Supplementary material
for
Integrative analysis reveals unique structural and functional features of the Smc5/6 complex

You Yu[1]*, Shibai Li[2]*, Zheng Ser[3,4]*, Tanmoy Sanyal[5,6]*, Koyi Choi[2], Bingbing Wan[2,#], Huihui Kuang[8], Andrej Sali[5,6,7], Alex Kentsis[3], Dinshaw J. Patel[1] and Xiaolan Zhao[2]

1, Structural Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA

2, Molecular Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA

3, Molecular Pharmacology Program, Tow Center for Developmental Oncology, Department of Pediatrics, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA

4, Tri-Institutional PhD Program in Chemical Biology, New York, NY 10065, USA

5, Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, CA 94158, USA

6, Quantitative Biosciences Institute, University of California San Francisco, San Francisco, CA, 94158, USA

7, Department of Pharmaceutical Chemistry, University of California, San Francisco, San Francisco, CA94158, USA.

8, Simons Electron Microscopy Center, New York Structural Biology Center, New York, NY, 10027, USA

Current address: Key Laboratory of Systems Biomedicine and Collaborative Innovation Center of Systems Biomedicine, Shanghai Center for Systems Biomedicine, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai 200240, China

* These authors contributed equally to this work.

Corresponding authors: zhaox1@mskcc.org (XZ) and pateld@mskcc.org (DJP)

Supplemental Methods

Yeast strains and genetic methods and plasmids. All yeast strains are derivatives of W1588-4C, a *RAD5* derivative of W303 (*MATa ade2-1 can1-100 ura3-1 his3-11,15 leu2-3,112 trp1-1 rad5-535*). For expressing the Smc5/6 holo-complex, genes encoding its eight subunits were driven by the galactose inducible promoters and the constructs were integrated into the yeast genome. For other *in vivo* assays, at least two strains per genotype were examined in each experiment, and only one is listed in Table S2. Standard procedures were used for cell growth, media preparation, epitope tagging at endogenous loci and spot assays. Plasmids used are listed in Table S3, respectively. Standard yeast two-hybrid assay procedure was used (1). Plasmids containing the Gal4 activation domain (AD) and Gal4 DNA binding domain (DB) with or without the fusion of genes encoding the proteins of interest were transformed in the two-hybrid host strain pJ69-4. The resultant transformants were mixed for each pair of plasmids and spotted on plates containing SC-Trp-Leu (selection of plasmids), SC-Trp-Leu-His with or without 3AT (report the expression of the *HIS3* reporter), and SC-Trp-Leu-Ade (report the expression of the *ADE2* reporter) media. Pictures were taken after plates were incubated at 30 °C for 48 h. Nse5 and Nse6 mutations were generated by site directed mutagenesis by PCR. All constructs were sequenced to verify the correct mutations.

Smc5/6 holo-complex purification. Cells harboring over-expression constructs for the subunits of the Smc5/6 complex were growth at 30 ºC in YP-GL (YP + 2% glycerol / 2% lactic acid) media till log phase. Proteins expression was induced by adding 2% galactose for 4 hours. Harvested cells were resuspended in buffer E (45 mM Hepes-KOH pH 7.6, 10% glycerol, 0.02% NP40) supplemented with 100 mM NaCl, 1 mM DTT, Protease inhibitor cocktail (Sigma), and cOmpleteTM Ultra EDTA free protease inhibitor (Roche), and frozen dropwise in liquid nitrogen before breakage by using a freezer mill (SPEX CertiPrep 6850 Freezer/Mill). Cell powder was resuspended with buffer E supplemented with 300 mM NaCl and 1 mM DTT before centrifuged for 30 minutes at 40,000 rpm to remove debris. The supernatant was supplemented with 2 mM $CaCl_2$ and incubated with calmodulin resin for 2 h at 4 ºC. After washing the resins with 10 bed volume of buffer E supplemented with 300 mM NaCl, 2 mM $CaCl_2$ and 1 mM DTT, proteins were eluted using the same buffer without $CaCl_2$ and supplemented with 1 mM EDTA and 2 mM EGTA. Peak fractions were pooled and subjected to gel-filtration on a Superose 6 Increase column. Peak fractions were collected and snap frozen for storage.

Negative stain EM of the Smc5/6 holo-complex with added ATPγS. Purified SMC5/6 holo-complex was crosslinked at a concentration of 0.007 mg/ml with BS3 at a molar ratio of 1:600 in buffer E supplemented with 200 mM NaCl, 1 mM ATPγS, and 2 mM $MgCl_2$ for 1 hour at 4°C. The reaction was quenched using 50 mM Tris-HCl pH 7.5. The Smc5/6 holo-complex was then applied to glow discharged 300 mesh copper grid with continuous carbon film coated. After 30 seconds, the grid was blotted, washed and stained three times using 2% w/v Uranyl Acetate. The dried grid was loaded into a JEOL JEM-1230 electron microscope operated at an acceleration voltage of 80 KV. Images were recorded at a calibrated magnification of ×60000, yielding the pixel size of 4.0Å at specimen level. 13,773 particles were manually picked from 592 images and 2D averaging and classification were performed by using Relion 3.0(2).

Purification of the Nse5-6 complex and the Nse1-3-4 complex. For cryo-EM studies, Nse5 and Nse6 were expressed from pRSF-Duet-1, in which Nse6 was fused with an His6-SUMO tag. The Nse5-Nse6 complex was expressed in BL21 (DE3) cells after induction for 18 hr at 16°C. Cells were lysed in lysis buffer (25 mM Tris-HCl, pH 8.0, 200 mM NaCl, 2 mM β-mercaptoethanol) supplemented with 1 mM phenylmethylsulfonyl fluoride and homogenized using EmulsiFlex-C3 (Avestin). After removing debris by centrifugation at 22,000 rpm for 1 h, supernatant was loaded onto HisTrap Fast flow column and washed with lysis buffer supplemented with 40 mM imidazole. The Nse5-6 complex was eluted with lysis buffer supplemented with 400 mM imidazole. Peak fractions were concentrated and purified over a Superdex 200 Increase column equilibrated with 25 mM Tris-HCl, pH 7.5, 200 mM NaCl, 2 mM β-mercaptoethanol, 0.002% w/v n-Dodecyl-β-D-maltoside. Peak fractions were collected and snapped frozen for further analyses.

For DNA binding assay, Nse5 expressed from pRSF-Duet-1 was fused with N-terminal His6-SUMO tag followed by a Ulp1 cleavage site, while Nse6 was expressed from pGEX-6P-1 in which it was fused with an N-terminal GST tag. Nse5 and Nse6 were co-expressed and purified as described above, except that eluted proteins from the HisTrap Fast flow column were incubated with the Ulp1 protease to cleave off the His-SUMO tag during dialysis at 4°C overnight against the lysis buffer. The His-SUMO tag was removed using an HisTrap FF column. The flow-through was loaded on GstTrap Fast flow column and target proteins were eluted using the lysis buffer supplemented with 15 mM reduced glutathione. The peak fractions were concentrated and purified using a Superdex 200 Increase column equilibrated in the lysis buffer. Peak fractions were collected and snapped frozen for further analyses.

Nse1, Nse3, and Nse4 were expressed from pETDuet-1, pET-28a, and pGEX-6P-1, respectively. Nse3 was fused with an N-terminal His6 tag following a TEV protease site, and Nse4 was fused with an N-terminal GST tag. The expression and purified procedures were the same to that described above for the Nse5-6 complex. In brief, expression plasmids were co-transformed into BL21 (DE3) cells and and co-expressed sample was induced for 18 hours at 16°C. The complex was first purified by an HisTrap FF column, then dialysis at 4°C overnight against the lysis buffer in the presence of the TEV protease. The cleaved His6 tag was removed via applying the dialyzed protein solution onto a HisTrap FF column. The flow-through was further purified using GstTrap Fast flow column and then Superdex 200 Increase column as described above. All expression vectors were purchased from Novagen and columns from GE Healthcare.

