Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

Single-molecule investigations of isolated cohesin domains

Marvin Sebastian Freitag
aus
Hildesheim, Deutschland

2023

Erklärung

Diese Dissertation wurde im Sinne von §7 der Promotionsordnung vom 28. November 2011 von Herrn Prof. Dr. Johannes Stigler betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, den 20. September 2023

_____

Marvin Sebastian Freitag

Dissertation eingereicht am: 21. September 2023

1. Gutachter: Prof. Dr. Johannes Stigler

2. Gutachter: Prof. Dr. Julian Stingele

Mündliche Prüfung am: 23. November 2023

This thesis has been prepared from December 2017 to September 2023 in the laboratory of Prof. Dr. Johannes Stigler at the Gene Center of the Ludwig-Maximilians-Universität München, Germany.

This is a cumulative thesis based on the following publications:

Freitag, M.*, D. Kamp*, M. Synakewicz, and J. Stigler. 2021. Identification and correction of miscalibration artifacts based on force noise for optical tweezers experiments. *J. Chem. Phys.* 155:175101, https://doi.org/10.1063/5.0063690

* These authors contributed equally.

Freitag, M., S. Jaklin, F. Padovani, E. Radzichevici, S. Zernia, K.M. Schmoller, and J. Stigler. 2022. Single-molecule experiments reveal the elbow as an essential folding guide in SMC coiled-coil arms. *Biophys J*. 121:4702–4713, https://doi.org/10.1016/j.bpj.2022.10.017

Contents

Abstract

The inheritance of genetic information depends on the duplication and transfer of deoxyribonucleic acid (DNA) molecules from parental cells to daughter cells (1). Therefore, DNA undergoes various transitions from different physical appearances throughout the cell cycle including highly compacted forms (2). To achieve these special requirements, protein complexes of the essential SMC (structural maintenance of chromosomes) family are utilized by prokaryotes and eukaryotes (1). Using their ability to interact with and manipulate DNA, SMC complexes are responsible for the genome organization and are involved in fundamental DNA-based phenomena like replication, DNA repair and sister chromatid cohesion (3–5).

Throughout the last decades evidence accumulated that the genome organization is based on a dynamic process called DNA loop extrusion mediated by the SMC complexes (6). It was found that eukaryotic SMCs can adopt ring-like shapes (O-shape) as well as collapsed forms (B-shape) via their elongated coiled-coil domain (7, 8). Consecutive transitions between these conformations enable the motor activity of SMCs and result in the growth of a DNA loop (9). Simultaneously acting complexes extrude the DNA fiber into a compacted and well-organized form (10). Loop lengths were found to be in the order of ten-thousands of base pairs and even more, which enable local contacts over far genomic distances (11).

In the main part of this thesis, the architecture of the SMC coiled coils is analyzed by using the isolated coiled-coil domain of Psm3 from *Schizosaccharomyces pombe (S. pombe)* as a model system (12). Using unfolding experiments the thermodynamic stability profile of the coiled coils was determined. The SMC coiled coils turned out to be a relatively unstable fold compared to other coiled coils, which could be attributed to the segmented composition. The coiled-coil segments were found to fold independently and reliably, which could be distorted by mutations *in vitro* and *in vivo*.

The unfolding experiments were performed on a commercially available optical tweezers set-up and are among the first high-resolution data obtained from such an instrument. Hence, a key aspect of this thesis involved the setup of the instrument, the reliable production of suitable samples and finally the reproducible acquisition of high-quality measurements. A bottleneck of the reproducibility is the calibration of the instrument (13). The second part of this thesis focusses on the identification and correction of miscalibration and non-linearity effects and paves the way for future experiments with high accuracy and precision.

The third part of this thesis addresses the SMC head domains, which possess the ATPase activity and create a DNA binding site upon ATP binding. Since loop extrusion was found to be dependent on ATP hydrolysis (9, 14), a detailed understanding of the conformational changes inside the SMC head domains will shed light on the unresolved questions facing the mechanism behind loop extrusion. To initialize future experiments, the purification of Smc1 and Smc3 head domains was established and their engagement and activity as well as their oligomeric states were analyzed.

I Biological introduction

This chapter will describe the biological role of SMC complexes for the three-dimensional arrangement of eukaryotic genomes and will introduce the architecture and regulation of SMCs.

Inside the nucleus, the DNA is highly compacted and shows a multi-layered organization (Fig. 1). The DNA fiber is wrapped around histone octamers to form nucleosomes (15). Furthermore, histones contain long tail regions, which can self-interact to build-up nucleosomal chains (16). Tightly packed nucleosomes correlate with epigenetic marks and form heterochromatin, which is typically characterized as inactive or silenced (17). On the contrary, euchromatin is supposed to be more accessible and active (18). The positioning of nucleosomes is further dependent on chromatin remodelers (19).

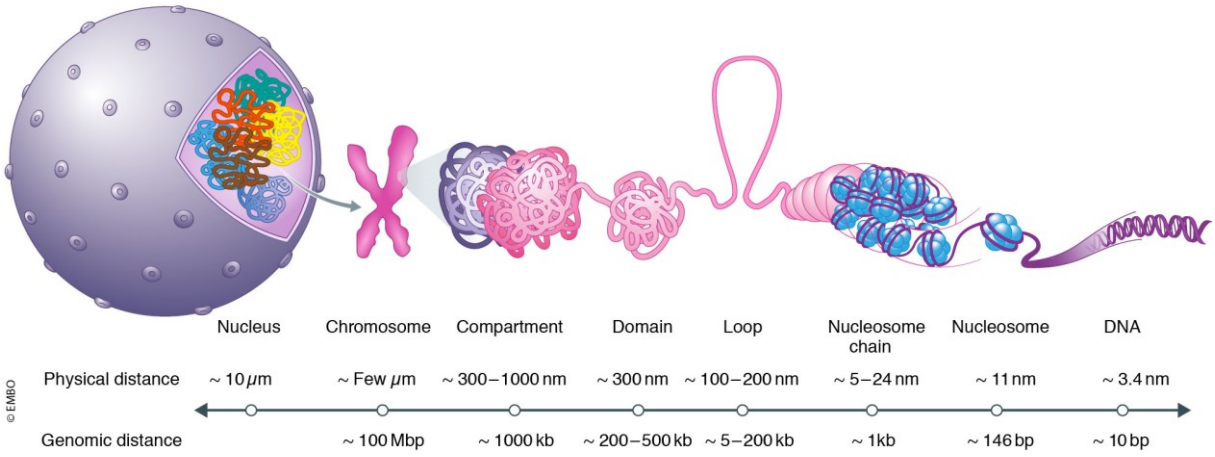

Figure 1: Genome organization. Inside the nucleus of eukaryotes, the genome is organized in different levels of compaction and chromosomal folding. Chromosomes contain distinct compartments. Compartments further consist of topological associated domains (TADs), which are regions of increased self-interaction and are based on DNA loops. DNA loops were recently found to be dependent on SMC complexes. Furthermore, DNA is wrapped around histones to form nucleosomes, which can arrange into nucleosomal chains. Adapted from (20).

Besides nucleosomal packing, DNA is condensed by chromosomal loops, which bridge DNA sequences over far genomic distances (1). Furthermore, genomic looping results in topological associated domains (TADs), which show an increased level of self-interactions (21). Domains lead to the compartmentalization of chromosomes, where transcriptionally active (A) and gene-rich compartments are differentiated from transcriptionally inactive (B) and gene-poor compartments (11).

The chromosomes undergo various transformations during the cell cycle. During interphase, the chromosomal conformation is dynamic, because loops are frequently formed and released (22). Interphase chromosomes are the least condensed and therefore accessible for e.g. transcription (23). During S-phase, the chromosomes are replicated and sister chromatids are physically held together (1, 2). Moreover, mitotic chromosomes exhibit nested loop structures, which lead to chromosome

condensation (24). Finally, sister chromatids segregate during anaphase, which ensures the end of each cell cycle (4).

I.1 The cohesin complex – a member of the SMC protein family

SMC proteins form multi-subunit complexes, which can be found in all kingdoms of life (1). Whether small prokaryotic genomes or multiple eukaryotic chromosomes, DNA molecules need to be accessible for transcription, replication or repair and simultaneously be highly compacted to fit into cells (21). This interplay between DNA accessibility and packing is organized by the members of the essential SMC complex family, which are postulated to dynamically form and resolve DNA loops (1, 21). DNA loops bridge distant regions in the genome, which provide physical contacts necessary for gene regulation (25). The development of chromosome conformation capture techniques identified numerous nested loops inside higher-order compartments of eukaryotic genomes, which were termed topologically associated domains (TADs) (11). Within TADs the amount of genomic self-interactions are upregulated, while interactions between different compartments are reduced (26).

The mechanism behind TAD formation and DNA looping was unknown until recent experiments have suggested the SMC complex cohesin as the mediator of DNA loops in interphase chromosomes (27) and the SMC complex condensin as the mediator of DNA loops in mitotic chromosomes (24). In a simplified view, the ring-shaped SMCs load onto chromosomes (Fig. 2 A,B) and extrude DNA through their lumen, which results in DNA loops of several kilobases in size (Fig. 2 C,D). Several complexes transform the linear DNA molecule into a compacted form. Since loop extrusion was found to be dependent on ATP hydrolysis (28, 29), a novel motor activity of SMCs was concluded (10, 30).

Furthermore, the SMC complexes share a conserved construction scheme from prokaryotes to eukaryotes (1), which will be introduced in chapter I.2. SMCs consist of elongated coiled-coil domains, which can perform large-scale conformational changes (31) and which will be explained in more detail in chapter I.3. Cohesin is highly regulated throughout the cell cycle (32), which will be highlighted in chapter I.4. After replication, the sister chromatids are entrapped inside cohesin's ring structure (Fig. 2 E), which gets later proteolytically opened to release the sister chromatids for segregation (Fig. 2 F) (2). Malfunction or impairment of this process frequently lead to severe defects like aneuploidy or even cell death (33). Further, cohesinopathies, like Cornelia de Lange syndrome or Roberts syndrome, are genetic disorders associated with mutations in cohesin or its regulators and show a wide range of deleterious phenotypes (34).

To summarize: SMC complexes, like cohesin, are responsible for the genome organization by the extrusion of DNA loops (DNA tethering in cis), which will be reviewed in chapter I.5. Additionally, cohesin is responsible for sister chromatid cohesion (DNA tethering in trans), which are essential processes and their molecular basis is after decades of research still not fully understood.

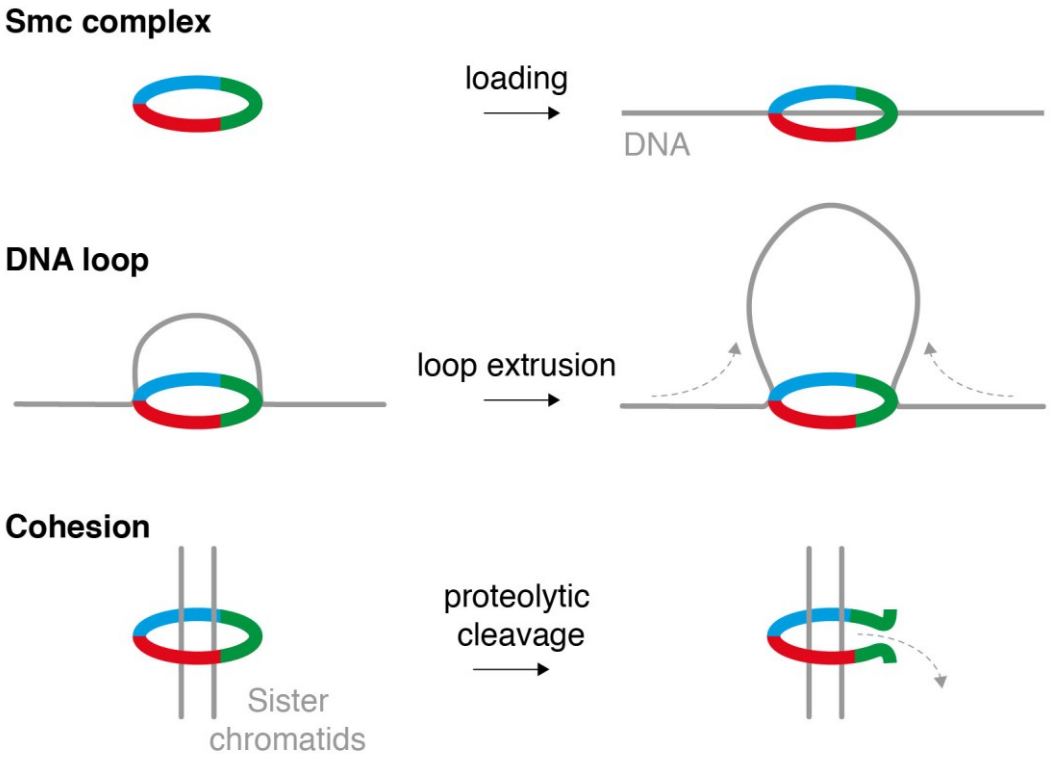

Figure 2: Chromosome organization by the SMC complexes. Top: Cartoon of a ring-like multi-subunit SMC complex, like cohesin, which can be loaded onto DNA. Middle: SMCs were found at the boundaries of DNA loops. Using ATP hydrolysis, the SMC complex extrudes DNA through its lumen in a processive manner, which increases the DNA loop length. It is still under debate, if single SMCs extrude loops symmetrically (as drawn) or asymmetrically. Bottom: After S-phase, the two sister chromatids are physically held together inside the cohesin complex (sister chromatid cohesion). In anaphase, cohesin is proteolytically opened by separase, which releases the sister chromatids.

I.2 Conserved SMC architecture

SMC proteins are present throughout all kingdoms of life (1). While prokaryotic SMC complexes like Smc-ScpAB (e.g. *Bacillus subtilis, B. subtilis*) or MukBEF (e.g. *Escherichia coli*, *E. coli*) consist of homodimeric SMC proteins (35), eukaryotic SMCs evolved into heterodimeric proteins and additionally, into different flavors of complexes, namely cohesin (Smc1/3), condensin (Smc2/4) and Smc5/6 (21). Cohesins are responsible for sister chromatid cohesion (32). Condensins are responsible for chromosome condensation during mitosis (10). The lack of an epithet for Smc5/6 is due to its diverse and poorly understood functions in DNA repair (3).

All members of the SMC family share a characteristic construction (Fig. 3, top left). Two SMC proteins dimerize at their hinge domains, which are connected via elongated, antiparallel coiled coils (SMC arms) of conserved length to the ATP-binding (ABC)-like ATPase head domains (1). Two active sites are generated by dimerization of the individual head domains upon ATP binding. This process is termed

head engagement and creates a recently discovered DNA binding site on top of the dimerized heads (Fig. 3, top right) (36). Hence, monomeric SMC proteins cannot hydrolyze ATP as they lack complete active sites (37–39).

The SMC dimer is bridged by the kleisin subunit at the heads in an asymmetric way (1). The N-terminal domain of the kleisin binds to the coiled coils above one SMC head by forming a helix bundle. This interface is known as the neck gate (40), a potential entry or exit gate of DNA inside the SMC lumen (41). The following amino acids of the kleisin are intrinsically disordered and harbor binding motifs for SMC regulatory subunits (Hawks, HEAT proteins associated with kleisins or Kites, kleisin interacting winged-helix tandem elements), which can largely differ between SMC complexes (1, 3). Finally, the C-terminal domain of the kleisin binds to the cap of the other SMC head as a winged-helix domain (42), which result in the characteristic tripartite ring structure of SMC complexes (8).

The conserved architecture suggests a conserved working principle for all the SMC complexes (30). The neck gate can be found in prokaryotes and eukaryotes and its regulation is essential for the entrapment of DNA inside the SMC complexes (1, 40). Furthermore, the head domains and the regulatory subunits were found to form a globular domain (8). Recent structures of SMCs bound to DNA revealed the gripping state of the globular domain (Fig. 3, below) as a conserved feature of SMCs and SMC-like complexes (43–47). This DNA clamp enables the regulation of DNA binding during the ATPase cycle. Interestingly, each SMC utilizes different regulatory subunits (Hawks or Kites) to clamp DNA on top of the engaged head domains. Furthermore, DNA binding was described for bacterial and eukaryotic hinge domains (35, 48). A conserved working principle can rely on these aspects, but latest studies focusing on SMC components, like the coiled-coil domains, have found complex-specific features (31, 49), which could enable the use of these conserved properties in a divergent manner.

This thesis focusses on the yeast cohesin complex and uses the protein nomenclature of *Saccharomyces cerevisiae* (*S. cerevisiae*) if not stated otherwise. The SMC proteins of yeast cohesin are Smc1 (blue in Fig. 3) and Smc3 (red in Fig. 3). The corresponding kleisin subunit is called Scc1 (green in Fig. 3) and cohesin's Hawks are Scc3, Pds5 and the cohesin loader complex Scc2/Scc4 (beige in Fig. 3) (1).

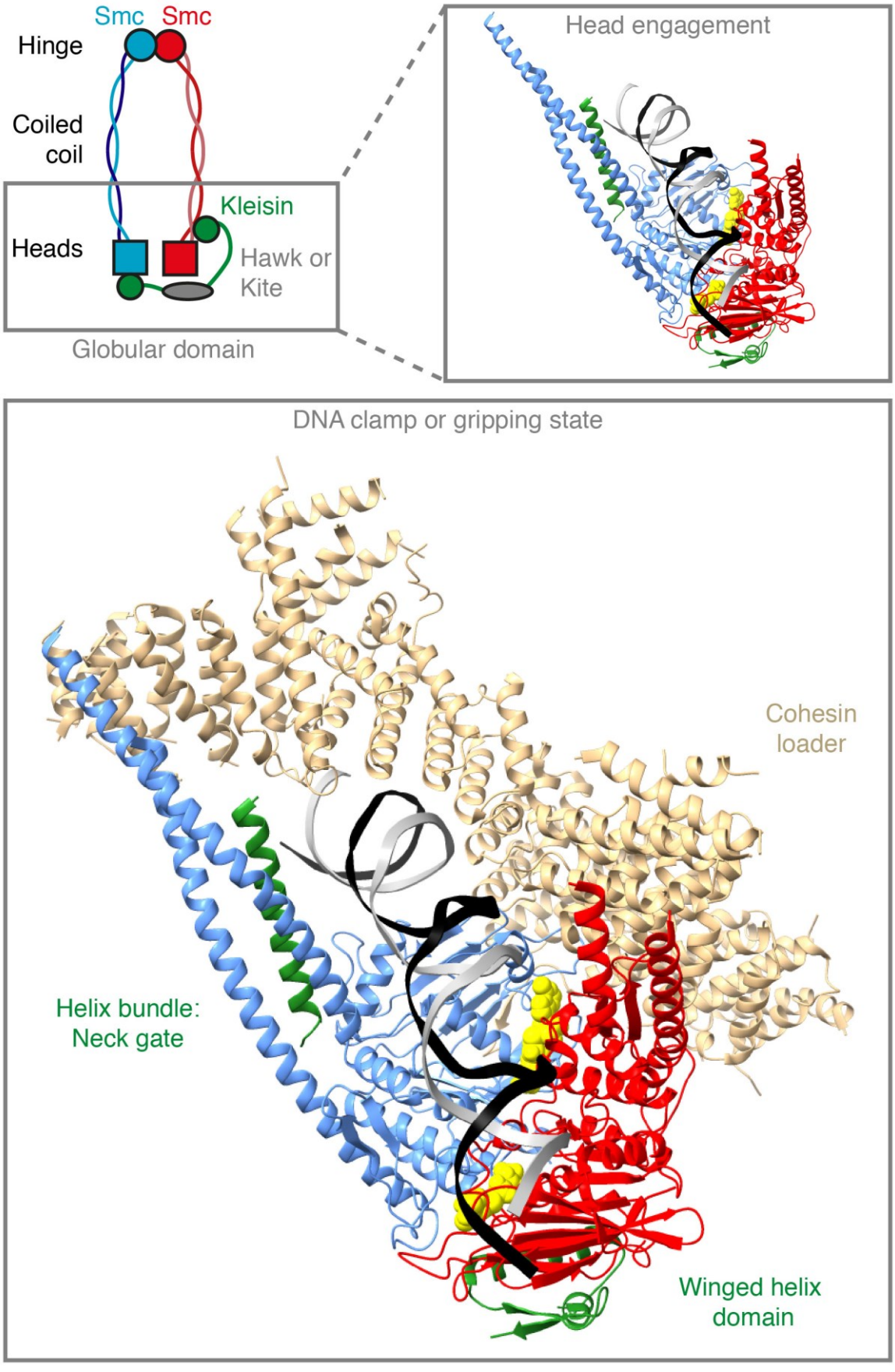
Smc
Smc
Hinge
Coiled coil
Kleisin
Heads
Hawk or Kite
Globular domain
Head engagement
DNA clamp or gripping state
Cohesin loader
Helix bundle: Neck gate
Winged helix domain

Figure 3: SMC architecture with focus on the globular domain. Top left: Two SMC proteins (blue and red) dimerize at their hinge domains. They further consist of elongated coiled coils and the head domains, where ATP can be hydrolyzed. A trimeric ring is formed by bridging the SMC proteins asymmetrically with the kleisin subunit (green), where additional regulatory subunits (Hawks or Kites) can bind and form the globular domain. Top right: The engagement of the SMC head domains is the dimerization upon ATP binding, which sandwiches two ATP molecules (yellow spheres) between the heads (shown by yeast cohesin, PDB: 6ZZ6). The resulting DNA (gray and black) binding site is located on top of the heads. Below: In the case of yeast cohesin, the loader complex (Scc2/Scc4) is the Hawk protein, which clamps DNA on top of the heads while the neck gate is closed. Similar structures of such a gripping state were resolved for fission yeast cohesin (PDB: 6YUF), human cohesin (PDB: 6WG3), yeast condensin (PDB: 7Q2Y), yeast Smc5/6 (PDB: 7TVE) and even MukBEF (PDB: 7NZ3), as well as bacterial MRN (PDB: 6S6V). This conformation represents an intermediate state, which was captured with point mutations in the active sites of both SMC proteins (EQ-mutants) that allowed ATP binding, but prevented ATP-hydrolysis and therefore head disengagement.

I.3 Features of SMC coiled coils

Coiled coils (CC) are an ubiquitous folding feature and can cover up to 10% of the proteome of an organism (50). Their function ranges from spacing binding sites (e.g. in restriction enzymes (51)), over mediating macromolecular assemblies (e.g. in the intermediary filament protein vimentin (52)) to propagating conformational changes (e.g. in the motor protein dynein (53)).

SMC proteins contain elongated, anti-parallel CCs of a conserved length (~ 50 nm) with domains at both ends (54), which indicate that a spacing purpose is likely. However, SMC CCs show, like dynein, a higher sequence conservation compared to other CC domains (50), which suggests a role in propagating conformational changes. The SMC complexes were found to adopt different CC conformations ranging from straight rods with a closed CC alignment to open, ring-like forms visualized with early imaging studies (7, 55) using atomic force microscopy (AFM) or rotary shadowing electron microscopy (EM). Crystallographic approaches could only resolve fractions of the CCs, if any (35, 40, 48, 56). In line, the estimated flexibility of the SMC arms (persistence length of 2-4 nm) (7) is ten times higher than the CC domain of the related MRN complex (57), which adopts closed rod-shape conformations (58).

Improved approaches using high-speed AFM under liquid conditions deepened the early observations and a nomenclature based on the alphabet were introduced for the different CC conformations (8, 59). Rods were called “I-shape”, rings were called “O-shape”. Moreover, collapsed (B-shape) conformations were described for the first time (Fig. 4, top) (31). In the B-shaped forms, the SMC arms fold onto each other at a novel bending point, which was consequently called the “elbow”. The bended elbow shrinks the distance between the hinge and the heads and could enable novel interactions between these domains, which was confirmed by the Cryo-EM structures of the cohesin and condensin holo-complexes

(Fig. 4, bottom) (60, 61). Additionally, frequent transitions between O- and B-shapes were revealed even for dimeric condensin proteins in the absence (7) and presence of ATP and DNA (8, 59). For the cohesin and condensin complex, DNA binding at the hinge domain and at the globular domain was observed, which identifies the O-shape/B-shape transition as the movement of two DNA binding sites relatively to each other (8), which is a prerequisite of loop extrusion. Single-molecule experiments confirmed that bending is independent of ATP (59), which suggests a passively bending motion highly depending on the elasticity of the elbow feature (12). Notably, I-shapes were sparsely populated, whereas the prokaryotic Smc-ScpAB (49) and Smc5/6 (62) preferentially adopts rod-shapes with aligned CCs (Fig. 4). Taken together, the imaging of SMC CCs revealed their extremely flexible polymer properties together with a portfolio of dynamic conformations, which encouraged more mechanisitic models (30, 63) for loop extrusion as described in chapter I.1.

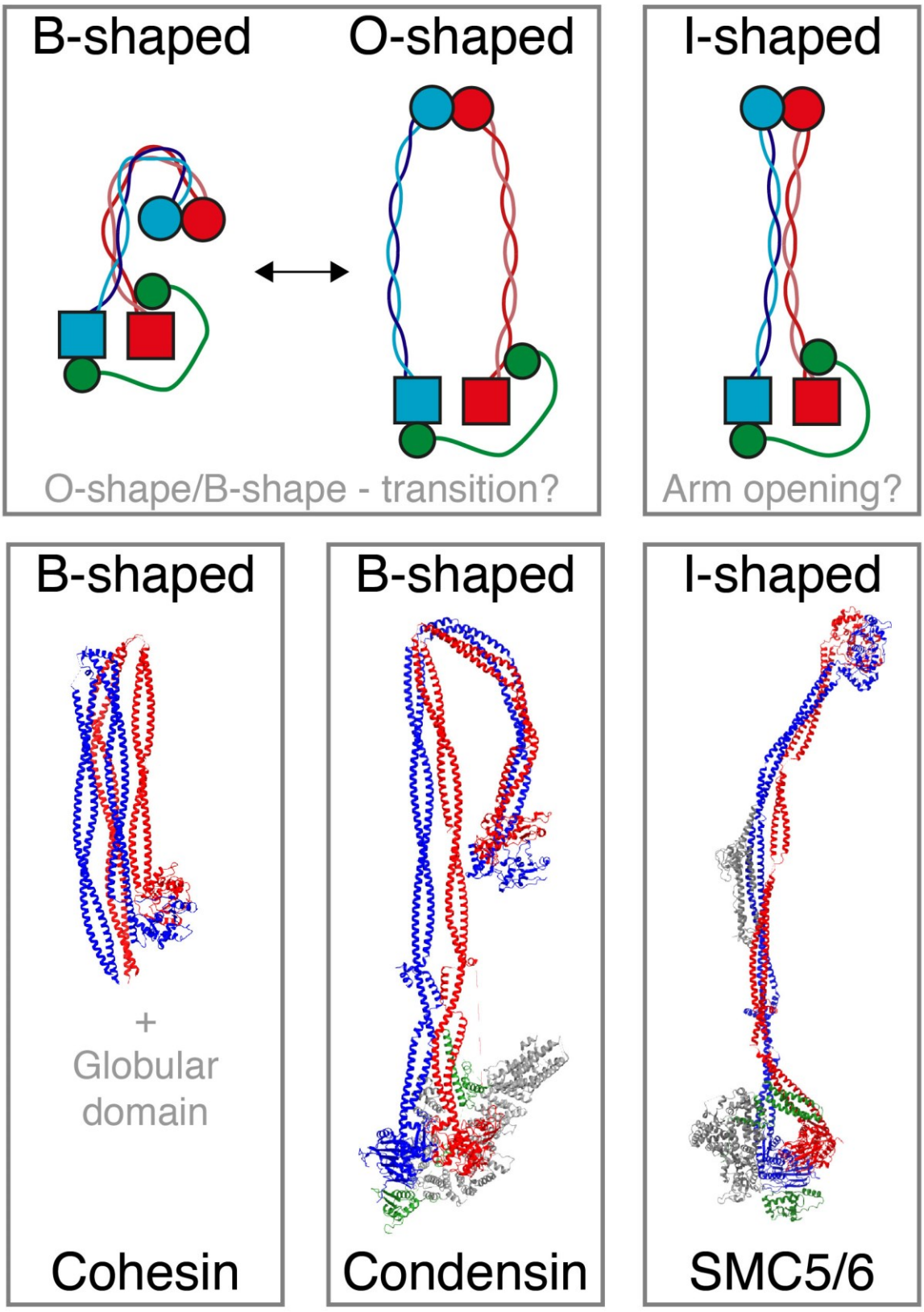

Figure 4: Cartoon and models of the SMC shapes. The B-shape of SMCs is a consequence of the bending at the elbow feature located in the middle segment of the CC domain. Frequent transitions between collapsed conformations and circular O-shapes were reported and could drive the SMC motor activity. Additionally, I-

shapes are preferentially found for some SMCs. DNA binding inside the I-shape could be achieved by a gradual opening of the aligned SMC arms. Bottom: Mid-resolution structures of the holo-complexes of the eukaryotic SMC complexes (yeast cohesin; PDB: 7OGT, yeast condensin; PDB: 6YVU and yeast Smc5/6; PDB: 7QCD), which enable the interpretation of the overall shape of the respective SMC complex, but detailed information about the coiled-coil kinks remain concealed.

Although the Cryo-EM structures of the SMC holo-complexes give fundamental insights about their overall shape, the lack of atomic resolution hinders a clear interpretation at critical features of the CC domains. The conformational flexibility of SMC complexes could be a consequence of the non-resolved parts of the CC domains, which were also observed as kinks in the mentioned imaging studies. In the solved crystal structures of the SMC arms (31, 35, 40, 64), kinks can be frequently found (Fig. 5), whereas their significance and the correlation between kinks and the observed CC flexibility is mostly unknown. For example, in the CC of MukB at least four so-called “knuckles” were identified (65), which are small globular folds that interrupt the CC (Fig. 5). However, only one is relevant for CC bending (31).

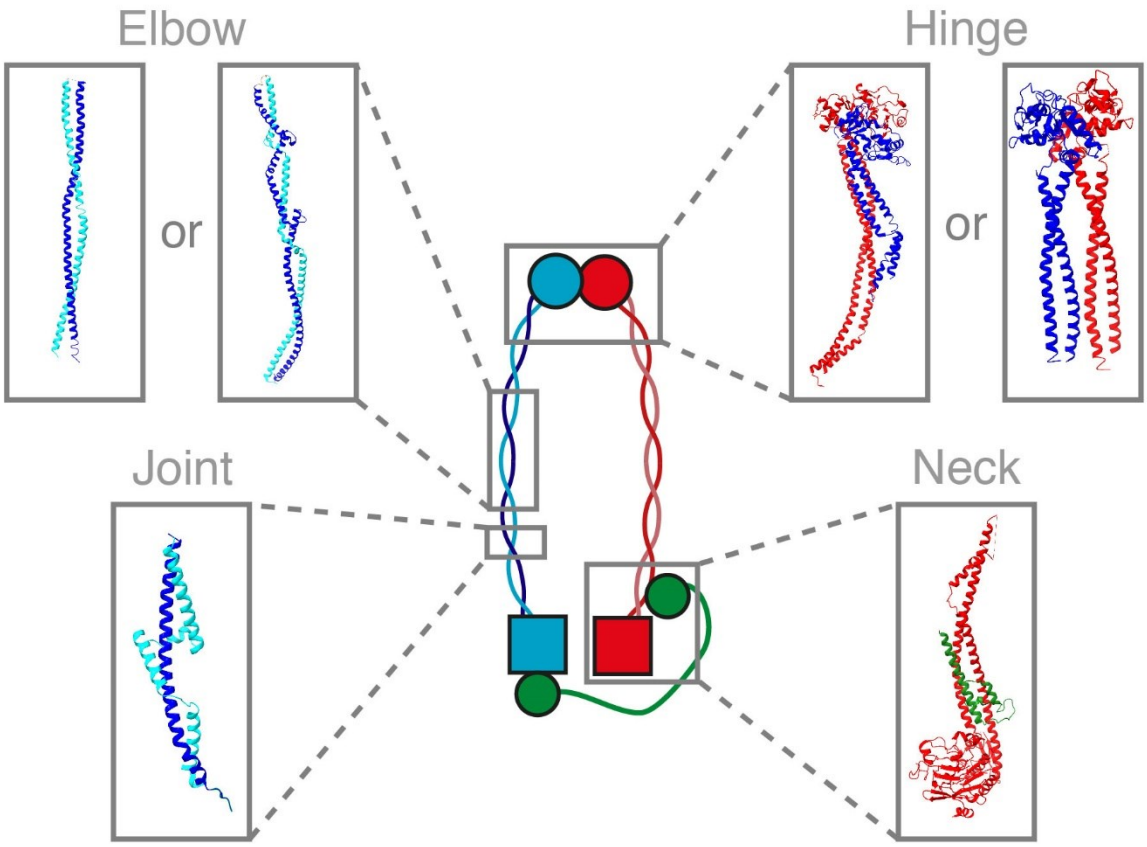

Figure 5: Structures of SMC coiled-coil fragments. The coiled coils emerging from the hinge domain are straight in the crystal structure of a homodimeric bacterial SMC (PDB: 4RSJ), whereas the eukaryotic coiled coils contain kinks and the hinge folds in a more asymmetric way (PDB: 4RSI). The elbow is located in the middle segment for cohesin, condensin or MukB (PDB: 6H2X), which enables bended shapes, whereas Smc-ScpAB (PDB: 5NNV) is nearly straight in that region, which would be

in agreement with a rod-like shape. The joint (PDB: 5NMO) is a conserved feature above the neck gate and provides local flexibility above the head domains. This features includes a straight helix on one coiled-coil strand (dark blue) surrounded by interrupted helices on the other strand (light blue). The neck gate (PDB: 4UX3) consists of a four-helix bundle including the N-terminus of the kleisin subunit (green) and is located between the joint and the head domain.

For unresolved CC regions, prediction software can be consulted (66–68). CC domains can fold in a parallel or anti-parallel manner, while they generally follow a heptad repeat pattern $(abcdefg)_n$ (50). In a canonic CC motif, the amino acids “a” and “d” are hydrophobic (Fig. 6 left) and build up the hydrophobic core. The flanking amino acids stabilize the fold via electrostatics, which results in the so-called knobs-into-holes packing (50). CC prediction software can evaluate, if the amino acid sequence of a given protein follows these guidelines and predicts a propensity along the sequence (67, 69, 70).

Besides canonical CC motifs, deviations can be frequently found in natural occurring proteins (50). CCs can assemble in different geometries (trimers, tetramers or higher oligomers) (71) and discontinuities can interrupt the heptad repeat pattern (72), which results in a reduced CC propensity (31). GCN4 contains a canonic CC fold (73), which correlates with high propensity values (Fig. 6 right). In contrast, OMP100 contains a linker insertion in between CC stretches (72). Consequently, the propensity profile at the insertion causes a drop (gray line in Fig. 6), but several other minima can be found in the profile, which do not affect the CC fold. An identification of the linker insertion solely based on the propensity profile is not possible. This shows the ambiguity of the sequence-based prediction software, which can identify canonical CCs, but have their limitations at non-canonical CC features, which SMC proteins seems to have (31, 65).

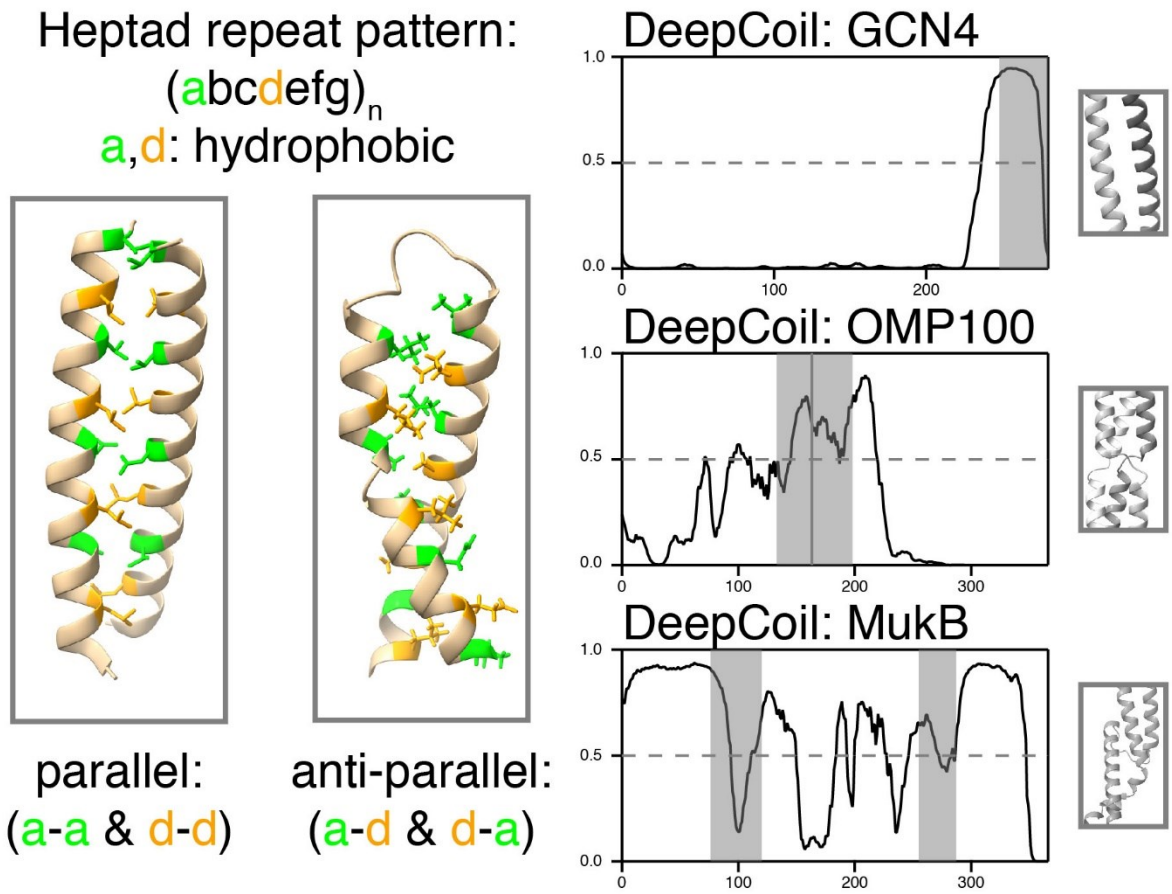

Figure 6: Construction of coiled coils. Left: Coiled coils generally follow a heptad repeat pattern and can fold in a parallel or an anti-parallel way. Right: Sequence-based coiled-coil propensity profiles of selected coiled-coil bearing proteins using DeepCoil (dashed line, significance threshold of 0.5). The gray highlighted regions correspond to the crystal structure on the right. For GCN4 (PDB: 2ZTA), the propensity profile identifies a coiled-coil fold at the C-terminus of the protein. The high score value correlates with a canonical coiled-coil fold, which was confirmed by crystallization. For OMP100 (PDB: 5APP), a linker insertion was found in the crystal structure, which causes a drop in coiled-coil propensity (gray line). It cannot be concluded that this protein has an insertion at the indicated position solely based on the propensity profile as several other minima are present in the profile. For the anti-parallel SMC coiled coils, like MukB (PDB: 6H2X), the N- and C-terminal strands of the same protein interact with each other (two gray bars). Therefore, the exact pairing must be known to allow an unambiguous interpretation of the propensity profile. In addition, several extended minima and maxima contradict the notion of a continuously elongated coiled-coil motif.

