting error for individual single-exponential fits was greater than the standard deviation in rate constants between independent replicates, and so error bars represent the fitting error averaged from three independent experiments.

Table S1. Highest cross-correlation coefficients (CCC) obtained by docking the apo-SpyCas9 alpha-helical domain crystal structure as a rigid body into the apo-SpyCas9 and spyCas9:RNA:DNA EM-derived alpha-helical lobes using SITUS.

	apo-SpyCas9 EM density		SpyCas9:RNA:DNA EM density	
	α-helical lobe	**α-helical lobe opposite hand**	**α-helical lobe**	**α-helical lobe opposite hand**
α-helical lobe crystal structure	0.74	0.70	0.83	0.74

Table S2. List of nucleic acid reagents used in this study

#	Description	Sequence (5'-3')
1	tracrRNA (nts 15-87)	GGACAGCAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUU
2	Targeting crRNA	GUGAUAAGUGGAAUGCCAUGGUUUUAGAGCUAUGCUGUUUUG
3	55-bp DNA substrate, non-target strand[a]	GAGTGGAAGGATGCCAGTGATAAGTGGAATGCCATGTGGGCTGTCAAAATTGAGC
4	55-bp DNA substrate, target strand[a]	GCTCAATTTTGACAGCCCACATGGCATTCCACTTATCACTGGCATCCTTCCACTC
5	Br-dU_1 containing 55 nt DNA substrate, non-target strand[a]	GAGTGGAAGGATGCCAGTGATAAGTGGAATGCCATG(Br-dU_1)GGGCTGTCAAAATTGAGC
6	Br-dU_2 containing 55 nt DNA substrate, target strand[a]	GCTCAATTTTGACAGCCC(Br-dU_2)CATGGCATTCCACTTATCACTGGCATCCTTCCACTC
7	reverse complement for # 6[a]	GAGTGGAAGGATGCCAGTGATAAGTGGAATGCCATGAGGGCTGTCAAAATTGAGC
8	Br-dU_3 containing 55 nt DNA substrate, non-target strand[a]	GAGTGGAAGGATGCCAGTGATAAGTGGAATGCCATGTGG(Br-dU_3)CTGTCAAAATTGAGC
9	reverse complement for #8[a]	GCTCAATTTTGACAGACCACATGGCATTCCACTTATCACTGGCATCCTTCCACTC
10	tracrRNA_ext[b]	GGACAGCAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUUGCUCGUGCGC
11	Biotinylated DNA oligo to hybridize to tracrRNA_ext[b]	Biotin-TTGCGCACGAGCAAA
12	Non-targeting crRNA (control, Fig. 7b)	GACGCAUAAAGAUGAGACGCGUUUUAGAGCUAUGCUGUUUUG
13	3'-Biotinylated DNA, non-target strand[c]	GAGTGGAAGGATGCCAGTGATAAGTGGAATGCCATG**TGG**GCTGTCAAAATTGAGC-Biotin
14	3'-Biotinylated DNA, target strand[c]	GCTCAATTTTGACAGCCCACATGGCATTCCACTTATCACTGGCATCCTTCCACTC-Biotin
15	ssDNA template for transcribing tracrRNA[c]	AAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATGCTGT**CCTATAGTGAGTCGTATTA**
16	ssDNA template for transcribing tracrRNA_ext[c]	GCGCACGAGCAAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATGCTGT**CCTATAGTGAGTCGTATTA**
17	Oligo for preparing double-stranded T7 promoters for *in vitro* transcription	TAATACGACTCACTATA

[a] The protospacer is depicted in red. The PAM is underlined.

[b] Nucleotides hybridizing between the tracrRNA_ext and biotin-DNA oligo are in blue.

[c] The reverse complement of the T7 promoter is indicated in **bold**.