***In vitro* Nse5 and Nse6 interaction assays.** Wild-type and mutant His-Nse5 and GST-Nse6 proteins were co-expressed as described above. Cell lysate was incubated with 1 ml Nickel-Chelating resin (Thermo Fisher) and washed by using 15 ml lysis buffer supplemented with 40 mM imidazole. His-Nse5 and associated GST-Nse6 proteins were eluted using the lysis buffer supplemented with 400 mM imidazole. Elution was further incubated with 0.5 ml Glutathione Agarose (Thermo Scientific) at 4°C for 2 hours. The glutathione beads were washed by 10 ml lysis buffer and eluted by 1 ml lysis buffer with 15 mM reduced glutathione before examining by SDS-PAGE. For wild-type and examined mutants, equal amounts of cell lysates were used in pull down tests. Nse5 and Nse6 levels in the extracts varied among constructs, but this does not explain the amount of recovered Nse5-Nse6 complex as explained in the Figure S3 legend.

Cryo-EM analyses of the Nse5-6 complex. Purified Nse5-6 complex (0.3 mg/mL) was applied onto glow-discharged UltrAuFoil 300 mesh R1.2/1.3 grids (Quantifoil). Grids were blotted for 1.5 s at 4 °C, 100% humidity, and flash frozen in liquid ethane using a FEI Vitrobot Mark IV. All images were collected on a FEI Titan Krios electron microscope operated at an acceleration voltage of 300 kV with a Gatan K3 camera with a 1.064 Å pixel size. Movies were recorded in super-resolution mode at an electron dose rate of 20 e^{-}/pixel/s with a total exposure time of 3 s, for an accumulated electron dose of 53 $e^{-}/Å^{2}$. Intermediate frames were recorded every 0.075 s for a total number of 40 frames. Motion correction was performed with MotionCor2 (3), and contrast transfer function parameters were estimated by Ctffind4 (4). All other steps of image processing were performed by RELION 3.0 (2). After Laplacian-of-Gaussian auto-picking from 1,800 images and multi-rounds of 2D classification, a total of 657,200 particles were selected for 3D classification. Initial model was generated using 3D INITIAL MODEL from Relion 3.0 (2). Particles corresponding to the best class with the highest-resolution features were selected and subjected to the second round of 3D classification. One of 3D classes with good secondary structural features and the corresponding 18,8986 particles were polished using RELION, yielding an electron microscopy map with a resolution of 3.2 Å after 3D auto-refinement and postprocessing. All reported map resolutions are from gold-standard refinement procedures with the Fourier shell correlation cutoff being 0.143 criterion after post-processing by applying a soft mask. De novo model-building of Nse5-6 atomic structure was performed manually based on the cryo-EM density map using COOT4 (5). The model was then refined against the cryo-EM density map using phenix.real space_refine by applying geometric and secondary structure restraints (6). All figures were prepared by PyMol (https://www.pymol.org) or UCSF Chimera (7). More details of data collection, image processing and model building are shown in Table S1. Sequence alignment was generated by PSI-BLAST, allowing maximal 95% identity between sequences and minimal 25% identity for homologs using the ConSurf Server (8). Conservation surface figures were presented from PyMol (https://www.pymol.org), and individual residues were colored by using the consurf_new.py script.

Electrophoretic Mobility shift assay (EMSA). ssDNA oligo 5’ -TGTCGCATAGTGTAGTC GGTCTTGTTCGGTCATAGCTCATCGTGG -3’ was used for DNA binding assay. Annealing this oligo with the one with reverse sequence produced dsDNA for *in vitro* assays. ssDNA was 5’ labeled with fluorescein. Protein-DNA binding reactions contained 5 pmol dsDNA or 1 pmol ssDNA with various concentrations of proteins. Reaction mixtures were incubated for 2 hours on ice before resolved on DNA Retardation Gel (Invitrogen) in 0.5X TBE buffer. dsDNA was stained by SYBR stain (Invitrogen) for 1 hour before scanning gel using Bio-Rad Gel image system. ssDNA binding assay gel was scanned using the Typhoon FLA-9500 imager (GE healthcare).

Cross-linking mass spectrometry. Purified Smc5/6 complex was cross-linked with DSSO (Thermo) (9) or CDI (Sigma Aldrich) (10) for 60 min at room temperature, followed by quenching with 20 mM Tris. Cross-linked proteins were purified by filter-aided sample preparation (11) and diluted 5-fold in urea buffer (8 M urea, 50 mM ammonium bicarbonate, pH 8.0) before applying to the filter unit, as previously described. (12) After centrifugation, the filter membrane was washed twice with 200 µl urea buffer then incubated with 200 µl of 100 mM dithiothreitol in urea buffer for 20 min at room temperature before centrifugation at 14,000 g for

20 min. Proteins were then incubated with 200 µl of 100 mM iodoacetamide in urea buffer for 20 min in the dark before centrifugation at 14,000 g for 20 min. The membrane was washed with urea buffer and 50 mM ammonium bicarbonate, pH 8. Proteins retained on the membrane were then digested with LysC at 37°C for 4 hours followed by trypsin at 37°C for 18 hours. Digested peptides were collected and purified using C18 Macrospin columns (Nest Group). The purified peptides were concentrated by vacuum centrifugation and stored at -20°C before analysis. For analysis by LC-MS, peptides were resuspended in 0.1% formic acid in water (v/v) and 0.6 µg of peptide was injected for analysis. Cross-linked peptides were separated by reverse phase nanoflow liquid chromatography (EKspert nanoLC 425, Ekisgent) coupled to the Orbitrap Fusion mass spectrometer (Thermo) as previously described (12), with a 120 min gradient from 2 to 30% acetonitrile over 105 min and from 30 to 50% acetonitrile over 15min.

For mass spectrometry, a CID-MS2/HCD-MS2 method was used (12). Precursor ion spectra were recorded at 400-1800 m/z with 60,000 m/z Orbitrap resolution, automatic gain control target of 1 $x10^5$ ions and maximum injection time of 50 ms. Precursor ions with 3-10 positive charge were selected for MS2 fragmentation with dynamic exclusion of 60 secs after 1 scan and isolation window of 2 Th. For CID-MS2, precursor ion spectra were recorded in the Orbitrap with resolution 30,000 m/z, automatic gain control target of 5.0 $x10^4$ and maximum injection time of 100 ms at normalized collision energy of 30%. For HCD-MS2, precursor ion spectra were recorded in the Orbitrap with resolution 30,000 m/z, automatic gain control target of 5.0 $x10^4$ and maximum injection time of 120 ms at normalized collision energy of 30%.

Mass spectrometry data analysis. MS data were analyzed using XlinkX (13, 14) against a database containing the sequences of Smc5/6 complex subunits. Up to 4 missed cleavages were allowed, with cross-link modifications to K, S, T, Y and fixed modifications: carbamidylation (C, +57.0214 Da) and variable modifications: oxidation (M, +15.9949 Da), and deamidation (N,Q, +0.9840 Da). For DSSO treated samples, following variable modifications were set: DSSO (+158.0037 Da), DSSO Tris (+279.0776 Da), DSSO hydrolyzed (+176.0143 Da). For CDI treated samples, following variable modifications were set: CDI (+25.9792 Da), CDI Tris (+147.0531 Da). Precursor mass tolerance was set at 10 ppm with fragment mass tolerance of 20 ppm. Percolator FDR rate was set to 1%. Matches with XlinkX scores less than 30 and 40 were removed for DSSO and CDI treated samples, respectively. Remaining spectra were manually inspected to remove ambiguous assignments, which lack reporter ions for either peptide A or peptide B, or if there was insufficient b- or y- ions to specifically assign peptide backbone sequence.

Circular plots were generated using xiVIEW (15). Atomic resolution structures were visualized with UCSF Chimera (7) and cross-links were mapped using Xlink analyzer (16). Graphs and kernel probability density functions were plotted using Origin 2019 (Microcal). Mass spectrometry raw files and search results are publicly available through the ProteomeXchange data repository via the PRIDE (17) database with the data set identifier PXD023164.