The propensity profiles of SMCs often contain drops along the sequence, which was previously reviewed (65). While the “joint” turned out to be a conserved feature (64), it was speculated based on the propensity profiles that all SMCs could have an elbow in the middle segment of their CC domains (31).

Latest studies (49, 62) including the Cryo-EM structure of the Smc5/6 holo-complex (Fig. 4) contradict this notion.

Emerging tools in protein structure prediction, like AlphaFold2, can perform a structure predictions with an unachieved accuracy relying on existing structures in the protein data bank (74–76). Unfortunately, the lack of solved structures containing kinked CCs results in a low prediction confidence at these sites for SMC proteins (12). To get detailed information about these features, different approaches than structural determination or structural prediction are necessary.

I.4 Regulation of cohesin throughout the cell cycle

The activities of the SMC family members are highly regulated by different subunit and post-translational modifications (1). As an example, the interplay of different regulatory subunits with cohesin throughout the cell cycle will be outlined in the following section. Since the fine-tuned regulation of condensin in chromosome condensation and Smc5/6 in DNA repair varies sharply from cohesin's cell-cycle regulation and are not the focus of this thesis, I refer to recent reviews (1, 3).

In early G1, the yeast cohesin complex lacks the kleisin subunit Scc1 (32) and the Smc1/3 dimers cannot bind to chromosomes (Fig. 7). In the late G1-phase, Scc1 is produced, which results in the full complex assembly and the loading of cohesin onto chromosomes (77). Essential for the loading reaction is the loader complex Scc2/Scc4, which gets recruited to centromeres beforehand (78). Therefore, cohesin gets preferentially loaded at centromeres, but translocates with an ATP-dependent mechanism, most likely loop extrusion (Fig. 2), into the pericentromeric region (14).

During S-phase, the acetyltransferase Eco1 travels with replications forks via PCNA (79) and acetylates conserved lysines in the Smc3 head domain (80), which results in the establishment of sister chromatid cohesion (Fig. 2,7). Here, the loader complex dissociates and instead Pds5 associates with cohesin (81), because Pds5 preferentially binds to acetylated Smc3 head domains (82) and additionally shields the lysines from deacetylation by Hos1 (83). These cohesin complexes are termed cohesive cohesins (21) and were found to topologically entrap the sister chromatids (84). Consequently, the topological entrapment results from an opening of the ring structure followed by a DNA entry and a re-closing of the ring structure. It is still under debate, whether the hinge dimerization domain (66) or the neck gate (40) is the interface, which allows DNA entry (41).

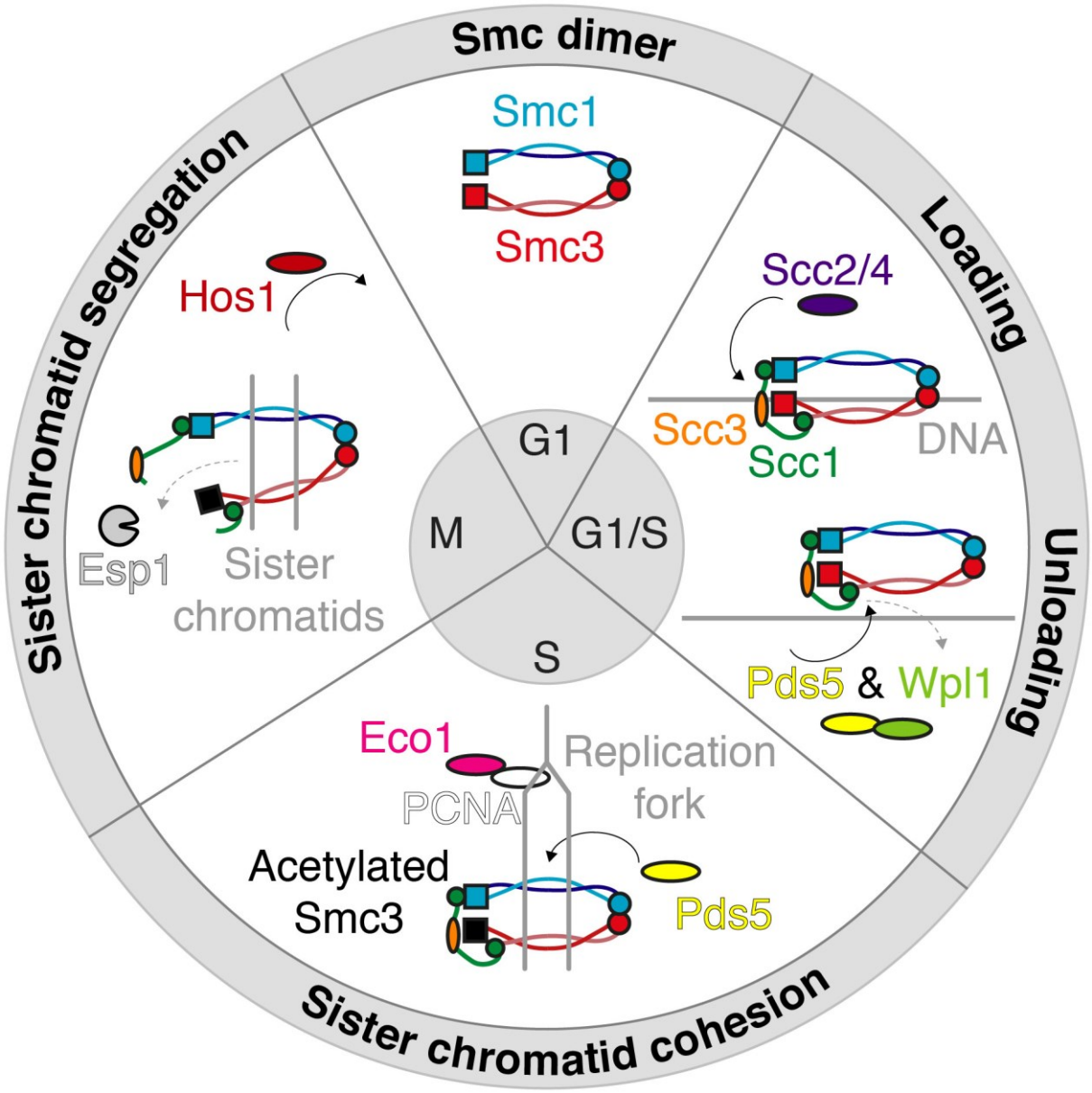

Figure 7: The transformation of yeast cohesin during the cell cycle. In G1, cohesin comprises only dimerized Smc1 and Smc3 proteins and is not bound to the chromosomes. The expression of the kleisin Scc1 enables the binding of Scc3 and Scc2/4 to the complex and results in the loading of cohesin to chromosomes. A pathway including Pds5 and Wpl1 removes cohesin from chromosomes. During S-phase, cohesin gets cohesive by the acetylation of conserved lysine residues in the Smc3 head domain by Eco1, which is bound via PCNA to replication forks. Pds5 associates to cohesin and protects the acetylation mark. Cohesive cohesin topologically entraps the sister chromatids through G2-phase. Cohesion is maintained during mitosis until Esp1 is activated and proteolytically opens the ring structure, which allows sister chromatid segregation. Deacetylation by Hos1 allows the re-cycling of the Smc1/3 dimers.

In anaphase, the sister chromatids are attached over the kinetochores to the spindle apparatus and the cohesin rings resist the force generated by the spindle apparatus (85) until they get proteolytically

opened by the separase Esp1 at the kleisin subunit (1), which releases the sister chromatids and enables chromosome segregation (Fig. 2,7). Beforehand, Esp1 activity is blocked by the securin Pds1, which gets degraded after the spindle checkpoint (32). Additionally, the cleaved kleisin gets completely degraded and the deacetylase Hos1 removes cohesin's acetylation (86), which allows the re-cycling of the resulting Smc1/3 dimer in the next cell cycle (32).

In addition, the turnover of non-cohesive cohesin on chromosomes during interphase is regulated by Wpl1 (87), which associates with the cohesin complex in the presence of Pds5 (88, 89). It was found that Wpl1 destabilizes the neck gate, which results in an opening of this interface and hence to the exit of entrapped DNA from the cohesin complex (40, 90). Consequently, the neck gate was suggested to be the DNA exit gate (88). The acetylation of Smc3 counteracts this Wpl1-dependent unloading reaction (90) and ensures the maintenance of sister chromatid cohesion between S-phase and anaphase (32). Interestingly, the human ortholog of Wpl1 causes the removal of acetylated cohesins during prophase at regions distal to the centromeres (91), which reflects the miscellaneous regulation of SMC complexes.

I.5 How can SMCs extrude DNA loops?

In the last chapter the transition from dynamic cohesin, which can be loaded and unloaded from DNA (14, 77, 91), to cohesive cohesin, which can persist cohesion in oocytes for decades, was retraced (92). While cohesion was already described (2, 93), the role of dynamic cohesin was strongly debated until first Hi-C experiments found the compartmentalization of interphase chromosomes (11), which suggested that dynamic cohesin complexes are able to extrude chromosomal loops. Recently, *in vitro* reconstitution experiments confirmed that loop extrusion (LE) is a conserved phenomenon of eukaryotic SMC complexes (9, 94, 95).

During the high-throughput chromosome conformation capture (Hi-C) procedure (11), chromosomal DNA of a population of cells is fixed by proximity-ligation, which preserves contacts of genomic regions. After sequencing, ligated DNA sequences can be identified, which revealed contact frequencies of adjacent DNA sequences illustrated in contact maps (Fig. 8). In Hi-C contact maps, chromosomal loops of kilobase to megabase pair length were found that lead to a compartmentalization of the genome into TADs (Fig. 8). The depletion of the human cohesin kleisin subunit Rad21 resulted in a complete loss of chromosomal loops in Hi-C maps identifying cohesin as the origin of chromosomal loops (27). Hi-C of single cells (96) revealed that contact maps are a result of the population averaging, because certain chromosomal loops cannot be found in each cell at each time. Chromosomal folding is therefore a dynamic process, in which DNA loops are constantly extruded and resolved, which is mediated by cohesin and its regulators (21).

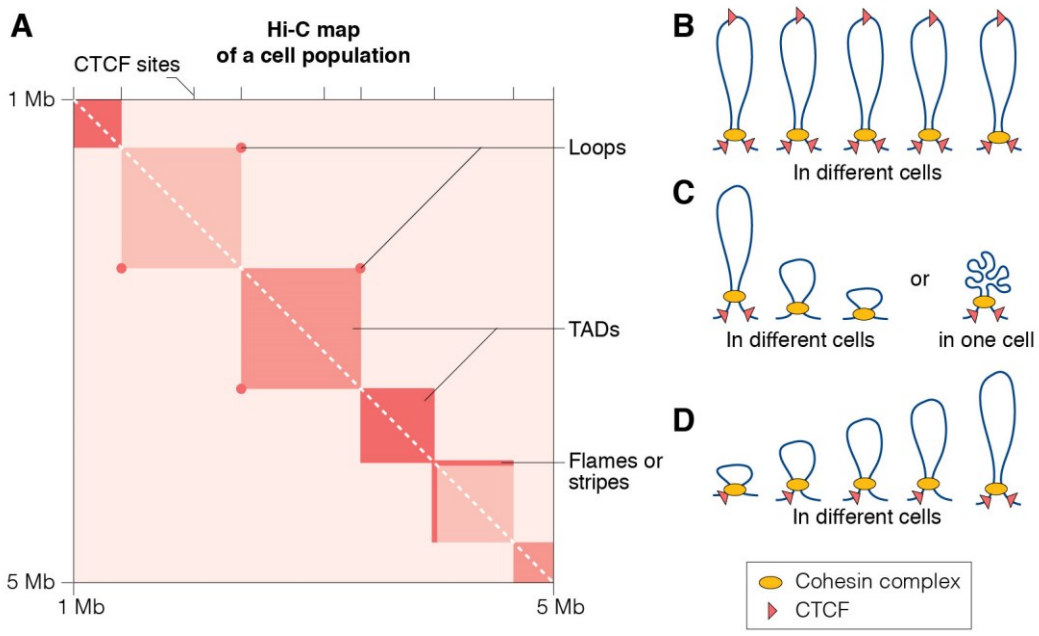

Figure 8: Cartoon of a typical Hi-C map and corresponding chromosomal loops. A) Schematic Hi-C map of a human chromosome including positions of CTCF sites. By averaging over a cell population, interactions of a given DNA sequence with its neighboring DNA sequences can be investigated. The brightness of the matrix elements indicate the observed contact frequencies and features like dots, triangles and flames or stripes can be observed. B) Dots represent chromosomal loops, which can be found in several cells. Dots correlate with converged CTCF sites in human genomes. C) Triangles indicate topologically associating domains (TADs), in which sequences can interact with each other and are insulated from the surrounding. TADs result from loops, which are about to be extruded and represent a phenomenon of the averaging over different cells or TADs result from long loops in one cell, which are able to interact at multiple positions with themselves. TADs are insulated by CTCFs. D) Flames or stripes indicate that a specific DNA sequence interact frequently with multiple sequences inside a TAD. This can be the result of anchored cohesins, which extrudes DNA asymmetrically until they encounter the next converging CTCF. Adapted from (21).

One of these regulators is human CTCF (97), a transcription factor, which is found at contact boundaries in Hi-C experiments if the orientation of CTCF binding sites face towards the loop center (98). Hence, CTCF is a LE barrier and can block or at least pause cohesin-mediated LE (99). Consequently, the depletion of CTCF reduces the genome-wide accumulation of chromatin loops at CTCF sites and reduces the insulation of TADs (100). The depletion of WAPL, the human ortholog of Wpl1, which is responsible for the removal of human cohesin from chromosomes, results in increased loop lengths even beyond converging CTCF sites (101, 102), which clearly demonstrated that LE is happening *in vivo*. Additionally, depletion of CTCF and WAPL resulted in “cohesin islands”, which are accumulations of cohesin complexes at convergently transcribed genes (103). The increased residence time of cohesin

on chromosomes (WAPL depletion) and the absence of the LE barrier (CTCF depletion) verified that the transcription machinery can push cohesin complexes (104), which resulted in the observed cohesin islands (21). Moreover, the replisome was shown to push cohesin complexes as well (5, 105). Recently, LE was found to be involved in other genomic processes, like V(D)J-recombination (106) or double-strand-break repair (107). To summarize: The *in vivo* experiments confirmed that cohesin mediates LE and organizes the genome into compartments of high self-interactions. LE is restricted by genomic barriers and utilized in different genomic processes to bring different DNA sequences in close proximity.

Recently, the *in vivo* experiments were supported by *in vitro* reconstitution experiments (9, 94, 95), where the formation of DNA loops could be observed in real-time. Although the hypothesis that LE structures the genome was proposed even before the first Hi-C experiments were performed (4), a detailed mechanistic description of the LE process is still missing. Through the years, the notion of a simple ring-like SMC complex, which hauls DNA through its lumen (Fig. 2), turned out to be not detailed enough. First, single-molecule observations by Stigler et al. and Davidson et al. found that cohesin cannot be fully circular at least at all times (108, 109), because cohesin cannot bypass obstacles smaller than its observed diameter. Second, the transitions between O-shaped and B-shaped complexes (Fig. 4) suggest that large-scale conformational changes mediated by the coiled-coil domain happen during LE (8, 59). Lastly, the fast extrusion rate (1 - 2 kbp/s) (9) combined with the large step sizes of condensin (110) and cohesin (99) per ATP cycle (20 - 60 nm, which can be larger than their diameter) contradicts the notion of a static ring-like SMC complex. Hence, LE models relying on a base-by-base translocation mechanism along the DNA “track” (111) similar to e.g. helicases or polymerases are unlikely, which implies that loop extruding SMC complexes own an unique and new motor activity (6). To better understand this process, I summarize briefly the observations of the *in vivo* reconstitution experiments, discuss the theoretical requirements for such a process and conclude with open and unanswered questions.

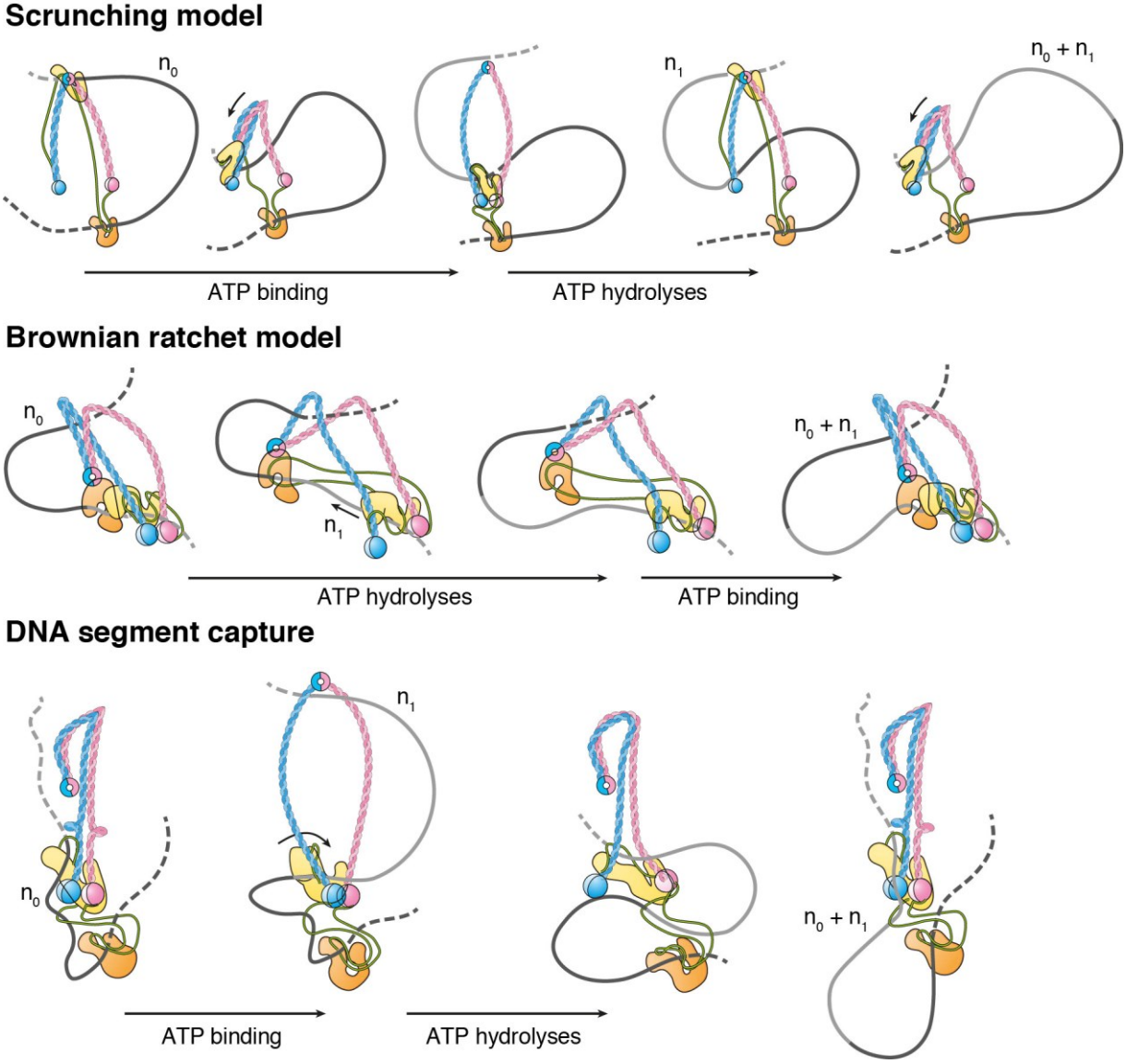

Figure 9: Models of SMC-mediated loop extrusion. Top: The scrunching model suggests that the DNA is initially bound at the hinge, potentially clamped by a Hinge-Hawk module and a safety-belt anchor at the other site, which captures an initial loop n_0. ATP binding and the bending of the coiled coils result in the gripping state conformation, while the DNA is transferred from the hinge to the globular domain. ATP hydrolysis resolves the gripping state and drives the SMC in the initial state, where the next loop n_1 is captured and later added to the first loop ($n_0 + n_1$) during the next ATP cycle. Middle: The Brownian ratchet model assumes that ATP hydrolysis detaches the Hinge-Hawk module from the gripping state to bind to a second position at the DNA. The disengagement of the ATPase head domains is bridged by the other Hawk subunit, which anchors the initial loop. The Hinge-Hawk module feeds DNA to the globular domain driven by Brownian motion and locked by the ATP bound state, where the Hinge-Hawk module can be bound to the globular domain. Bottom: The DNA segment capture model suggests that the Hawk subunits form individual pores with the tripartite SMC ring and capture an initial loop. The open conformation binds DNA at the hinge and transfers the DNA by the alignment

of the coiled coils towards the globular domain. The disengagement of the heads allows the transfer of DNA into the next pore, which results in the DNA segment capture. Figure adopted from (6).

Observations of loop extrusion reconstitution experiments

Besides the conserved and extraordinarily low stalling force of LE (0.5-1 pN) (9, 94, 95), the symmetry and the conformational states are different between the SMC complexes. Yeast condensin performs LE asymmetrically as monomers, while two individual condensins can form nested loops (Z-Loops), which can reel in DNA symmetrically (9). Monomeric cohesins were first reported to extrude loops symmetrically (94), which was later clarified. Further experiments suggest that cohesin reels in DNA asymmetrically, but switches the direction of extrusion rapidly, which leads to the notion of a symmetric process (99). Single-peptide chain SMC complexes were further reported to extrude loops (94, 112), which suggests that DNA is bound non-topologically because no interface is needed to be opened during LE. In line, SMCs can reel in large obstacles attached to the DNA (up to 200 nm) during LE, which indicates that the complexes “dangle” along the DNA and therefore are able to evade roadblocks (113). This is in contrast to cohesive cohesin, which topologically entrap DNA (114) and which would be blocked by these large obstacles. Smc5/6 was recently reported to translocate unidirectionally along DNA as a monomer, but starts to extrude DNA symmetrically as a dimer (95). Thereby, the eukaryotic SMC complexes share the ability to extrude loops and share a conserved construction scheme, but in a detailed consideration, they perform LE processes in an individual way (6), which is a challenge for mechanistic modeling approaches (30, 63, 115).

Modeling of loop extrusion

A reasonable LE model that agrees with the vast number of observations must address high demands. First, SMC complexes need multiple DNA binding sites. Second, these binding sites must move relatively towards each other. Third, alongside this movement, DNA must be bound (at least) at two binding sites, transferred and released (at least) at one binding site and finally re-bound at a distal position to start the next cycle. Furthermore, directionality and processivity must be guaranteed to accomplish functional motor proteins, which can extrude DNA loops of the observed length (6). Since LE was reported to dependent on ATP hydrolysis (9, 14), the motor cycle must be coupled to the ATPase cycle, while individual steps of the motor cycle can be favored by the energy gained through ATP hydrolysis.

Interestingly, the SMC domain, which provides the transient DNA binding site, is not identified yet (30). The hinge domain has a weak interaction with DNA and could initially bind to the DNA (35, 48). The transfer of DNA towards the heads could be achieved via the coiled-coil bending dynamics (O-shape / B-shape transition) (8, 31) and DNA could be released if it encounters a binding site with higher affinity, probably the globular domain. This mechanism is the basis of the scrunching model (110) (Fig. 9). Alternatively, the ATPase heads can engage upon ATP binding to form a DNA binding site (36) and disengage after ATP hydrolysis, which would release DNA. This would provide a transient DNA binding site coupled to the ATPase cycle. Otherwise, the transient DNA binding site could be provided by a regulatory subunit, like the cohesin loader complex, which was found to clamp DNA while being in

contact with either the hinge or the heads in an ATP and DNA dependent manner (59), which is postulated in the Brownian ratchet model (115) (Fig. 9).

The movement of the binding sites and the transfer of DNA can be achieved by the coiled-coil bending dynamics at the elbow (see above) for cohesin and condensin, whereas Smc5/6 is not reported to have an elbow (62). Therefore, another mechanism likely based on the dynamic alignment of the coiled coils as described in the DNA segment capture model could ensure the DNA transfer from the hinge towards the globular domain (116).

Directionality and processivity

Directionality of condensin- and cohesin-mediated LE could be achieved by the restriction of the elbow bending movement towards a preferred side, which was described for the human cohesin complex (59). In line, cohesin was recently reported to extrude loops asymmetrically, but can switch the direction of LE (6, 99). Yeast condensin extrudes asymmetrically (9), which is a result of the static anchoring at Ycg1, the so-called safety-belt (117, 118). Depletion of the safety-belt results in the slippage of DNA loops, while increased salt concentrations led to unidirectionally translocating complexes (119). Both effects are a result of poor anchoring (9, 119) and affect the directionality or the processivity of LE. Furthermore, human condensin I and II are found to extrude DNA symmetrically or asymmetrically (120).

Ycg1 is most similar to Scc3 (1) and was found to bind DNA distal to the globular domain (112). In contrast, Scc3 was reported to stably form a DNA clamp with the hinge (45, 115). Recently, cohesin was investigated in the context of CTCF encounters (99), which could establish the static anchoring of cohesin-mediated LE by binding to cohesin via Scc3 (121). During CTCF encounters, growing loops could either shrink, be stalled or change their direction of extrusion (99). Additionally, CTCF was found to be an imperfect LE barrier. Further experiments are needed to shed light on the cause of directionality, anchoring and the differences between condensin- and cohesin-mediated LE.

In addition, the processivity of LE is still puzzling. How can SMCs move along DNA with that enormous speed and perform these large step sizes in a repeatable manner? One interesting observation is that EQ-condensin, which is able to bind, but not hydrolyze ATP, can perform only a single extrusion step in a magnetic tweezers experiment, which was also found to be reversible (110). This suggests that ATP-binding after the capture of DNA at the hinge allows the movement of the Hinge-DNA clamp to the globular domain, where this clamp can bind. Although bound to the globular domain, the DNA is apparently not released or transferred from the Hinge-DNA clamp, which must be a consequence of the EQ-mutation. This indicates that following LE steps are dependent on ATP hydrolysis and perform the DNA transfer. Interestingly, the step size of the single extrusion step is comparable with the conformational change of the O-shape/B-shape transition (8) and this single step is reversible, which suggests that the Hinge-DNA clamp binds at the globular domain with a certain dissociation rate in the B-shape conformation. However, also other subunits ensure processive loop growth, as their depletion impedes LE (59, 94). In summary, the interplay of SMC domains with their regulatory subunits during the ATPase cycle must guarantee processivity, but critical aspects about the different states of the ATPase cycle and LE are not well understood.

Taken together, the fragile regulation of the SMC head domains is the key element to understand the LE phenomenon. The heads ensure the ATPase turnover (56), which is linked to the directed transfer of DNA during LE in the complex (6). Furthermore, the SMC coiled-coil domains are a crucial component of the LE process as they facilitate the observed large-scale conformational changes (54) and additionally contribute to the topological entrapment of sister chromatids inside the cohesin lumen (114). Detailed structural information of the SMC coiled coils are still elusive, as the coiled coils are inherent flexible (7) and most structures of SMC domains contain only partially resolved coiled coils, if any (48, 56). To gain further understanding of the construction and stability of the SMC arms, I chose a large stretch of the fission yeast Smc3 (Psm3) coiled coil as a model system for unfolding studies (12). These experiments were performed with a high-resolution optical tweezers set-up, which will be explained in chapter II. The power of the miscalibration correction procedure, which allows measurements with high accuracy and precision, will be presented in chapter III.1. The results of the coiled-coil unfolding experiments are reported in chapter III.2. The purification of active SMC head domains as single-peptide chains will be the focus of chapter III.3. In chapter IV, the discussion will focus on the differences between cohesive and loop extruding cohesin complexes and on the loop extrusion models, which are based on the indications listed above.

II Methodological introduction

During my first year as a PhD student in 2018, Arthur Ashkin was awarded with the Nobel Prize for the invention of optical tweezers, which motivated me to investigate biological systems with this remarkable technique. This chapter will focus on the principle of optical trapping, force spectroscopy of proteins using advanced polymer models, calibration and sources of miscalibration of optical tweezers and lastly the emergent tool of mass photometry to study protein sizes and oligomeric states of proteins.

II.1 Optical tweezers

Following the anecdote of Ashkin (122), he discovered that living objects can be trapped inside a laser focus by accident in 1987. He forgot to switch off a laser, which was focused in a water bath. The next day, Ashkin found bacteria trapped alive inside the laser focus. Ashkin explained this phenomenon with radiation pressure and ray optics (123). Dielectric particles with a refraction index greater than the surrounding medium exposed to highly focused laser light experience a scatter force along the laser beam due to the transfer of momentum from the incoming photons. Additionally, a gradient force in the direction of highest laser intensity acts on the dielectric particle (124). Closer to the laser focus (thicker red line in Fig. 10), the photons will transfer more momentum to the particle as photons from a distal laser beam (thinner red line in Fig. 10), which results in a net force (black arrow). Therefore, particles displaced from the trap center will be pushed towards the region of highest laser intensity by the gradient force. Considering both forces, the bead will be trapped closely to the laser focus, where both forces cancel out.

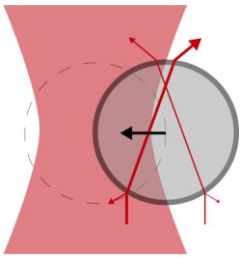

Figure 10: Optical trapping. If a dielectric particle is displaced from the trapping position (dashed lines), the scatter force along the laser beam and the gradient force, which acts towards the region of highest laser intensity, drive the bead back towards the trapping position (black arrow).

II.2 Force spectroscopy using a dumbbell approach

Since the discovery of Ashkin, a variety of applications have emerged that provided insights into microscale systems ranging from single atoms (125) and single proteins (126–128) to living cells (129) and phase-separated droplets (130). In this thesis, optical beads, whose surfaces are chemically modified to allow site-specific attachment of biomolecules, were trapped within a dual-laser beam optical tweezers set-up (12). Proteins were tethered between the beads using DNA handles and a dumbbell

geometry. This approach allows the observation of un- and refolding events with high spatio-temporal resolution (131). Folding events were recorded by actively pulling on tethered molecules while recording force-distance information or by passively recording folding events over time at constant trap separations (132).

In detail, the 1 μm-sized silica beads were coated with either streptavidin or anti-digoxigenin. To prevent photo damage, DNA handles of 510 base pairs served as stiff spacers to transduce the force response of the tethered molecules to the beads. The DNA handles were equipped with compatible moieties (Dual-biotin or Dual-digoxigenin) at one end and a 34 nucleotide overhang at the other end. Finally, the protein of interest was attached to complementary single-stranded oligonucleotides, which can be hybridized to the DNA handles. The resulting linkage of bead-handle-protein-handle-bead reminds of a dumbbell (Fig. 11).

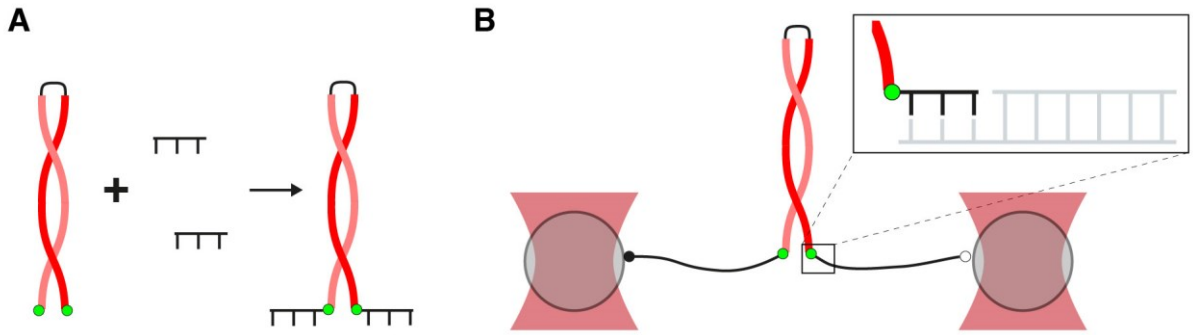

Figure 11: Sample preparation for optical tweezers. (A) A purified protein (red cartoon) is attached to oligonucleotides (black) via engineered tags (green). The attachment generates protein-DNA chimeras suitable for optical tweezers manipulation experiments. (B) The protein-DNA chimera is hybridized with DNA handles via the single-stranded overhang of the handles and the complementary oligonucleotides. The handles are tethered between optical beads with the biotin:streptavidin or the digoxigenin:anti-digoxigenin interaction. Figure adapted from (12).

II.3 Polymer models

Most biomolecules are macromolecules and consist of covalently attached monomers or subunits. In case of DNA or proteins these monomers are nucleotides or amino acids, respectively, which are lined up as long strings. The stretching of these macromolecules can be described using various models including the versatile worm-like chain (WLC) model (133), which accounts for restrictions in the movement of subunits and introduces a short-range correlation parameter called persistence length p. The WLC model is used to describe unfolded polypeptide chains and can be interpolated:

$$\mathrm{F}(\xi_p) = \frac{k_\mathrm{B}T}{p_p}\left(\frac{1}{4}\left(1-\frac{\xi_p}{L_p}\right)^{-2} - \frac{1}{4} + \frac{\xi_p}{L_p}\right) \quad (1)$$

with force F, unfolded protein extension ξ_p, thermal energy $k_\mathrm{B}T$, protein persistence length p_p and protein contour length L_p.

The contour length L is the length of the macromolecule. For proteins, each amino acid contributes 0.365 nm to the contour length (134), which follows along the peptide bonds. For DNA, each base pair contributes 0.34 nm to the contour length (135), which follows along the sugar backbone. During the stretching of DNA (Fig. 12) the elastic response cannot be neglected and is taken into account using the extensible worm-like chain (eWLC) model (136):

$$\mathrm{F}(\xi_D) = \frac{k_\mathrm{B}T}{p_D}\left(\frac{1}{4}\left(1-\frac{\xi_D}{L_D}+\frac{F}{K}\right)^{-2}-\frac{1}{4}+\frac{\xi_D}{L_D}-\frac{F}{K}\right) \tag{2}$$

with DNA extension ξ_D, DNA persistence length p_D, DNA contour length L_D, and the stretch modulus K.