Supplementary References

1. B. Wiedenheft, S. H. Sternberg, J. A. Doudna, *Nature* **482**, 331–338 (2012).

2. S. Al-Attar, E. R. Westra, J. van der Oost, S. J. J. Brouns, *Biol. Chem.* **392**, 277–289 (2011).

3. M. P. Terns, R. M. Terns, *Curr Opin Microbiol* **14**, 321–327 (2011).

4. R. Sorek, C. M. Lawrence, B. Wiedenheft, *Annu Rev Biochem* **82**, 237–266 (2013).

5. R. Barrangou *et al.*, *Science* **315**, 1709–1712 (2007).

6. S. J. J. Brouns *et al.*, *Science* **321**, 960–964 (2008).

7. K. S. Makarova *et al.*, *Nat Rev Microbiol* **9**, 467–477 (2011).

8. M. Jinek *et al.*, *Science* **337**, 816–821 (2012).

9. T. Karvelis *et al.*, *RNA biology* **10** (2013).

10. G. Gasiunas, R. Barrangou, P. Horvath, V. Siksnys, *Proc Natl Acad Sci US A* **109**, E2579–86 (2012).

11. T. R. Sampson, S. D. Saroj, A. C. Llewellyn, Y.-L. Tzeng, D. S. Weiss, *Nature*, 1–5 (2013).

12. K. S. Makarova, L. Aravind, Y. I. Wolf, E. V. Koonin, *Biol Direct* **6**, 38 (2011).

13. K. Chylinski, A. Le Rhun, E. Charpentier, *RNA biology* **10**, 726–737 (2013).

14. P. Mali, K. M. Esvelt, G. M. Church, *Nat Methods* **10**, 957–963 (2013).

15. P. Mali *et al.*, *Science* **339**, 823–826 (2013).

16. L. Cong *et al.*, *Science* **339**, 819–823 (2013).

17. M. Jinek *et al.*, *elife* **2**, e00471 (2013).

18. L. S. Qi *et al.*, *Cell* **152**, 1173–1183 (2013).

19. L. A. Gilbert *et al.*, *Cell* **154**, 442–451 (2013).

20. P. Mali *et al.*, *Nat Biotechnol* (2013), doi:10.1038/nbt.2675.

21. Z. Hou *et al.*, *Proc Natl Acad Sci US A* (2013), doi:10.1073/pnas.1313587110.

22. K. M. Esvelt *et al.*, *Nat Methods* (2013), doi:10.1038/nmeth.2681.

23. J. M. Berger, S. J. Gamblin, S. C. Harrison, J. C. Wang, *Nature* **379**, 225–232 (1996).

24. R. Sapranauskas *et al.*, *Nucleic Acids Res* **39**, 9275–9282 (2011).

25. J. E. Garneau *et al.*, *Nature* **468**, 67–71 (2010).

26. Y. Zhang *et al.*, *Mol Cell* **50**, 488–503 (2013).

27. I. Fonfara *et al.*, *Nucleic Acids Res* (2013), doi:10.1093/nar/gkt1074.

28. K. M. Górecka, W. Komorowska, M. Nowotny, *Nucleic Acids Res* (2013), doi:10.1093/nar/gkt769.

29. W. Yang, *Nat Struct Mol Biol* **15**, 1228–1231 (2008).

30. W. Yang, *Quart. Rev. Biophys.* **44**, 1–93 (2011).

31. W. Yang, J. Y. Lee, M. Nowotny, *Mol Cell* **22**, 5–13 (2006).

32. I. Ivančić Baće, J. A. L. Howard, E. L. Bolt, *J Mol Biol* **422**, 607–616 (2012).

33. B. W. Shen, M. Landthaler, D. A. Shub, B. L. Stoddard, *J Mol Biol* **342**, 43–56 (2004).

34. M. Radermacher, T. Wagenknecht, A. Verschoor, J. Frank, *J Microsc* **146**, 113–136 (1987).

35. P. Chacón, W. Wriggers, *J Mol Biol* **317**, 375–384 (2002).

36. M. M. Jore *et al.*, *Nat Struct Mol Biol* **18**, 529–536 (2011).

37. C.-G. Yang *et al.*, *Nature* **452**, 961–965 (2008).

38. Y. Qi *et al.*, *Nature* **462**, 762–766 (2009).

39. C. M. Crenshaw *et al.*, *Journal of Biological Chemistry* **287**, 24916–24928 (2012).

40. J. Song, O. Rechkoblit, T. H. Bestor, D. J. Patel, *Science* **331**, 1036–1040 (2011).

41. D. G. Sashital, B. Wiedenheft, J. A. Doudna, *Mol Cell* **46**, 606–615 (2012).

42. E. Semenova, M. Nagornykh, M. Pyatnitskiy, I. I. Artamonova, K. Severinov, *FEMS Microbiol. Lett.* **296**, 110–116 (2009).