For cross-link midpoint analysis, the Smc5 and Smc6 sequence was transformed into a normalized amino acid coordinate system in a similar fashion as reported by Burmann et al (18). The N-terminal and C-terminal head regions were converted to a normalized coordinate of 1, while the hinge region was converted to a normalized coordinate of 252 for Smc5 and Smc6. The N-terminal coiled-coil and C-terminal coiled-coil amino acids were then mapped using a linear interpolation function between the coordinate of the head and hinge region (1-252 for Smc5 and Smc6). The cross-linking distance based on the normalized coordinate was calculated and the probability based on kernel distribution of the distance was plotted. Cross-link distance for Smc1 and Smc3 were obtained from Burmann et al. (18)

Smc5 and Smc6 protein sequences from six different organisms were aligned using Probcons algorithm (19, 20) in Jalview (21) with default settings. These organisms include *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Homo sapiens*, *Mus musculus*, *Xenopus laevis*, and *Caenorhabditis elegans*. The conservation score per residue was calculated using the Shenkin algorithm (22) on default settings. The binned conservation score (average score of 10 amino acid bins) was plotted over the sequence for Smc5 and Smc6 *S. cerevisiae* strain.

Integrative structure modeling of Smc5/6-Nse2/5/6 complex. Integrative structure modeling proceeded through the standard four-stage protocol (23–28): (i) gathering data, (ii) representing constituent subunits and translating the data into spatial restraints, (iii) structural sampling to produce an ensemble of alternate structural models that satisfy the restraints, and (iv) analyzing and validating structural ensemble and data (Figure S4a).

Gathering data: Comparative models of Smc5 hinge (485-633 a. a.) and Smc6 hinge (506-692 a. a.) were built automatically using the SWISS-MODEL webserver (29) based on the structure of the corresponding regions of the fission yeast Smc5 and 6 (PDB ID: 5mg8)(30). Comparative models of Smc5 head (42-204, 950-1093 a. a.) and Smc6 head (80-232, 988-1114 a. a.) regions were built using MODELLER (31), based on two different structures of the budding yeast cohesin head regions (PDB IDs: 6qpw and 6zz6) (32, 33). Backbones for Smc5 and 6 coiled-coils were designed through optimization of Crick parameters (34), by minimizing steric strain, using the biomolecular-design software package ISAMBARD (35). We also used the structures of Nse2 bound to the middle part of Smc5 arm (PDB: 3HTK) (36) and Nse5-6 determined by single particle cryo-electron microscopy in this study.

Subunit representation and translating data into spatial restraints: To achieve sufficient computational efficiency of structural sampling, we used a coarse-grained one-residue-per-bead (rpb) representation of the complex. Regions absent from crystal structures, comparative models, and designed coil-coils were represented as flexible strings of beads at resolutions of 10-rpb for Smc5/6 and Nse2 and 20-rpb for Nse5/6. Regions lacking structural and crosslink data were represented at higher coarseness: 50-rpb for Smc5:1-41 a. a., 80 rpb for Smc6:1-79 a. a., 30-rpb for Nse5:1-21 a. a. and 120-rpb for Nse5:442-556 a. a.. Non-coiled-coil regions with X-ray structures or comparative models were treated as rigid bodies. Several regions of the Smc5/6 coiled-coil were treated as rigid bodies and included Smc5: 208-266, 272-303, 365-388, 398-459, 653-714, 715-738, 851-880, 885-946 a. a., and Smc6: 236-294, 298-310, 311-430, 436-501, 699-760, 770-882, 897-916, 923-984 a. a.. CL-MS data were used to construct Bayesian distance restraints that restricted the distance between cross-linked residues (30 Å for DSSO and 20 Å for CDI) (37, 38). Excluded volume restraints were applied to prevent steric clashes (24, 39), while sequence connectivity restraints were enforced through a harmonic upper bound on the distance between consecutive beads with threshold distance equal to four times the sum of the van der Waals radii of connected beads (24, 38, 39)

Structural sampling of alternate models to produce an ensemble of structures that satisfy the restraints: Sampling started with randomized initial positions and orientations of rigid bodies and positions of the flexible strings of beads. 8,000,000 alternate models were generated from 100 independent runs of replica exchange Gibbs sampling based on the Metropolis Monte Carlo algorithm (38, 40), where each Monte Carlo step consisted of a series of random rotations and translations of flexible beads and rigid bodies.

Analysis and validation of the structural ensemble and data: Model validation followed four steps (28, 41, 42). First, 8,000,000 models were filtered to select 2,920,264 good scoring models. Second, these good scoring models were divided into two independent samples and clustered according to bead RMSD, and the sampling precision was estimated through a statistical hypothesis testing pipeline detailed in references (28, 42). Briefly, sampling precision is defined as the largest allowed RMSD between the cluster centroid and model within any cluster in the finest clustering, for which each sample contributes models proportionally to its size (considering both significance and magnitude of the difference), and for which a sufficient proportion of all models occur in sufficiently large clusters (42). The sampling precision for our integrative modeling of the pentameric Smc5/6 complex is 3.3 nm. Third, model precision was evaluated by re-clustering the good scoring models at a threshold of 5 nm bead RMSD. Out of the seven clusters that were produced, only the most populated cluster containing 88.7% of all the good scoring models was retained for further analysis and the remaining six clusters were ignored. The precision of the model was defined as the average bead RMSD of all models in the top cluster, from the model representing the centroid of that cluster (42). The model precision for our Smc5/6-Nse2/5/6 model is 3.8 nm, and the entire structural ensemble was depicted as 3D localization probability densities of the head, hinge, arm regions of Smc5/6, Nse2 and the Nse5/6 subcomplex, around the centroid model (Figure 4A left panel). Fourth, structure models were tested for satisfaction of input information, i.e. our CL-MS data. A crosslink is considered satisfied if the corresponding Cα-Cα distance in any of the models in the ensemble is less than 30 Å for DSSO (43) or 20 Å for CDI crosslinkers (10). The ensemble satisfied an overall 97.6% of all the crosslinks. Cross-link violations observed mostly in Smc6 and were small (<3%), and can be rationalized as the net effect of factors like experimental uncertainty, local structural fluctuations, coarse-grained representation of the model and/or finite structural sampling.The structural ensemble was also validated using data not directly used in modeling, namely the Smc5/6 arm length. This length was calculated to be 31.9 nm based the negative stain EM images (Figure 1D), and was largely in agreement with the value of 27 ± 3 nm computed from the structural ensemble as measured by the distance between Cα atoms of Smc5:523 a. a. and Smc6:886 a. a.

Detection of *in vivo* protein sumoylation. In most cases, total sumoylated proteins were pulled down before immuno-blotting to examine specific substrates (44). In brief, yeast SUMO (Smt3) was tagged with His8 tag at its N-terminus and expressed from its endogenous promoter (45). Cells were grown in YPD to reach log phase and treated with 0.03% MMS for 2 hours before harvest. Cells extracts prepared by 55% TCA precipitation were dissolved in Buffer A (6 M Guanidine HCl, 100 mM sodium phosphate pH 8.0, 10 mM Tris-HCl pH 8.0) and incubated overnight with Ni-NTA resin with additional 0.05% Tween 20 and 4.4 nM imidazole. Resins were washed twice with Buffer A containing 0.05% Tween 20 and four times with Buffer C (8 M urea, 100 mM sodium phosphate pH 6.3, 10 mM Tris-HCl pH 6.3) containing 0.05% Tween 20. HU buffer (8 M urea, 200 mM Tris-HCl pH 6.8, 1 mM EDTA, 5% SDS, 0.1% bromophenol blue, 1.5% DTT, 200 mM imidazole) was used to elute proteins. Ponceau S stain was used to assess equal loading. For Pol2, cells were treated as above and harvested. Proteins were extracted by the RIPA buffer (50 mM Tris-HCl pH 7.4, 5 mM EDTA, 150 mM NaCl, 1.25% Triton-X 100, 1x Sigma protease inhibitor cocktail, 40 mM NEM). Immuno-precipitation was carried out by incubating with anti-HA antibody conjugated on beads at 4 °C for 2-6 hours. Beads were washed with the RIPA buffer supplemented with 0.1% SDS and eluted with the protein loading buffer. Unmodified form of Pol2 was used for assessing equal loading.