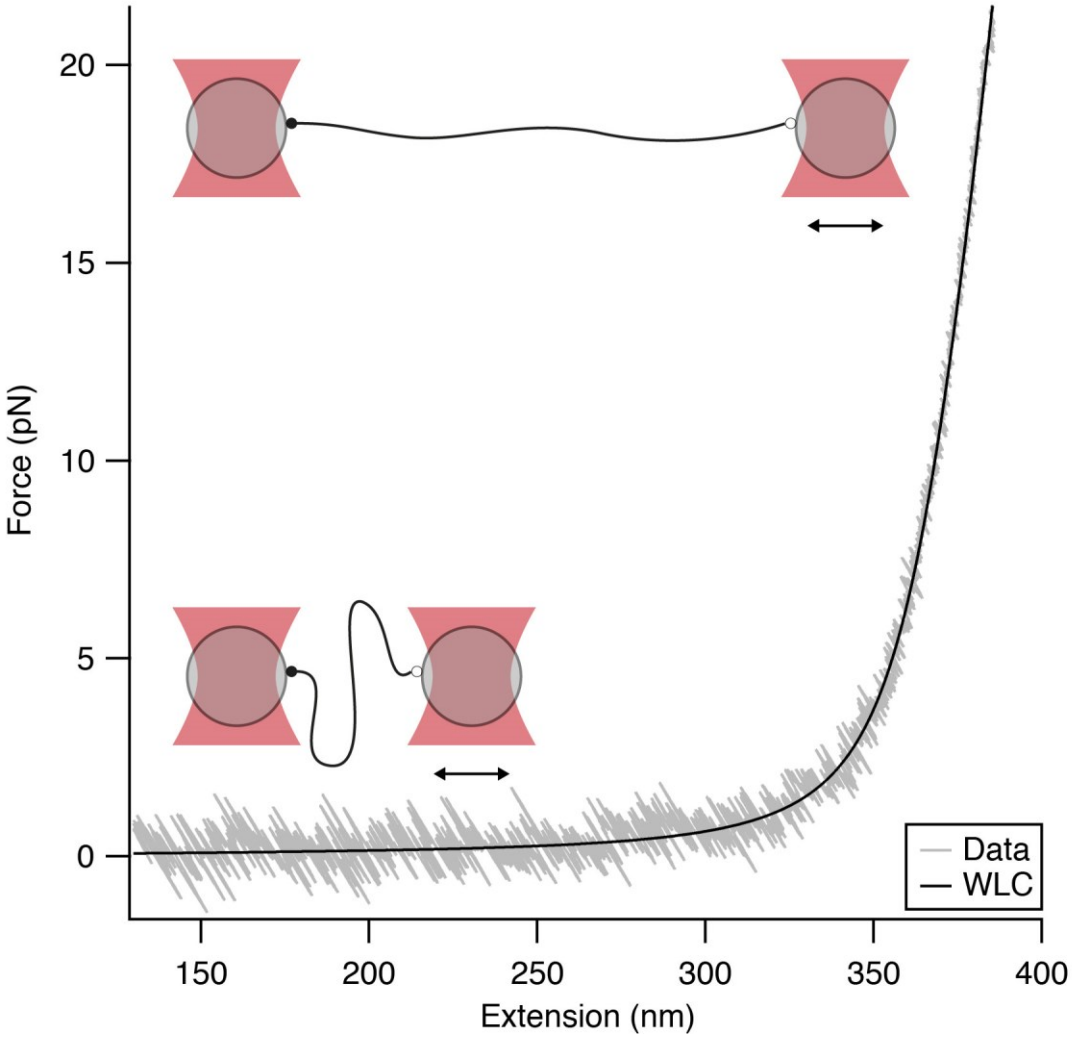

Figure 12: Force-extension curve. DNA is tethered between beads in the optical tweezers set-up and stretched. The force response (gray) is mainly driven by entropy and can be described with the extensible worm-like chain model (black). At low extensions the DNA is flexible and can adopt a wide range of conformations. During the stretching, the force is a consequence of the reduced number of available

conformations at the reduced end-to-end distance. This leads to a high amount of force noise for small forces and the force noise decreases for increasing force values. This correlation will be used in chapter III.1 to identify miscalibration effects.

During the stretching of a dumbbell, which contains a chimeric molecule (DNA-protein-DNA), the force response is dominated by the DNA until the protein unfolds (Fig. 13 A). Then, the contribution of the unfolded polypeptide chain can be described with a WLC model in series with the DNA handles' response (12, 127, 128).

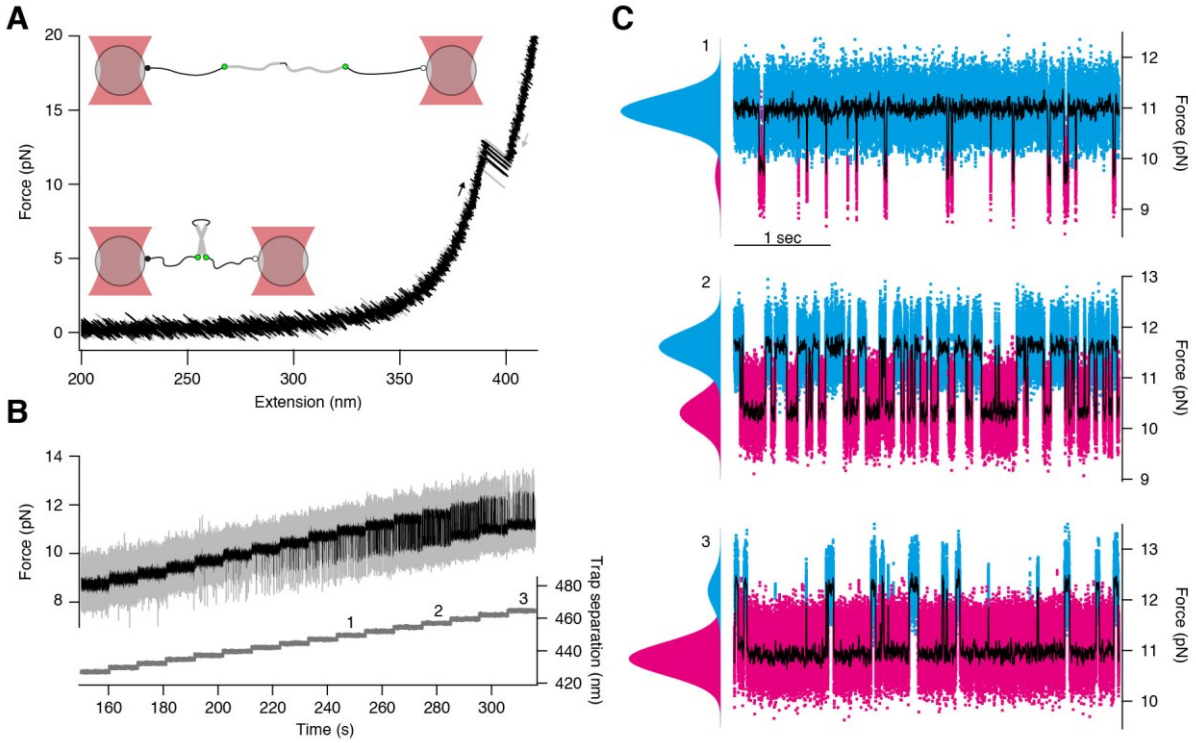

Figure 13: Unfolding pattern of a single protein. (A) Force-extension curve (100 nm/s) of a single protein tethered between the trapped beads in the optical tweezers set-up. During the pulling (black line) and the relaxing (gray line) transitions between the folded protein and the unfolded peptide-chain can be observed at 10-15 pN. (B) The protein was held at a constant trap separation for 10 seconds and the trap separation was increased in steps of 2.5 nm. At low tensions, the protein stays folded, whereas at higher tensions (marked with 1, 2 and 3) the protein spends more and more time in the unfolded state. Full bandwidth data: gray, smoothed data: black. (C) Zoom of the individual constant distance traces (1 - 2 - 3). Hidden Markov modeling identified the folded state (cyan) and the unfolded state (magenta). The observed transitions reveal the folding kinetics. Due to the increasing tether tension, the shift in equilibrium of the transition can be used to determine the folding energy of the protein.

Besides actively pulling on the tether, the traps can be held at a constant trap separation for a given time (Fig. 13 B,C). These force-time traces show the equilibrium fluctuations of the protein under tension

(128, 132) and can be analyzed with hidden Markov modeling (137). The analysis of the force-dependent state probabilities $P_i(F_i)$ using a global fit estimates the energy of the folded protein:

$$P_i(F_i) = \left(1 + \sum_{j \neq i} exp\left(\frac{-\Delta G_{ij}^{0} - \Delta G_{ij}^{sys}(F_i, F_j)}{k_B T} \right) \right), \quad (3)$$

where ΔG_{ij}^{sys} is the energy stored in the bead deflection (assuming a Hookean spring model), the stretching of the DNA handles (integral over the eWLC model) and the stretching of the unfolded polypeptide (integral over the WLC models). Therefore, ΔG_{ij}^{sys} can be determined based on the recorded data (132), whereas the equilibrium energy ΔG_{ij}^{0} can be determined by fitting (138). Furthermore, force-dependent rate models and contour length transformations can be further considered as described in (127). Accordingly, the energy barrier heights $\Delta G^{\dagger}$ can be determined using:

$$\frac{\Delta G^{\dagger}}{k_B T} = \ln\left(\frac{k_0}{A}\right), \qquad A = 1.2 * 10^4 \, s^{-1} \quad (4)$$

with k_0 as the extrapolated zero-force folding rate. The Arrhenius-factor A gives the frequency of folding attempts of the unfolded polypeptide chain as estimated by Gebhardt et al. (132). With these formula the energy landscape of the SMC coiled coil was estimated in chapter III.2.

II.4 Calibration of an optical tweezers set-up

In optical tweezers experiments, the raw data is initially detected as voltage signals, which are converted into forces and distances by a set of calibration parameters (139). For the distance conversion factor, the piezo-based movement of a trapped bead can be compared with the distance in the bright-field. The force detection is based on the measurement of bead deflections with a quadrant-photodiode. Hence, the voltage signal is first converted with the sensitivity (nanometer per volt) into distances and further with the trap stiffness (piconewton per nanometer) into actual forces. For the force conversion factors, the power spectral density of a diffusing bead in the trap potential can be considered (140).

In addition, correction factors for the force detection account for the crosstalk between the lasers (141). To estimate the amount of this effect, one bead is needed to be trapped and displaced, while the other trap is empty (Fig. 14). If the displacement is visible on the other detector, the relative crosstalk must be determined and later subtracted from the raw data. A displacement can be achieved by applying a flow in the microfluidic system of the utilized Lumicks C-Trap. The same crosstalk effect acts additionally on the power spectral density calibration. Hence, a bead pair must be calibrated before and after the release of one bead and the comparison determines the influence of the neighboring bead on the calibration parameters. Both crosstalk measurements must be repeated for the second trap.

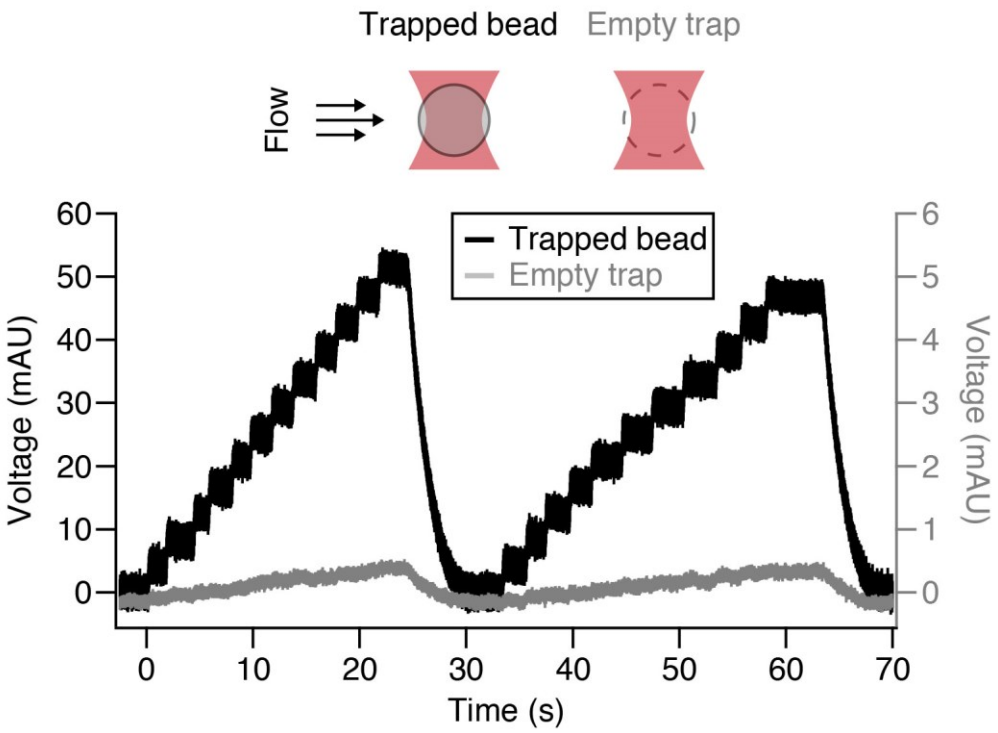

Figure 14: Estimating the crosstalk using microfluidics. Crosstalk between the lasers distorts accurate optical tweezers measurements and should be corrected. Therefore, a bead is caught in one trap and displaced with flow in the microfluidics, whereas the other trap remains empty. The deflection of the trapped bead is measured as a voltage signal (black line) and the change in signal of the other trap (gray line, note that the right axis is ten times zoomed in for visualization) is monitored. The ratio between both signals is the crosstalk from the first trap to the second trap and likely results from a depolarization of the first trap in the optical path. Typical values were below 5%.

Since optical tweezers are considered to be precise and accurate single-molecule techniques, miscalibration effects can introduce deviations in the highly processed data sets, which can manifest into incorrect conclusions and statements (13). Sources of error can be thermal drift, external noise or incorrect calibration or conversion factors. Additionally, the optical trap potentials are modeled as Hookean springs (137), which can affect the analysis of the data at extreme forces (non-linearity effects). Chapter III.1 will focus on these issues.

II.5 Mass photometry

The development of mass photometry in 2018 provided a novel and label-free technique to measure the mass of single-molecules in solution (142, 143). The samples are placed on a coverslip and illuminated by a powerful laser through a microscope objective (Fig. 15). The laser light is refracted at the coverslip and additionally back-scattered by the molecules in solution. The resulting low-intensity interference pattern is detected with a camera.

The refractive index of biomolecules varies only minimally. Therefore, the contrast is closely correlated with the mass of the biomolecules and largely unaffected by the shape or conformation of the molecule (143). Hence, mass photometry estimates the mass of a molecule with high accuracy and was utilized to analyze ligand binding, filament assemblies or the integrity of macromolecular complexes (59, 143). In chapter III.3 the oligomeric state of different SMC head domain constructs is analyzed with mass photometry.

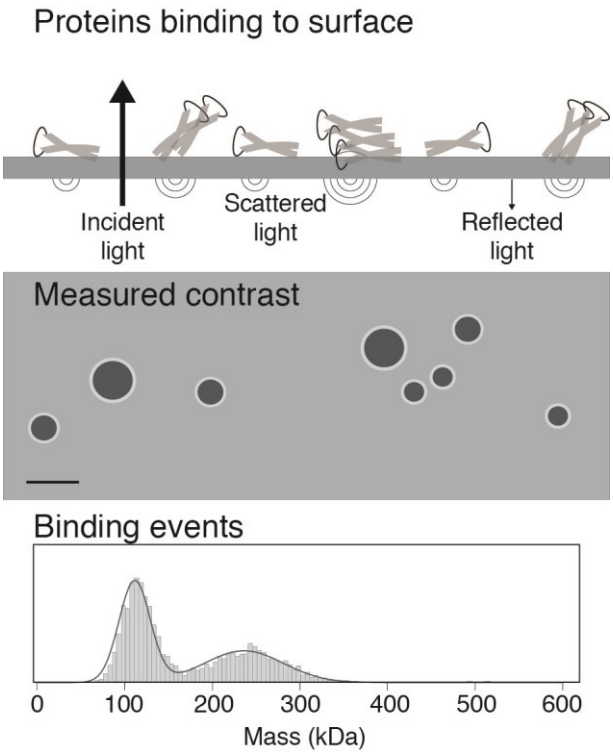

Figure 15: Principle of mass photometry. Top: Proteins in solution immobilize at the glass surface and scatter incident laser light. Middle: The differential interferometric scattering pattern of the proteins is measured for a given time (~ 1 minute). Dots represent individual proteins and vary in contrast for different oligomeric states. Bottom: Histogram of a typical measurement, which was calibrated to convert observed contrast values into actual masses. This graphic was kindly provided by Sarah Zernia (unpublished).

III Results

III.1 Identification and correction of miscalibration artifacts based on force noise for optical tweezers experiments

Freitag, M., D. Kamp, M. Synakewicz, and J. Stigler. 2021. Identification and correction of miscalibration artifacts based on force noise for optical tweezers experiments. *J. Chem. Phys.* 155:175101. (13)

Summary:

This project was initiated following the data collection of Synakewicz et al. (144), which showed discrepancies between the data sets of the same construct recorded over a large time span in a custom-built optical tweezers set-up. Within a measurement cycle the data was sound, whereas between measurement cycles the exact linker parameters were not matching, while similar folding features could be observed. Upon further investigation, it was discovered that in between measurement cycles the optical tweezers set-up underwent different hardware modifications including recalibrations of the instrument.

After estimating the effect of apparent calibration parameters (detector sensitivity and trap stiffness) on linker parameters, the deflection noise was identified as an indicator for calibration errors. Modeling of the filter cascade including filtering and subsampling, which is applied on real experimental data sets, enabled the calculation of theoretical expected noise levels. By matching measured to expected noise levels, correction factors could be estimated, which reliably removed the miscalibration originated from hardware modifications of the experimental data sets. In addition, the effects of non-harmonic trap potentials could be modeled and removed in an analogous way.

Based on our results, miscalibration and non-linearities can be identified and removed in force-distance data generated with optical tweezers, which could improve the reproducibility of these single-molecule experiments and the comparability between different data sets and different instruments.

Author contributions:

I created the bridged DNA construct and measured this construct over a wide force range in the optical tweezers set-up. I created a procedure to evaluate the full trapping potential of a single trap. I was involved in discussions about the data and writing of the paper. I contributed significantly to the revision process.

The Journal of Chemical Physics ARTICLE scitation.org/journal/jcp

Identification and correction of miscalibration artifacts based on force noise for optical tweezers experiments

Cite as: J. Chem. Phys. 155, 175101 (2021); doi: 10.1063/5.0063690
Submitted: 16 July 2021 • Accepted: 15 October 2021 •
Published Online: 1 November 2021

Marvin Freitag,[1] Dieter Kamp,[1] Marie Synakewicz,[2,a)] and Johannes Stigler[1,b)]

AFFILIATIONS

[1] Gene Center, Ludwig-Maximilians-University, Munich 81377, Germany
[2] Department of Pharmacology, University of Cambridge, Cambridge CB2 1PD, United Kingdom

[a)] **Present address:** Department of Biochemistry, University Zürich, 8057 Zürich, Switzerland.
[b)] **Author to whom correspondence should be addressed:** stigler@genzentrum.lmu.de

ABSTRACT

Single-molecule force spectroscopy using optical tweezers continues to provide detailed insights into the behavior of nanoscale systems. Obtaining precise measurements of their mechanical properties is highly dependent on accurate instrument calibration. Therefore, instrumental drift or inaccurate calibration may prevent reaching an accuracy at the theoretical limit and may lead to incorrect conclusions. Commonly encountered sources of error include inaccuracies in the detector sensitivity and trap stiffness and neglecting the non-harmonicity of an optical trap at higher forces. Here, we first quantify the impact of these artifacts on force-extension data and find that a small deviation of the calibration parameters can already have a significant downstream effect. We then develop a method to identify and remove said artifacts based on differences in the theoretical and measured noise of bead fluctuations. By applying our procedure to both simulated and experimental data, we can show how effects due to miscalibration and trap non-linearities can be successfully removed. Most importantly, this correction can be performed post-measurement and could be adapted for data acquired using any force spectroscopy technique.

I. INTRODUCTION

Force spectroscopy by optical tweezers has provided detailed insights into the mechanics and thermodynamics of nanoscale biological systems. Its high stability and precision have enabled the deciphering of energetic folding models in the folding of proteins and nucleic acids[1–7] and in concurrent binding of ligands,[8,9] as well as the observation of conformational changes in enzymes[10] and the stepping of molecular motors[11–13] or of enzymatic complexes.[14,15] In a typical optical tweezers experiment, the force signal is determined from the bead displacements from the trap centers. The detectors record the bead deflection from the traps and output a signal in units of voltage that is then translated into units of length. To obtain quantitative forces from the bead deflection, the trap stiffness has to be accurately calibrated. Typically, the two necessary parameters, i.e., the detector sensitivity β (in units of nm/V) and the trap stiffness k (in units of pN/nm), are determined using the equipartition theorem by measuring the thermal noise of bead fluctuations.[16,17] When performed carefully, the calibration procedure allows for a determination of β and k with errors of <1%.[17] In practice, however, the estimated accuracy of the calibration is often much worse and can exceed ≈10%. Possible sources for these larger uncertainties are diverse but mostly arise from experimental limitations, e.g., by introducing statistical errors due to a finite acquisition time of the free bead motion during calibration or systematic errors due to wrongly assumed values for temperature, viscosity, or bead size. In a dual-bead assay, further complications arise from beam depolarization, causing crosstalk between the two traps, which is typically removed using an additional calibration step.[18] If these crosstalk parameters are not determined accurately or change over time (e.g., due to instrumental drift), their error propagates into all subsequent measurements. Finally, it is possible, in practice, that

The Journal of Chemical Physics ARTICLE scitation.org/journal/jcp

calibration data are not available for each individual bead pair or that the user must rely on a previously determined average.

Here, we show that miscalibration of β and k can, furthermore, cause significant misestimation of the tether persistence length or the stretch modulus by more than 100%, which, consequently, may complicate downstream analyses. We derive a framework for calculating the expected deflection noise from equilibrium force-extension curves (FECs) and use this information to identify, quantify, and reverse miscalibration artifacts. Furthermore, we introduce a force and noise model for non-harmonic trap potentials to account for trap softening at higher forces. Finally, we test our method by applying it to simulated and experimental datasets of double stranded DNA (dsDNA) and proteins coupled to dsDNA tethers.

II. RESULTS

In a dual-trap optical tweezers setup [the inset of Fig. 1(a); also see Fig. S1], the signal x_V, representing the deflection of a bead away from the trap center, is recorded for each trap in units of voltage. The deflection is then converted to a deflection in units of length, $x = x_V\beta$, using the calibration factor β. After this, the signal is converted to a force using the trap stiffness, k, and $F = kx$. In the case of miscalibration, the true calibration factors β and k are replaced by apparent calibration factors $\beta^{app} = \beta\beta^{\dagger}$ and $k^{app} = kk^{\dagger}$, respectively. Here, $\beta^{\dagger}$ and $k^{\dagger}$ indicate how much the calibration factors deviate from the perfectly calibrated case, in which $\beta^{\dagger} = k^{\dagger} = 1$.

In the following derivations, we assume that a tethered biomolecule undergoing conformational transitions is being stretched at quasi-equilibrium, i.e., the trap distance, d, is changed at speeds slower than the kinetic timescale of the system such that all components can equilibrate at any given trap distance. In particular, this requires that the pulling must be slower than the slowest timescale (e.g., folding/unfolding transitions) in the molecule of interest. For a given trap distance, d_i, we, therefore, define the average force as $F = \langle F \rangle_{d_i}$ and the corresponding deflection noise as $\sigma = \sqrt{Var(x)_{d_i}}$. To unambiguously determine the trap distance d, defined here such that $d = 0$ when the bead surfaces touch, we employ a correlation-based strategy outlined in Fig. S1.

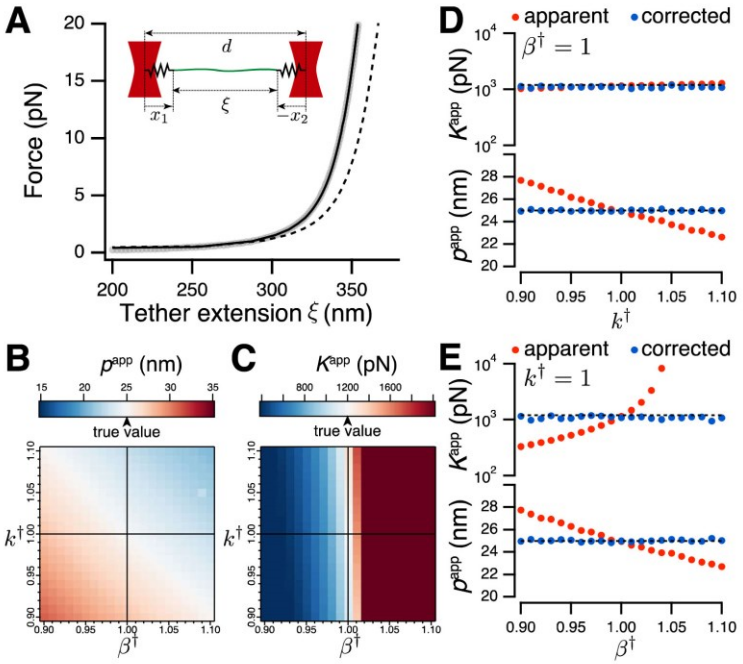

FIG. 1. Effects of miscalibration using simulated data. (a) Apparent force-extension curve (FEC) of a single dsDNA tether, obtained in the presence of miscalibration (gray points), fitted to an eWLC model (black line) and true FEC for comparison (dashed line). Inset: relevant lengths in a two-bead dumbbell assay (also see Fig. S1). [(b) and (c)] Apparent persistence length, p^{app}, and stretch modulus, K^{app}, obtained from miscalibrated data. [(d) and (e)] Apparent stretch modulus and persistence length obtained from FECs when miscalibration is present in either (d) the spring constant (and $\beta^{\dagger} = 1$) or (e) the sensitivity (and $k^{\dagger} = 1$). The red and blue points correspond to values before and after noise-based correction of miscalibration, respectively.

The Journal of Chemical Physics ARTICLE scitation.org/journal/jcp

A. Effects of miscalibration on tether parameters

We first simulated force-extension curves (FECs) of a dsDNA tether (L = 360 nm, p = 25 nm, and K = 1200 pN), modeled as an extensible worm-like chain (eWLC)[19] in the presence of miscalibration (see Sec. IV for further details on simulation). The tether extension ξ is given by the difference of the trap distance d and both bead deflections $x_1, -x_2$ [Fig. 1(a), inset]. A representative apparent force-extension curve [FEC, gray points in Fig. 1(a)] shows slight deviations from the expected behavior (dashed line). However, the shape of the apparent FEC still resembles the true FEC, and it is still well-fitted by using an eWLC model [black line in Fig. 1(a)]. As such, FECs that result from a miscalibrated trapped bead are almost indistinguishable from non-miscalibrated FECs, and inaccurate calibration parameters are hard to identify.

We then systematically varied the miscalibration factors $\beta^\dagger$ and $k^\dagger$ to determine their effect on the extracted eWLC fit parameters. In the case of no miscalibration ($\beta^\dagger = k^\dagger = 1$), fits to the apparent FEC reproduced the true, initial parameters [Figs. 1(b) and 1(c)]. An underestimation of the trap stiffness ($k^\dagger < 1$) generally led to an overestimation of the persistence length and a slight underestimation of the stretch modulus of the tether [Fig. 1(d)]. The effects were more dramatic when the sensitivity was not accurately determined ($\beta^\dagger \neq 1$): whereas inaccuracy of sensitivity had a similar effect on the persistence length as that of the trap stiffness, only a 10% error in the sensitivity led to misestimation of the stretch modulus by more than 100% [Fig. 1(e)].

B. Identifying miscalibration in force data

While it is almost impossible to determine inaccuracies in calibration from fits of eWLC models to FECs alone, we reasoned that they might become apparent when the noise of the bead deflection signal, $\sigma = \sqrt{Var(x)}$, is considered in addition to the average force. The force noise can be quantified from either passive mode [constant trap separation in distinct steps, Fig. 2(a)] or active mode measurements [active pulling, Fig. 2(b)] and generally decreases with increasing force [Fig. 2(c)]. Since $x^{\mathrm{app}} = x_V\beta \times \beta^\dagger = x \times \beta^\dagger$, the bead deflection will only be affected by errors in sensitivity. In contrast, the accuracy of the force depends on the amount of miscalibration present in both sensitivity and trap stiffness as $F^{\mathrm{app}} = k^{\mathrm{app}}x^{\mathrm{app}} = kx \times \beta^\dagger k^\dagger$. Here, we show that, under perfectly calibrated conditions, the noise level σ can be calculated directly from an FEC $F(\xi) := \langle F \rangle_{d_i}\left(\langle \xi \rangle_{d_i}\right)$, but that there is a discrepancy between the measured and calculated noise levels under miscalibration conditions.

In the absence of miscalibration, the expected deflection noise at infinite bandwidth is

$$\sigma^{\circ}_{\mathrm{calc}} = \sqrt{\frac{k_B T}{k_{\mathrm{eff}}}}, \tag{1}$$

where $k_{\mathrm{eff}} = k_c + k_L$ is the effective tether stiffness of the system, $k_c = (1/k_1 + 1/k_2)^{-1}$ is the combined stiffness of the two traps, and $k_L = \frac{\partial F(\xi)}{\partial \xi}\Big|_{\xi_{\mathrm{eq}}}$ is the tether spring constant (see the supplementary material for further details).

We performed Langevin dynamics simulations of FECs to show that the expected noise, σ_{calc} [Eq. (1)], is equal to the measured noise σ_{meas} in the absence of miscalibration [red line and gray points in Fig. 2(d); see Sec. IV for details]. However, in the presence of miscalibration, we observed significant differences between the expected noise and the measured noise [red line and gray points in Fig. 2(e)]. Even small miscalibration artifacts are often readily identifiable in the noise, and therefore, we can use the coincidence of σ_{calc} and σ_{meas} as an indicator for miscalibration in force-extension data.

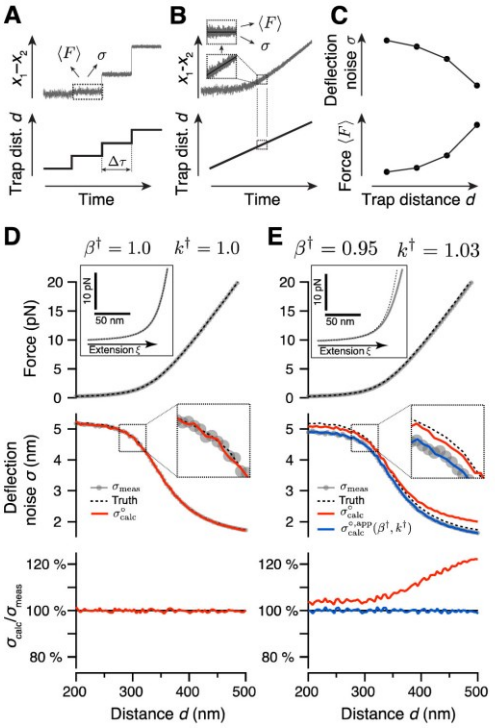

FIG. 2. Identifying calibration errors in simulated data. (a) Acquisition of the mean force $F = \langle F \rangle$ and the deflection noise σ from passive mode data. (b) Acquisition of the mean force and deflection noise from ramp experiments after eliminating the slope for each FEC bin. (c) Determining sufficient distance-depending $\langle F \rangle_{d_i}$ and σ_{d_i} values required for the correction procedure. (d) Perfectly calibrated scenario ($\beta^\dagger = k^\dagger = 1$) with no deviation from the calculated and measured noise. Top panel: simulated force-distance curve (FDC). Inset: corresponding FEC. Middle panel: measured noise σ_{meas} (gray data points) and expected noise calculated according to Eq. (1) (red lines). Bottom panel: ratio of calculated and measured noise. (e) Effects of miscalibration ($\beta^\dagger = 0.95$ and $k^\dagger = 1.03$). Top panel: simulated force-distance curve (FDC). Inset: corresponding FEC. Middle panel: measured noise σ_{meas} (gray data points), expected noise based on the apparent FEC calculated according to Eq. (1) (red lines), and calculated noise according to Eq. (10) taken miscalibration into account. Bottom panel: ratio of calculated and measured noise for the apparent and corrected noise. In all panels, the dashed lines indicate the true quantities (i.e., input parameters).

The Journal of Chemical Physics ARTICLE scitation.org/journal/jcp

C. A compliance model for force noise

Having recognized that the difference of the experimental and calculated noise levels allows us to identify the presence of miscalibration, we next set out to devise a theoretical framework for calculating the expected noise in a realistic two-bead system and how it varies in the presence of miscalibration.

To this end, we considered the continuous-time Langevin equation of the bead deflections $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ in a bead–tether–bead system [see the inset of Fig. 1(a)],

$$\boldsymbol{\epsilon}(t) = -\boldsymbol{\mu}^{-1}\dot{\boldsymbol{x}} - \boldsymbol{\kappa}\boldsymbol{x}, \tag{2}$$

where

$$\boldsymbol{\mu} = \begin{pmatrix} 1/\gamma & 1/\Gamma \\ 1/\Gamma & 1/\gamma \end{pmatrix} \quad \text{and} \quad \boldsymbol{\kappa} = \begin{pmatrix} k_1 + k_L & -k_L \\ -k_L & k_2 + k_L \end{pmatrix} \tag{3}$$

describe the friction interactions of the beads and the spring constants in the system, respectively.[20] Here, k_1 and k_2 are the Hookean spring constants of the two traps, k_L is the stiffness of the tether at extension ξ, γ is the Stokes friction coefficient of a bead, and Γ describes the hydrodynamic coupling between the beads (see the supplementary material).

Thermal fluctuations are white and, with means and covariances, can be determined by

$$\langle \boldsymbol{\epsilon}(t) \rangle = 0, \qquad \langle \boldsymbol{\epsilon}(t) \otimes \boldsymbol{\epsilon}(t') \rangle = 2k_B T \boldsymbol{\mu}^{-1} \delta(t - t'), \tag{4}$$

where "⊗" is the outer product. This system of equations can be solved to produce the power spectral density (PSD) of the combined bead deflection,[20] of which a detailed derivation can be found in Ref. 21. In brief, we diagonalize the matrix $\boldsymbol{\mu\kappa}$ such that $\boldsymbol{A}^{-1}\boldsymbol{\mu\kappa}\boldsymbol{A} = \boldsymbol{\lambda}$, where $\boldsymbol{\lambda}$ is a diagonal matrix of the eigenvalues of $\boldsymbol{\mu\kappa}$.[21] The transformation matrix $\boldsymbol{A}$ can be normalized such that $\boldsymbol{v}\boldsymbol{A}\boldsymbol{A}^t = \boldsymbol{\mu}$, and hence, the covariance matrix of the experimental coordinates becomes

$$\langle \boldsymbol{x}(t) \otimes \boldsymbol{x}(t') \rangle = k_B T \boldsymbol{A} \boldsymbol{\lambda}^{-1} \exp\left(-\boldsymbol{\lambda}|t - t'|\right) \boldsymbol{A}^t, \tag{5}$$

which can be used to derive the PSD matrix

$$\boldsymbol{P}_{\boldsymbol{x}}(\nu) = 2k_B T \boldsymbol{A}\left(\boldsymbol{\lambda}^2 + (2\pi\nu)^2 \mathbb{1}\right)^{-1} \boldsymbol{A}^t. \tag{6}$$

The PSD along the experimental bead deflection coordinate $x_1 - x_2$ is

$$P_{x_1 - x_2}(\nu) = 2k_B T \boldsymbol{\eta}^t \boldsymbol{A}\left(\boldsymbol{\lambda}^2 + (2\pi\nu)^2 \mathbb{1}\right)^{-1} \boldsymbol{A}^t \boldsymbol{\eta}, \tag{7}$$

with $\boldsymbol{\eta}^t = (1, -1)$ in the absence of miscalibration. In the more general case, in which both traps may be miscalibrated by $\beta_1^\dagger$ and $\beta_2^\dagger$, $\boldsymbol{\eta}^t = \left(\beta_1^\dagger, -\beta_2^\dagger\right)$. An explicit form of the experimental PSD can be found in Eq. (S5). Finally, the noise of the combined bead deflection is

$$\sigma_{\text{calc}} = \sqrt{\int_{-\infty}^{\infty} P_{x_1 - x_2}(\nu)\, \mathrm{d}\nu} = \sqrt{2\int_0^{\infty} P_{x_1 - x_2}(\nu)\, \mathrm{d}\nu}. \tag{8}$$

If both trap stiffnesses are similar ($k_1 \approx k_2 \approx 2k_c$), the PSD becomes

$$P_{x_1 - x_2}(\nu) \approx \frac{k_B T \gamma}{(k_c + k_L)^2 + \nu^2\gamma^2\pi^2} = \frac{k_B T \gamma}{k_{\text{eff}}^2 + \nu^2\gamma^2\pi^2}, \tag{9}$$

i.e., a Lorentzian, and we recover the noise relation expected from the equipartition theorem, $\sigma_{\text{calc}} = \sqrt{k_B T/(k_c + k_L)} = \sqrt{k_B T/k_{\text{eff}}}$ [cf. Eq. (1)].

D. Filtering and aliasing

Equation (8) describes the noise only at infinite bandwidth, which is experimentally inaccessible. In a real experiment, the expected noise is modified by a filter cascade that depends on the instrumental setup [Fig. 3(a)]. Possible filters include parasitic filtering by using detectors with a heterogeneous frequency response,[16] anti-aliasing filters that are applied before signal sampling,[22] and/or down-sampling before signal storage. The effect of filtering can be accounted for using

$$\sigma_{\text{calc}}^{\text{filt}} = \sqrt{2\int_0^{\infty} P_{x_1 - x_2}(\nu)\, \mathcal{G}(\nu)^2\, \mathrm{d}\nu}, \tag{10}$$

where $\mathcal{G}(\nu)$ is the frequency-dependent filter gain. We list several commonly used models of filters in the supplementary material. The calculated noise $\sigma_{\text{calc}}^{\text{filt}}$ [Eq. (10)] differs from the previously derived approximation for $\sigma_{\text{calc}}^{\circ}$ [Eq. (1)] in that it correctly describes the filtering and sampling-induced noise modifications that are present in all experimental data.