43. E. Semenova *et al.*, *Proc Natl Acad Sci US A* **108**, 10098–10103 (2011).

44. B. Wiedenheft *et al.*, *Nature* **477**, 486–489 (2011).

45. R. H. J. Staals *et al.*, *Mol Cell* **52**, 135–145 (2013).

46. C. Rouillon *et al.*, *Mol Cell* **52**, 124–134 (2013).

47. M. Spilman *et al.*, *Mol Cell* **52**, 146–152 (2013).

48. T. Sinkunas *et al.*, *EMBO J* **32**, 385–394 (2013).

49. W. Kabsch, *Acta Crystallogr D Biol Crystallogr* **66**, 125–132 (2010).

50. C. Vonrhein, E. Blanc, P. Roversi, G. Bricogne, *Methods Mol Biol* **364**, 215–230 (2007).

51. T. Terwilliger, *J Synchrotron Radiat* **11**, 49–52 (2004).

52. P. Emsley, P. Emsley, K. Cowtan, K. Cowtan, *Acta Crystallogr D Biol Crystallogr* **60**, 2126–2132 (2004).

53. P. V. Afonine *et al.*, *Acta Crystallogr D Biol Crystallogr* **68**, 352–367 (2012).

54. H. R. Powell, O. Johnson, A. G. W. Leslie, *Acta Crystallogr D Biol Crystallogr* **69**, 1195–1203 (2013).

55. P. Evans, *Acta Crystallogr D Biol Crystallogr* **62**, 72–82 (2006).

56. G. N. Murshudov, A. A. Vagin, E. J. Dodson, *Acta Crystallogr D Biol Crystallogr* **53**, 240–255 (1997).

57. C. Suloway *et al.*, *J Struct Biol* **151**, 41–60 (2005).

58. M. Hohn *et al.*, *J Struct Biol* **157**, 47–55 (2007).

59. G. Tang *et al.*, *J Struct Biol* **157**, 38–46 (2007).

60. N. A. Baker, D. Sept, S. Joseph, M. J. Holst, J. A. McCammon, *Proc Natl Acad Sci USA* **98**, 10037–10041 (2001).

61. H. Ashkenazy, E. Erez, E. Martz, T. Pupko, N. Ben-Tal, *Nucleic Acids Res* **38**, W529–33 (2010).

62. B. Wiedenheft *et al.*, *Structure* **17**, 904–912 (2009).

63. T. R. Schneider, G. M. Sheldrick, *Acta Crystallogr D Biol Crystallogr* **58**, 1772–1779 (2002).

64. P. D. Adams *et al.*, *Acta Crystallogr D Biol Crystallogr* **66**, 213–221 (2010).

65. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, *Acta Crystallogr D Biol Crystallogr* **66**, 486–501 (2010).

66. I. W. Davis *et al.*, *Nucleic Acids Res* **35**, W375–83 (2007).

67. P. Roepstorff, J. Fohlman, *Biomed. Mass Spectrom.* **11**, 601 (1984).

68. P. H. Zwart *et al.*, *Methods Mol Biol* **426**, 419–435 (2008).

69. A. J. McCoy *et al.*, *J Appl Crystallogr* **40**, 658–674 (2007).

70. S. P. Mallick, B. Carragher, C. S. Potter, D. J. Kriegman, *Ultramicroscopy* **104**, 8–29 (2005).

71. A. M. Roseman, *J Struct Biol* **145**, 91–99 (2004).

72. S. H. W. Scheres, R. Núñez-Ramírez, C. O. S. Sorzano, J. M. Carazo, R. Marabini, *Nat Protoc* **3**, 977–990 (2008).

73. N. R. Voss *et al.*, *J Struct Biol* **169**, 389–398 (2010).

74. G. C. Lander *et al.*, *J Struct Biol* **166**, 95–102 (2009).

75. N. R. Voss, C. K. Yoshioka, M. Radermacher, C. S. Potter, B. Carragher, *J Struct Biol* **166**, 205–213 (2009).

76. M. van Heel, G. Harauz, E. V. Orlova, R. Schmidt, M. Schatz, *J Struct Biol* **116**, 17–24 (1996).

77. J. Frank *et al.*, *J Struct Biol* **116**, 190–199 (1996).

78. B. Wiedenheft *et al.*, *Proc Natl Acad Sci US A* **108**, 10092–10097 (2011).

79. S. J. Ludtke, P. R. Baldwin, W. Chiu, *J Struct Biol* **128**, 82–97 (1999).

80. E. F. Pettersen *et al.*, *J Comput Chem* **25**, 1605–1612 (2004).

81. G. D. Pintilie, J. Zhang, T. D. Goddard, W. Chiu, D. C. Gossard, *J Struct Biol* **170**, 427–438 (2010).

82. K. Katoh, D. M. Standley, *Methods Mol Biol* **1079**, 131–146 (2014).

83. P. Gouet, E. Courcelle, D. I. Stuart, F. Métoz, *Bioinformatics* **15**, 305–308 (1999).

84. L. Holm, C. Sander, *Trends Biochem Sci* **20**, 478–480 (1995).

85. M. A. Larkin *et al.*, *Bioinformatics* **23**, 2947–2948 (2007).

86. W. Wriggers, S. Birmanns, *J Struct Biol* **133**, 193–202 (2001).