Supplemental Figure Legends

Fig. S1. CL-MS data of the Smc5/6 complex and Smc5 and 6 sequence conservation. (*A*) Schematic showing typical SMC protein domain and fold. (*B*) Protein sequence coverage of the subunits of the Smc5/6 holo-complex based on MS analyses. Bars are labeled with average sequence coverage (%) from 3 replicates cross-linked by DSSO and 3 replicates cross-linked by CDI. Error bars indicate standard deviation across all 6 replicates. (*C*) A representative gel picture for examining the crosslinked ATPγS-bound Smc5/6 holo-complex. (*D*) Circular plot showing crosslinks for ATPγS-bound Smc5/6 holo-complex by both DSSO and CDI crosslinker. Intra-protein cross-links are colored purple. Inter-protein cross-links are colored green. (*E*) Mapping DSSO and CDI crosslinks to the Nse2-Smc5 structure [PDB ID: 3HTK, (36)] and histogram depicting Cα-Cα distances for mapped crosslinks. (*F*) Plot of conservation score for Smc5 and Smc6 sequence with bin size of 10 amino acid. Horizontal dotted line represents average conservation score across the whole protein sequence.

Fig. S2. Cryo-EM reconstruction of the Nse5-6 complex. (*A*) Size exclusion chromatogram (SEC) of Nse5-Nse6 complex. Peak fractions were labeled as line segments on the top. (*B*) Analysis of Nse5-Nse6 complex peak fractions by SDS-PAGE. A representative Coomassie staining picture is shown. Band corresponding to each subunit is indicated. (*C*) Workflow of cryo-EM image processing for the Nse5-Nse6 complex. Examples of cryo-EM 2D classification results of the Nse5-Nse6 complex are shown on the top. (*D*) Global Fourier Shell Correlation (FSC) curve of the Nse5-Nse6 complex. The curve for the two half datasets is in blue and that for the refined model versus the cryo-EM map is in red. The overall cryo-EM map resolution is 3.2 Å with FSC set at 0.143. (*E*) Angular distribution plot of final 3D EM map for the Nse5-Nse6 complex. (*F*) Final 3D reconstructed map of the Nse5-Nse6 complex colored according to local resolution. The resolution of the majority of the map is 3 Å, with relatively poor density seen at the Nse5 C-terminal domain.

Fig. S3. Functional and structural features of the Nse5-6 complex. (*A*) Analysis of purified Nse5-6 and Nse1-3-4 complexes by SDS-PAGE. Representative Coomassie stained gel pictures are shown. Band corresponding to each subunit is indicated. (*B*) *In vitro* DNA binding assay results. Fluorescein labeled single stranded (ss) and double stranded (ds) DNA were mixed with increasing levels of proteins at protein:DNA molar ratios from 1:1 to 1:8 as indicated. Reactions mixtures were analyzed on gel. Representative images of scanned fluorescent signals are shown. (*C*) Examination of Nse5 and Nse6 proteins in the crude extract of bacterial cells induced for their expression. The presence of Nse5 and Nse6 proteins was detected by immunoblotting using anti-GST and anti-His antibodies. Labeling is as Figure 3C and 3D. Most mutant proteins were well expressed, except when the Nse5 and Nse6 mutants affecting the interface region I were combined. Each region I mutant alone showed moderate expression levels, yet disrupted the

Nse5-6 complex formation (Fig. 3C). Note that moderate protein levels can fully support the Nse5-6 complex formation as shown by the Nse6 region III mutant. (*D*) Surface presentation of the Nse5 and Nse6 structures showing residues supporting their interaction at the interface region I-III and their conservation scores. (*E*) Sequence alignment of Nse5 and Nse6 in their interaction region I, II and III. Four *Saccharomyces* species were examined. *Scer: Saccharomyces cerevisiae*, *Scas*: *Saccharomyces castellii*, *Sbay*: *Saccharomyces bayanus*, *Spar*: *Saccharomyces paradoxus*. Residues involved in Nse5 and Nse6 interaction are colored red. Identical and similar amino acids are indicated by black and grey circles, respectively, while non-conserved amino acids are indicated by open circles below the sequences. (*F*) Mapped DSSO and CDI crosslinks in the structure of the Nse5-Nse6 complex, and histogram depicting Cα-Cα distances for the mapped crosslinks.

Fig. S4. Integrative model of the Smc5/6-Nse2/5/6 complex. (*A*) Four-stage scheme of integrative modeling, of the Smc5/6-Nse2/5/6 complex. (*B*) Distribution of Cα-Cα distances of crosslinked residues for DSS (left subplot) and CDI (right subplot) with the threshold distances of 30 Å (for DSSO) and 20 Å (CDI) marked with dotted vertical lines. The overall crosslink satisfaction is 97.6%.

Fig. S5. Examination of Nse5's SUMO binding motifs. (*A*) Nse5 sites that match SIM consensus sequences. These sequences were assessed for their accessibility for SUMO binding and locations relative to the Nse5-Nse6 binding interface. Mutations made for testing SUMO binding are indicated in red. (*B*) Summary of yeast two-hybrid results of Nse5 mutations in effecting SUMO binding. **(***C***)** Surface presentation of the Nse5 structure showing conservation scores. (*D*) The effects of *nse5-sim6,9* mutation on sumoylation of non-Smc5/6 substrates. Nse2 substrates, including the Holliday junction dissolution complex, Sgs1-Top3-Rmi1, and the DNA polymerase Pol2, were maintained. Sgs1-Top3-Rmi1 sumoylation was examined as in Figure 4D. HA-tagged Pol2 was immunoprecipitated and its sumoylated form was detected using anti-SUMO antibody in immunoblotting, while unmodified Pol2 was detected using the anti-HA antibody.

Supplemental References

1. N. Dhingra, L. Wei, X. Zhao, Replication protein A (RPA) sumoylation positively influences the DNA damage checkpoint response in yeast. *J. Biol. Chem.* **294**, 2690–2699 (2019).
2. S. H. W. Scheres, RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
3. S. Q. Zheng, *et al.*, MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
4. A. Rohou, N. Grigorieff, CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
5. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
6. P. D. Adams, *et al.*, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *International Tables for Crystallography*, 539–547 (2012).
7. E. F. Pettersen, *et al.*, UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
8. H. Ashkenazy, *et al.*, ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44**, W344–50 (2016).
9. A. Kao, *et al.*, Development of a novel cross-linking strategy for fast and accurate identification of cross-linked peptides of protein complexes. *Mol. Cell. Proteomics* **10**, M110.002212 (2011).
10. C. Hage, C. Iacobucci, A. Rehkamp, C. Arlt, A. Sinz, The First Zero-Length Mass Spectrometry-Cleavable Cross-Linker for Protein Structure Analysis. *Angew. Chem. Int. Ed Engl.* **56**, 14551–14555 (2017).
11. J. R. Wiśniewski, A. Zougman, N. Nagaraj, M. Mann, Universal sample preparation method for proteome analysis. *Nature Methods* **6**, 359–362 (2009).
12. Z. Ser, P. Cifani, A. Kentsis, Optimized Cross-Linking Mass Spectrometry for in Situ Interaction Proteomics. *J. Proteome Res.* **18**, 2545–2558 (2019).
13. F. Liu, D. T. S. Rijkers, H. Post, A. J. R. Heck, Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry. *Nat. Methods* **12**, 1179–1184 (2015).
14. F. Liu, P. Lössl, R. Scheltema, R. Viner, A. J. R. Heck, Optimized fragmentation schemes and data analysis strategies for proteome-wide cross-link identification. *Nat. Commun.* **8** 15473 (2017).

15. M. Graham, C. Combe, L. Kolbowski, J. Rappsilber, xiView: a common platform for the downstream analysis of crosslinking mass spectrometry data. *bioRxiv* 561829; doi: https://doi.org/10.1101/561829 .

16. J. Kosinski, *et al.*, Xlink Analyzer: software for analysis and visualization of cross-linking data in the context of three-dimensional structures. *J. Struct. Biol.* **189**, 177–183 (2015).

17. Y. Perez-Riverol, *et al.*, The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **47**, D442–D450 (2019).

18. F. Bürmann, *et al.*, A folded conformation of MukBEF and cohesin. *Nat. Struct. Mol. Biol.* **26**, 227–236 (2019).

19. C. B. Do, ProbCons: Probabilistic consistency-based multiple sequence alignment. *Genome Research* **15**, 330–340 (2005).

20. U. Roshan, Multiple sequence alignment using Probcons and Probalign. *Methods Mol. Biol.* **1079**, 147–153 (2014).

21. A. M. Waterhouse, J. B. Procter, D. M. A. Martin, M. Clamp, G. J. Barton, Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).