E. Reversal of miscalibration

Miscalibration artifacts can be reversed when $\beta^\dagger$ and $k^\dagger$ are known. Under miscalibration, the apparent deflection is $x^{\text{app}} = x\beta^\dagger$. Similarly, the true force and combined spring constant are replaced by $F^{\text{app}} = F\beta^\dagger k^\dagger$ and $k_c^{\text{app}} = k_c k^\dagger$, respectively. Consequently,

$$k_c = \frac{k_c^{\text{app}}}{k^\dagger} \quad \text{and} \quad k_L = \frac{\partial F}{\partial \xi} = \frac{\partial F^{\text{app}}}{\partial \xi}\frac{1}{\beta^\dagger k^\dagger} = \frac{\partial F^{\text{app}}}{\partial\left(d - \frac{F}{k_c}\right)}\frac{1}{\beta^\dagger k^\dagger} = \frac{\partial F^{\text{app}}}{\partial\left(d - \frac{F^{\text{app}}}{k_c^{\text{app}}\beta^\dagger}\right)}\frac{1}{\beta^\dagger k^\dagger}, \tag{11}$$

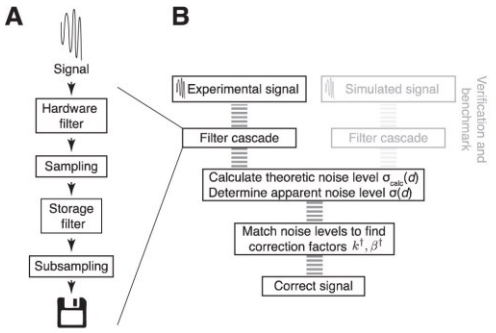

FIG. 3. The data processing of experimental data. (a) Flowchart of the signal acquisition and filtering process in typical experiments. (b) The proposed correction procedure to reverse miscalibration artifacts, including simulations for verification.

The Journal of Chemical Physics ARTICLE scitation.org/journal/jcp

which is only applicable in the case of identical traps, i.e., $k_1 = k_2$, $k_1^\dagger = k_2^\dagger = k^\dagger$, and $\beta_1^\dagger = \beta_2^\dagger = \beta^\dagger$. In the more general case of all these values being different between traps, the miscalibration factors in Eq. (11) become

$$k^\dagger = \frac{k_c^{\mathrm{app}}}{k_c} = \frac{k_1^\dagger/k_1^{\mathrm{app}} + k_2^\dagger/k_2^{\mathrm{app}}}{1/k_1^{\mathrm{app}} + 1/k_2^{\mathrm{app}}}, \tag{12}$$

$$\beta^\dagger = \frac{x^{\mathrm{app}}}{x} = \frac{\beta_1^\dagger k_1^\dagger k_2^{\mathrm{app}} + \beta_2^\dagger k_2^\dagger k_1^{\mathrm{app}}}{k_1^\dagger k_2^{\mathrm{app}} + k_2^\dagger k_1^{\mathrm{app}}}. \tag{13}$$

The apparent noise in the presence of miscalibration $\sigma_{\mathrm{calc}}^{\mathrm{app}}(\beta_1^\dagger, \beta_2^\dagger, k_1^\dagger, k_2^\dagger)$ is then calculated using Eq. (10) and $\boldsymbol{\eta}^t = \left(\beta_1^\dagger, -\beta_2^\dagger\right)$, $k_1 = k_1^{\mathrm{app}}/k_1^\dagger$, and $k_2 = k_2^{\mathrm{app}}/k_2^\dagger$. For example, in the simplified case of identical traps and in the limit of infinite bandwidth, the calculated noise becomes

$$\sigma_{\mathrm{calc}}^{\circ,\mathrm{app}}(\beta^\dagger, k^\dagger) = \beta^\dagger \sqrt{k_{\mathrm{B}}T}\left(\frac{k_c^{\mathrm{app}}}{k^\dagger} + \frac{\partial F^{\mathrm{app}}}{\partial\left(d - \frac{F^{\mathrm{app}}}{k_c^{\mathrm{app}}\beta^\dagger}\right)}\frac{1}{k^\dagger\beta^\dagger}\right)^{-\frac{1}{2}}. \tag{14}$$

Next, we performed simulations to verify that the calculated noise $\sigma_{\mathrm{calc}}^{\mathrm{app}}(\beta_1^\dagger, \beta_2^\dagger, k_1^\dagger, k_2^\dagger)$ in different filtering scenarios, indeed, coincides with the measured apparent noise σ_{meas}. Therefore, we can find the correction factors of miscalibration $\beta^\dagger$ and $k^\dagger$ by matching the measured and calculated noise using

$$\left\{\hat{\beta}^\dagger, \hat{k}^\dagger\right\} = \underset{\beta^\dagger, k^\dagger}{\operatorname{argmin}} \sum\left(\sigma_{\mathrm{calc}}^{\mathrm{app}}(\beta^\dagger, k^\dagger) - \sigma_{\mathrm{meas}}\right)^2, \tag{15}$$

for identical traps, or in the general case,

$$\left\{\hat{\beta}_1^\dagger, \hat{\beta}_2^\dagger, \hat{k}_1^\dagger, \hat{k}_2^\dagger\right\} = \underset{\beta_1^\dagger, \beta_2^\dagger, k_1^\dagger, k_2^\dagger}{\operatorname{argmin}} \sum\left(\sigma_{\mathrm{calc}}^{\mathrm{app}}(\beta_1^\dagger, \beta_2^\dagger, k_1^\dagger, k_2^\dagger) - \sigma_{\mathrm{meas}}\right)^2. \tag{16}$$

A solution to Eqs. (15) and (16) can be found by non-linear least squares.

To verify the effectiveness of the proposed correction procedure [Fig. 3(b)], we simulated FECs with varying values for $\beta^\dagger$ and $k^\dagger$ and applied Eq. (16) to estimate the miscalibration factors $\hat{\beta}^\dagger$ and $\hat{k}^\dagger$. We were then able to calculate the corrected forces and bead deflections $F = F^{\mathrm{app}}/\left(\hat{\beta}^\dagger \hat{k}^\dagger\right)$ and $x = x^{\mathrm{app}}/\hat{\beta}^\dagger$, leading to the recovery of the true FEC [dashed line in Fig. 1(a)] and, therefore, the true eWLC fit parameters [blue points in Figs. 1(d) and 1(e)].

F. Practical notes on accuracy and bias

The post-measurement correction of data is only unbiased when both the experimental noise $\sigma_{\mathrm{meas}}(d)$ and the compliance of the tether $k_L(d)$ are accurate. We tested the data recovery performance using two realistic, but non-ideal experimental scenarios: an FEC and its corresponding noise can be sourced either from discrete steps of the trap distance d [Fig. 2(a)] or from an equilibrium force ramp in which the noise is estimated based on fluctuations of the deflection relative to a fitted baseline [Fig. 2(b)]. In both situations, there is a trade-off between accurately estimating $\sigma_{\mathrm{meas}}(d)$ (requiring few, but long bins) and $k_L(d) = \frac{\partial F}{\partial \xi}$ (usually determined by finite differences, i.e., requiring many short bins). In the scenario of a 360 nm tether, simulations show that $\sigma_{\mathrm{meas}}(d)$ can be calculated to sufficient accuracy if the number of points in a bin is $\gtrsim$300 (i.e., $\gtrsim$10 ms at 30 kHz, Fig. S2). For data with positive compliance, k_L is adequately estimated if an FDC consists of $\gtrsim$50 data points. In practice, a single curve pulled at 500 nm/s is sufficient.

A further potential source of bias arises from variations in bead size. In our derivations above, we assumed that both beads have the same size and, therefore, the same Stokes friction coefficient. Large differences ($\gtrsim$30%) in bead size will, indeed, affect the PSD. However, a typical optical tweezers setup includes a bright-field camera to visualize trapped beads, and hence, unusual bead sizes and shapes are immediately apparent. In contrast, small differences in bead size are likely to remain undetected, and therefore, we performed simulations in which we assumed a variability in the bead diameter of 10%. The uncertainty in $\beta^\dagger$ increased only slightly from 1.8% to 2.0%, whereas the uncertainty in $k^\dagger$ remained unaffected at 2.6%. Hence, small differences in the bead diameter are negligible, and the assumption of representing two beads with the same diameter is met.

We note that the fluctuation noise model only describes the thermodynamic fluctuation noise of the mechanical system itself. Significant external noise (for example, electronic noise or mechanical vibrations) will increase the apparent signal noise, usually in a narrow frequency range, and may lead to erroneous results if unidentified. It is, therefore, still important to inspect the signal PSDs for excess noise.

G. Showcase: Energy estimation of a simulated equilibrium two-state folder

Quasi-equilibrium fluctuations of proteins have previously been used to determine their free energy of folding.[23,24] Here, we simulated the effect of miscalibration artifacts on a protein with an unfolded contour length of 30 nm and a free energy $\Delta G = -16.1\, k_{\mathrm{B}}T$. The protein modeled was connected to a tether of length 360 nm, and the trap distance was increased in 80 steps from 200 to 500 nm. Similar to real experiments, trajectories were 8-pole Bessel filtered at 75 kHz, acquired at 150 kHz, and down-sampled using a 5-point boxcar filter before storage. Figure 4(a) shows FECs with different, random miscalibration of trap stiffness and sensitivity. The

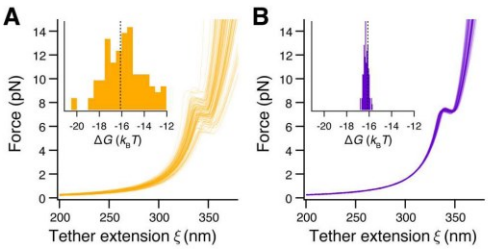

FIG. 4. Correcting simulated traces of an equilibrium two-state folder. (a) Simulated FECs with random miscalibration and different calibration parameters. (b) Traces after miscalibration correction. Insets: transition energies obtained from the trajectories. The dashed lines indicate true values.

The Journal of Chemical Physics ARTICLE scitation.org/journal/jcp

trajectories scatter significantly, and consequently, there are large variations in the obtained free energy. After miscalibration correction, the FECs collapse onto the same curve [Fig. 4(b)] and the precision of free energy estimation increases significantly, effectively removing the substantial artifacts.

H. Data recovery in a non-harmonic trap potential

Until this point, the traps were treated as Hookean springs as this is the simplest approximation mathematically for any attractive potential and is widely used for optical traps. However, optical traps become decidedly non-Hookean at higher forces, especially when measuring with silica beads that have a lower refractive index than polystyrene beads. Therefore, it is necessary to account for deviations from the ideal case of a harmonic trap potential.

In the low-force regime, the trap potential is well-approximated by a constant trap stiffness k,

$$U_{\mathrm{H}}(x|k) = \frac{1}{2}kx^2, \quad k > 0, \tag{17}$$

resulting in the linear dependence of force on the deflection, x. However, a realistic trap potential does not extend to infinity, and consequently, the trap must soften. Therefore, we consider the following descriptive model of an attractive trap potential that has the same stiffness at the equilibrium point but softens as the deflection increases [Figs. 5(a) and S3]:

$$U_{\mathrm{N}}(x|k,w) = -\frac{4kw^2}{\pi^2}\cos\left(\frac{\pi x}{2w}\right), \tag{18}$$

$$F_{\mathrm{N}}(x|k,w) = -\frac{\partial U_{\mathrm{N}}(x)}{\partial x} = -\frac{2kw}{\pi}\sin\left(\frac{\pi x}{2w}\right), \tag{19}$$

where w describes the harmonicity of the potential and k is the stiffness around the equilibrium point $x = 0$. A fully Hookean trap is recovered as $w \to \infty$. In addition to the softening of the trap, for back-focal plane detection, where the trapping laser also detects the bead deflection, also the deflection signal is affected. Hence, when the instrument becomes nonlinear at high force and this is unknown to the experimenter (i.e., they work with a Hookean trap assumption), they will observe a bent FEC, i.e., wrongly interpret it as a softening of the tether [Fig. 5(b)].

The apparent deflection signal is lower than the true deflection to the same extent that a realistic force is lower than a Hooke force,

$$x^{\mathrm{app}}(x) = x\frac{|F_{\mathrm{N}}(x)|}{kx} = \frac{2w}{\pi}\sin\left(\frac{\pi x}{2w}\right). \tag{20}$$

Consequently, the apparent force experienced by beads in a softening trap in a setup with back-focal plane detection,

$$\left|F^{\mathrm{app}}\right| = kx^{\mathrm{app}} = |F|, \tag{21}$$

is, indeed, the true force, and only the bead deflections themselves have to be adjusted.

To correct for non-harmonicity of the traps, we can obtain the true deflection of trap i (with $i = 1, 2$ in a dual-trap assay) by inverting Eq. (19),

$$x_i(F|k_i,w_i) = -\frac{2w_i}{\pi}\arcsin\left(\frac{\pi F}{2k_iw_i}\right), \tag{22}$$

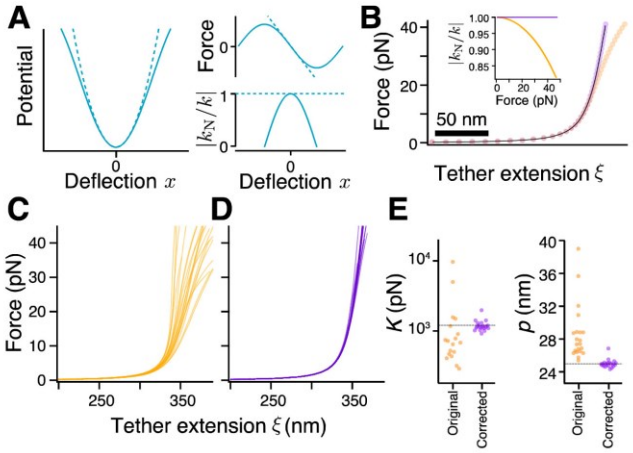

FIG. 5. Correcting effects of non-harmonic trap potentials on simulated data. (a) Comparison of a realistic softening trap potential (continuous line) to the Hookean harmonic approximation (dashed line) and corresponding force and local trap stiffness profiles. See Fig. S3 for an experimental estimate of the trap non-harmonicity. (b) Simulation of an eWLC tether in a non-harmonic trap (orange) and in a Hookean harmonic trap (purple). The continuous black lines correspond to the true parameters. Inset: the trap potential softens as the force increases. (c) Simulated FECs with non-harmonic traps including miscalibrated sensitivity and trap stiffness. The non-harmonicities and miscalibration factors were chosen randomly for each trace. (d) Resulting FECs after noise-based correction. (e) Comparison of the stretch modulus (left) and the persistence length (right) derived from eWLC fits to apparent and corrected traces. The dashed lines indicate the true value.

The Journal of Chemical Physics ARTICLE scitation.org/journal/jcp

and an FEC can thus be corrected using the transformations

$$F \to F^{\mathrm{app}}, \qquad \xi \to d - |x_1| - |x_2|, \tag{23}$$

which, in the more general case of also including inaccuracies in the zero-deflection calibration factors (i.e., $k_i^\dagger \neq 1, \beta_i^\dagger \neq 1$), become

$$F \to \frac{F^{\mathrm{app}}}{\beta^\dagger k^\dagger}, \tag{24}$$

$$\xi \to d - \frac{2w_1}{\pi} \arcsin \psi_1 - \frac{2w_2}{\pi} \arcsin \psi_2. \tag{25}$$

Here, ψ_i of trap i is defined as

$$\psi_i = \frac{\pi F^{\mathrm{app}} k_i^\dagger}{2\beta^\dagger k^\dagger k_i^{\mathrm{app}} w_i}. \tag{26}$$

In summary, an FEC recorded on an instrument with non-harmonic and miscalibrated traps can be corrected when the user has access to the values of $k_i^\dagger$, $\beta_i^\dagger$, and w_i. As described above, the miscalibration parameters can be extracted by matching using the theoretical and experimental noise profiles [Eq. (16)]. Here, we set $k_L = \partial F/\partial \xi$ [Eqs. (24) and (25)], describe the softening of the traps with the transformation

$$k_i \to \left| \frac{\partial}{\partial_x} F_{\mathrm{N}}\left(x_i \middle| k_i^{\mathrm{app}}/k_i^\dagger, w_i\right) \right| = \frac{k_i^{\mathrm{app}}}{k_i^\dagger} \sqrt{1 - \psi_i^2}, \tag{27}$$

and take into account the additional noise reduction due to the softening traps [$\mathrm{Var}(f(x)) \approx (f'(U(x)))^2 \mathrm{Var}(x)$],

$$\beta_i^\dagger \to \left| \frac{\partial}{\partial_x} x_i^{\mathrm{app}}\left(x_i \middle| k_i^{\mathrm{app}}/k_i^\dagger, w_i\right) \beta_i^\dagger \right| = \beta_i^\dagger \sqrt{1 - \psi_i^2}. \tag{28}$$

We verified in simulations that we can efficiently restore true FECs from miscalibrated non-harmonic FECs for a wide range of miscalibration factors and trap non-harmonicities [Figs. 5(c)–5(e)].

I. Application to experimental data

So far, we have reported results based on simulated datasets. To address the performance and limitations of the proposed correction methods, we next applied it to experimental data.

1. Stretching of short DNA handles

We first examined a simple construct consisting of only two short dsDNA handles, each 545 bp in length, that were bridged by a dimerized oligonucleotide (see Sec. IV for details). These tethers were stretched using 1 μm silica beads, and traps set to each have a stiffness of $k_{1,2} \approx 0.3$ pN/nm. One of our first observations was that the experimental FECs often deviated from the expected dsDNA behavior at high force [orange points in Fig. 6(a)]. Our hypothesis that this deviation was caused by a softening of the traps instead of a softening of the DNA tether was confirmed by an analysis of the noise: When we calculated the expected noise from the apparent tether stiffness assuming harmonic traps, we found significant deviations from the experimentally determined noise [Fig. 6(b)]. Next, we used our correction method to identify and account for trap non-harmonicity such that the calculated noise matched the experimental noise [Fig. 6(c)]. The resulting corrected FEC followed the expected eWLC behavior without the artificial curvature at high forces [purple points in Fig. 6(a)]. When applied to data of multiple dsDNA molecules, our correction procedure robustly removed variations at both low and high forces [Figs. 6(d) and 6(e)].

2. Long-term instrumental drift in the force response of a tandem-repeat protein

In a common experimental setup, a biomolecule of interest is tethered between dsDNA handles. Here, we chose to reanalyze published data of CTPRrv5, a tetratricopeptide repeat protein with five helix-turn-helix repeats.[25] FECs were recorded on the same instrument at different times over the course of almost three years. When

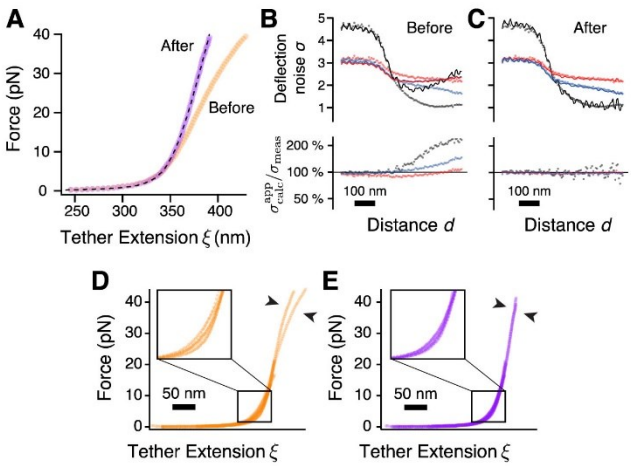

FIG. 6. Correcting experimental data of short DNA handles. (a) Experimental FEC before (orange) and after correction (purple). Dashed line: eWLC fit. [(b) and (c)] Experimental noise (points) and calculated noise from Eq. (10) (lines) for individual traps (red: x_1 and blue: x_2) and the differential signal of both traps (black: $x_1 - x_2$) before and after correction. Upper graph: absolute values. Lower graph: ratio of calculated and experimental noise. (d) Multiple aligned experimental FECs of short dsDNA handles. (e) Same FECs after noise-based correction of miscalibrations and non-harmonicities. The arrows and blow-ups highlight the removal of non-Hookean trap effects.

The Journal of Chemical Physics ARTICLE scitation.org/journal/jcp

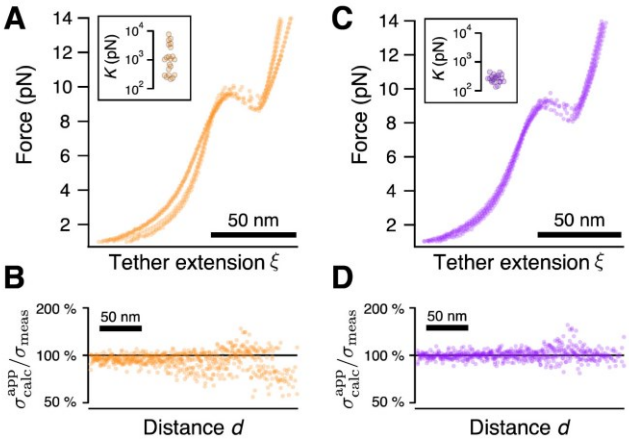

FIG. 7. Correcting long-term instrumental drift in CTPRv5 data. (a) Overlays of experimental FECs recorded before and after instrument reconfiguration. (b) Corresponding ratio of the expected and experimental noise. (c) Overlay of experimental FECs after noise-based correction, including both trap non-harmonicities and miscalibrations. (d) Corresponding ratio of the expected and experimental noise after correction. Insets: fitted stretch moduli of the handles.

overlaying FECs from different time periods, we noticed minor but noticeable differences in their shapes [Fig. 7(a)]. In our previous study, we developed a model to fit equilibrium FEC to extract the mechanical parameters of the dsDNA linker and the energetic parameters of the protein. After fitting this model to our data, we found that the resulting values for the stretch modulus of the linker scattered significantly between different molecules [the inset of Fig. 7(a)]. We verified that the observed heterogeneity was not a consequence of physical modifications to the protein or dsDNA linker using, e.g., mass spectrometry and electrophoresis. Instead, we were able to trace back larger variations in the data to hardware modifications of the instrument between measurement cycles. Therefore, we compared the expected and experimental noise for a molecule [Fig. 7(b)] measured before and after one particular instrument modification and observed that the noise ratio, $\sigma^{\mathrm{app}}_{\mathrm{calc}}/\sigma_{\mathrm{meas}}$, differed significantly. This led us to assume that the physical shape of the trap potential and/or the calibration factors changed due to the instrument reconfiguration. After applying our correction procedure, all FECs superimposed clearly [Fig. 7(c)], the deflection noise fulfilled the theoretical thermodynamics expectations [Fig. 7(d)], and the scatter was removed from the fitted stretch modulus [the inset of Fig. 7(c)].

III. DISCUSSION AND SUMMARY

Force spectroscopy experiments have been used successfully and extensively to characterize the thermodynamics and kinetics of nanoscale biological systems. However, the accuracy and precision of such measurements strongly depend on the correct calculation of parameters that describe all mechanical components present in the system. Particularly in systems in which the energy of a conformational change within the biomolecule is much smaller than the energy stored in the stretched handles, it is crucial to describe the tether properties accurately and precisely. As with any other type of experiment, artifacts can be present in force measurements but are difficult to identify solely from force signals (Fig. 1). Here, we showed that small inaccuracies in instrument calibration lead to large discrepancies in the tether properties. Consequently, as miscalibration artifacts are propagated, they also influence the thermodynamic description of the system under investigation and, hence, may alter down-stream conclusions.

The methodological framework presented here provides researchers with tools to identify and correct these artifacts by comparing and fitting theoretical and experimental noise fluctuations. This procedure is not limited to dual-beam optical tweezers but can also be adapted to other types of nanomechanical experiments, such as atomic force microscopy, acoustic force spectroscopy, and magnetic tweezers, where the relevance of registering noise fluctuations has also been recognized by other groups.[26] Furthermore, it can be used to verify and re-calibrate data post-analysis even if they were previously recorded with uncertain calibration parameters. We would like to point out that analyzing noise fluctuations can also be an integral part of characterizing the system of interest,[26] and hence, our method can be streamlined with other thermodynamic investigations. In summary, we anticipate that our work will aid in the reduction and removal of commonly encountered calibration artifacts and hope that such a method has the potential to enable a world-wide benchmark study of the custom-built and rapidly increasing number of commercial instruments.

IV. METHODS

A. Simulations

Simulations of the tether fluctuations at a trap distance d were performed by integrating the time-discretized Langevin equations of the bead displacements [Eq. (2)] as described previously,[23]

$$\Delta x_i = \frac{\Delta t}{\gamma_i}\left(-k_i x_i + F(\xi) + \sqrt{2k_{\mathrm{B}}T\gamma_i/\Delta t}\,\Gamma_i(t)\right), \tag{29}$$

where $\xi = d - x_1 - x_2$, $\Delta t = 1 \times 10^{-8}$ s, and Γ_i describes the uncorrelated noise with $\langle\Gamma_i(t)\rangle = 0, \langle\Gamma_i(t)\Gamma_i(t')\rangle = \delta(t-t')$. Note that the sign of x_2 is opposite to the description in the main text.

The Journal of Chemical Physics ARTICLE scitation.org/journal/jcp

The force response of the tether was modeled by using an extensible worm-like chain model,[19]

$$F(\xi) = \frac{k_B T}{p}\left(\frac{1}{4(1-\xi/L)^2} - \frac{1}{4} + \frac{\xi}{L} - \frac{F}{K}\right). \quad (30)$$

For simulations of equilibrium transitions, an additional worm-like chain compliance was introduced.[27] Its contour length was determined using an additional Monte Carlo step, and the kinetics were chosen such that transitions occurred at a much faster timescale than that of pulling. Unless indicated otherwise, trajectories were filtered using an 8-pole Bessel filter at 75 kHz, sampled at 150 kHz, subjected to a five-point boxcar filter, and stored after down-sampling to 30 kHz, thereby mimicking typical experimental conditions. Folding free energies of simulated FDCs were then obtained using an equilibrium model as described previously.[23]

B. Experiments

1. Sample preparation

Bridged DNA oligos were generated by linking two 3′-maleimide-modified oligos overnight at 4° C using DTT at an equimolar ratio. Linked oligos were subsequently purified by gel filtration using a S200 increase 10/300 GL size exclusion column (GE Healthcare) equilibrated in 25 mM Tris-HCl pH7.0, 150 mM NaCl. CTPRrv5-oligo chimeras were generated using cysteine–maleimide- or ybbR-CoA-based coupling strategies, which are described in detail elsewhere.[25] Oligo-dimers were hybridized to 545 bp dsDNA handles modified with dual-biotin/dual-digoxigenin modifications (Biomers, custom synthesis), and protein-oligo chimeras were hybridized as described previously.[28] Measurements were performed in 25 mM (short DNA) or 50 mM (CTPRrv5) Tris-HCl pH 7.5, 150 mM NaCl using the glucose-oxidase oxygen scavenging system [0.65% (w/v) glucose (Sigma), 13 U ml^{-1} glucose oxidase (Sigma), and 8500 U ml^{-1} catalase (Calbiochem)].

2. Data acquisition

Data of short dsDNA handles were acquired on a LUMICKS C-trap at a sampling rate of 78.125 kHz. Parasitic filtering due to detection using QPD devices was modeled as described previously.[29] Signals were passed through an on-board anti-aliasing filter (National Instruments) prior to storage, and its gain was modeled according to manufacturer publications.

CTPRrv5 data were acquired on a custom-built optical tweezers instrument, and detection filtering by using QPD devices was modeled again as described.[29] Data were passed through an 8-pole Bessel filter set to 50 kHz, sampled at 100 kHz, and boxcar-filtered down to 20 kHz to save storage space.

In both cases, the correction procedure modeled the respective filter cascade.

SUPPLEMENTARY MATERIAL

See the supplementary material for detailed information on deflection noise, PSDs, hydrodynamic coupling, and commonly used filter models, as well as additional figures about the length convention in a two-bead dumbbell assay, the estimator accuracy, and the extent of non-harmonic trap potentials.

ACKNOWLEDGMENTS

We thank Andreas Weißl and members of the Stigler and Hopfner labs for critical reading of the manuscript. M.S. acknowledges the funding from the BBSRC Doctoral Training Partnerships, Cambridge, and a travel grant provided by the Non-globular Protein Network (COST Action BM1405-39176). J.S. acknowledges the support from the LMU Center for Nanoscience CeNS, a DFG Emmy Noether grant (No. STI673/2-1), and an ERC Starting Grant (No. 758124).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

M.F. and D.K. contributed equally to this work.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

An implementation of the correction procedure (Correction of Tweezers Calibration Factor, CTCF) is available as a Python script on http://github.com/StiglerLab/CTCF.

REFERENCES

[1] M. T. Woodside, W. M. Behnke-Parks, K. Larizadeh, K. Travers, D. Herschlag, and S. M. Block, "Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins," Proc. Natl. Acad. Sci. U. S. A. **103**, 6190–6195 (2006).

[2] K. Neupane, D. A. N. Foster, D. R. Dee, H. Yu, F. Wang, and M. T. Woodside, "Direct observation of transition paths during the folding of proteins and nucleic acids," Science **352**, 239–242 (2016).

[3] A. Mossa, J. M. Huguet, and F. Ritort, "Investigating the thermodynamics of small biosystems with optical tweezers," in *Proceedings of the international conference Frontiers of Quantum and Mesoscopic Thermodynamics FQMT'08* [Phys. E **42**, 666–671 (2010)].

[4] A. Mukhortava, M. Pöge, M. S. Grieb, A. Nivina, C. Loot, D. Mazel, and M. Schlierf, "Structural heterogeneity of attC integron recombination sites revealed by optical tweezers," Nucleic Acids Res. **47**, 1861–1870 (2018).

[5] J. Stigler, F. Ziegler, A. Gieseke, J. C. M. Gebhardt, and M. Rief, "The complex folding network of single calmodulin molecules," Science **334**, 512–516 (2011).

[6] M. Jahn, J. Buchner, T. Hugel, and M. Rief, "Folding and assembly of the large molecular machine Hsp90 studied in single-molecule experiments," Proc. Natl. Acad. Sci. U. S. A. **113**, 1232–1237 (2016).

[7] J. M. Huguet, C. V. Bizarro, N. Forns, S. B. Smith, C. Bustamante, and F. Ritort, "Single-molecule derivation of salt dependent base-pair free energies in DNA," Proc. Natl. Acad. Sci. U. S. A. **107**, 15431–15436 (2010).

[8] D. Bauer, S. Meinhold, R. P. Jakob, J. Stigler, U. Merkel, T. Maier, M. Rief, and G. Žoldák, "A folding nucleus and minimal ATP binding domain of Hsp70 identified by single-molecule force spectroscopy," Proc. Natl. Acad. Sci. U. S. A. **115**, 4666–4671 (2018).

[9] J. Stigler and M. Rief, "Calcium-dependent folding of single calmodulin molecules," Proc. Natl. Acad. Sci. U. S. A. **109**, 17814–17819 (2012).

[10] B. Pelz, G. Žoldák, F. Zeller, M. Zacharias, and M. Rief, "Subnanometre enzyme mechanics probed by single-molecule force spectroscopy," Nat. Commun. **7**, 10848 (2016).

[11]S. C. Kuo and M. P. Sheetz, "Force of single kinesin molecules measured with optical tweezers," Science **260**, 232–234 (1993).
[12]J. T. Finer, R. M. Simmons, and J. A. Spudich, "Single myosin molecule mechanics: Piconewton forces and nanometre steps," Nature **368**, 113–119 (1994).
[13]A. Ramaiya, B. Roy, M. Bugiel, and E. Schäffer, "Kinesin rotates unidirectionally and generates torque while walking on microtubules," Proc. Natl. Acad. Sci. U. S. A. **114**, 10894–10899 (2017).
[14]W. Cheng, S. G. Arunajadai, J. R. Moffitt, I. Tinoco, and C. Bustamante, "Single–base pair unwinding and asynchronous RNA release by the hepatitis C virus NS_3 helicase," Science **333**, 1746–1749 (2011).
[15]M. Righini, A. Lee, C. Cañari-Chumpitaz, T. Lionberger, R. Gabizon, Y. Coello, I. Tinoco, and C. Bustamante, "Full molecular trajectories of RNA polymerase at single base-pair resolution," Proc. Natl. Acad. Sci. U. S. A. **115**, 1286–1291 (2018).
[16]K. Berg-Sørensen and H. Flyvbjerg, "Power spectrum analysis for optical tweezers," Rev. Sci. Instrum. **75**, 594–612 (2004).
[17]S. Tolić-Nørrelykke, E. Schäffer, J. Howard, F. Pavone, F. Jülicher, and H. Flyvbjerg, "Calibration of optical tweezers with positional detection in the back focal plane," Rev. Sci. Instrum. **77**, 103101 (2006).
[18]Y. von Hansen, A. Mehlich, B. Pelz, M. Rief, and R. R. Netz, "Auto- and cross-power spectral analysis of dual trap optical tweezer experiments using Bayesian inference," Rev. Sci. Instrum. **83**, 095116 (2012).
[19]M. D. Wang, H. Yin, R. Landick, J. Gelles, and S. M. Block, "Stretching DNA with optical tweezers," Biophys. J. **72**, 1335–1346 (1997).
[20]J. R. Moffitt, Y. R. Chemla, D. Izhaky, and C. Bustamante, "Differential detection of dual traps improves the spatial resolution of optical tweezers," Proc. Natl. Acad. Sci. U. S. A. **103**, 9006–9011 (2006).
[21]R. Zwanzig, *Nonequilibrium Statistical Mechanics* (Oxford, 2001).
[22]H. Yu, X. Liu, K. Neupane, A. N. Gupta, A. M. Brigley, A. Solanki, I. Sosova, and M. T. Woodside, "Direct observation of multiple misfolding pathways in a single prion protein molecule," Proc. Natl. Acad. Sci. U. S. A. **109**, 5283–5288 (2012).
[23]G. Zoldák, J. Stigler, B. Pelz, H. Li, and M. Rief, "Ultrafast folding kinetics and cooperativity of villin headpiece in single-molecule force spectroscopy," Proc. Natl. Acad. Sci. U. S. A. **110**, 18156–18161 (2013).
[24]P. Ringer, A. Weißl, A.-L. Cost, A. Freikamp, B. Sabass, A. Mehlich, M. Tramier, M. Rief, and C. Grashoff, "Multiplexing molecular tension sensors reveals piconewton force gradient across talin-1," Nat. Methods **14**, 1090–1096 (2017).
[25]M. Synakewicz, R. S. Eapen, A. Perez-Riba, D. Bauer, A. Weissl, G. Fischer, M. Hyvönen, M. Rief, L. S. Itzhaki, and J. Stigler, "Consensus tetratricopeptide repeat proteins are complex superhelical nanosprings," bioRxiv:10.1101/2021.03.27.437344 (2021).
[26]A. Stannard, M. Mora, A. E. M. Beedle, M. Castro-López, S. Board, and S. Garcia-Manyes, "Molecular fluctuations as a ruler of force-induced protein conformations," Nano Lett. **21**, 2953–2961 (2021).
[27]C. Bustamante, J. F. Marko, E. D. Siggia, and S. Smith, "Entropic elasticity of lambda-phage DNA," Science **265**, 1599–1600 (1994).
[28]M. Synakewicz, D. Bauer, M. Rief, and L. S. Itzhaki, "Bioorthogonal protein-DNA conjugation methods for force spectroscopy," Sci. Rep. **9**, 13820–13910 (2019).
[29]K. C. Neuman and S. M. Block, "Optical trapping," Rev. Sci. Instrum. **75**, 2787 (2004).