22. P. S. Shenkin, B. Erman, L. D. Mastrandrea, Information-theoretical entropy as a measure of sequence variability. *Proteins* **11**, 297–313 (1991).

23. D. Russel, *et al.*, Putting the pieces together: integrative modeling platform software for structure determination of macromolecular assemblies. *PLoS Biol.* **10**, e1001244 (2012).

24. F. Alber, *et al.*, Determining the architectures of macromolecular assemblies. *Nature* **450**, 683–694 (2007).

25. A. Sali, *et al.*, Outcome of the first wwPDB hybrid/integrative methods task force workshop. *Structure* **23**, 1156–1167 (2015).

26. S. J. Kim, *et al.*, Structure and functional anatomy of the nuclear pore complex. *Biophys. J.* **114**, 372a (2018).

27. D. Schneidman-Duhovny, R. Pellarin, A. Sali, Uncertainty in integrative structural modeling. *Curr. Opin. Struct. Biol.* **28**, 96–104 (2014).

28. D. J. Saltzberg, *et al.*, Using *Integrative Modeling Platform* to compute, validate, and archive a model of a protein complex structure. *Protein Sci.* **30**, 210-261 (2021).

29. A. Waterhouse, *et al.*, SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).

30. A. Alt, *et al.*, Specialized interfaces of Smc5/6 control hinge stability and DNA association.

Nat. Commun. **8**, 14011 (2017).

31. A. Sali, T. L. Blundell, Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815 (1993).

32. K. W. Muir, Y. Li, F. Weis, D. Panne, The structure of the cohesin ATPase elucidates the mechanism of SMC-kleisin ring opening. *Nat. Struct. Mol. Biol.* **27**, 233–239 (2020).

33. J. E. Collier, *et al.*, Transport of DNA within cohesin involves clamping on top of engaged heads by Scc2 and entrapment within the ring by Scc3. *Elife* **9** e59560 (2020).

34. G. Grigoryan, W. F. DeGrado, Probing designability via a generalized model of helical bundle geometry. *J. Mol. Biol.* **405**, 1079–1100 (2011).

35. C. W. Wood, *et al.*, ISAMBARD: an open-source computational environment for biomolecular analysis, modelling and design. *Bioinformatics* **33**, 3043–3050 (2017).

36. X. Duan, *et al.*, Structural and functional insights into the roles of the Mms21 subunit of the Smc5/6 complex. *Mol. Cell* **35**, 657–668 (2009).

37. J. P. Erzberger, *et al.*, Molecular Architecture of the 40S□eIF1□eIF3 Translation Initiation Complex. *Cell* **158**, 1123–1135 (2014).

38. Y. Shi, *et al.*, Structural characterization by cross-linking reveals the detailed architecture of a coatomer-related heptameric module from the nuclear pore complex. *Mol. Cell. Proteomics* **13**, 2927–2943 (2014).

39. M.-Y. Shen, A. Sali, Statistical potential for assessment and prediction of protein structures. *Protein Sci.* **15**, 2507–2524 (2006).

40. R. H. Swendsen, J. S. Wang, Replica Monte Carlo simulation of spin glasses. *Phys. Rev. Lett.* **57**, 2607–2609 (1986).

41. M. P. Rout, A. Sali, Principles for integrative structural biology studies. *Cell* **177**, 1384–1403 (2019).

42. S. Viswanath, I. E. Chemmama, P. Cimermancic, A. Sali, Assessing Exhaustiveness of Stochastic Sampling for Integrative Modeling of Macromolecular Structures. *Biophys. J.* **113**, 2344–2353 (2017).

43. E. D. Merkley, *et al.*, Distance restraints from crosslinking mass spectrometry: mining a molecular dynamics simulation database to evaluate lysine-lysine distances. *Protein Sci.* **23**, 747–759 (2014).

44. H. D. Ulrich, A. A. Davies, In vivo detection and characterization of sumoylation targets in Saccharomyces cerevisiae. *Methods Mol. Biol.* **497**, 81–103 (2009).

45. L. Wei, X. Zhao, A new MCM modification cycle regulates DNA replication initiation. *Nat.*

Struct. Mol. Biol. **23**, 209–216 (2016).

Dataset S1. CL-MS data for the Smc5/6 holo-complex
Dataset S2. Coordinate distance for Smc5 and Smc6 crosslinks
Dataset S3. Conservation scores for Smc5 and Smc6

Table S1. Cryo-EM data collection, processing, and validation statistics

Sample	**Nse5-Nse6 complex (EMID–23517, PDB 7LTO)**
Data collection	
Microscope	Titan Krios
Detector	Gatan K3
Automation software	SerialEM
Nominal magnification	22,500
Calibrated magnification	47,262
Voltage (kV)	300 kV
Total dose ($e^-/Å^2$)	53
Dose rate (e^-/pixel/s)	20
Number of frames collected	40
Defocus range (µm)	-1.0 to -2.5
Pixel size (Å)	1.064
Collected Micrographs	3,600
Selected Micrographs	1,800
Reconstruction	
Initially autopicked particles	1,787,057
Particles used for classification	657,200
Particles in the final map	188,916
Symmetry	C1
Resolution	
FSC 0.5 (unmasked/masked, Å)	3.4/3.3
FSC 0.143 (unmasked/masked, Å)	3.2/3.2
Map sharpening B factor ($Å^2$)	-95.45
Model composition	
Protein residues	605
Nonhydrogen atoms	4785
Validation	
MolProbity	2.64
Clash score	9.05
Map Correlation Coefficient	0.78
R.m.s. deviations	
Bond lengths (Å)	0.004
Bond angles (°)	0.656
Ramachandran plots	
Favored (%)	91.08
Allowed (%)	8.92
Outliers (%)	0
Rotamer outliers (%)	1.0

Table S2. Yeast strains used in this study

Name	**Genotype**
T2146-1	*ADE2 pep4::kanMX NSE5-PGAL1,10-NSE6-3XFLAG::LEU2 MMS21-PGAL1,10-SMC5-CBP::TRP1 Nse3-PGAL1,10-NSE4::URA3 Nse1-PGAL1,10-SMC6::HIS3*
X8459-3B	*SMC5-TAP::TRP1 8His-SMT3::TRP1*
X8459-3D	*SMC5-TAP::TRP1 8His-SMT3::TRP1 nse5-T141,L142,V433,V434,I435,L438A -AID-3FLAG::HIS3*
X8457-4D	*SMC6-TAP::URA3 8His-SMT3::TRP1*
X8457-8B	*SMC6-TAP::URA3 8His-SMT3::TRP1 nse5-T141,L142,V433,V434,I435,L438A -AID-3FLAG::HIS3*
X8458-7B	*NSE4-13myc::HIS3 8His-SMT3::TRP1*
X8458-14B	*NSE4-13myc::HIS3 8His-SMT3::TRP1 nse5-T141,L142,V433,V434,I435,L438A -AID-3FLAG::HIS3*
X8460-1B	*POL2-3HA::KAN*
X8460-1C	*POL2-3HA::KAN nse5- T141,L142,V433,V434,I435,L438A -AID-3FLAG::HIS3*
X6684-2B	*SGS1-9myc::KAN 8His-SMT3::TRP1*
X8453-3D	*SGS1-9myc::KAN 8His-SMT3::TRP1 nse5-T141,L142,V433,V434,I435,L438A -AID-3FLAG::HIS3*
X6185-2B	*TOP3-TAP::HIS3 8His-SMT3::TRP1*
X8456-4B	*TOP3-TAP::HIS3 8His-SMT3::TRP1 nse5-T141,L142,V433,V434,I435,L438A -AID-3FLAG::HIS3*
X6160-6A	*RMI1-TAP::HIS3 8His-SMT3::TRP1*
X8454-5D	*RMI1-TAP::HIS3 8His-SMT3::TRP1 nse5-T141,L142,V433,V434,I435,L438A-AID-3FLAG::HIS3*
X5134-13C	*Nse5-AID-3FLAG::HIS3*
T2050-1	*nse5-T141,L142,V433,V434,I435,L438A-AID-3FLAG::HIS3*
T2205-14	*nse5-Y93,R97A-AID-3FLAG::HIS3/+*
T2206-1	*nse5-Y93,R97,R102A-AID-3FLAG::HIS3/+*
X8596-1	*Smc5-TAP::TRP1/+ Nse5-AID-3FLAG::HIS3/+*

Table S3. Plasmids used in this study

Name	Vector information
	pOAD
pXZ220	pOAD-SUMO
pXZ166	pOAD-Nse6
	pOBD
pXZ199	pOBD-Nse5
pXZ631	pOBD-Nse5-sim6
pXZ654	pOBD-Nse5-sim9
p251	pOBD-Nse5-sim1
p252	pOBD-Nse5-sim2
p253	pOBD-Nse5-sim3
p254	pOBD-Nse5-sim5
p256	pOBD-Nse5-sim7
p270	pOBD-Nse5-sim10
Y43	pETDuet-1: MCS1:Nse1
Y44	pET28a:His-TEV-Nse3
Y53	pGEX-6P-1:GST-Nse4
Y080	pRSFDuet-1: MCS1: 6His-SUMO-Nse6 MCS2: Nse5
Y139	pRSFDuet-1: MCS1: 6His-SUMO-Nse5
Y140	pGEX-6P-1: GST-Nse5

Figure S1

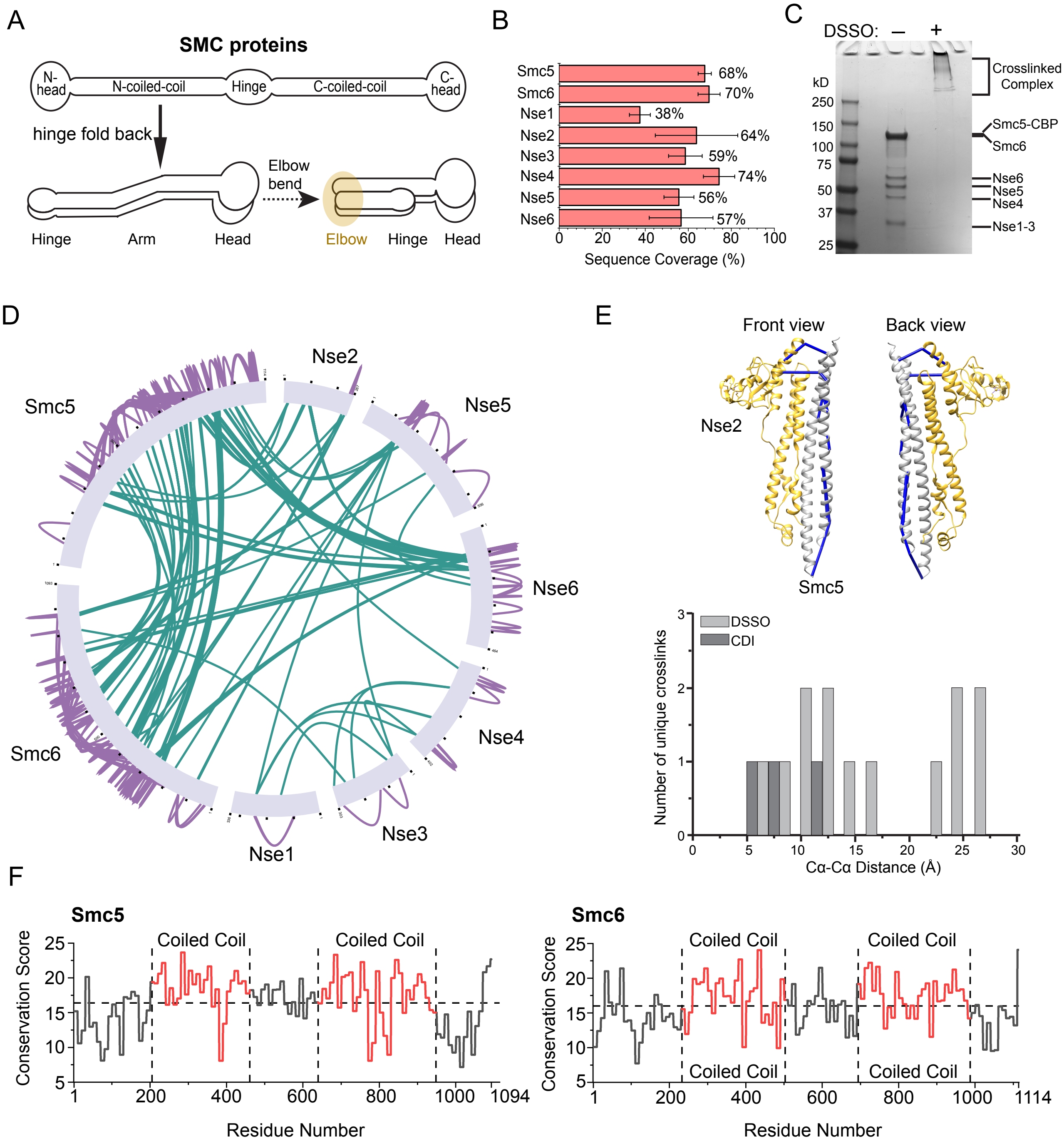

A
SMC proteins
N-head
N-coiled-coil
Hinge
C-coiled-coil
C-head
hinge fold back
Elbow bend
Hinge
Arm
Head
Elbow
Hinge
Head
B
Smc5 68%
Smc6 70%
Nse1 38%
Nse2 64%
Nse3 59%
Nse4 74%
Nse5 56%
Nse6 57%
0 20 40 60 80 100
Sequence Coverage (%)
C
DSSO: – +
kD
250
150
100
75
50
37
25
Crosslinked Complex
Smc5-CBP
Smc6
Nse6
Nse5
Nse4
Nse1-3
D
Smc5
Nse2
Nse5
Nse6
Nse4
Nse3
Nse1
Smc6
E
Front view
Back view
Nse2
Smc5
DSSO
CDI
Number of unique crosslinks
0 1 2 3
0 5 10 15 20 25 30
Cα-Cα Distance (Å)
F
Smc5
Coiled Coil
Coiled Coil
Conservation Score
5 10 15 20 25
1 200 400 600 800 1000 1094
Residue Number
Smc6
Coiled Coil
Coiled Coil
Coiled Coil
Coiled Coil
Conservation Score
5 10 15 20 25
1 200 400 600 800 1000 1114
Residue Number