Supplementary information for: Identification and correction of miscalibration artifacts based on force noise for optical tweezers experiments

Marvin Freitag,[1, a)] Dieter Kamp,[1, a)] Marie Synakewicz,[2] and Johannes Stigler[1, b)]

[1)] *Gene Center, Ludwig-Maximilians-University, 81377 Munich, Germany*

[2)] *Department of Pharmacology, University of Cambridge, Cambridge CB2 1PD, United Kingdom. Present address: Department of Biochemistry, University Zürich, 8057 Zürich, Switzerland*

[a)] These authors contributed equally

[b)] Address correspondence to: stigler@genzentrum.lmu.de

I. DERIVATION OF DEFLECTION NOISE AT INFINITE BANDWIDTH

In the approximation of infinite bandwidth (i.e. without taking into account filtering), the Hamiltonian describing the combined bead deflection $x = x_1 + x_2$ at a given trap distance d is

$$\mathcal{H}(x) = \frac{1}{2}k_c x^2 + \int_0^{d-x} F(\xi)\ \mathrm{d}\xi \tag{S1}$$

where $k_c = \left(\frac{1}{k_1} + \frac{1}{k_2}\right)^{-1}$ is the combined spring constant of both traps. We now Taylor-expand $\mathcal{H}(x)$ around the equilibrium point $x_{\mathrm{eq}} = d - \xi_{\mathrm{eq}}$:

$$\mathcal{H}(x - x_{\mathrm{eq}}) \approx \mathrm{const} + \frac{1}{2}\left(k_c + \left.\frac{\partial F(\xi)}{\partial \xi}\right|_{\xi_{\mathrm{eq}}}\right)(x - x_{\mathrm{eq}})^2 = \mathrm{const} + \frac{1}{2}k_{\mathrm{eff}}(x - x_{\mathrm{eq}})^2 \tag{S2}$$

with $k_{\mathrm{eff}} = k_c + k_L$, and the spring constant of the tether defined as $k_L = \left.\frac{\partial F(\xi)}{\partial \xi}\right|_{\xi_{\mathrm{eq}}}$. Finally, according to the equipartition theorem, the expected noise at full bandwidth is

$$\sigma^{\circ}_{\mathrm{calc}} = \sqrt{\frac{k_{\mathrm{B}}T}{k_{\mathrm{eff}}}}. \tag{S3}$$

II. EXPLICIT FORM OF THE POWER SPECTRAL DENSITY IN THE PRESENCE OF MISCALIBRATION

As described in the main text, the power spectral density (PSD) of the bead deflection signal in the presence of miscalibration $P_{x_1-x_2}(\nu)$ can be calculated from the Langevin equation of the bead fluctuations (eq. (7)). Using $\boldsymbol{\eta}^t = (\beta_1^\dagger, -\beta_2^\dagger)$ it can be written in the form of a rational expression

$$P_{x_1-x_2}(\nu) = \frac{2k_{\mathrm{B}}T\,\gamma\Gamma\ (n_0 + n_2\nu^2)}{d_0 + d_2\nu^2 + d_4\nu^4}, \tag{S4}$$

with coefficients

$$\begin{aligned}
n_0 =& \left(\Gamma^2 - \gamma^2\right)\left(\Gamma\beta_1^{\dagger 2}k_2^2 + 2\gamma\beta_1^\dagger\beta_2^\dagger k_1 k_2 + \Gamma\beta_2^{\dagger 2}k_1^2\right) \\
&+ 2k_L(\beta_1^\dagger - \beta_2^\dagger)(\gamma + \Gamma)(\gamma - \Gamma)^2(\beta_1^\dagger k_2 - \beta_2^\dagger k_1) \\
&+ 2k_L^2(\beta_1^\dagger - \beta_2^\dagger)^2(\gamma + \Gamma)(\gamma - \Gamma)^2 && \text{(S5)} \\
n_2 =& 4\pi^2\gamma^2\Gamma^2\left(\Gamma\left(\beta_1^{\dagger 2} + \beta_2^{\dagger 2}\right) - 2\beta_1^\dagger\beta_2^\dagger\gamma\right) && \text{(S6)} \\
d_0 =& \left(\gamma^2 - \Gamma^2\right)^2\left(k_L(k_1 + k_2) + k_1 k_2\right)^2 && \text{(S7)} \\
d_2 =& 4\pi^2\gamma^2\Gamma^2\left(2\gamma^2 k_1 k_2 + 2k_L(\gamma - \Gamma)^2(k_1 + k_2) + 4k_L^2(\gamma - \Gamma)^2 + \Gamma^2 k_1^2 + \Gamma^2 k_2^2\right) && \text{(S8)} \\
d_4 =& 16\pi^4\gamma^4\Gamma^4. && \text{(S9)}
\end{aligned}$$

III. HYDRODYNAMIC COUPLING

A thorough description of friction in an optical tweezers system is provided in ref..[1] In brief, the friction of two beads of radius a suspended in solution at a center-to-center

distance R is frequency (ν)-dependent and parameterized by the inverse penetration depth of vortex diffusion $\alpha = \sqrt{-2\pi i\nu\rho/\eta}$, where ρ and η are the density and shear viscosity of the fluid, respectively. The friction term of the beads themselves is given by the Stokes solution $\gamma_{1,2} = \gamma_0\left(1 + \alpha a + \alpha^2 a^2/9\right)$, where $\gamma_0 = 6\pi\eta a$.[2] Since a closed form solution of the cross-friction does not exists, we employ here the approximation by Saarloos and Mazur[3]

$$\Gamma = \gamma_0 \frac{12\alpha^2 e^{\alpha R} R^3}{3a^4\alpha^5 R^2 + 36a\left(e^{\alpha R} - \alpha R - 1\right) - 4a^3\alpha^2\left(5 - 2e^{\alpha R} + 5\alpha R\right)}. \tag{S10}$$

IV. FILTER AND DATA ACQUISITION MODELS

Alterations to the signal such as sub-sampling and filtering can affect the experimentally determined noise. Generally, signal modifications can occur at various stages of the data acquisition process. During acquisition, signals incur hardware-related filtering, such as parasitic filtering in detectors[4] or anti-aliasing filters.[5] Data are typically acquired at high frequencies, and therefore filtered and down-sampled before storage. Figure 3A shows a typical chain of events.

Although the actual filter properties depend on the particular type of experiment, their effects on the sampled signal can be accurately modeled by a chain of filters. Here we provide explicit equations for the expected noise of several filter models.

A. Subsampling and aliasing

Aliasing effects can occur whenever data are sub-sampled at a lower frequency. The aliased PSD P_{ss} that is sub-sampled at a frequency ν_s is[6]

$$P_{\mathrm{ss}}(\nu) = P(\nu) + \sum_{k=1}^{\infty}\left[P(k\nu_s - \nu) + P(k\nu_s + \nu)\right]. \tag{S11}$$

B. Bessel filter

The gain of a Bessel filter of order N is[7]

$$\mathcal{G}(\nu) = \left|\frac{\theta_N(0)}{\theta_N\left(i\nu/(\nu_{3\mathrm{db}}c)\right)}\right|, \tag{S12}$$

where

$$\theta_N(\nu) = \sum_{k=0}^{N}\frac{(2N-k)!}{2^{N-k}k!(N-k)!}\nu^k \tag{S13}$$

describe reverse Bessel polynomials and c is chosen such that $\mathcal{G}(\nu_{3\mathrm{db}}) = \sqrt{2}/2$.

C. Butterworth filter

The gain of an N-pole Butterworth filter is[8]

$$\mathcal{G}(\nu) = \frac{1}{\sqrt{1 + \left(\frac{\nu}{\nu_{3\mathrm{db}}}\right)^{2N}}}. \tag{S14}$$

D. Boxcar filter

The gain of a boxcar or moving-average filter of width N (i.e. N data points used to re-calculate a single one) of a trajectory with sampling rate ν_s is

$$\mathcal{G}(\nu) = \frac{1}{N}\left|\frac{\sin\left(\pi N \nu/\nu_s\right)}{\sin\left(\pi\nu/\nu_s\right)}\right|. \tag{S15}$$

REFERENCES

[1]Y. von Hansen, A. Mehlich, B. Pelz, M. Rief, and R. R. Netz, "Auto- and cross-power spectral analysis of dual trap optical tweezer experiments using Bayesian inference," Review of Scientific Instruments **83**, 095116 (2012).

[2]G. G. Stokes, "On the effect of internal friction of fluids on the motion of pendulums," Proc. Cambridge Philos. Soc **9** (1851).

[3]W. Van Saarloos, P. Mazur, and 1983, "Many-sphere hydrodynamic interactions: II. Mobilities at finite frequencies," Physica A **120A**, 77–102 (1983).

[4]K. Berg-Sørensen, E. J. G. Peterman, T. Weber, C. Schmidt, and H. Flyvbjerg, "Power spectrum analysis for optical tweezers. II: Laser wavelength dependence of parasitic filtering, and how to achieve high bandwidth," Review of Scientific Instruments **77**, 063106 (2006).

[5]K. Berg-Sørensen and H. Flyvbjerg, "Power spectrum analysis for optical tweezers," Review of Scientific Instruments **75**, 594–612 (2004).

[6]J. W. Kirchner, "Aliasing in 1/fαnoise spectra: Origins, consequences, and remedies," Physical review E **71**, 262–16 (2005).

[7]G. Bianchi and R. Sorrentino, *Electronic Filter Simulation & Design* (McGraw-Hill Education, 2007).

[8]S. Butterworth, "On the Theory of Filter Amplifiers," Experimental Wireless **7**, 536–541 (1930).

V. SUPPLEMENTARY FIGURES

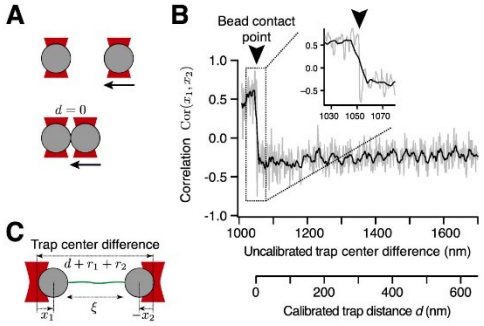

FIG. S1. Procedure to find the trap distance offset. (A) Prior to forming a tether, the beads are slowly approached and the correlation of their deflection signals is monitored. (B) The correlation signal is close to zero (slightly negative due to hydrodynamic coupling) while the beads do not touch and can move independently in their respective traps, but the correlation signal becomes positive as soon as the bead surfaces touch. The distance d is chosen such that d=0 applies at the bead contact point. (C) Relevant lengths (cf. fig. 1A inset) of a two-bead dumbbell assay with non-zero bead radii r_1 and r_2.

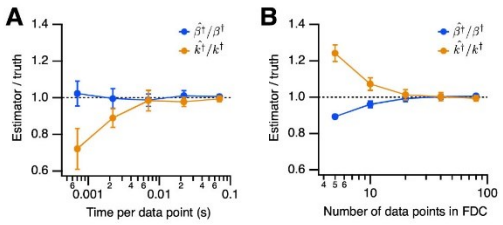

FIG. S2. Estimator accuracy and bias in 30 kHz data. (A) Application of the method for filtered FDCs with random miscalibration of 80 data points with varying time per data point. (B) Application for filtered FDCs with random miscalibration of 70 ms time per data point and varying number of data points. Error bars represent standard deviations.

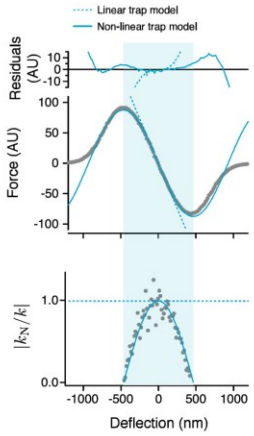

FIG. S3. Comparing models for a non-harmonic trap potential. A bead was immobilized in trap #1 and the apparent force signal was recorded while scanning over the trapped position with trap #2 (gray points). The apparent force signal was then fitted close to the center $x = 0$ to a model describing a harmonic (Hookean) trap (dashed line) and a non-harmonic trap (continuous line, see text). Bottom: The local trap stiffness is well-described by the non-harmonic trap model over the entire shaded area, while the Hookean model shows the expected systematic deviations.

III.1.1 Appendix

In this Appendix, I will present the characterization of samples suitable for optical tweezers experiments using the dumbbell geometry, which was introduced in chapter II.2 and utilized in chapter III.1 and III.2. This chapter represents a trouble-shooting guide for efficient unfolding experiments, which dependent on the reliable finding of single tethers between bead pairs.

After protein purification, an efficient attachment of DNA oligonucleotides (oligos) to the protein of interest (Fig. 11, reviewed in (145)) is essential for unfolding studies using the dumbbell geometry. The success of the attachment can be analyzed by size-exclusion chromatography or SDS-PAGE, which can be found in chapter III.2. Note that some unfavorable attachment strategies (NHS-based coupling or maleimide-based coupling at too high pH-values) can result in more than two oligos attached to the protein of interest, which inhibits efficient optical tweezers experiments by unspecified tether geometries. This effect can be easily detected within the handle hybridization test (Fig. 16).

For this, the oligos attached at the protein of interest (oligo-POI-oligo), can be further hybridized with the single-stranded overhangs of the DNA handles, which are generated by a polymerase chain reaction (PCR). The reverse-primer used for the handle PCR incorporates an abasic site, which causes the single-stranded overhang of the PCR product. The efficiency of the handle hybridization can be determined by mass shifts in an agarose gel (Fig. 16). Therefore, different concentrations of the oligo-POI-oligo chimera are added to a fixed amount of the handle PCR product. The condition, which gives the highest relative shift, will be later used in the optical tweezers experiment to maximize the chance of getting single tethers. The outcome of the handle hybridization test will depend on the sample purity and concentration, as well as the diffusion properties of the components, which can vary for different proteins, other handle lengths or other single-stranded overhangs (146). I recommend to do a hybridization test for each attachment. In Fig. 16, a typical handle hybridization test is shown for the oligo-DTT-oligo construct used in chapter III.1. The concentrations, which result in good conjugation yields, were found to differ by one order of magnitude, so it is sufficient to screen for concentration magnitudes in the handle hybridization test.

The DNA handles include 5’ moieties of either biotin or digoxigenin, which are supposed to bind in the optical tweezers experiment to streptavidin beads or anti-digoxigenin beads, respectively. Normally used handles utilize forward-primers with internal modifications and modification at the 5’ end (127). Besides the advantages of multiple binding sites of these handles, these moieties are 11 nucleotides apart, which can lead to a discrepancy of 7.5 nm assuming possible tethering geometries. To avoid this, I improved the tethering geometry by using forward-primers bearing dual-biotin or dual-digoxigenin, which both contain two moieties at the 5’ end and no further internal modifications. This approach reduced the distance between the moieties to a minimum and ensured well-defined contour lengths of the DNA handles (see chapter III.1).

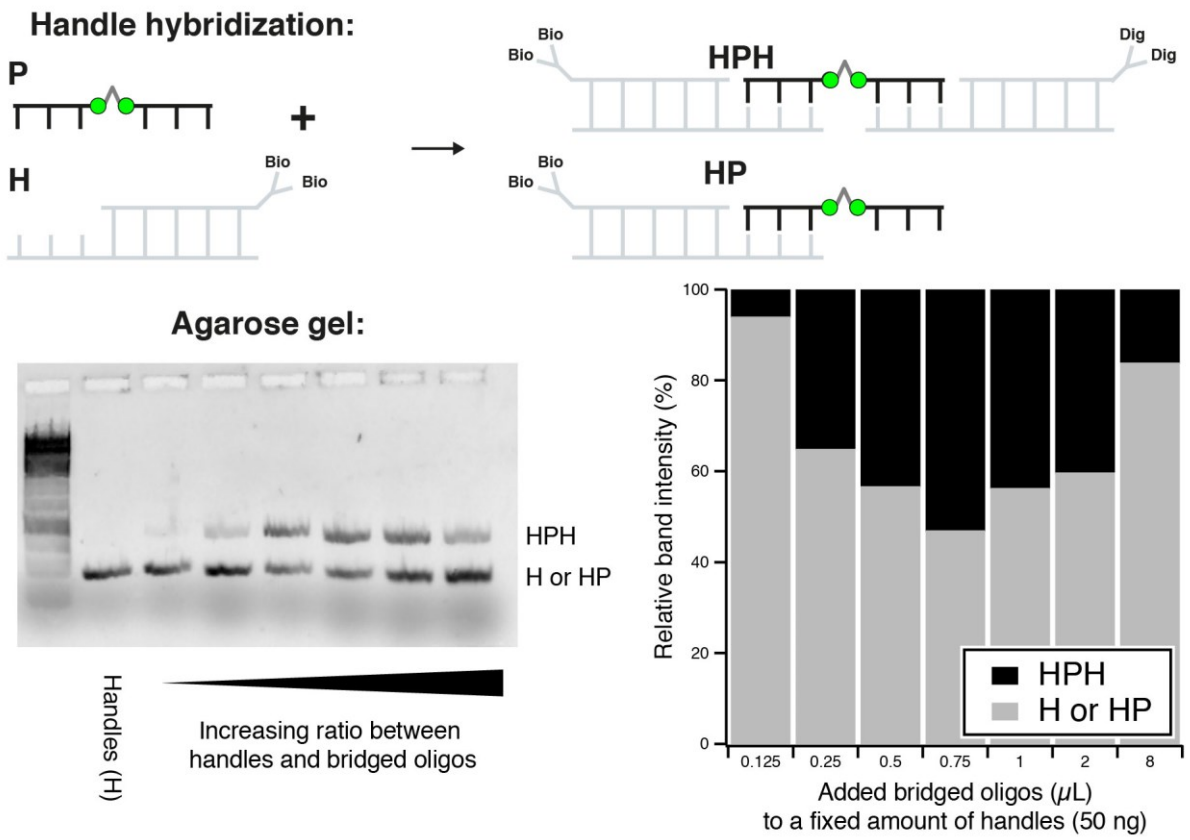

Figure 16: Handle hybridization test. Oligos bearing 3' modification, which are bridged by a bi-functional molecule (P, green dots and gray connection could represent a protein of interest with two engineered attachment-tags or DTT, which contains two thiols) can be hybridized to two handles (H) leading to by-products (HP, H or P) and successful conjugates (HPH). The newly established handles contain 5' dual-biotin (bio) or 5' dual-digoxigenin (dig) modifications. The efficiency of this stochastic labeling approach can be evaluated from relative band intensities in an agarose gel. The observed sweet spot, which allows optimal conjugation yields of the bridged oligos (DTT-bridged 3'-maleimide oligos elute from size-exclusion at 12 mAU at UV_{280}), is roughly one order of magnitude wide for 50 ng of the utilized handles (510 bp).

In the dumbbell geometry, a single tether should be formed between the conjugated handle-protein-handle hybrid (HPH) and the two beads (anti-digoxigenin (A) and streptavidin (S)), while the appearance of multi-tethers should be avoided.

After the ideal ratio between P and H was found (Fig. 16), the established protocol relies further on the pre-incubation of the HPH mixture with one kind of beads (e.g. fluorescent A-beads), which will generate AHPH. The optimal ratio between HPH and A will be influenced by the amount of beads in the stock solution and the surface density of the beads, which can vary between different batches. To avoid multi-tethers, it is rational to choose a ratio between HPH and A for the pre-incubation step, by which most beads will remain empty. Hence, the few bead pairs, which give tethers, will bear only single-tethered molecules and it is unlikely that additional tethers will be picked up during consecutive pulling attempts.

Additionally, a bad yield at the oligo attachment step, insufficient purification of the reaction mixture or a bad yield at the handle hybridization step will affect the optical tweezers measurements negatively by the tethering of by-products. As a second step, the APHP solution will be mixed with the other bead type (S) in a final dilution containing an oxygen scavenger system (147), while the ratio between the beads should be close to a one to one ratio. This solution (APHP + S) is then applied to the passivated and intensively washed measuring chamber of the optical tweezers.

During the measurement, two beads are caught in the two traps of the optical tweezers. The bead type can be distinguished if fluorescent beads or beads of different size are used. Next, a bead calibration (preferable before the measurement) is required and the recording of an undisturbed approach of the beads for the baseline correction, which is relevant for the data analysis. After this, a tether can be picked-up during the repeated approaching and touching of the beads.

The utilized optical tweezers instrument possesses vertically aligned objectives, whereby the measuring chamber is mounted horizontally. Therefore, the silica beads tend to sink down throughout the measurement, which hampers the adjustment of the amount of beads per chamber. The final number of beads should be in the order of one to two beads per field of view, so that during a measurement no further bead or particle will fall into the traps, which will end the measurement.

In summary, unfolding studies in optical tweezers rely on the careful adjustment of several ratios. The efficiency of the measurements is highly dependent on a good yield by the oligo attachment, which eases following steps and enables the frequent formation of single tethers in the assay.

III.2 Single-molecule experiments reveal the elbow as an essential folding guide in SMC coiled-coil arms

Freitag, M., S. Jaklin, F. Padovani, E. Radzichevici, S. Zernia, K.M. Schmoller, and J. Stigler. 2022. Single-molecule experiments reveal the elbow as an essential folding guide in SMC coiled-coil arms. *Biophys J*. 121:4702–4713. (12)

Summary:

In this publication, we measured the un- and refolding pattern of the isolated Psm3 coiled-coil domain using optical tweezers. Two series of crosslinking experiments revealed the amino acid pairing of the coiled coils, which enabled us to match unfolding features to positions in the coiled coil. The coiled coils unfold at comparatively low forces while three obligatory intermediates were observed, which are caused by a stutter, the elbow and an undescribed feature, which we termed “separator”. The truncation of the construct confirmed that SMC arms possess a modular arrangement of flexible segments.

Furthermore, we observed that mutations of the elbow region caused frequently appearing misfolded configurations, which comprised non-native alignments of adjacent segments. Hence, the elbow can be considered as a guide for proper coiled-coil folding. Additionally, *in vivo* studies revealed that alteration of the elbow in budding yeast causes size and growth defective phenotypes, which are indicators for mitosis defects and malfunctioning cohesin complexes. This highlights the crucial role of finely tuned elbow mechanics in driving large-scale conformational changes of the cohesin complex.

Author contributions:

I designed, cloned, purified and coupled all constructs with the help of Sigrun Jaklin in the wet lab. I performed all measurements at the optical tweezers by myself and together with Johannes Stigler discussed and analyzed the data. I wrote the paper with Sarah Zernia, Johannes Stigler and input from all authors.

Biophysical Journal
Article

Single-molecule experiments reveal the elbow as an essential folding guide in SMC coiled-coil arms

Marvin Freitag,[1] Sigrun Jaklin,[1] Francesco Padovani,[2] Ecaterina Radzichevici,[2] Sarah Zernia,[1] Kurt M. Schmoller,[2,3] and Johannes Stigler[1,*]
[1]Gene Center Munich, Ludwig-Maximilians-Universität München, Munich, Germany; [2]Institute of Functional Epigenetics, Helmholtz Zentrum München, Neuherberg, Germany; and [3]German Center for Diabetes Research (DZD), Neuherberg, Germany

ABSTRACT Structural maintenance of chromosome (SMC) complexes form ring-like structures through exceptional elongated coiled-coils (CCs). Recent studies found that variable CC conformations, including open and collapsed forms, which might result from discontinuities in the CC, facilitate the diverse functions of SMCs in DNA organization. However, a detailed description of the SMC CC architecture is still missing. Here, we study the structural composition and mechanical properties of SMC proteins with optical tweezers unfolding experiments using the isolated Psm3 CC as a model system. We find a comparatively unstable protein with three unzipping intermediates, which we could directly assign to CC features by crosslinking experiments and state-of-the-art prediction software. Particularly, the CC elbow is shown to be a flexible, potentially non-structured feature, which divides the CC into sections, induces a pairing shift from one CC strand to the other and could facilitate large-scale conformational changes, most likely via thermal fluctuations of the flanking CC sections. A replacement of the elbow amino acids hinders folding of the consecutive CC region and frequently leads to non-native misalignments, revealing the elbow as a guide for proper folding. Additional in vivo manipulation of the elbow flexibility resulted in impaired cohesin complexes, which directly link the sensitive CC architecture to the biological function of cohesin.

SIGNIFICANCE The detailed understanding of the molecular mechanisms of SMC complexes is important to identify the origin of chromosomal defects. The coiled-coil (CC) domains were recently found to undergo large-scale conformational changes, which enable diverse functions of these proteins. However, detailed structural information of the CCs is still elusive due to their high flexibility. By measuring mechanical responses, we identified thermodynamically relevant features in the CC that structure the CC in segments. One of these features, the elbow, additionally ensures proper alignment of flanking CC sections and can be viewed as a folding guide. Furthermore, in vivo manipulation of the elbow resulted in impaired complex function, showing the relevance of the CC features for SMC mechanics.

INTRODUCTION

Structural maintenance of chromosome (SMC) complexes are essential for the organization and regulation of chromosomes. They comprise a conserved architecture consisting of two SMC subunits forming a heterodimer bridged by a flexible kleisin unit (1) (Fig. 1 *A*). This tripartite ring is able to entrap DNA and form loops, leading to DNA compaction, chromosome segregation, and gene regulation (2). Each SMC protein consist of long anti-parallel coiled-coils (CCs), flanked by a hetero-dimerization domain (hinge) and an ATP-binding domain (head) at opposing ends (3). Recently, electron microscopy (EM) and imaging atomic force microscopy (AFM) studies have revealed that SMC complexes can adopt a wide range of shapes, ranging from open circles (O shaped) to collapsed (B shaped) or half-collapsed conformations (4–6). This started a discussion as to whether these conformational changes are relevant for the loop extrusion mechanism.

Single-molecule studies on the conformation of the SMC complex cohesin during DNA binding have shown that cohesin can bypass small obstacles bound to DNA, but not obstacles larger than 20 nm (7,8). This is surprising as the cohesin ring has a diameter of up to 50 nm (9). In addition, cohesin can be pushed by even smaller motors, such as T7 RNA polymerase or FtsK, while being topologically or pseudo-topologically bound to DNA (7,8). These findings suggest that collapsed conformations are preferred over

Submitted May 6, 2022, and accepted for publication October 12, 2022.
*Correspondence: stigler@genzentrum.lmu.de
Editor: Michael T. Woodside.
https://doi.org/10.1016/j.bpj.2022.10.017

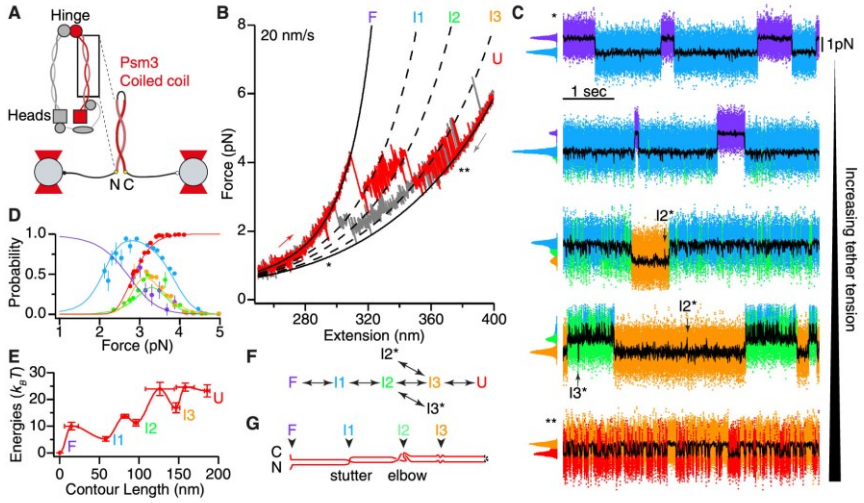

FIGURE 1 Measuring the thermodynamic stability profile of Psm3 CCs with optical tweezers. (*A*) Cartoon representation of the unfolding experiments: a large single-peptide stretch of the anti-parallel CC (D245-T505 and G658-K908) was attached to DNA handles and tethered between optically trapped beads for manipulation. (*B*) A force-extension curve of the WT construct shows a typical unfolding (red curve) and refolding trace (gray curve) containing transitions over the three obligatory intermediates (I1, I2, and I3) to the unfolded protein (U). Asterisks indicate the minimal and maximal distances used for passive mode measurements (see *C*). (C) Passive mode force versus time traces recorded at constant trap distances, demonstrating transition kinetics and the rarely populated, non-productive states I2* and I3*. Color code as in (*B*). With increased tether tension the protein populates higher energetic states. The minimal extension (single asterisk) corresponds to an average applied force of 2.9 pN, the maximal extension (double asterisk) corresponds to an average force of 4.1 pN. (*D*) Global energy fit over the force-dependent state probabilities. Error bars indicate standard deviations and result from bootstrapping of the corresponding passive mode trace (see section "materials and methods"). Color code as in (*B*). (*E*) Interpolated energy landscape consolidating contour lengths and energies of the folding states. The energy barriers are taken from force-dependent transition kinetics (Fig. S3), n = 5 molecules, error bars indicate standard deviations. (*F*) Folding network for the WT construct showing a zipper-like unfolding pattern with two non-productive misfolded states. (*G*) Identification of structural features from folding intermediates, containing the unfolded ends, the stutter, the elbow, and a possible feature located at the third intermediate (Fig. S1). To see this figure in color, go online.

fully circular ones, as also confirmed by recent cryogenic electron microscopy (cryo-EM) structures of holo complexes (10–12). However, a dynamic transition between O and B shapes is likely and was reported for condensin (4,6) and cohesin (13) in recent high-speed AFM studies. Hence, SMC CC domains have to undergo large-scale conformational changes to allow this transition.

Canonical CCs follow a heptad repeat pattern $(abcdefg)_n$, where "a" and "d" denote hydrophobic amino acids that promote pairing of the CC strands by constructing a hydrophobic core (14). CC prediction software (15,16) can be used to analyze this repeat pattern and create CC propensity profiles to identify deviations. It was found that all tested SMC proteins exhibit discontinuities in their CC domains (5,17), which provide flexibility and could enable large-scale conformational changes. Crystallography studies together with crosslink data identified two general discontinuities for bacterial and eukaryotic SMC proteins (5,18), which were named "joint" (17,19) and "elbow"(5). The joint is located near the SMC heads and increases local CC flexibility (11,20,21), while the elbow functions as a turning point, where the CCs are able to bend and fold back onto themselves. Therefore, the complex folding of the CCs is able to mediate novel interactions between the hinge and head domains, inspiring different loop extrusion models (13,22,23). Besides the identification of the features, little is known about the mechanical properties of SMC CCs, which would provide more detailed information about the capabilities of the SMC complexes in DNA organization.

Optical tweezers allow the study of protein folding features with high temporal and spatial resolution on a single-molecule basis. A precise CC unzipping pattern reveals the internal CC construction and the position of certain folding features, which appear as unfolding intermediates during protein stretching (24–26). Additionally, folding kinetics can be used to obtain the stability profile along the CC (27). For example, optical tweezers revealed that the CCs of vimentin contain several CC sections with varying stability (28). Therefore, an unfolding study of SMC CCs could pinpoint structurally and thermodynamically relevant

features based on the protein unfolding pattern and specify the stability of found CC sections.

Here, we study the structural composition and mechanical properties of SMC proteins with optical tweezers unfolding experiments using the non-canonical Psm3 CC as a model system. We find a relatively unstable protein compared with other CCs. We show that the selected CC region unfolds through three obligatory intermediates: a stutter, the SMC elbow, and a not-predicted feature at the hinge-proximal site, and is divided by these features into four sections. Misalignments of the wild-type (WT) CC strands are rarely observed. However, replacing the elbow induces frequently appearing metastable and non-productive misfoldings, which highlights the elbow as a proper guide for further CC alignment. Furthermore, in vivo experiments using changes in cell volume as an indicator for impaired cohesin complexes revealed the elbow flexibility as crucial for cohesin's biological function.

MATERIALS AND METHODS

All reagents were purchased from New England Biolabs (NEB, Ipswich, MA), Merck (Darmstadt, Germany), or Thermo Fisher Scientific (Waltham, MA) unless otherwise stated.

Protein expression and sample preparation

A pET28a-based plasmid containing the CC construct was transformed in Rosetta (DE3). Cells were grown at 37°C to optical density 600 (OD_{600}) = 0.4–0.6. The expression was induced by 0.4 mM IPTG and carried out at 18°C overnight. Harvested cells were resuspended in lysis buffer (25 mM Tris-HCl pH 7.5, 500 mM NaCl, 10 mM imidazole) supplemented with 1 mM PMSF and 0.1 mg/mL lysozyme, sonicated, and centrifuged. The supernatant was loaded onto a HisTrap FF-columns (GE Healthcare, Chicago, IL), washed with 10 CV lysis buffer, and eluted in lysis buffer containing 200 mM imidazole. Pooled fractions were gel filtrated using a Sephacryl S300 (GE Healthcare) or Superdex S200 (GE Healthcare) column. Eluted proteins were equilibrated in measuring buffer (25 mM Tris-HCl pH 7.5, 150 mM NaCl) for ybbR-tagged proteins or in measuring buffer at pH 7.2 for proteins with terminal cysteines.

From 5 to 15 μM proteins containing terminal ybbR-tags (WT, ΔEB, ΔEB_N, ΔEB_C) were coupled with 3′-coenzyme A (CoA)-oligos (Biomers, Ulm, Germany) in 2 molar excess with twofold excess of Sfp (made in-house, plasmids were kindly provided by Prof. H.E. Gaub, LMU Munich) and 10 mM $MgCl_2$ for 120 min at room temperature and reduced with 0.5 mM TCEP for 30 min before purifying with a size-exclusion column (Superdex S200 or Yarra 3000; Phenomex, Torrance, CA). The TR construct containing terminal cysteines (10–20 μM) was reduced with 0.5 mM TCEP 30 min before coupling to 3′-maleimide-oligos (Biomers, Ulm, Germany) in at least 2.5-fold molar excess. After 120 min at room temperature or an overnight reaction at 4°C, the mix was loaded onto the size-exclusion column. Fractions containing the protein coupled to two oligos were identified as described in (29) by SDS-PAGE stained first with SYBR Gold and afterward with SimplyBlue (Fig. S2).

DNA handles containing a dual-biotin or dual-digoxigenin modification and the single-stranded DNA (ssDNA) overhang were produced by PCR (30) and suitable ratios of handles to protein-DNA-chimeras were determined as reported earlier (31). Streptavidin-coated and anti-digoxigenin-coated silica beads were functionalized from carboxyl-functionalized 1-μm beads (Bangs Laboratories, Fischers, IN) as reported previously (32). Handles and protein-DNA chimeras were incubated with anti-digoxigenin beads and mixed with streptavidin beads in a passivated custom-made chamber, as described previously (33). Measurements were performed in measuring buffer containing the glucose oxidase oxygen scavenging system (0.65% w/v glucose, 13 U/mL glucose oxidase, and 8500 U/mL catalase).

Cysteine-based crosslinking of CCs

BMOE-crosslinks were made analogous to (34). The proteins (5 μM) were reduced for 60 min on ice, then BMOE (20 mM stock solution in DMSO) was added 1:20 and the reaction was quenched with excess of β-mercaptoethanol. Additionally, proteins were labeled with a fluorescent CoA-LD555 dye (Lumidyne, New York, NY, custom synthesis) according to the above-stated coupling procedure using the CoA-dye instead of CoA-oligo.

Prediction software and replacement of the elbow feature

Discontinuities in the Psm3 heptad pattern were identified using CC prediction software (15,16). The amino acids of the N-terminal Psm3 CC (F385-W394) were replaced with the amino acid sequence QMQRINSEISD for the corresponding ΔEB mutants. Additionally, the heptad repeat of the C-terminal CC was continued by deletion of P780 for Psm3 for the ΔEB mutants.

In addition, the CC propensity profile was generated with DeepCoil (35). N-terminal CC and C-terminal CC were analyzed separately due to length limitations of DeepCoil. The AlphaFold model was derived from the AlphaFold Protein Structure Database (36,37). Models for the WT elbow region and the ΔEB elbow region were generated using ColabFold with I337-I406 and N824–L760 of Psm3 linked by GGSGGSGGSGGS for the WT construct and the same amino acid sequence with the above-described mutations for the ΔEB construct (36,38–41). Illustrations were made with ChimeraX (42).

Strain construction, microscopy, and cell volume measurements

For the Smc3 CC, the replacement mutations were designed analogously to the Psm3 mutations. The N-terminal section (F393-W402) was replaced with QMQRINSEISD and the replacement of Q791, E792, and F793 on the C-terminal CC to threonine continued the Smc3 heptad repeat pattern. The single mutations only contained the respective ΔEB mutants.

All yeast strains used in this work are based on W303 and were constructed using standard methods. Full genotypes are listed in the Supporting Materials (Table S3). The DNA used to transform each strain was obtained by PCR amplification of plasmids listed in the Supporting Materials (Table S4). The PCR product was then integrated into the *SMC3* endogenous locus. All plasmids carry the C terminus of (mutated or WT) *SMC3* with the *CglaTRP1* marker. The correctness of the mutated *SMC3* sequence was confirmed by Sanger sequencing.

Cell cultures (4 mL) were grown at 30°C in synthetic complete medium containing 2% glucose (SCD) for exactly 24 h. The OD_{600} (measured with the Thermo Fisher Scientific, NanoDrop 2000 spectrophotometer) was maintained below 1.0 through appropriate dilutions during culture growth. Cell volume distributions were measured with a Coulter counter (Beckman Coulter, Brea, CA, Z2 Particle Counter). The measurements are based on at least four independent biological replicates. Each biological replicate consisted of two technical replicates. We note that, to reproduce our results, it is important to measure cell volume after 24 h of growth in SCD. This is because we noticed that the size distribution of the strains carrying the single N-terminal mutation shifted toward WT-like distribution when keeping the cells in exponential growth ($OD < 1.0$) for longer than 24 h. To exclude an artifact due to contamination, we re-plated cultures after 72 h of growth

in SCD, and picked single clones to repeat the experiment. After 24 h of growth, we found a size phenotype consistent with our original result shown in Fig. 5.

For microscopy, cells were grown with the same conditions as for the Coulter counter measurements. Next, wells of chambered coverslips (μ-Slice 8 Well, ibidi) were covered with 200 μL of 1 mg/mL (in water) Concavalin A and incubated for 10 min. The wells were subsequently washed twice with water and twice with SCD medium. Then, cells were sonicated for 10 s and 200 μL of the sonicated suspension was transferred to the coated well and left to stabilize for about 5 min, after which the medium was removed and the unbound cells were washed away twice with SCD medium. Finally, the well was covered with 200 μL of SCD medium and mounted onto the microscope. Images were taken on a Zeiss LSM 800 microscope (Carl Zeiss, Oberkochen, Germany, Zen 2.3, blue edition software) with an Axiocam 506 camera using a Plan-Apochromat 63×/1.4 Oil DIC objective. The cells were imaged in bright-field mode. The images were then segmented with YeaZ (43) embedded in the bioimage analysis software Cell-ACDC (44), which also automatically estimates the cell volume from 2D segmentation masks.

Optical trap setup and measurement modes

Optical tweezers experiments were performed on a LUMCIKS C-trap (LUMICKS, Amsterdam, the Netherlands) with embedded smooth motion update by active pulling on the tethered molecule with a fixed pulling and relaxation speed. In passive mode, the molecule was held at a constant trap distance in distinct steps. Data were recorded at a sampling rate of 78.125 kHz and downsampled by a factor of 4 for additional analysis.