Figure S2

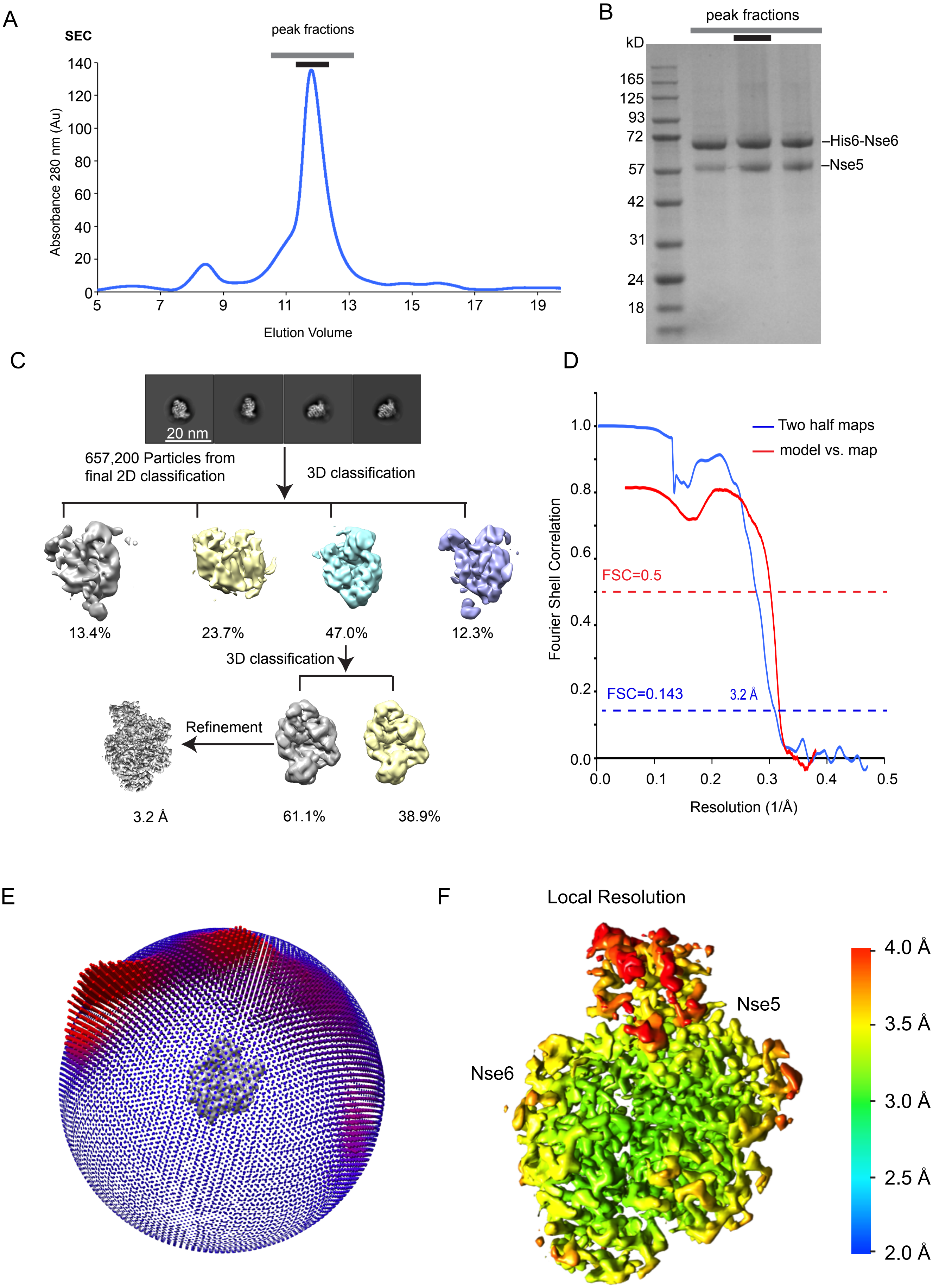

A
SEC
peak fractions
Absorbance 280 nm (Au)
140
120
100
80
60
40
20
0
5
7
9
11
13
15
17
19
Elution Volume
B
peak fractions
kD
165
125
93
72
57
42
31
24
18
–His6-Nse6
–Nse5
C
20 nm
657,200 Particles from
final 2D classification
3D classification
13.4%
23.7%
47.0%
12.3%
3D classification
Refinement
3.2 Å
61.1%
38.9%
D
Two half maps
model vs. map
Fourier Shell Correlation
1.0
0.8
0.6
0.4
0.2
0.0
FSC=0.5
FSC=0.143
3.2 Å
0.0
0.1
0.2
0.3
0.4
0.5
Resolution (1/Å)
E
F
Local Resolution
Nse5
Nse6
4.0 Å
3.5 Å
3.0 Å
2.5 Å
2.0 Å

Figure S3

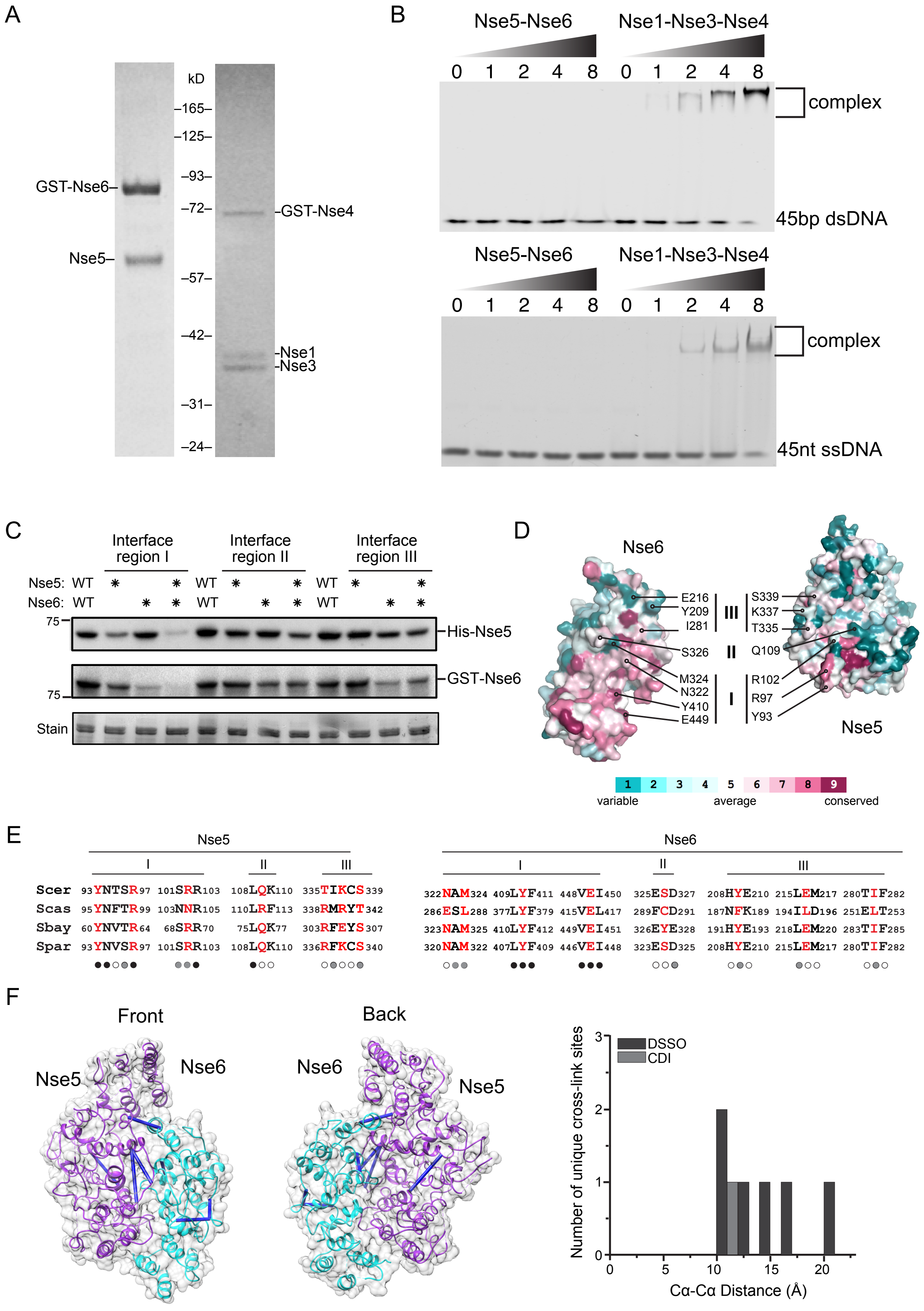
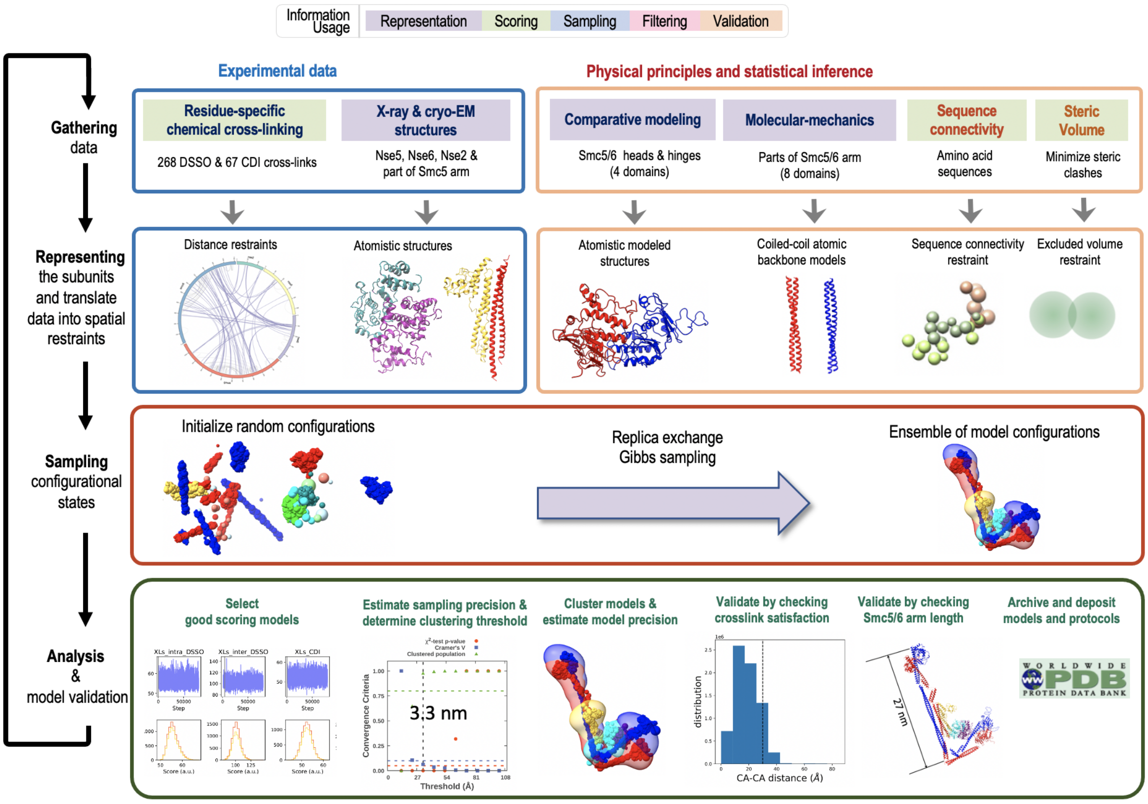