Polymer models and length coordinates

Force-extension curves recorded by active pulling on the tethered molecule follow the stretching response of the DNA handles, which was modeled with the extensible worm-like chain eWLC model:

$$F(\xi_D) = \frac{k_B T}{p_D}\left(\frac{1}{4}\left(1-\frac{\xi_D}{L_D}+\frac{F}{K}\right)^{-2}-\frac{1}{4}+\frac{\xi_D}{L_D}-\frac{F}{K}\right) \quad (1)$$

with DNA extension ξ_D, force F, thermal energy $k_B T$, DNA persistence length p_D, DNA contour length L_D, and the stretch modulus K.

The additional force-extension of an unfolded protein was modeled by using a worm-like chain (WLC) model in series with the DNA handles' response (above):

$$F(\xi_p) = \frac{k_B T}{p_p}\left(\frac{1}{4}\left(1-\frac{\xi_p}{L_p}\right)^{-2}-\frac{1}{4}+\frac{\xi_p}{L_p}\right) \quad (2)$$

with unfolded protein extension ξ_p, protein persistence length p_p and protein contour length L_p.

The curves were fitted using a temperature T of 296 K, the DNA stretch modulus K was fixed to 800 pN/nm, and the protein persistence length p_p was fixed to 0.7 nm. Resulting DNA persistence lengths were typically $p_D \sim 10$–20 nm and resulting DNA contour lengths L_D were ∼350–370 nm.

Total free-energy estimation based on the Crooks fluctuation theorem

The Crooks fluctuation theorem is applicable for "stochastic, microscopically reversible dynamic" (45) and provides an estimation of the total free energy $\Delta G^0_{F\text{-}U}$ from non-equilibrium force versus distance curves (46):

$$\frac{P(W)}{P(-W)} = e^{\frac{W-\Delta G^0}{k_B T}} \quad (3)$$

with the work distribution during unfolding $P(W)$ and the work distribution during folding $P(-W)$. The work was determined as the area below a force versus distance curve and the contributions from the DNA handles were subtracted with respect to the linker models (27).

Equilibrium energies, barrier heights, rate models, and contour length transformation

Passive mode measurements were analyzed as previously described (28,47). In brief, the different passive mode traces were modeled using hidden Markov modeling (48). The errors of the force-dependent state probabilities were gained from bootstrapping from each individual constant-distance trace with 200 iterations (28) , and resulting standard deviations are shown. The equilibrium energies ΔG^0_{ij} between folding states i,j were calculated by global fits to an equilibrium model:

$$P_i(F_i) = \left(1+\sum_{j\neq i} exp\left(\frac{-\Delta G^0_{ij}-\Delta G^{sys}_{ij}\left(F_i,F_j\right)}{k_B T}\right)\right), \quad (4)$$

where $P_i(F_i)$ is the force-dependent state probability and ΔG^{sys}_{ij} is the energy contribution resulting from the bead deflection from the trap center (assuming a Hookean spring model), the stretching of the DNA handles, and the stretching of the unfolded polypeptide (integrals over eWLC and WLC models, see above) and can be determined based on the recorded data (27), whereas the equilibrium energy ΔG^0_{ij} can be determined by fitting (for a detailed description, please see (49)).

The force-dependent rate models with missed events correction and the contour length transformations were done for the protein-dumbbell system as described in (47), and the energy barrier heights were determined to construct an interpolated energy landscape using

$$\frac{\Delta G^{\dagger}}{k_B T} = \ln\left(\frac{k_0}{A}\right),\ A = 1.2\times 10^4\ s^{-1} \quad (5)$$

with energy barrier height $\Delta G^{\dagger}$, extrapolated zero-force folding rate k_0 and an Arrhenius-factor A, which estimates the frequency of folding attempts of the construct (27).

Classification of force-extension curves for the ΔEB constructs

For classification of the ΔEB F-I1 folding transition, the rupture of the initial DNA stretch was monitored in 500 nm/s force-extension curves of all constructs. If the rupture value was below 2 pN, corresponding force-extension curves lacking a folded F-I1 section were classified as non-native-like pulling traces. Traces were then transformed into contour space (50), where the population of non-productive, metastable states was clearly distinguishable from WT-like folds due to their longer residence time. Consequently, pulling traces showing intermediates resisting higher forces than 8.5 pN, which were never reached for the WT construct at that pulling speed, were classified as misaligned configurations. The ΔEB analysis includes 11 molecules with at least 20 cycles summing up to a total of 436 cycles of 500 nm/s, ΔEB_N analysis includes five molecules with 769 cycles of 500 nm/s, and ΔEB_C analysis includes seven molecules with 486 cycles

of 500 nm/s. Errors were given as standard deviations of the single-molecule distribution.

Simulation of force-extension curves from interpolated energy landscapes

The Monte Carlo simulations were performed as described previously (28,51). In brief, in the modeled system two trapped beads are linked with eWLC handles to the protein. The transition probabilities were taken from the force-dependent rates (Table S2; Fig. S3). The bead deflection and protein contour length were updated at each time step ($\Delta t = 10^{-5}$ s) and sampled at 20 kHz, mimicking experimental conditions. Force-extension curves were generated for the WT and TR construct. To generate force-extension traces for ΔEB, the WT energy landscape was iteratively adjusted such that the resulting force-extension curves matched the experimental WT-like force-extension curves of ΔEB (Fig. S6 *B*). To achieve this, the I2 state had to be eliminated and the resulting energy of the I3 state and the energy barrier of the I1-I3 transition were adjusted.

RESULTS

The Psm3 CC shows low-force unfolding over three intermediates

To study the mechanics of SMC CC domains, we engineered a single-peptide construct of the isolated CC region of the fission yeast cohesin subunit Psm3 (Fig. 1 *A*). Lacking high-resolution structural information about the entire CC, we identified endpoints of a suitable construct relying on crosslinking studies (52) and CC prediction software (15,16,35). In detail, the head-proximal endpoints (D245-K908) of Psm3 were found from sequence alignment of human Smc3 crosslinks. Using prediction software (Fig. S1 *A*), the CC fold was verified and spans at least to T505 and G658 at the hinge-proximal site. The utilized construct containing the selected stretches was equipped with terminal ybbR-tags, which served as attachment points for dsDNA handles (Figs. 1 *A*, S1 *A*, and S2). A 16-amino-acid (aa) flexible linker replaced the hinge region between T505 and G658. We then manipulated this construct in an optical tweezers instrument using a zipper geometry (Fig. 1 *A*).

Force-extension curves recorded at a pulling speed of 20 nm/s showed first unfolding events between ∼2 and 4 pN. This initial transition is followed by subsequent un- and refolding transitions with three obligatory intermediates close to equilibrium (red curve in Fig. 1 *B*). We used polymer models (53) to determine the length of the folded CC sections and found a length of 185.9 ± 3.4 nm for the unfolded construct U (red). All errors represent standard deviations, unless stated otherwise. This corresponds to 509 ± 10 aa, 19 aa shorter than the 528 aa contained in the construct. This suggests that the head-proximal amino acid of the CC construct does not contribute to the stably folded part of the CC, since no further unfolding event corresponding to an unfolding of these 19 aa was observed. The intermediates, which we termed I1 (blue), I2 (green), and I3 (orange in Fig. 1 *B*), consistently appeared at contour lengths of 58.0 ± 1.3 nm, 96.0 ± 2.9 nm and 146.5 ± 4.9 nm, respectively (see Table S1). The construct readily refolded into the native conformation during relaxation (gray curve in Fig. 1 *B*).

To characterize the exchange between the intermediates in greater detail, we next recorded equilibrium fluctuation traces in passive mode with a constant trap distance for each trace (27). Different applied tensions revealed the kinetics and energetics of the protein folding transitions (Fig. 1 *C*). At low distances (single asterisk in Fig. 1 *B* and *C*), the protein was most present in the folded configuration F (native state, purple), while partial unfolding into the I1 intermediate state was observed. At higher tensions, the construct was most present in the I1 intermediate, while first transitions into the I2 state could be resolved. Following traces demonstrate further unfolding from I2 into the I3 state. Finally, the population of states shifted toward U for higher trap distances (double asterisk in Fig. 1 *B* and *C*). From these traces, the stability profile of the Psm3 CC can be revealed. First, we used hidden Markov models (48) to identify the force-dependent populations (Fig. 1 *A*) and rate constants (Fig. S3 *A*). Next, we applied mechanical linker models to determine the thermodynamic energies of the states (27). The resulting energy of the fully folded CC is $\Delta G^0_{\text{F-U}} = (23.2 \pm 2.3)\ k_BT$. The intermediates are at 5.2 ± 0.9 k_BT, 11.2 ± 1.2 k_BT and 16.9 ± 1.9 k_BT measured from the native state (see Table S1), which agrees with the total free energy derived from the Crooks fluctuation theorem (45) of 19.1 ± 2.8 k_BT based on non-equilibrium force-distance curves (Fig. S4 *A* and *B*). Furthermore, we determined the barrier heights from the kinetics of the states by using the extrapolated zero-force unfolding rates (Fig. S3 *A*) and assumed an Arrhenius factor of 1.2×10^4 s^{-1} (27) (see Table S2). This allowed us to generate a putative energy landscape for the unzipping of the Psm3 CC (Fig. 1 *E*), which revealed faster kinetics for the I1-I2 section compared with the F-I1 or I2-I3 section due to the lower energy barrier height between I1 and I2. The I3-U energy profile is comparatively flat after the energy barrier. However, the profile in this region could be masked by the introduction of the hinge-replacing flexible linker. By analyzing ΔG^0 per contour length for each intermediate stretch, we found that the energy stored inside this CC construct is quite uniformly distributed and averages to 0.12 k_BT/nm (Table S1; Fig. S4 *C*), which means that the Psm3 CC is a thermodynamically less stable fold compared with other CC proteins (25–28).

Generally, folding and unfolding transitions proceeded sequentially in the scheme F-I1-I2-I3-U, suggesting that the Psm3 CC folds and unfolds in a zipper-like fashion. However, a detailed analysis of many molecules also revealed rarely populated and short-lived off-pathway states, which we called I2* (mean dwell time, $\tau_{I2}^* = 5.2 \pm 4.7$ ms) and I3* ($\tau_{I3}^* = 4.0 \pm 2.4$ ms), respectively (Fig. 1 *C* and *F*). Note that their contour lengths of

98.7 ± 3.4 nm and 146.6 ± 5.1 nm are indistinguishable from the on-pathway states I2 and I3. However, because these I2* and I3* are significantly less stable than I2 and I3, respectively ($\Delta G^0_{I2\text{-}I2}{}^* = 8.4 \pm 2.2\ k_BT$, $\Delta G^0_{I3\text{-}I3}{}^* = 5.5 \pm 2.3\ k_BT$) and always refolded into the on-pathway states, we suspect that they represent misaligned configurations of the I2-I3 section. Taken together, we found for the Psm3 CC a general zipper-like transition pattern between F and U, with occasional off-pathway deviations (Fig. 1 *G*).

Crosslinking corroborates unfolded amino acids in the head-proximal region

To solve the conundrum of the missing 19 aa in the folded conformation (see above) and unambiguously determine the pairing of the CC in the head-proximal region, we created a series of cysteine point mutations to precisely identify adjacent amino acids in the folded conformation by BMOE crosslinking experiments (34) (Fig. 2 *A*). Based on CC prediction software (15,16) we chose the most head-proximal C-terminal "d" amino acid of the CC (D906) as an anchor point and tested its association to seven "a" amino acids on the N-terminal strand (R246, N253, F260, I267, I274, L281, K288). We expressed and purified seven proteins containing double cysteine mutations and tested whether these constructs were crosslinked by the homo-bifunctional zero-length crosslinker BMOE. The efficiencies were evaluated by gel electrophoresis and fluorescence imaging (Figs. 2 *B* and S5).

We found a significant crosslinking efficiency of D906C to R246C, N253C, F260C, and I267C (Fig. 2 *B*, top panel). Surprisingly, the range of identified pairs exceeds the length of BMOE (8.0 Å). We conclude that the region around D906 must be flexible and hence not in a well-folded CC (17). This agrees with the observation of the missing 19 aa in pulling experiments (see above). To identify the actual start of the well-folded CC section, we performed an additional series of crosslink experiments with an amino acid (L892) further inside the CC as an anchor point (Fig. 2 *B*, bottom panel). In contrast to D906, L892C only paired with two amino acids on the N-terminal strand (F260C and I267C), which is compatible with the BMOE length and consistent with a well-folded CC close to L892. We conclude that the head-proximal CC region of our construct is well folded until I267 and L892 but is likely unfolded around R246 and D906 (Fig. 2 *C*).

Intermediates occur at non-canonical CC features

We next sought to use the length information obtained from our measurements to pinpoint the structural location of the intermediate states to features inside the CC. This assignment is straightforward if the CC adopts a regular fold without discontinuities. However, structural and biochemical studies have indicated that SMC CCs deviate from a regular CC fold (5,17) and may contain several non-canonical features such as loops or linker regions (54), which lead to shifts in pairing along the hydrophobic core of the protein (Fig. S1 *A*). In our tweezers experiments, we measure the number of amino acids that free up during an unfolding event independently of a symmetric or shifted pairing of the CC, which results in ambiguous assignments of folding features (Fig. S1 *B*).

We therefore used computational prediction software to identify non-canonical features of the CC. First, we studied the predicted CC propensities by DeepCoil (35,55,56) to identify possible separated CC sections (Fig. S1 *A*). The minima in the profile indicate three or four non-structured linkers between sections of continuous CC (Fig. S1 *A*). In addition, PCOILS (16) and MARCOIL (15), which estimate the CC heptad repeat pattern, showed two sites with shifts in CC pairing for the selected Psm3 CC region. Next, we used the atomic model of the whole Psm3 protein predicted by AlphaFold (36,37) and identified the elbow and a stutter as non-canonical CC features of the selected Psm3 region. In the model, the stutter is located close to P831, while the elbow is predicted as non-alpha helical linkers on both CC strands, located at F385-T387 and S773-D779 (Fig. S1 *C*, top), which is close to the elbow predicted by crosslinks K388-Q398 and D779-L789 (18). Interestingly,

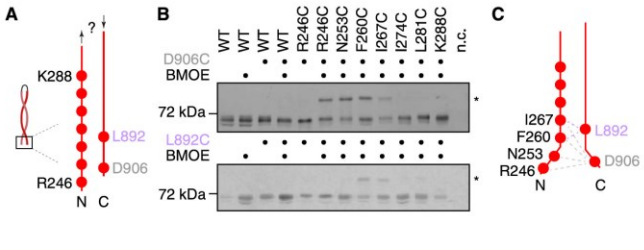

FIGURE 2 Crosslinking of terminal amino acids reveals fraying at the termini. (*A*) Cysteine-based BMOE crosslinking was performed for the anchor amino acids D906 (gray) and L892 (light purple), which are located inside the predicted hydrophobic core and at the C-terminal end of the WT construct. Seven positions inside the hydrophobic core of the N-terminal end, ranging from R246 to K288, were tested as possible crosslink partners. (*B*) BMOE-crosslinked proteins bearing denoted cysteine point mutations were additionally labeled with CoA-dyes via their ybbR-tags and further investigated by SDS-PAGE and fluorescent imaging. Top, D906 crosslinks; bottom, L892 crosslinks. WT, C-terminal single mutations, controls without BMOE, and negative control (n.c.) lacking a CC construct display no crosslinks (only lower band or no band, respectively), whereas certain double mutations show a high fraction of crosslinked proteins (higher band, asterisks). A quantitative analysis of the crosslink efficiency can be found in Fig. S5. (*C*) Summary of the crosslink series. Pairing partners of D906 (gray) and L892 (light purple) indicated with dashed lines. To see this figure in color, go online.

the prediction programs agree on a pairing shift between the CC strands of two amino acids at the stutter and five amino acids at the elbow.

To assign our measured intermediate contour lengths to paired amino acids in the protein, we used the pairing at the CC termini identified by our crosslinking experiments (Fig. 2 *B*) to find the folded state F and extrapolate further CC alignment. The measured contour length of the folded construct (F – U: 509 ± 10 aa) significantly exceeds the crosslink-based minimal paired amino acids I267 and L892, which includes only 488 aa (238 aa from the N-terminal coil I267-T505, 16 aa from the linker, and 234 aa from the C-terminal coil G658-L892). Therefore, the folded state starts around S257 and I903, which agrees with the pairing of I267-L892 and includes 509 aa (F: S257-I903 ± 10 aa). We hence used the measured number of unfolded amino acids to locate the intermediates in the protein with respect to the predicted pairing shifts (I1, I337-N824 ± 4 aa; I2, D392-M775 ± 8 aa; I3, R461-D706 ± 14 aa; also see Fig. S1 *C*, bottom). Interestingly, the stutter and the elbow lie close to the detected intermediates I1 and I2.

A truncated CC shows regular elbow folding

In contrast to I1 and I2, which coincided with sites of predicted non-canonical CC features, the intermediate I3 was clearly identified in pulling experiments but not predicted. We therefore wondered if I3 folding is a necessary prerequisite for productive folding and designed a truncation mutant TR, where the hinge-proximal region was shortened to G445 and T733, eliminating I3 (Fig. 3 *A*). Pulling cycles of TR showed I1 and I2 features at the same contour lengths as for WT (Fig. 3 *B*), but lacked I3 features. In addition, passive-mode measurements revealed that the population and energetics of I1 and I2 are identical to WT (Figs. 3 *C–E* and S3 *B*). Interestingly, the I2-U_{TR} section showed faster kinetics than the I2–I3 section of the WT construct. Furthermore, TR did not misfold into the intermediates I2* and I3*, which both require a functional I3 according to the stated on-pathway folding network (Fig. 1 *F*). We conclude that Psm3 folds consecutively and proper F-I1 and I1-I2 folding does not require a folded I3.

Replacing the elbow results in rare non-productive, metastable folds

Since the elbow is a central part of SMC CCs and allows collapsed conformations of SMC complexes (13,22), we wondered how the CC will be aligned without the elbow. Recent findings showed that altered SMC CC length and even point mutations of aromatic amino acids inside the elbow region can be lethal (5,20). Therefore, we first identified conserved aromatic amino acids in the N-terminal elbow region of Psm3 and found a non-helical sequence that causes a heptad discontinuity. We replaced this elbow sequence with a suitable CC section on the N-terminal CC strand (ΔEB_N). For the C-terminal strand, a deletion of P780 was sufficient to complete the heptad repeat pattern (ΔEB_C). Furthermore, a combination of ΔEB_N and ΔEB_C excludes the elbow on both strands (ΔEB), and possesses the same number of amino acids as WT. In accordance with our design, prediction software confirmed the improved CC propensity and indicates a straightening of the CCs in the double mutant ΔEB (Fig. 4 *A*).

In contrast to WT, we found that all mutants (ΔEB, ΔEB_N, ΔEB_C) displayed more diverse force-extension curves. Stretching traces often started at intermediate contour lengths, with many traces exhibiting unexpectedly high unfolding forces, indicating that, during relaxation, the mutants frequently misfolded. The misfolded configurations occurred in 2% ± 2% of the traces for ΔEB_N and in 3% ± 2% of the traces for ΔEB_C, whereas, in the double mutation ΔEB, we observed 17% ± 8% misfolded configurations (Fig. 4 *B*). This suggests an additive effect of both mutations on the CC folding. Hence, we characterized the misfolding of the ΔEB mutant in more detail. In 84% ± 7% of ΔEB pulling traces recorded at 500 nm/s we found a folded F-I1 section similar to WT and no intermediate comparable with the WT I2 state. The remaining pulling traces exposed configurations that had not folded into the native state F during previous relaxation (Fig. S6 *C*, blue). Therefore, these configurations can be considered as misfolded kinetic traps. However, here, the consecutive relaxation (Fig. S6 *C*, gray) showed refolding into the native state, suggesting that the protein is able to frequently exchange between folded and misfolded states. For a closer look, we display representative cycles at 20 nm/s showing unzipping of folded and misfolded configurations (Fig. S6 *B* and *C*). Next, we focused on the novel ΔEB-specific intermediates that ruptured at unexpectedly high forces (Fig. 4 *B*, blue traces). Here, the protein adopts half-folded conformations, which are non-productive as further pairing of unfolded, terminal amino acids is not observed. In contrast to WT and TR, these configurations of ΔEB still resisted high forces at 20 nm/s (Fig. S6 *C*), which concealed information about the consecutive I3-U section. A closer look by reducing the pulling speed to 2 nm/s revealed various misaligned configurations of the I1-I3 section, which are clearly different from the WT I2 state and are followed by the familiar WT-like I3-U transitions (Fig. S6 *D*). Interestingly, the misfolded states exchanged over the I3 intermediate (arrowhead in Fig. S6 *E*) and not over the fully unfolded protein U. We conclude that the heptad discontinuity induced by the Psm3 elbow, which is present in WT but has been removed in ΔEB, ensures a native-like alignment of the head-proximal CCs, and that the replacement of the elbow by a continuous CC section frequently leads to non-native configurations. The elbow, albeit not being a canonical CC feature, is therefore an

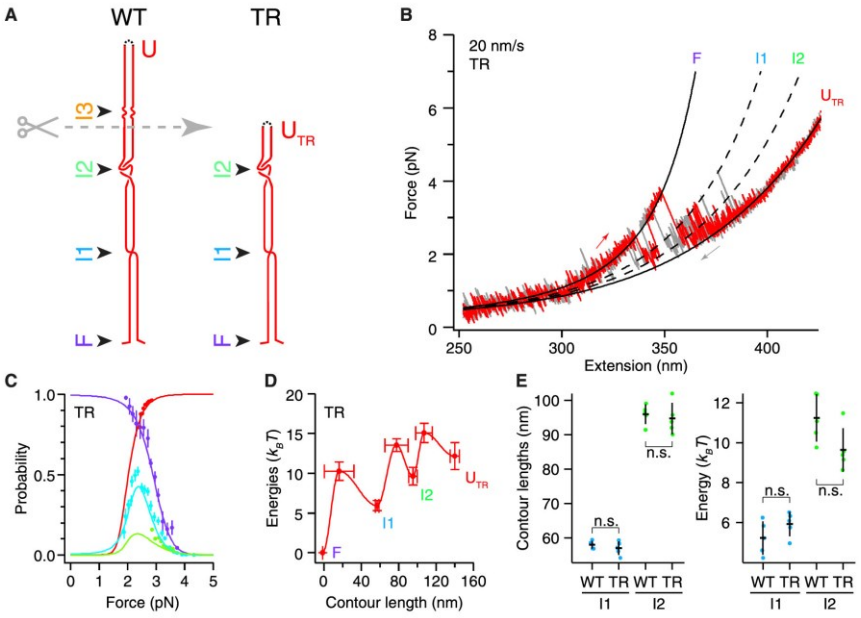

FIGURE 3 Unzipping of a truncated Psm3 CC. (*A*) A truncation of the WT construct removes I3. (*B*) Force-extension curve of the truncated construct. Pull, red curve; relaxation, gray curve. Two folding intermediates (I1 and I2) and the unfolded protein (U_{TR}) were identified. (*C*) Global energy fit of the force-dependent state probabilities; error bars indicate standard deviations and result from bootstrapping of the corresponding passive mode trace (see section "materials and methods"). Color code as in (*B*). (*D*) Energy landscape constructed from the contour lengths and energies of the folding states. Barrier heights were estimated from force-dependent transition kinetics (Fig. S3 *B*), n = 5 molecules, error bars indicate standard deviations. (*E*) Comparison of the measured contour length and energies of the intermediates I1 (blue) and I2 (green) found for the WT and the TR construct, n = 5, error bars indicate standard deviations; n.s., not significant based on a Student's *t*-test. To see this figure in color, go online.

integral component of the Psm3 CC region and ensures proper folding of the CC sections.

To validate if the elbow replacement had a stabilizing effect on the CC, we were interested in the energy landscape of the WT-like ΔEB configuration. Unfortunately, passive-mode measurements of the ΔEB construct were ambiguous to analyze since transitions between native-like and misfolded configurations, the variety of misfolded states, and their longevity masked the folding network, which was also reported for a misfolding hairpin (57). This prevented us from directly obtaining an energy landscape of ΔEB. Instead, we used an approach based on Monte Carlo simulations. Simulated force-extension curves matched experimental data for WT and TR (Fig. S7 *A–F*). We then generated a likely ΔEB landscape by altering the WT landscape around the region of the elbow replacement to match simulated and experimental force-extension curves (Fig. S7 *G–I*). This analysis showed that the ΔEB elbow replacement resulted in a removal of I2 and a destabilization of the I1–I3 section by ∼7 k_BT.

Replacing the elbow in vivo causes cell growth defects

To test the influence of the elbow feature on the cohesin complex in a biological context, we introduced similar mutations of the elbow regions into the budding yeast cohesin subunit Smc3. As cohesin plays a major role in cell cycle progression and chromatin organization and because mitosis defects often manifest in increased cell volume (58), we chose cell volume measurements as a sensitive read-out for biological cohesin function. An elbow replacement corresponding to ΔEB showed significantly increased cell volume compared with the WT strain or a control, where the WT *SMC3* sequence was re-integrated in the same locus using an analogous construct to exclude unspecific effects (Fig. 5 *A*). We next sought to pinpoint the phenotype of the replacement to the individual CCs and repeated the experiment with only N- or C-terminal mutations (ΔEB_N and ΔEB_C, respectively). To our surprise, we found that ΔEB_N had an even stronger phenotype than the double mutation, whereas ΔEB_C was not significantly different from

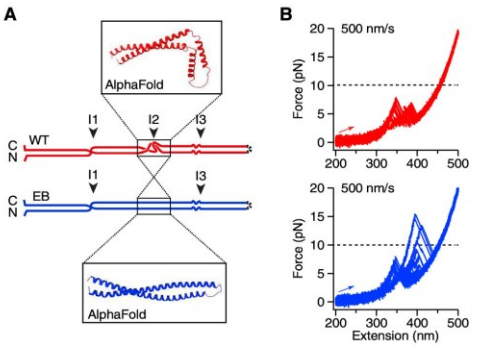

FIGURE 4 Unzipping of the elbow replacement mutant ΔEB. (*A*) The ΔEB mutant (blue) is constructed from WT (red) by replacing the elbow at I2 with a continuous heptad repeat pattern that encodes a canonical CC. The AlphaFold prediction created with ColabFold indicates a straightening of the ΔEB CC and the removal of the elbow. (*B*) Comparison of 20 overlaid pulling traces from the WT construct (red) and the ΔEB construct (blue) at a pulling speed of 500 nm/s. The WT construct follows the described folding pathway with moderate unfolding forces. The ΔEB mutant shows altered, halfway-folded conformations, which resist unexpected high forces >10 pN (dashed lines to guide the eye, see also Figs. S6 and S7 for a detailed comparison between WT and ΔEB). To see this figure in color, go online.

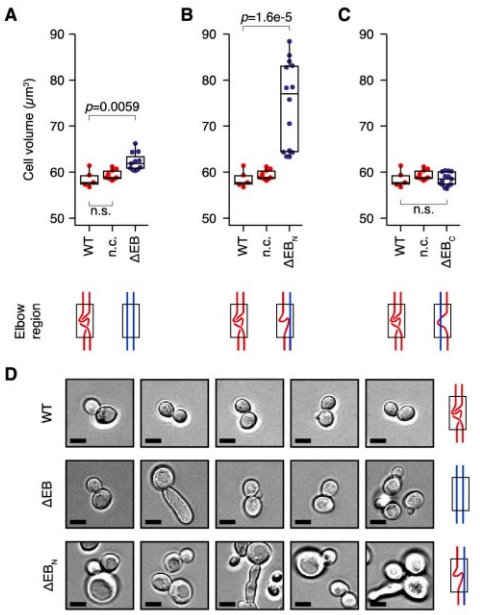

FIGURE 5 In vivo investigation of budding yeast elbow replacement mutations (ΔEB) for Smc3 via cell volume analysis. The *SMC1* gene was unmodified in all strains. (*A*) Average yeast cell volumes for WT, negative control, in which the targeted *SMC3* gene and the following untranslated region is replaced by the WT sequence (n.c., two independent clones), and the double mutation containing N- and C-terminal ΔEB mutations in Smc3 (two independent clones). The double mutation causes a significantly increased cell volume. Lower part: corresponding cartoons of the mutated (blue) or unmodified elbow region (red). (*B*) A more severe phenotype was found for the single N-terminal elbow replacement (ΔEB_N, three independent clones). Lower part: cartoons as in (A). (*C*) The single C-terminal elbow replacement mutation (ΔEB_C, four independent clones) does not influence the cell volume. Lower part: cartoons as in (A). *p*-values based on Student's two-sample *t*-test. n.s., not significant. (*D*) Exemplary microscopy images of yeast cells carrying WT, ΔEB, and ΔEB_N mutations in the *SMC3* gene. Compared with WT, the mutants ΔEB and ΔEB_N show cell size and shape defects. Scale bar in all images, 5 μm (see also Fig. S8). To see this figure in color, go online.

WT (Fig. 5 *A* and *C*). To conclude, while cohesin tolerates a small mutation that continues the heptad in the Smc3 C-terminal CC, the heptad continuation of the Smc3 N-terminal CC results in severe size and shape defects (Figs. 5 *D* and S8). Interestingly, a simultaneous continuation of both CCs partially rescues the phenotype, but still shows a cell volume increase consistent with a mitosis defect.

DISCUSSION

Here, we have characterized a long region of Psm3's CC as a model protein for eukaryotic SMC proteins by optical tweezers with supporting in vitro and in vivo experiments. We identified structural features from their mechanical response, which uncovered the arrangement of the CC.

CCs are a ubiquitous and frequently occurring folding motif of proteins, where the amino acids follow a heptad repeat pattern (14). Besides facilitating oligomerization, CCs are known to serve as molecular spacers that separate functional domains of proteins. While the length of SMC CCs is conserved in eukaryotes (17), which is consistent with a role as spatial separators, the relatively high sequence conservation (14) compared with spacer CCs and the transitions between different large-scale conformations (O and B shape; Fig. S9) contradict a function as rigid spacers and suggest a more subtle role within SMC mechanics.

SMC CCs were found to contain kinks in rotary-shadow experiments (4,59), and recent crystal structures of SMC CCs exhibit interrupted helices (5,12,60), which suggest the presence of different segments in the CC. Numerous short breaks were already found in other elongated CCs, such as MRN (61), Myosin II (62), or Golgins (63). However, a segmented architecture does not necessarily result in high flexibility, as MRN (64) or Myosin II (65) show dramatically higher persistence length than SMC CCs (4). Our unfolding experiments supports the notion of a modular arrangement revealing that the Psm3 discontinuities are structurally and thermodynamically relevant features. In addition, the free energies per length of the individual segments (0.1–0.2 k_BT/nm) (Figs. 1 *E* and 3 *D*) are significantly lower than the reported values for canonical CC proteins, such as vimentin (0.54 k_BT/nm) (28) or GCN4 (1.20

k_BT/nm) (26,27). Therefore, the modular arrangement, but also the weak fold of the CC segments, promotes flexible polymer properties and facilitates the different conformations of SMC complexes.

SMC complexes contain an ATPase domain; hence, it stands to reason that CC bending at the elbow is an active process triggered by ATP hydrolysis and was already suggested for condensin (22). For cargo-carrying motor proteins such as dynein (52), conformational changes were found to be promoted by ATP-cycle-dependent helix sliding of the CC. These proteins contain, exactly like SMC proteins, kinked CCs attached to ATPase domains. In SMC CCs, different heptad pairings inside the hydrophobic core could lock the SMC complex in either the straight CC conformation or in the bent CC conformation. The transition between these conformations would then be promoted by helix sliding, whose energetic cost is estimated to be in the order of ~2–3 k_BT (66), which could be afforded by ATP hydrolysis (67). However, our crosslinking experiments of the Psm3 CC showed no indications of a flexible heptad pairing (Fig. 2 *B*), which would be a prerequisite of helix sliding. In addition, passive mode traces did not show typical features of staggered helices, which were reported for GCN4 (26). In summary, our data support a model where elbow bending is driven by passive thermal fluctuations (13) of the CC sections identified in this study.

A passive CC bending mechanism critically depends on a structurally flexible elbow, but recent cryo-EM densities lack resolution at this crucial component (11,12). The simplest possible feature would be a kink, such as in GreA (68), or a loop, such as in Ndc80 (69) (Fig. S10 *A*), that is only present on one CC strand. Since AlphaFold and CC propensity profiles revealed a non-helical linker on both CC strands (Fig. S1 *A*), it is more likely that the elbow forms an entangled linker (e.g., as in Omp100 (70) or KKT4 (71); see Fig. S10 *B*) or a "knuckle motif," as in bacterial SMC or SMC-like proteins, such as MukB (5,17), RecN (72), or SbcCD (73) (Fig. S10 *C*), both of which are consistent with the found pairing shift of five amino acids (Fig. S1 *C*). Replacing the elbow by creating a continuous canonical heptad repeat pattern destabilized the CC (Fig. S7). Intuitively, one would expect that a continuation of the regular heptad repeat stabilizes the CC. However, a reason for destabilization might be that the flexible linkers in the elbow are required for CC strand movement and proper alignment, while a regular heptad continuation locks them in an unfavorable configuration. Without the separating effect of the elbow linkers, these misaligned configurations then propagate over the elbow region and result in non-native alignments of the amino acids close to I1, which might result in the metastable intermediates found in the ΔEB unzipping traces (Fig. S4 *C*). In contrast to continuous short CCs such as GCN4 that need to completely unfold to remove misfolds before starting a new folding attempt (26), unfolding to the previous intermediate is sufficient for Psm3 CC (Fig. S6 *E*). Taken together, the elbow guides proper folding of the entire Psm3 CC and prevents misfolding. As other long CCs, such as Golgins (63) or MRN (61), also contain interrupted helices, we propose that the modular composition is a general feature of elongated CCs to reliably enable correct folding and functional proteins.

In in vivo studies, alterations of the CC length (20) and even point mutations inside the elbow (5) were found to lead to lethality. We observed severe size and shape phenotypes of the N-terminal replacement mutant (ΔEB_N), which were tempered by the addition of the C-terminal deletion (Fig. 5). This indicates that the altered cohesin complexes are still partially functional, but their ability to perform the stated large-scale conformational changes seemed to be significantly restricted by the alteration in elbow flexibility. Maybe the mutated CC arrangement is furthermore unfavorable for additional binding partners. Hence, the elbow is not only a folding guide but also likely ensures proper cohesin function.

CONCLUSIONS

Using optical tweezers, we have found four relatively unstable sections inside the Psm3 CC. In contrast to many other CCs, and especially to homodimers, the non-canonical elbow feature comprises non-structured linkers, which induces a pairing shift from one CC strand to the other, enabling potentially unique CC mechanics. We found that the elbow is a key feature inside Psm3s CC, which ensures proper CC arrangement and further complex functions. As discontinuities are common CC features, these findings might translate to other elongated CCs and especially to other SMC or SMC-like proteins. Additionally, this study provides new details for SMC arm flexibility, which could be used to improve future modeling approaches of the SMC loop extrusion mechanism.

SUPPORTING MATERIAL

Supporting material can be found online at https://doi.org/10.1016/j.bpj.2022.10.017.

AUTHOR CONTRIBUTIONS

Conducting experiments, M.F., F.P., E.R., and S.J.; data analysis, M.F., F.P., E.R., K.M.S., and J.S.; concept of the study, J.S.; writing, M.F., S.Z., and J.S. with input from all authors.

ACKNOWLEDGMENTS

J.S. and K.M.S. acknowledge support by the LMU Center for NanoScience CeNS. J.S. is supported by a DFG Emmy Noether grant STI673/2-1 and an ERC starting grant 758124.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Haering, C. H., J. Löwe, …, K. Nasmyth. 2002. Molecular architecture of SMC proteins and the yeast cohesin complex. *Mol. Cell.* 9:773–788.
2. Davidson, I. F., and J.-M. Peters. 2021. Genome folding through loop extrusion by SMC complexes. *Nat. Rev. Mol. Cell Biol.* 22:445–464.
3. Yatskevich, S., J. Rhodes, and K. Nasmyth. 2019. Organization of chromosomal DNA by SMC complexes. *Annu. Rev. Genet.* 53:445–482.
4. Eeftens, J. M., A. J. Katan, …, C. Dekker. 2016. Condensin smc2-smc4 dimers are flexible and dynamic. *Cell Rep.* 14:1813–1818.
5. Bürmann, F., B.-G. Lee, …, J. Löwe. 2019. A folded conformation of MukBEF and cohesin. *Nat. Struct. Mol. Biol.* 26:227–236.
6. Ryu, J.-K., A. J. Katan, …, C. Dekker. 2020. The condensin holocomplex cycles dynamically between open and collapsed states. *Nat. Struct. Mol. Biol.* 27:1134–1141.
7. Davidson, I. F., D. Goetz, …, F. Uhlmann. 2016. Rapid movement and transcriptional re-localization of human cohesin on DNA. *EMBO J.* 35:2671–2685.
8. Stigler, J., G. Ö. Çamdere, …, E. C. Greene. 2016. Single-molecule imaging reveals a collapsed conformational state for DNA-bound cohesin. *Cell Rep.* 15:988–998.
9. Lee, H., H. Noh, and J.-K. Ryu. 2021. Structure-function relationships of SMC protein complexes for DNA loop extrusion. *BioDesign.* 9:1–13.
10. Bürmann, F., L. F. H. Funke, …, J. Löwe. 2021. Cryo-EM structure of MukBEF reveals DNA loop entrapment at chromosomal unloading sites. *Mol. Cell.* 81:4891–4906.e8.
11. Lee, B.-G., F. Merkel, …, C. H. Haering. 2020. Cryo-EM structures of holo condensin reveal a subunit flip-flop mechanism. *Nat. Struct. Mol. Biol.* 27:743–751.
12. Petela, N. J., A. Gonzalez Llamazares, …, K. A. Nasmyth. 2021. Folding of cohesin's coiled coil is important for Scc2/4-induced association with chromosomes. *Elife.* 10:e67268.
13. Bauer, B. W., I. F. Davidson, …, J.-M. Peters. 2021. Cohesin mediates DNA loop extrusion by a "swing and clamp" mechanism. *Cell.* 184:5448–5464.e22.
14. Truebestein, L., and T. A. Leonard. 2016. Coiled-coils: the long and short of it. *Bioessays.* 38:903–916.
15. Delorenzi, M., and T. Speed. 2002. An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. *Bioinformatics.* 18:617–625.
16. Gruber, M., J. Söding, and A. N. Lupas. 2006. Comparative analysis of coiled-coil prediction methods. *J. Struct. Biol.* 155:140–145.
17. Waldman, V. M., T. H. Stanage, …, M. G. Oakley. 2015. Structural mapping of the coiled-coil domain of a bacterial condensin and comparative analyses across all domains of life suggest conserved features of SMC proteins. *Proteins.* 83:1027–1045.
18. Higashi, T. L., P. Eickhoff, …, F. Uhlmann. 2020. A structure-based mechanism for DNA entry into the cohesin ring. *Mol. Cell.* 79:917–933.e9.
19. Diebold-Durand, M.-L., H. Lee, …, S. Gruber. 2017. Structure of full-length SMC and rearrangements required for chromosome organization. *Mol. Cell.* 67:334–347.e5.
20. Bürmann, F., A. Basfeld, …, S. Gruber. 2017. Tuned SMC arms drive chromosomal loading of prokaryotic condensin. *Mol. Cell.* 65:861–872.e9.
21. Vazquez Nunez, R., Y. Polyhach, …, S. Gruber. 2021. Gradual opening of Smc arms in prokaryotic condensin. *Cell Rep.* 35:109051.
22. Takaki, R., A. Dey, …, D. Thirumalai. 2021. Theory and simulations of condensin mediated loop extrusion in DNA. *Nat. Commun.* 12:5865.
23. Nomidis, S. K., E. Carlon, …, J. F. Marko. 2022. DNA tension-modulated translocation and loop extrusion by SMC complexes revealed by molecular dynamics simulations. *Nucleic Acids Res.* 50:4974–4987, gkac268.
24. Bornschlögl, T., G. Woehlke, and M. Rief. 2009. Single molecule mechanics of the kinesin neck. *Proc. Natl. Acad. Sci. USA.* 106:6992–6997.
25. Gao, Y., S. Zorman, …, Y. Zhang. 2012. Single reconstituted neuronal SNARE complexes zipper in three distinct stages. *Science.* 337:1340–1343.
26. Xi, Z., Y. Gao, …, Y. Zhang. 2012. Single-molecule observation of helix staggering, sliding, and coiled coil misfolding. *Proc. Natl. Acad. Sci. USA.* 109:5711–5716.
27. Gebhardt, J. C. M., T. Bornschlögl, and M. Rief. 2010. Full distance-resolved folding energy landscape of one single protein molecule. *Proc. Natl. Acad. Sci. USA.* 107:2013–2018.
28. Ramm, B., J. Stigler, …, M. Rief. 2014. Sequence-resolved free energy profiles of stress-bearing vimentin intermediate filaments. *Proc. Natl. Acad. Sci. USA.* 111:11359–11364.
29. Mukhortava, A., and M. Schlierf. 2016. Efficient formation of site-specific protein–DNA hybrids using copper-free click chemistry. *Bioconjugate Chem.* 27:1559–1563.
30. Freitag, M., D. Kamp, …, J. Stigler. 2021. Identification and correction of miscalibration artifacts based on force noise for optical tweezers experiments. *J. Chem. Phys.* 155:175101.
31. Synakewicz, M., D. Bauer, …, L. S. Itzhaki. 2019. Bioorthogonal protein-DNA conjugation methods for force spectroscopy. *Sci. Rep.* 9:13820.
32. Tych, K., and G. Žoldák. 2019. Stable substructures in proteins and how to find them using single-molecule force spectroscopy. *In* Protein Supersecondary Structures: Methods and Protocols. A. E. Kister, ed Springer, pp. 263–282.
33. Synakewicz, M., R. S. Eapen, …, J. Stigler. 2022. Unraveling the mechanics of a repeat-protein nanospring: from folding of individual repeats to fluctuations of the superhelix. *ACS Nano.* 16:3895–3905.
34. Soh, Y.-M., J. Basquin, and S. Gruber. 2021. A rod conformation of the Pyrococcus furiosus Rad50 coiled coil. *Proteins.* 89:251–255.
35. Ludwiczak, J., A. Winski, …, S. Dunin-Horkawicz. 2019. DeepCoil—a fast and accurate prediction of coiled-coil domains in protein sequences. *Bioinformatics.* 35:2790–2795.
36. Jumper, J., R. Evans, …, D. Hassabis. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature.* 596:583–589.
37. Varadi, M., S. Anyango, …, S. Velankar. 2022. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* 50:D439–D444.
38. Mirdita, M., K. Schütze, …, M. Steinegger. 2022. ColabFold: making protein folding accessible to all. *Nat. Methods.* 19:679–682.
39. Mirdita, M., M. Steinegger, and J. Söding. 2019. MMseqs2 desktop and local web server app for fast, interactive sequence searches. *Bioinformatics.* 35:2856–2858.
40. Mirdita, M., L. von den Driesch, …, M. Steinegger. 2017. Uniclust databases of clustered and deeply annotated protein sequences and alignments. *Nucleic Acids Res.* 45:D170–D176.
41. Mitchell, A. L., A. Almeida, …, R. D. Finn. 2020. MGnify: the microbiome analysis resource in 2020. *Nucleic Acids Res.* 48:D570–D578.
42. Pettersen, E. F., T. D. Goddard, …, T. E. Ferrin. 2021. UCSF ChimeraX : structure visualization for researchers, educators, and developers. *Protein Sci.* 30:70–82.
43. Dietler, N., M. Minder, …, S. J. Rahi. 2020. A convolutional neural network segments yeast microscopy images with high accuracy. *Nat. Commun.* 11:5723.
44. Padovani, F., B. Mairhörmann, …, K. M. Schmoller. 2022. Segmentation, tracking and cell cycle analysis of live-cell imaging data with Cell-ACDC. *BMC Biol.* 20:174.

45. Crooks, G. E. 1999. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Phys. Rev. E.* 60:2721–2726.

46. Collin, D., F. Ritort, …, C. Bustamante. 2005. Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies. *Nature.* 437:231–234.

47. Bauer, D., S. Meinhold, …, G. Žoldák. 2018. A folding nucleus and minimal ATP binding domain of Hsp70 identified by single-molecule force spectroscopy. *Proc. Natl. Acad. Sci. USA.* 115:4666–4671.

48. Stigler, J., and M. Rief. 2012. Hidden Markov analysis of trajectories in single-molecule experiments and the effects of missed events. *ChemPhysChem.* 13:1079–1086.

49. Žoldák, G., J. Stigler, …, M. Rief. 2013. Ultrafast folding kinetics and cooperativity of villin headpiece in single-molecule force spectroscopy. *Proc. Natl. Acad. Sci. USA.* 110:18156–18161.

50. Puchner, E. M., G. Franzen, …, H. E. Gaub. 2008. Comparing proteins by their unfolding pattern. *Biophys. J.* 95:426–434.

51. Stigler, J., F. Ziegler, …, M. Rief. 2011. The complex folding network of single calmodulin molecules. *Science.* 334:512–516.

52. Huis in 't Veld, P. J., F. Herzog, …, J.-M. Peters. 2014. Characterization of a DNA exit gate in the human cohesin ring. *Science.* 346:968–972.

53. Wang, M. D., H. Yin, …, S. M. Block. 1997. Stretching DNA with optical tweezers. *Biophys. J.* 72:1335–1346.

54. Hons, M. T., P. J. Huis In 't Veld, …, F. Herzog. 2016. Topology and structure of an engineered human cohesin complex bound to Pds5B. *Nat. Commun.* 7:12523.

55. Gabler, F., S.-Z. Nam, …, V. Alva. 2020. Protein sequence analysis using the MPI bioinformatics toolkit. *Curr. Protoc. Bioinformatics.* 72:e108.

56. Zimmermann, L., A. Stephens, …, V. Alva. 2018. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J. Mol. Biol.* 430:2237–2243.

57. Alemany, A., A. Mossa, …, F. Ritort. 2012. Experimental free-energy measurements of kinetic molecular states using fluctuation theorems. *Nat. Phys.* 8:688–694.

58. Schmoller, K. M. 2017. The phenomenology of cell size control. *Curr. Opin. Cell Biol.* 49:53–58.

59. Anderson, D. E., A. Losada, …, T. Hirano. 2002. Condensin and cohesin display different arm conformations with characteristic hinge angles. *J. Cell Biol.* 156:419–424.

60. Soh, Y.-M., F. Bürmann, …, S. Gruber. 2015. Molecular basis for SMC rod formation and its dissolution upon DNA binding. *Mol. Cell.* 57:290–303.

61. Park, Y. B., M. Hohl, …, Y. Cho. 2017. Eukaryotic Rad50 functions as a rod-shaped dimer. *Nat. Struct. Mol. Biol.* 24:248–257.

62. Heissler, S. M., A. S. Arora, …, K. Chinthalapudi. 2021. Cryo-EM structure of the autoinhibited state of myosin-2. *Sci. Adv.* 7:eabk3273.

63. Ishida, R., A. Yamamoto, …, N. Nakamura. 2015. GM130 is a parallel tetramer with a flexible rod-like structure and N–terminally open (Y-shaped) and closed (I-shaped) conformations. *FEBS J.* 282:2232–2244.

64. de Jager, M., J. van Noort, …, C. Wyman. 2001. Human rad50/mre11 is a flexible complex that can tether DNA ends. *Mol. Cell.* 8:1129–1135.

65. Schwaiger, I., C. Sattler, …, M. Rief. 2002. The myosin coiled-coil is a truly elastic protein structure. *Nat. Mater.* 1:232–235.

66. Gomez, D., Y. Gavrilov, and Y. Levy. 2019. Sliding mechanism at a coiled-coil interface. *Biophys. J.* 116:1228–1238.

67. Rosing, J., and E. C. Slater. 1972. The value of ΔG° for the hydrolysis of ATP. *Biochim. Biophys. Acta.* 267:275–290.

68. Stebbins, C. E., S. Borukhov, …, S. A. Darst. 1995. Crystal structure of the GreA transcript cleavage factor from Escherichia coli. *Nature.* 373:636–640.

69. Ciferri, C., S. Pasqualato, …, A. Musacchio. 2008. Implications for kinetochore-microtubule attachment from the structure of an engineered Ndc80 complex. *Cell.* 133:427–439.

70. Hartmann, M. D., C. T. Mendler, …, B. Hernandez Alvarez. 2016. α/β coiled coils. *Elife.* 5:e11861.

71. Ludzia, P., E. D. Lowe, …, B. Akiyoshi. 2021. Structural characterization of KKT4, an unconventional microtubule-binding kinetochore protein. *Structure.* 29:1014–1028.e8.

72. Pellegrino, S., J. Radzimanowski, …, J. Timmins. 2012. Structural and functional characterization of an SMC-like protein RecN: new insights into double-strand break repair. *Structure.* 20:2076–2089.

73. Käshammer, L., J.-H. Saathoff, …, K.-P. Hopfner. 2019. Mechanism of DNA end sensing and processing by the mre11-rad50 complex. *Mol. Cell.* 76:382–394.e6.

Supplementary information for

Single-molecule experiments reveal the elbow as an essential folding guide in SMC coiled coil arms

Marvin Freitag[1], Sigrun Jaklin[1], Francesco Padovani[2], Ecaterina Radzichevici[2], Sarah Zernia[1], Kurt M. Schmoller[2,3] and Johannes Stigler[1,*]

[1] Gene Center Munich, Ludwig-Maximilians-Universität München, Munich, Germany
[2] Institute of Functional Epigenetics, Helmholtz Zentrum München, Neuherberg, Germany
[3] German Center for Diabetes Research (DZD), Neuherberg, Germany
[*] To whom correspondence should be addressed: stigler@genzentrum.lmu.de

This file contains:

- Supplementary figures S1–S10
- Supplementary tables ST1–ST4
- Construct list

Supplementary figures

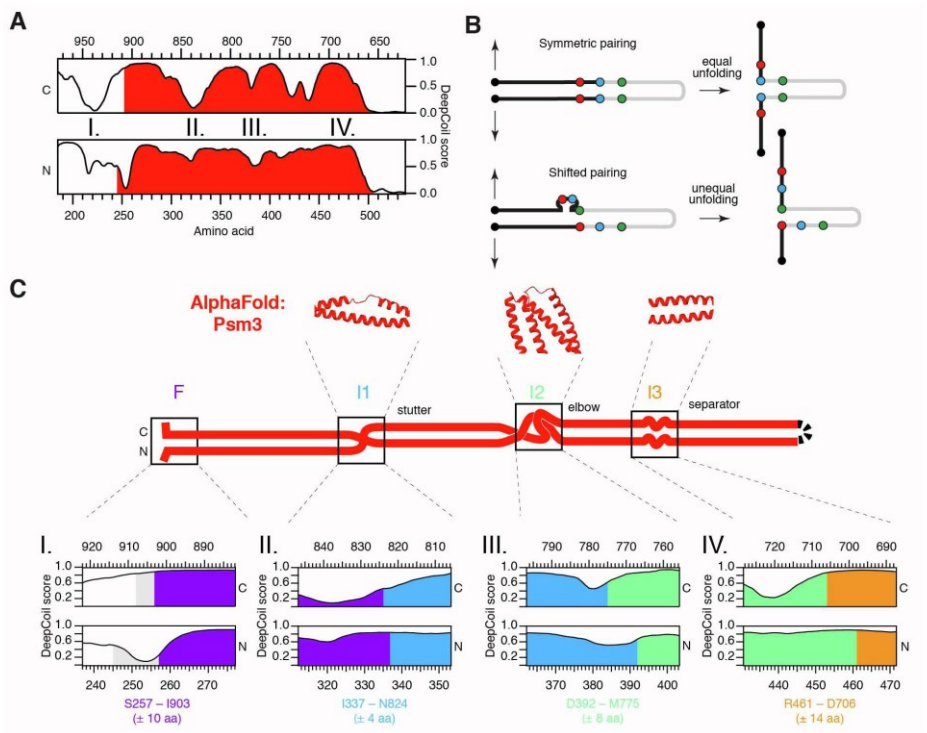

Figure S1. Localization of non-canonical coiled coil features and folding intermediates in the Psm3 coiled coil. **(A)** The DeepCoil profile shows the selected wild-type amino acid sequence from Psm3 (D245-T505 and G658-K908) in red and contains various potential non-canonical coiled coil features. **(B)** Unfolding of coiled coils with a canonical, symmetric pairing enables a straightforward assignment of contour lengths towards amino acids (top). Consequently, the unfolded length (black line) would be distributed equally to the coiled coil strands and agree with the true pairing of blue positions. In contrast, coiled coils containing non-canonical features, like loops or linkers, can result in a shifted pairing between the coiled coil strands, which easily lead to misinterpretations (bottom). The black lines represent the same observed contour length as in the symmetric case. A naïve interpretation of the unfolded length would wrongly conclude that in the protein the blue positions are paired. **(C)** Cartoon of the Psm3 coiled coil architecture highlights different folding features. First (I.), the construct starts with unfolded amino acids (grey) until amino acids are paired at the observed folded state F (purple). Second (II.), the first unfolding intermediates I1 (blue) localizes with a predicted stutter around P831. Here, the CC prediction has an elongated valley. The corresponding AlphaFold prediction shows a short linker within that coiled coil strand, which results in a pairing shift of two amino acids. Third (III.), the second unfolding intermediate I2 (green) is located at the elbow of Psm3, where the corresponding coiled coil prediction shows a minimum. The AlphaFold prediction suggests two short linkers on both coiled coils

at (F385–T387) & (S773–D779) inducing a five amino acids pairing shift from one CC strand to the other. Forth (IV.), the third intermediate I3 (orange) is not correlated to a predicted feature in AlphaFold or in DeepCoil. Based on the unfolding of the WT construct (Fig. 1) and the ΔEB construct (Fig. 4, Fig. S6) this feature splits the coiled coil into the I2-I3 and I3-U sections and was termed “separator”.

3

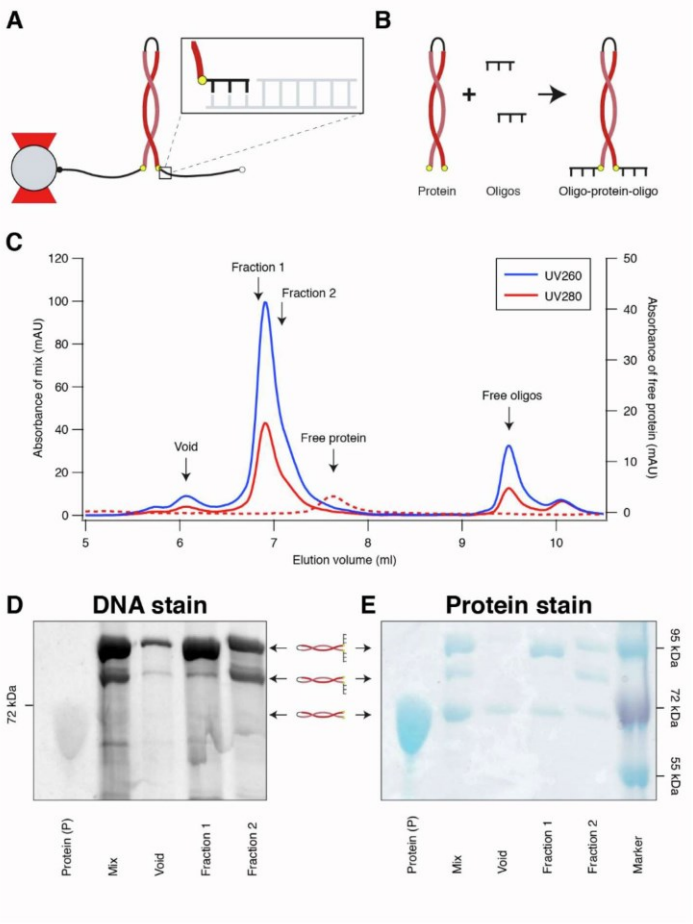

Figure S2. Analysis of the successful addition of 3'-modified oligonucleotides to terminally-tagged Psm3 coiled coils. **(A)** For efficient protein-DNA conjugation, 3'-modified oligonucleotides (oligos) were attached to protein-tags in a site-specific geometry and annealed to DNA handles for further manipulation in the optical tweezers assay. **(B)** CoA-oligos were attached via the magnesium-dependent Sfp reaction to doubly ybbR-tagged proteins (WT, ΔEB, ΔEB_N, ΔEB_C) or maleimide-oligos were attached to reduced terminal cysteines (TR), see Methods. **(C)** A size-exclusion chromatogram of a typical reaction mix of (B) shows a void peak with small amounts of aggregates, a peak with reacted oligo-protein-oligo chimeras (Fraction 1 and 2) and unreacted (free) oligos. Additionally, a run with just the free protein was added for comparison (dashed line). **(D)** SDS-PAGE of the reaction mix and the size-exclusion fractions. The lanes contain unmodified protein, the reaction mix with protein, CoA-oligos, Sfp and magnesium chloride and selected fractions after size-exclusion chromatography. SYBR Gold staining (left) reflects the DNA content of the oligo-protein chimeras. Cartoons indicate the three different protein species in the sample. **(E)** Subsequent Coomassie-based protein staining of the same gel reflects

the protein content of the oligo-protein chimeras. Fraction 1 contains mostly the desired, doubly modified protein suitable for efficient optical tweezers experiments, whereas fraction 2 contains a mix of doubly modified, singly modified and free protein.

5

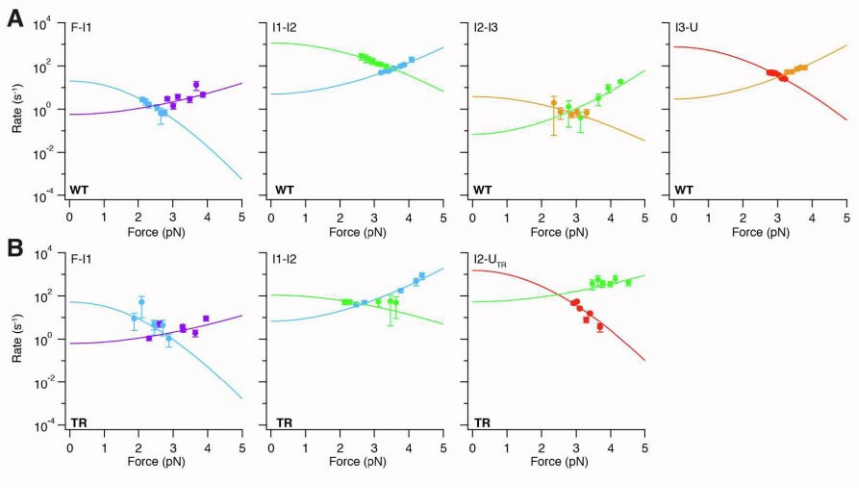

Figure S3. Force-dependent transition rates. Rates were extracted from at least three equilibrium fluctuation traces in passive mode with a constant trap distance for each trace. Error bars indicate the standard deviation of the state's life-time distribution within the corresponding trace. The data points were fitted to a model that describes the contribution due to stretching of the DNA and unfolded polypeptide, as described by polymer models (see Methods, Table ST2). **(A)** Plots for the WT on-pathway transitions. **(B)** Plots for the TR on-pathway transitions.

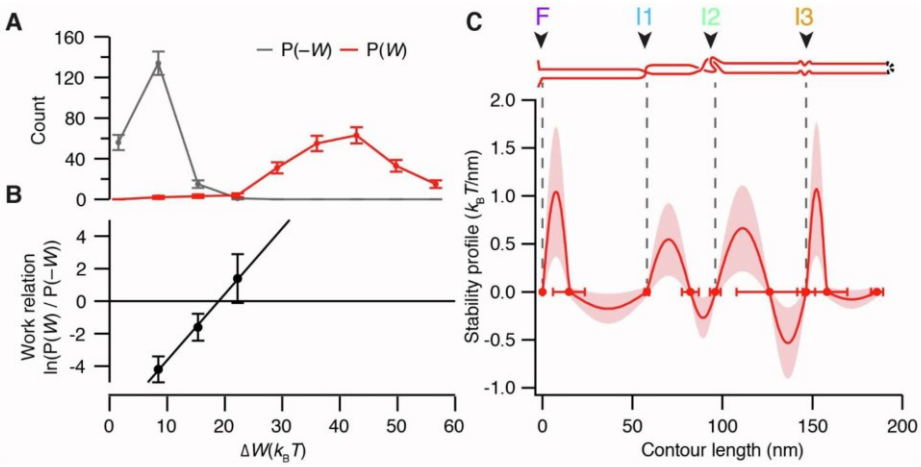

Figure S4. Work histogram and stability profile of WT. (**A**) Distribution of the unfolding (P(W), red, n = 206) and refolding (P(-W), grey, n = 206) work for 500 nm/s force-distance curves of 13 WT molecules. Error bars indicate the square root of the number of counts in each bin. (**B**) Based on the Crooks fluctuation theorem (45) the total free energy is given by the intersection at (19.1 ± 2.8) k_BT estimated by a linear fit of the work relation (ln(P(W)/P(-W))). (**C**) Estimated stability profile based on the interpolation of the fixpoints from table ST1, shaded area estimates the maximal error in y-direction based on ST1.

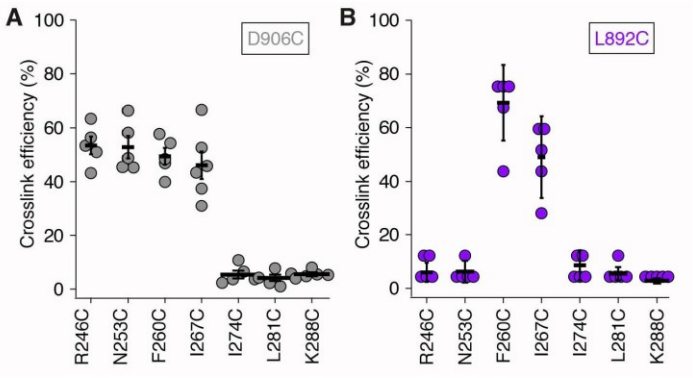

Figure S5. Quantitative analysis of the crosslink efficiencies. **(A)** The crosslink efficiency of the D906C series (grey) was determined based on the SDS-PAGE band intensities, n = 5, error bars indicate standard deviations. **(B)** The crosslink efficiency of the L892C series (purple) was determined based on the SDS-PAGE band intensities, n = 5, error bars indicate standard deviations.

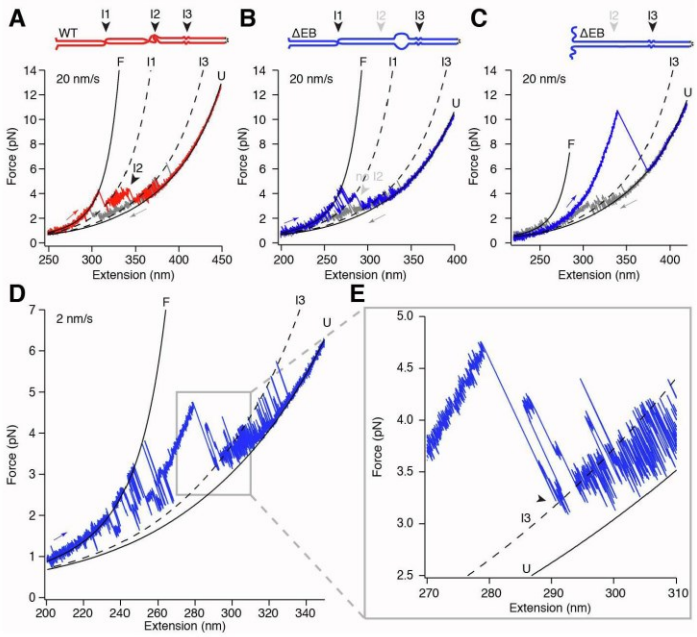

Figure S6. Exemplary force-extension traces of the elbow replacement mutant ΔEB. **(A)** Top: cartoon of the WT construct, bottom: force-extension curve of the WT construct (zoom-out of Fig. 1B) showing several I1-I2 transitions indicated by the arrowhead at a pulling speed of 20 nm/s. Pull: red, subsequent relaxation: grey **(B)** Top: cartoon of the ΔEB construct within a folded configuration at the F-I1 section containing an unstable or unpaired region before I3, bottom: force-extension curve of ΔEB showing a folded F-I1 section with consecutive unfolding into I3. We observed WT-like I3-U transitions but no WT-like I2 intermediate indicated by the arrowhead. Pull: blue, subsequent relaxation: grey **(C)** Top: cartoon of the non-productive metastable ΔEB configurations, which are off-pathway to the folding shown in (B), bottom: exemplary force-extension curve of the ΔEB construct with an unfolded F-I1 section and a metastable intermediate. During relaxation (grey), the protein folds into the native state F. Note that such transitions between the on- and off-pathway behavior occur frequently within one trace and unambiguous assignments are difficult at low tether tensions. **(D)** Force-extension curve of the ΔEB construct at a reduced pulling speed of 2 nm/s showing several misfolded configurations lasting up to 10 seconds and switches between the on- and off-pathway folding behavior. **(E)** Blow-up of (D). The protein escapes the kinetic trap via the I3 intermediate indicated by the arrowhead and not via the unfolded state U. Further native-like I3-U transitions are observed and are separated from the competing folds in the I1-I3 section.

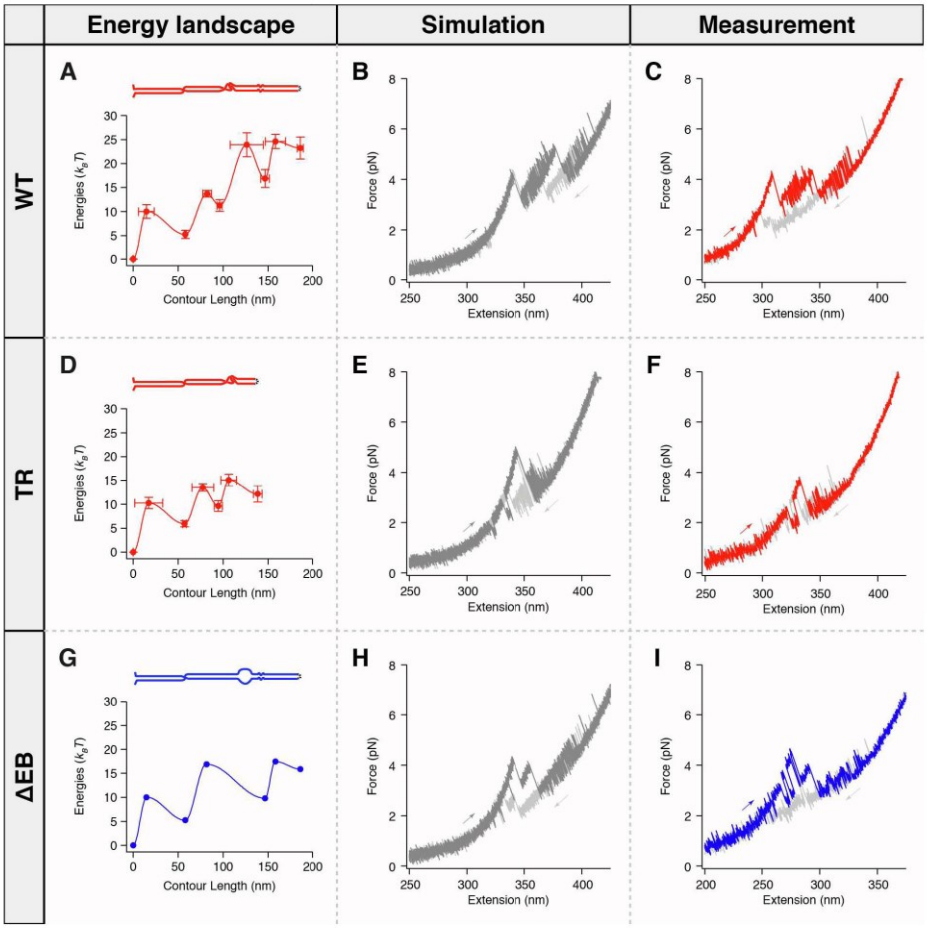

Figure S7. Validation of the energy landscape estimations for the WT, TR and ΔEB constructs. **(A)** top: construct cartoon, bottom: energy landscape of the WT construct (see Fig. 1E). **(B)** Simulated force-extension curve for the WT construct using the energy landscape in (A) showing good agreement with measurements (see C). Pull: dark grey, subsequent relaxation: light grey. **(C)** Measured force-extension curve of the WT construct (see Fig. 1B) at a pulling speed of 20 nm/s. Pull: red, subsequent relaxation: light grey. **(D)** top: construct cartoon, bottom: energy landscape of the TR construct (see Fig. 3D). **(E)** Simulated force-extension curve for the TR construct using the energy landscape in (D) showing good agreement with measurements (see F). Pull: grey, subsequent relaxation: light grey. **(F)** Measured force-extension curve of the TR construct (see Fig. 3B) at a pulling speed of 20 nm/s. Pull: red, subsequent relaxation: light grey. **(G)** top: construct cartoon, bottom: Estimated energy landscape for the ΔEB construct from matching simulated force extension curves (H) to experimental data (I). For details, see Methods. **(H)** Simulated force-extension curve for the ΔEB construct using the energy landscape estimation in (G) showing good agreement with measurements (I). Pull: grey, subsequent relaxation:

light grey. **(I)** Measured force-extension curve of the ΔEB construct (see Fig. S6B) at a pulling speed of 20 nm/s. Pull: blue, subsequent relaxation: light grey.

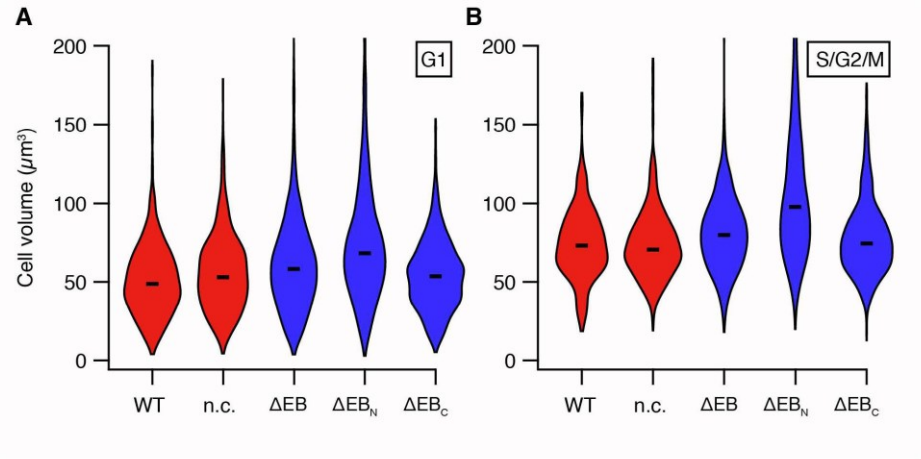

Figure S8. Cell size determination based on microscopy images. The distribution of cell sizes visualizes the size shift caused by the mutations on an individual cell level and verifies the data gained from the Coulter counter (Fig. 5). **(A)** Cell size in G1 phase (unbudded cells). **(B)** Cell size in S/G2/M phases (budded cells). Bars denote medians.

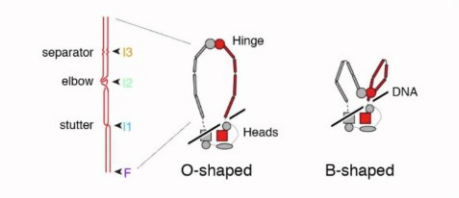

Figure S9. Cartoons showing cohesin's large-scale conformational changes. The cohesin complex undergoes conformational changes including open (O-shaped) and collapsed (B-shaped) conformations. The modular arrangement of the CC supports a model where elbow bending is driven by passive thermal fluctuations of the CC sections identified in this study.

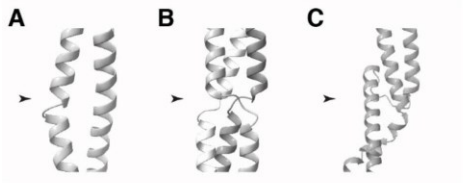

Figure S10. Discontinuities in coiled coils. **(A)** The kinked coiled coil of the transcription elongation factor GreA (PDB: 1GRJ) contains one additional amino acid in the heptad repeat, indicated by the arrow. **(B)** The coiled coil of Omp100 (PDB: 5APP) comprises a non-helical linker region, indicated by the arrow. **(C)** The so-called "knuckles" of the SMC protein MukB in the elbow region (PDB: 6H2X), indicated by the arrow, contain disrupted helices, where a hydrophobic cluster leads to enriched contacts between the coiled coil strands compared to a continuous coiled coil section.

Supplementary tables

Supplementary table ST1. Comparison of the contour lengths and energies of the WT and the TR construct resulting from passive mode measurements in optical tweezers unfolding experiments using a protein persistence length p_p of 0.7 nm, errors indicate standard deviations.

WT	Contour length	ΔG^0 [$k_B T$]	TR	Contour length	ΔG^0 [$k_B T$]
F-I1	(58.0 ± 1.3) nm (159 ± 4) aa	5.2 ± 0.9	F-I1	(57.0 ± 2.0) nm (156 ± 6) aa	5.9 ± 0.7
F-I2	(96.0 ± 2.9) nm (263 ± 8) aa	11.2 ± 1.2	F-I2	(94.7 ± 4.6) nm (260 ± 13) aa	9.6 ± 1.1
F-I3	(146.5 ± 4.9) nm (401 ± 14) aa	16.9 ± 1.9	F-U_{TR}	(138.4 ± 5.0) nm (379 ± 14) aa	12.1 ± 1.7
F-U	(185.9 ± 3.4) nm (509 ± 10) aa	23.2 ± 2.3			
F-I2*	(98.7 ± 3.4) nm (270 ± 10) aa	19.6 ± 1.9			
F-I3*	(146.6 ± 5.1) nm (401 ± 14) aa	22.4 ± 1.4			

Supplementary table ST2. Fitting parameters of the force-dependent folding rates for the WT and TR construct were used to create the energy landscape estimations using **Eq. (5)**, errors indicate standard deviations.