A
kD
GST-Nse6
Nse5
165
125
93
72
57
42
31
24
GST-Nse4
Nse1
Nse3
B
Nse5-Nse6
Nse1-Nse3-Nse4
0 1 2 4 8 0 1 2 4 8
complex
45bp dsDNA
Nse5-Nse6
Nse1-Nse3-Nse4
0 1 2 4 8 0 1 2 4 8
complex
45nt ssDNA
C
Interface region I
Interface region II
Interface region III
Nse5: WT * * WT * * WT * *
Nse6: WT * * WT * * WT * *
75
75
His-Nse5
GST-Nse6
Stain
D
Nse6
Nse5
E216
Y209
I281
S326
M324
N322
Y410
E449
III
II
I
S339
K337
T335
Q109
R102
R97
Y93
1 2 3 4 5 6 7 8 9
variable
average
conserved
E
Nse5
I
II
III
Scer 93YNTSR97 101SRR103 108LQK110 335TIKCS339
Scas 95YNFTR99 103NNR105 110LRF113 338RMRYT342
Sbay 60YNVTR64 68SRR70 75LQK77 303RFEYS307
Spar 93YNVSR97 101SRR103 108LQK110 336RFKCS340
●●○●● ●●● ●○○ ○●○○●
Nse6
I
II
III
322NAM324 409LYF411 448VEI450 325ESD327 208HYE210 215LEM217 280TIF282
286ESL288 377LYF379 415VEL417 289FCD291 187NFK189 194ILD196 251ELT253
323NAM325 410LYF412 449VEI451 326EYE328 191HYE193 218LEM220 283TIF285
320NAM322 407LYF409 446VEI448 323ESD325 208HYE210 215LEM217 280TIF282
○●● ●●● ●●● ○○● ○●○ ●○○ ○●○
F
Front
Back
Nse5
Nse6
Nse6
Nse5
DSSO
CDI
Number of unique cross-link sites
0
1
2
3
0
5
10
15
20
Cα-Cα Distance (Å)

Figure S4

A


Information Usage
Representation
Scoring
Sampling
Filtering
Validation
Gathering data
Representing the subunits and translate data into spatial restraints
Sampling configurational states
Analysis & model validation
Experimental data
Physical principles and statistical inference
Residue-specific chemical cross-linking
268 DSSO & 67 CDI cross-links
X-ray & cryo-EM structures
Nse5, Nse6, Nse2 & part of Smc5 arm
Comparative modeling
Smc5/6 heads & hinges (4 domains)
Molecular-mechanics
Parts of Smc5/6 arm (8 domains)
Sequence connectivity
Amino acid sequences
Steric Volume
Minimize steric clashes
Distance restraints
Atomistic structures
Atomistic modeled structures
Coiled-coil atomic backbone models
Sequence connectivity restraint
Excluded volume restraint
Initialize random configurations
Replica exchange Gibbs sampling
Ensemble of model configurations
Select good scoring models
XLs_intra_DSSO
XLs_inter_DSSO
XLs_CDI
Step
Score (a.u.)
Estimate sampling precision & determine clustering threshold
χ²-test p-value
Cramer's V
Clustered population
Convergence Criteria
Threshold (Å)
3.3 nm
Cluster models & estimate model precision
Validate by checking crosslink satisfaction
distribution
CA-CA distance (Å)
Validate by checking Smc5/6 arm length
27 nm
Archive and deposit models and protocols
WORLDWIDE PDB PROTEIN DATA BANK

B

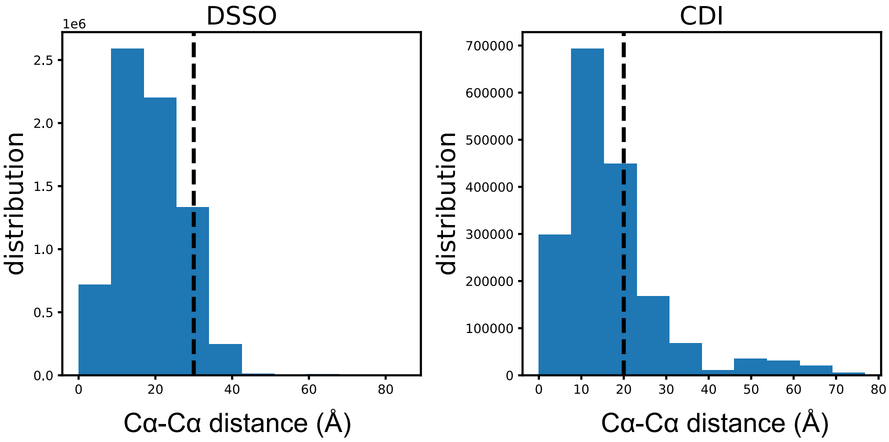

DSSO
1e6
2.5
2.0
1.5
1.0
0.5
0.0
distribution
0
20
40
60
80
Cα-Cα distance (Å)
CDI
700000
600000
500000
400000
300000
200000
100000
0
distribution
0
10
20
30
40
50
60
70
80
Cα-Cα distance (Å)

Figure S5

A

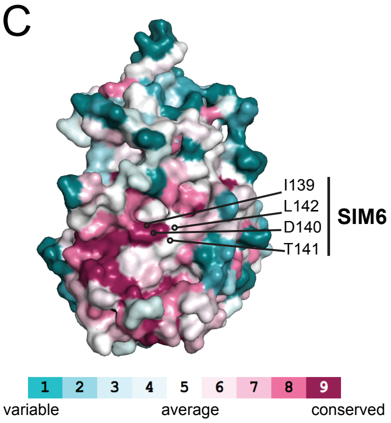
Potential Nse5 SIM sequences	Accessiblity by SUMO	Nse6 interface	Mutation
SIM1 $_{22}$VELT$_{25}$	**Yes**	**No**	AEAA
SIM2 $_{47}$VLLF$_{50}$	No	Yes	AAAF
SIM3 $_{68}$IILV$_{71}$	No	No	AAAV
SIM4 $_{112}$LAIL$_{115}$	No	Yes	NA
SIM5 $_{128}$LELL$_{131}$	No	No	AEAA
SIM6 $_{139}$IDTL$_{142}$	**Yes**	**No**	IDAA
SIM7 $_{246}$ILID$_{249}$	No	No	AAAD
SIM8 $_{414}$IPLL$_{417}$	No	No	NA
SIM9 $_{433}$VVILTLV$_{439}$	**Yes**	**No**	AAALTAV
SIM10 $_{476}$VLNI$_{479}$	NA	NA	AANA

B

	pOAD	
pOBD-Nse5	Nse6	SUMO
WT	++++	+++
SIM1m	++	–
SIM2m	+++	+
SIM3m	–	–
SIM5m	–	–
SIM6m	++++	–
SIM7m	+	–
SIM9m	++++	–
SIM10m	+++	++

++++: growth on -A and -H+3AT
+++: growth on -A and -H+3AT
++: growth on -H+3AT only
+: growth on -H only
–: no growth on dropout plate

C



D