WT	$\log_{10}(k^0_u$ [s^{-1}]), $\log_{10}(k^0_f$ [s^{-1}])	Δx_u [nm], Δx_f [nm]	TR	$\log_{10}(k^0_u$ [s^{-1}]), $\log_{10}(k^0_f$ [s^{-1}])	Δx_u [nm], Δx_f [nm]
F-I1	-0.13 ± 0.37 1.9 ± 0.5	-10.8 ± 4.1 39.6 ± 7.9	F-I1	-0.36 ± 0.47 1.8 ± 0.4	-13.7 ± 7.1 36.3 ± 14.0
I1-I2	0.43 ± 0.24 2.9 ± 0.2	-23.7 ± 3.9 13.5 ± 4.0	I1-I2	0.75 ± 0.29 2.4 ± 0.2	-24.9 ± 10.0 12.6 ± 7.2
I2-I3	-0.94 ± 0.95 0.55 ± 0.54	-23.3 ± 15.1 13.5 ± 10.6	I2-U_{TR}	1.5 ± 0.4 3.0 ± 0.4	-12.2 ± 5.2 26.5 ± 7.2
I3-U	0.71 ± 0.32 3.5 ± 0.6	-14.0 ± 7.2 30.3 ± 8.8			

Supplementary table ST3. Yeast strains used in this work. All strains are based on W303 and CglaTRP1 refers to the TRP1 gene of the organism C. glabrata. Strain WT_SMC3_ ΔEBC_UTR_Trp1 contains a cytosine point mutation in the UTR on position 33. For this strain, one out of five independent clones showed a dramatically increased cell volume larger than 150 μm^3, most likely caused by a secondary mutation. This clone was excluded from the analysis shown in **Fig. 5C**.

Name	Descriptive Name	Genotype	Origin
MMY116-2C	WT	*Mat α; ADE2*	Skotheim lab stock
FPY17AH-6,8	WT_SMC3_UTR_Trp1	*Mat α; ADE2, smc3::smc3-CglaTRP1*	This study
FPY17BH-1,6	SMC3_ΔEB$_N$_UTR_Trp1	*Mat α; ADE2, smc3::smc3-ΔEB$_N$-CglaTRP1*	This study
FPY17CH-1,4,41	SMC3_ΔEB$_C$_UTR_Trp1	*Mat α; ADE2, smc3::smc3-ΔEB$_C$-CglaTRP1*	This study
FPY17DH-2,36,37,38	SMC3_ΔEB_UTR_Trp1	*Mat α; ADE2, smc3::smc3-ΔEB-CglaTRP1*	This study

Supplementary table ST4. Primers

Name	Descriptive name	Sequence
FPP006	SMC3_3_R	CCGCTAGTGCTGCTTTAGGTAAGAAGAAGAAGCCAAGTGG TGGATTTGCATCATTAATAAAAGATTTCAAGAAAAAATGAGA ATTCGAGCTCGTTTAAAC
FPP007	SMC3_F	ATGTATATCAAAAGGGTGATAATTAAGG
FPP008	SMC3_seq_F	GCGACGCGTTAGGAAATGAT
FPP009	SMC3_check_R	TGACTCTAACTCCAGTTCGGAC
FPP010	SMC3_seq_R	CGAAGAGAAGTCTCAATGAGGT
FPP011	SMC3_seq_F_1	ACCAGTTCACATTGTATGATCG
FPP012	SMC3_seq_F_2	TGACGTCAATCAAAATCAAAGAA
FPP013	SMC3_seq_F_3	AGAAAAGCTGAATACAAATCGT
FPP014	SMC3_seq_F_4	AAAACAACAAAAGGTTAACGCG
FPP015	SMC3_seq_F_5	CTTTTCTTAAAGAAGTTGCCCG
FPP016	SMC3_seq_R_1	AATCAAAGTGGCCGTTTCTTCG
FPP017	SMC3_seq_R_2	TAAAGCATCCTCTGGCAGAAGC
FPP018	SMC3_overlapM3_F	CAAATGCAAAGAATCAATTCTGAA
FPP019	SMC3_overlapM4_F	GATTTGCTTACAGATAGCGAAC

Construct list

WT^{Psm3}: 6xHIS-ybbR-D245-T505-GGGSAGGSGSGSSGGS-G658-K908-ybbR

MGSSHHHHHHSSGDSLEFIASKLAGSDRIAALERNDDDSGAFIQREERIERIKAEITELNHSLELLRVEK
QQNDEDYTNIMKSKVALELQSSQLSRQIEFSKKDESSKLNILSELESKISEKENELSEILPKYNAIVSEAD
DLNKRIMLLKNQKQSLLDKQSRTSQFTTKKERDEWIRNQLLQINRNINSTKENSDYLKTEYDEMENEL
KAKLSRKKEIEISLESQGDRMSQLLANITSINERKENLTDKRKSLWREEAKLKSSIENVKDDLSRSEKAL
GTTMDRNTGGGSAGGSGSGSSGGSGYRDYRNSRLDAIKNVKTYQIKFSDLQESLEKCRSEIESFDQK
ITACLDDLQKAQLSLKQFERDHIPLKDELVTITGETTDLQESMHHKSRMLELVVLELHTLEQQANDLKS
ELSSEMDELDPKDVEALKSLSGQIENLSHEFDAIIKERAHIEARKTALEYELNTNLYLRRNPLKAEIGSD
NRIDESELNSVKRSLLKYENKLQIIKSSSSGLEEQMQRINSEISDKRNELESLEELQHEVATRIEQDAKG
SDSLEFIASKLA

Green highlighted amino acids were mutated to cysteines for the crosslinking experiments (**Fig. 2**). Red amino acids were exchanged in the ΔEB^{Psm3} construct (**Fig. 4**).

TR^{Psm3}: 6xHIS-KKCK-D245-G445-GGGSAGGSGSGSSGGS-T733-K908-KKCK

MGSSHHHHHHSSGKKCKGSDRIAALERNDDDSGAFIQREERIERIKAEITELNHSLELLRVEKQQNDE
DYTNIMKSKVALELQSSQLSRQIEFSKKDESSKLNILSELESKISEKENELSEILPKYNAIVSEADDLNKRI
MLLKNQKQSLLDKQSRTSQFTTKKERDEWIRNQLLQINRNINSTKENSDYLKTEYDEMENELKAKLSR
KKEIEISLESQGGGGSAGGSGSGSSGGSTGETTDLQESMHHKSRMLELVVLELHTLEQQANDLKSEL
SSEMDELDPKDVEALKSLSGQIENLSHEFDAIIKERAHIEARKTALEYELNTNLYLRRNPLKAEIGSDNRI
DESELNSVKRSLLKYENKLQIIKSSSSGLEEQMQRINSEISDKRNELESLEELQHEVATRIEQDAKGSK
KCK

ΔEB^{Psm3}: 6xHIS-ybbR-D245-Q384-QMQRINSEISD-I395-T505-GGGSAGGSGSGSSGGS-G658-D779-K781-K908-ybbR

MGSSHHHHHHSSGDSLEFIASKLAGSDRIAALERNDDDSGAFIQREERIERIKAEITELNHSLELLRVEK
QQNDEDYTNIMKSKVALELQSSQLSRQIEFSKKDESSKLNILSELESKISEKENELSEILPKYNAIVSEAD
DLNKRIMLLKNQKQSLLDKQSRTSQQMQRINSEISDIRNQLLQINRNINSTKENSDYLKTEYDEMENEL
KAKLSRKKEIEISLESQGDRMSQLLANITSINERKENLTDKRKSLWREEAKLKSSIENVKDDLSRSEKAL
GTTMDRNTGGGSAGGSGSGSSGGSGYRDYRNSRLDAIKNVKTYQIKFSDLQESLEKCRSEIESFDQK
ITACLDDLQKAQLSLKQFERDHIPLKDELVTITGETTDLQESMHHKSRMLELVVLELHTLEQQANDLKS
ELSSEMDELD_KDVEALKSLSGQIENLSHEFDAIIKERAHIEARKTALEYELNTNLYLRRNPLKAEIGSD
NRIDESELNSVKRSLLKYENKLQIIKSSSSGLEEQMQRINSEISDKRNELESLEELQHEVATRIEQDAKG
SDSLEFIASKLA

Red amino acids were exchanged in the ΔEB^{Psm3} construct (**Fig. 4**) and _ denotes the P780 deletion. ΔEB$^{Psm3}{}_{N}$ contains only the first and ΔEB$^{Psm3}{}_{C}$ contains only the second mutation.

ΔEB^{Smc3}:

MYIKRVIIKGFKTYRNETIIDNFSPHQNVIIGSNGSGKSNFFAAIRFVLSDDYSNLKREERQGLIHQGSG
GSVMSASVEIVFHDPDHSMILPSGVLSRGDDEVTIRRTVGLKKDDYQLNDRNVTKGDIVRMLETAGFS
MNNPYNIVPQGKIVALTNAKDKERLQLLEDVVGAKSFEVKLKASLKKMEETEQKKIQINKEMGELNSKL
SEMEQERKELEKYNELERNRKIYQFTLYDRELNEVINQMERLDGDYNNTVYSSEQYIQELDKREDMID
QVSKKLSSIEASLKIKNATDLQQAKLRESEISQKLTNVNVKIKDVQQQIESNEEQRNLDSATLKEIKSIIE
QRKQKLSKILPRYQELTKEEAMYKLQLASLQQKQRDLILKKGEYARQMQRINSEISDIHSEIEELKSSIQ
NLNELESQLQMDRTSLRKQYSAIDEEIEELIDSINGPDTKGQLEDFDSELIHLKQKLSESLDTRKELWRK
EQKLQTVLETLLSDVNQNQRNVNETMSRSLANGIINVKEITEKLKISPESVFGTLGELIKVNDKYKTCAE
VIGGNSLFHIVVDTEETATLIMNELYRMKGGRVTFIPLNRLSLDSDVKFPSNTTTQIQFTPLIKKIKYEPR
FEKAVKHVFGKTIVVKDLGQGLKLAKKHKLNAITLDGDRADKRGVLTGGYLDQHKRTRLESLKNLNES
RSQHKKILEELDFVRNELNDIDTKIDQVNGNIRKVSNDRESVLTNIEVYRTSLNTKKNEKLILEESLNAIIL
KLEKLNTNRTFAQEKLNTFENDLLTDSELSKEEKERLESLTKEISAAHNKLNITSDALEGITTTIDSLNAE
LESKLIPQENDLESKMSEVGDAFIFGLQDELKELQLEKESVEKQHENAVLELGTVQREIESLIAEETNNK
KLLEKANNQQRLLLKKLDNFQKSVEKTMIKKTTLVTRREELQQRIREIGLLPEDALVNDFSDITSDQLLQ
RLNDMNTEISGLKNVNKRAFENFKKFNERRKDLAERASELDESKDSIQDLIVKLKQQKVNAVDSTFQK
VSENFEAVFERLVPRGTAKLIIHRKNDNANDHDESIDVDMDAESNESQNGKDSEIMYTGVSISVSFNS
KQNEQLHVEQLSGGQKTVCAIALILAIQMVDPASFYLFDEIDAALDKQYRTAVATLLKELSKNAQFICTT
FRTDMLQVADKFFRVKYENKISTVIEVNREEAIGFIRGSNKFAEV

Red amino acids were exchanged in the ΔEB^{Smc3} construct compared to wild-type. ΔEB$^{Smc3}{}_{N}$ contains only the first and ΔEB$^{Smc3}{}_{C}$ contains only the second mutation (**Fig. 5**).

III.2.1 Appendix

In this appendix, I will describe the mass spectrometry-based analysis of the different bands, which occurred during the crosslinking series in chapter III.2. During SDS-PAGE, the crosslinked samples showed an additional band, which suggests that this band could be caused by the crosslinking. As the coiled-coil constructs used in chapter III.2 consist of single-peptide chains, a clear mass shift as in other studies (148) cannot be expected. To verify that the additional band corresponds to the intended crosslink with more certainty, I performed a peptide identification based on tryptic digestion and mass spectrometry (MS).

First, the bands were dissected and subjected to in-gel protein digestion with trypsin, which cleaves at lysine or arginine residues (Fig. 17). The resulting samples for MS contain a vast variety of different peptides. To identify individual peptides, a theoretically cleavage pattern (including missed cleavage events) is calculated based on the amino acid sequence, which generates a list of peptides, which can be compared with the hits of the MS spectrum and filtered for possible crosslinking candidates (Fig. 17). For BMOE, all peptides containing cysteines or lysines can be considered (149). To identify crosslinked peptides, the mass shift due to the crosslinker has to be added to the masses of both peptides. During the tryptic digestion (14h at pH 8.8), the maleimides of BMOE were hydrolyzed, which resulted in a combined mass shift of 256.07 Da (molecular formula C10H12N2O6) for crosslinked peptides.

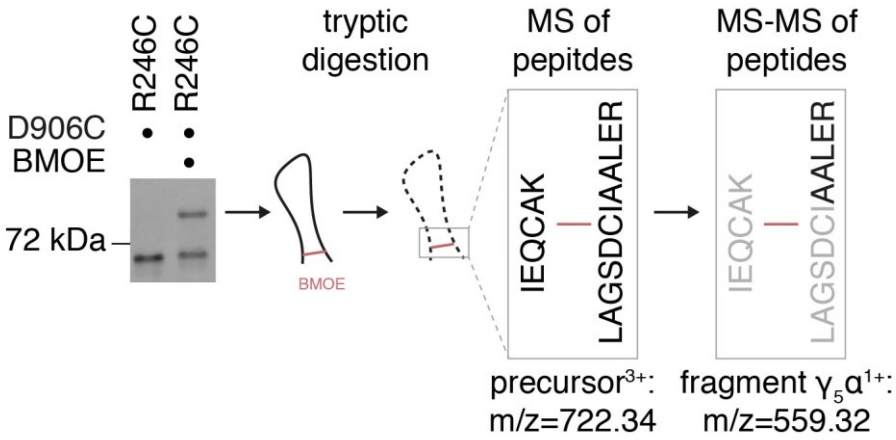

Figure 17: Peptide fragmentation. For the Psm3 R246C D906C construct, the SDS-PAGE bands were isolated, digested and identified using tandem mass spectroscopy (MS-MS). During the first mass detection abundant peptides (precursor) were successively selected for further fragmentation, followed by a second mass detection of the fragments (e.g. $\gamma_5\alpha^{1+}$, see Figure 18 for the corresponding MS-MS spectrum). A peptide can be identified with high confidence if the precursor mass and the mass of different fragments agree with a candidate of the theoretically cleavage pattern.

For the identification of crosslinked peptides a tandem MS approach (also called MS-MS) was used on a high-resolution instrument. First, masses of the digested peptides were determined. Then, abundant peptides were successively selected for further fragmentation by collision with nitrogen molecules (higher-energy C-trap dissociation: HCD) (150). Lastly, fragments were analyzed by a second mass

determination. If the precursor mass agrees with the MS spectrum of the fragments (Fig. 18), the detected peptide can be identified with high confidence using suitable analysis software (151). For the Psm3 R246C D906C construct, 6 precursors could be significantly identified, which contained the desired crosslink (FDR <1%, mass deviation 10.0 ppm).

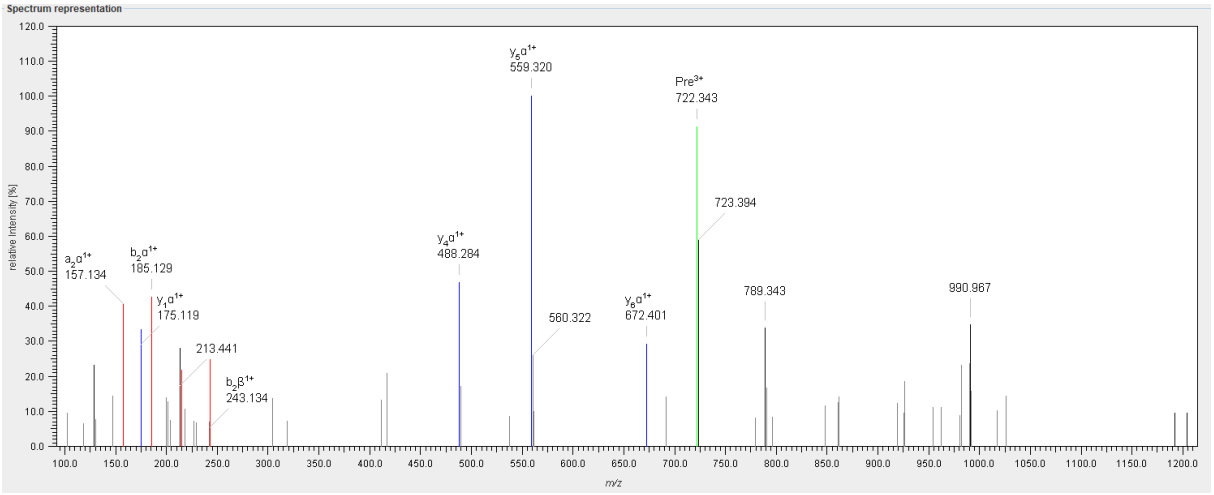

Figure 18: Representative MS-MS spectrum of a detected crosslinked peptide. The precursor (green, Pre^{3+}) is three-times positively charged. The corresponding fragments are highlighted in blue or red for the different peptides. The nomenclature indicates which amino acids are missing in the fragment (e.g. $\gamma_5\alpha^{1+}$ indicates that this fragment contains the last five amino acids of the α-peptide and is charged with one proton, see Figure 17).

In summary, the observed upper band represents the crosslinked peptide chain, which was verified by peptide identification using tryptic digestion and tandem mass spectroscopy. Although circular peptide chains would be expected to migrate faster than linear ones because of their shape (analogously to circular and linear DNA), it is likely that the crosslink inhibited the accumulation of SDS to the denatured protein, which would reduce the net charge of the molecule. This would further inhibit its mobility inside the gel and could lead to the observed separation of crosslinked and non-crosslinked proteins.

III.3 Purification and engagement of SMC head domain complexes as single-peptide chains

Freitag, M., J. Deplazes-Lauber, J. Probst, S. Zernia and J. Stigler (to be published).

Summary:

During the last years, the *in vitro* reconstitution of loop extrusion and the elbow bending motion of cohesin and condensin shifted the focus of the SMC field from the coiled coils and discussions about the functionality of rod-like or circular shapes towards the fragile regulation of interfaces and conformational changes during the ATPase cycle, which guarantee actively extruding complexes.

In this chapter, the purification strategy of individual head domains of cohesin as well as different fusion constructs with the respective kleisin domains as single-peptide chain constructs are described. The utilization of single-peptide chains facilitates the repeatable manipulation of interfaces in the optical tweezers assay by preventing the dissociation of binding partners during the measurement. A three-step purification procedure ensured pure proteins, which were able to form active ATPase complexes. Additionally, an analysis of the oligomeric state of individual head domains will be presented using size-exclusion chromatography, mass photometry and site-specific crosslinking.

Future unfolding experiments of these different constructs promise to provide information about the SMC head engagement, which is involved in loop extrusion. A second construct provides the opportunity to study the stability of the neck gate, which is involved in DNA release.

Author contribution:

I designed the single-peptide chain constructs of the SMC head domains and fusion constructs with the kleisin domains. I have established the purification strategy of these construct with the help of Joelle Deplazes-Lauber and Jeanny Probst. I performed the bulk experiments with the help of Sarah Zernia and wrote the manuscript with Sarah Zernia and Johannes Stigler.

III.3.1 Abstract

Since loop extrusion relies on multiple DNA binding sites, which move relatively towards each other, the engagement of the SMC head domains (SMC-HD) could likely provide a transient DNA binding site, which is formed during the ATP hydrolysis cycle. Single-molecule unfolding studies of engaged SMC-HDs promise to reveal the stability of the engagement interface and could characterize conformational changes inside the SMC-HDs during the ATPase cycle or interaction of binding partners. Therefore, the manufacturing of a single-peptide chain containing the Smc3-HD, the Smc1-HD and the C-terminal Scc1 domain (cScc1), which is necessary for engagement, will be presented in this chapter. During the purification, the Smc1-HD-cScc1 constructs turned out to form homodimers, which possess basal ATPase activity. The addition of ATP and the Smc3-HD led to the formation of Smc1/Smc3 heterodimers. The Smc3-HD constructs are monomeric proteins and equipped with a SpyTag, which can be coupled to the SpyCatcher domain of the Smc1-HD-cScc1 fusion construct to create an active single-peptide chain ATPase subcomplex. The linkage of the HDs resulted in an increased ATPase activity of the single-peptide chain, which suggests that the ATPase activity of cohesin is stimulated if the HDs are kept in close proximity. Further constructs, which introduced increased linker lengths between the domains of the fusion constructs, were found to form oligomeric structures. The observed oligomerization of the fusion constructs suggests that clustering of cohesin complexes can happen in the apo state, where no ATP is bound.

In addition, the purification of the neck gate (Smc3-HD-nScc1) constructs is described. The neck gate is thought to be relevant for the release of entrapped DNA from cohesin. Single-molecule experiments promise to reveal how sister chromatid cohesion and cohesin unloading is balanced. Although the expression of the Smc3-HD-nScc1 construct is low and truncated versions are occurring, the construct could be purified in a sufficient amount and purity from *E. coli*.

III.3.2 Introduction

The SMC-HDs are highly conserved and harbor the ATP binding sites (1). For yeast cohesin (Smc1/Smc3), the heterodimerization of the Smc3-HD with the Smc1-HD, which is also called engagement, is dependent on ATP binding and the presence of the C-terminal kleisin domain (cScc1), which primes the Smc1-HD for further interactions (37, 39). The engaged heads are able to hydrolyze ATP in the complex and as a truncated subcomplex (37, 59). Previously, the Smc1-HD-cScc1 subcomplex was reported to form homodimers (42), whose configuration reminds of the engaged Smc1/Smc3 heterodimer, but bears a reduced ATPase activity in comparison to the heterodimer and the full cohesin complex. In contrast, the Smc3-HD cannot form homodimers, which is independent of kleisin binding. Smc3-HD can bind, but not hydrolyze ATP on its own (40). Interestingly, the Smc2-HD of condensin (Smc2/Smc4) from *Chaetomium thermophilum* (*C. thermophilum*), which is closely related to cohesin's Smc3-HD, is not able to bind ATP (38). Only the engagement with the Smc4-HD bound to the corresponding C-terminal kleisin domain induces a conformational change in the Smc2-HD, which allows ATP binding and further hydrolysis. A homodimerization of condensin SMC-HDs was not found (38).

Bacterial SMC proteins, like MukB, form regularly homodimers. MukB dimers have a low ATPase activity, which can be boosted in the presence of either the N-terminal or the C-terminal MukF kleisin domain (152). The addition of both individual domains increases the ATPase activity further to levels of full MukBF complexes. The ATPase activity of full cohesin complexes is only slightly higher than the activity of the isolated Smc1/Smc3 heterodimer (37, 39, 153), which changes dramatically in the presence of DNA and the loader complex Scc2/Scc4 (154). This implies that the active sites inside cohesin's HDs need further conformational changes, in contrast to bacterial MukBF, to achieve a processive state. Additionally, the MukBEF complex is able to form monomeric complexes as well as dimeric complexes, which are reported to be the functional form (155). For eukaryotic SMC complexes, dimerization is still under debate. While loop extruding cohesin and condensin were reported to be monomeric complexes (9, 94), Smc5/6 was recently found to perform loop extrusion exclusively as dimers (95). Furthermore, clustering have been proposed for cohesin in the context of sister chromatid cohesion (156).

To conclude, the ATPase of SMC complexes is subject to a multi-layered regulation including the control of ATP binding, dimerization and further conformational changes, which propagate to the active sites to improve the ATP turnover (1, 37, 39). How these processes affect each other and enable loop extrusion is still not known in detail. Single-molecule unfolding studies of engaged HDs promise to reveal the influence of conformational changes on the HDs by e.g. ATP binding. Moreover, the stability of the engagement interface, which additionally forms a DNA binding site involved in loop extrusion, can be quantified to investigate the usage of energy gained by ATP hydrolysis. Hence, retracing the cascade of different conformational changes will lead to a mechanistic understanding of loop extrusion.

Besides engagement, the Smc1- and Smc3-HDs are involved in the regulation of the neck gate, which is the interface between the coiled coils above the Smc3-HD and the helical N-terminal kleisin domain (nScc1) (21). For isolated Smc3-HD-nScc1 subcomplexes, the interface was found to be closed (nScc1 bound to the Smc3-HD) (40), whereas after engagement with the Smc1-HD-cScc1 complex the neck gate was found to be opened (nScc1 not bound to the Smc3-HD) (37, 117). On the contrary, the neck gate is closed in the gripping state (44–46, 112), if DNA and loader are bound to the engaged SMC-HDs, which raises questions about the sequence of events. Do the heads engage first, which opens the neck gate and later the loader re-closes the neck gate in the gripping state? Or does the loader complex clamp the neck gate in the closed conformation, which is now able to persist the conformational changes induced by the engagement? Besides that, the potential opening and re-closing of the neck gate suggests for cohesive cohesin, in which the loader complex had dissociated and DNA is topologically bound inside the SMC lumen (1) that either the HDs cannot be engaged or that another protein, potentially Pds5, clamps the neck gate similarly to the loader complex in the closed conformation (61). Since latest research suggests that SMC complexes can perform loop extrusion without the opening of any interface (6), the opening of the neck gate must be avoided not only in cohesive cohesin but also during loop extrusion to prevent DNA release. To investigate the mechanical stability of this interface and track the opening and closure of the neck gate through the binding of the Smc1-HD, I created a single-peptide chain construct of the Smc3-HD-nScc1 subcomplex suitable for manipulations in the optical tweezers assay. This construct can be covalently bound to a Smc1-HD using a SpyCatcher

domain to investigate the regulation of the neck gate during engagement in the optical tweezers assay and additionally the influence of different Hawk proteins or nucleotides could be investigated.

In the following, I present the successful purification of SMC-HDs and the creation of single-peptide constructs, which allow manipulation in the optical tweezers instrument. In addition, I was able to prove the engagement of the Smc3-HD and the Smc1-HD-cScc1 subcomplex, which counteracts unwanted homodimerization of the Smc1-HD-cScc1 subcomplex, and I characterized the oligomeric state and the ATPase activity of different HD constructs.

III.3.3 Methods

Construct design

To generate the SMC-HD constructs, I obtained plasmids from the lab of Daniel Panne (37), which contained the Smc3-HD, the N- and C-terminal domains of Scc1 from budding yeast and the Smc1-HD from *C. thermophilum*, which can be expressed in *E. coli*. These constructs contain cysteine point mutations (Smc1-HD L1160C A1201C and Smc3-HD A1159C N1204C), which can be used to crosslink the engagement interface.

The Smc3-HD construct (highlighted in red) includes amino acids 2-262 linked with ESSKHPTSLVPRGSS to 971-1230 (40) with additionally added terminal ybbR-tags (DSLEFIASKLA) and a His-tag (HHHHHH) for purification. For a second construct the C-terminal ybbR-Tag was exchanged with a SpyTag003 (RGVPHIVMVDAYKRYK).

MGSSHHHHHHSSGDSLEFIASKLAGSMAYIKRVIIKGFKTYRNETIIDNFSPHQNVIIGCNGSGKSNFFA
AIRFVLSDDYSNLKREERQGLIHQGSGGSVMSASVEIVFHDPDHSMILPSGVLSRGDDEVTIRRTVGL
KKDDYQLNDRNVTKGDIVRMLETAGFSMNNPYNIVPQGKIVALTNAKDKERLQLLEDVVGAKSFEVKL
KASLKKMEETEQKKIQINKEMGELNSKLSEMEQERKELEKYNELERNRKIYQFTLYDRELNEVINQME
RLDGDYNNTVYSSESSKHPTSLVPRGSSDITSDQLLQRLNDMNTEISGLKNVNKRAFENFKKFNERRK
DLAERASELDESKDSIQDLIVKLKQQKVNAVDSTFQKVSENFEAVFERLVPRGTAKLIIHRKNDNANDH
DESIDVDMDAESNESQNGKDSEIMYTGVSISVSFNSKQNEQLHVEQLSGGQKTVCAIALILAIQMVDPA
SFYLFDEIDACLDKQYRTAVATLLKELSKNAQFICTTFRTDMLQVADKFFRVKYECKISTVIEVNREEAI
GFIRGSNKFAEVSDSLEFIASKLA

The neck gate construct includes an N-terminal SpyTag002 (VPTIVMVDAYKRYK), the sequence of the Smc3-HD construct (red), a C-terminal linker including a ybbR-tag, two Flag-tags (DYKDDDDK) and a His-tag followed by the N-terminal domain of Scc1 (amino acid 1-115, green) and an additional C-terminal ybbR-tag (adopted from (59)). This fusion construct generated a single-peptide construct of the Smc3-HD-nScc1 subcomplex, which contains attachment sites to manipulate the neck gate interface.

MAVPTIVMVDAYKRYKYIKRVIIKGFKTYRNETIIDNFSPHQNVIIGCNGSGKSNFFAAIRFVLSDDYSNL
KREERQGLIHQGSGGSVMSASVEIVFHDPDHSMILPSGVLSRGDDEVTIRRTVGLKKDDYQLNDRNV
TKGDIVRMLETAGFSMNNPYNIVPQGKIVALTNAKDKERLQLLEDVVGAKSFEVKLKASLKKMEETEQ
KKIQINKEMGELNSKLSEMEQERKELEKYNELERNRKIYQFTLYDRELNEVINQMERLDGDYNNTVYS
SESSKHPTSLVPRGSDITSDQLLQRLNDMNTEISGLKNVNKRAFENFKKFNERRKDLAERASELDESK

DSIQDLIVKLKQQKVNAVDSTFQKVSENFEAVFERLVPRGTAKLIIHRKNDNANDHDESIDVDMDAESN
ESQNGKDSEIMYTGVSISVSFNSKQNEQLHVEQLSGGQKTVCAIALILAIQMVDPASFYLFDEIDACLD
KQYRTAVATLLKELSKNAQFICTTFRTDMLQVADKFFRVKYECKISTVIEVNREEAIGFIRGSNKFAEVG
SDSLEFIASKLADYKDDDDKHMDYKDDDDKHHHHHHMAKLGDIEFIKVNKMVTENPQRLTVLRLATNK
GPLAQIWLASNMSNIPRGSVIQTHIAESAKEIAKASGSDDESGDNEYITLRTSGELLQGIVRVYSKQATF
LLTDIKDTLTKISMLFKTSQKMTSTVNGSDSLEFIASKLA

The Smc1-HD (2-242 linked with ESSKHPTSLVPRGSS to 1058-1264, blue) was cloned analogously to the Smc3-HD with terminal ybbR-tags and a His-tag. For another construct the C-terminal ybbR-tag was replaced with an NdeI restriction site and the SpyCatcher003 domain (HMVTTLSGLSGEQGPSGDMTTEEDSATHIKFSKRDEDGRELAGATMELRDSSGKTISTWISDGHVK
DFYLYPGKYTFVETAAPDGYEVATPIEFTVNEDGQVTVDGEATEGDAHT).

MGSSHHHHHHSSGDSLEFIASKLAGSGKLIRLELFNFKSYKGHHTLLFGDSYFTSIIGPNGSGKSNSM
DAISFVLGIKSSHLRSSNLRDLIYRGRVMKTSKIQDDGTTAPATNGDVNGYENGDAGDDEDTSQRTSR
NDPKTAWVMAVYEDDAGELHRWKRTITANGTSEYRINDRVVNAQQYNEALEKENILIKARNFLVFQG
DVEAIASQSPQDLTRLIEQISGSLEYKEEYERLEEEVRQATEEQAYKLQRRRAANSEIKQYMEQKSSS
KHPTSLVPRGSNPNLRAMDRLDHVRKQLEQTEQEFEASKAKLRQARESFQAVKQKRLELFNKAFTHI
QEQITHVYKELTRSEAYPLGGQAYLDIEEDTDTPFLSGVKYHAMPPCKRFRDMEHLSGGEKTMAALA
LLFAIHSYQPSPFFVLDEVDCALDNANVEKIKKYIREHAGPGMQFIVISLKPALFQASESLIGVYRDQEA
NTSRTLTLDLRKYRGSDSLEFIASKLA

The Smc1-HD-cScc1 subcomplex was cloned as a single-peptide chain, where a BamHI restriction site, the C-terminus of Scc1 (482-564, green) and a SnoopTag (KLGDIEFIKVNK) replaced the C-terminal ybbR-tag.

MGSSHHHHHHSSGDSLEFIASKLAGSGKLIRLELFNFKSYKGHHTLLFGDSYFTSIIGPNGSGKSNSM
DAISFVLGIKSSHLRSSNLRDLIYRGRVMKTSKIQDDGTTAPATNGDVNGYENGDAGDDEDTSQRTSR
NDPKTAWVMAVYEDDAGELHRWKRTITANGTSEYRINDRVVNAQQYNEALEKENILIKARNFLVFQG
DVEAIASQSPQDLTRLIEQISGSLEYKEEYERLEEEVRQATEEQAYKLQRRRAANSEIKQYMEQKSSS
KHPTSLVPRGSNPNLRAMDRLDHVRKQLEQTEQEFEASKAKLRQARESFQAVKQKRLELFNKAFTHI
QEQITHVYKELTRSEAYPLGGQAYLDIEEDTDTPFLSGVKYHAMPPCKRFRDMEHLSGGEKTMAALA
LLFAIHSYQPSPFFVLDEVDCALDNANVEKIKKYIREHAGPGMQFIVISLKPALFQASESLIGVYRDQEA
NTSRTLTLDLRKYRGSSKAIVQMAKILRKELSEEKEVIFTDVLKSQANTEPENITKREASRGFFDILSLAT
EGCIGLSQTEAFGNIKIDAKPALFERFIKLGDIEFIKVNK

An additional construct added the NdeI restriction site and the SpyCatcher003 domain (orange) between the C-terminus Scc1 domain and the SnoopTag, which results in the Smc1-HD-cScc1-SpyCatcher003 construct.

MGSSHHHHHHSSGDSLEFIASKLAGSGKLIRLELFNFKSYKGHHTLLFGDSYFTSIIGPNGSGKSNSM
DAISFVLGIKSSHLRSSNLRDLIYRGRVMKTSKIQDDGTTAPATNGDVNGYENGDAGDDEDTSQRTSR
NDPKTAWVMAVYEDDAGELHRWKRTITANGTSEYRINDRVVNAQQYNEALEKENILIKARNFLVFQG
DVEAIASQSPQDLTRLIEQISGSLEYKEEYERLEEEVRQATEEQAYKLQRRRAANSEIKQYMEQKSSS

KHPTSLVPRGSNPNLRAMDRLDHVRKQLEQTEQEFEASKAKLRQARESFQAVKQKRLELFNKAFTHI
QEQITHVYKELTRSEAYPLGGQAYLDIEEDTDTPFLSGVKYHAMPPCKRFRDMEHLSGGEKTMAALA
LLFAIHSYQPSPFFVLDEVDCALDNANVEKIKKYIREHAGPGMQFIVISLKPALFQASESLIGVYRDQEA
NTSRTLTLDLRKYRGSSKAIVQMAKILRKELSEEKEVIFTDVLKSQANTEPENITKREASRGFFDILSLAT
EGCIGLSQTEAFGNIKIDAKPALFERFIHMVTTLSGLSGEQGPSGDMTTEEDSATHIKFSKRDEDGREL
AGATMELRDSSGKTISTWISDGHVKDFYLYPGKYTFVETAAPDGYEVATPIEFTVNEDGQVTVDGEAT
EGDAHTKLGDIEFIKVNK

Additionally, versions with increased linker lengths (GSGGSGGSGGSGGSGS introduced at the BamHI restriction site and HNGGSGGSGGSGGSHM introduced at NdeI restriction site) were designed to account for the artifacts during SDS-PAGE, which created Smc1-HD-link-cScc1-link-SpyCatcher003.

MGSSHHHHHHSSGDSLEFIASKLAGSGKLIRLELFNFKSYKGHHTLLFGDSYFTSIIGPNGSGKSNSM
DAISFVLGIKSSHLRSSNLRDLIYRGRVMKTSKIQDDGTTAPATNGDVNGYENGDAGDDEDTSQRTSR
NDPKTAWVMAVYEDDAGELHRWKRTITANGTSEYRINDRVVNAQQYNEALEKENILIKARNFLVFQG
DVEAIASQSPQDLTRLIEQISGSLEYKEEYERLEEEVRQATEEQAYKLQRRRAANSEIKQYMEQKSSS
KHPTSLVPRGSNPNLRAMDRLDHVRKQLEQTEQEFEASKAKLRQARESFQAVKQKRLELFNKAFTHI
QEQITHVYKELTRSEAYPLGGQAYLDIEEDTDTPFLSGVKYHAMPPCKRFRDMEHLSGGEKTMAALA
LLFAIHSYQPSPFFVLDEVDCALDNANVEKIKKYIREHAGPGMQFIVISLKPALFQASESLIGVYRDQEA
NTSRTLTLDLRKYRGSGGSGGSGGSGGSGGSSKAIVQMAKILRKELSEEKEVIFTDVLKSQANTEPENI
TKREASRGFFDILSLATEGCIGLSQTEAFGNIKIDAKPALFERFIHNGGSGGSGGSGGSHMVTTLSGLS
GEQGPSGDMTTEEDSATHIKFSKRDEDGRELAGATMELRDSSGKTISTWISDGHVKDFYLYPGKYTF
VETAAPDGYEVATPIEFTVNEDGQVTVDGEATEGDAHTKLGDIEFIKVNK

Expression, purification, labeling and crosslinking

All SMC-HDs were expressed using pET28a-based plasmids. Cells were grown at 37°C to an OD_{600} of 0.4 - 0.6. The expression was induced by the addition of 0.4 mM IPTG and incubated at 18°C overnight.

For the SMC-HD constructs the pellet of 8 l culture for Smc3-HD or 2 l culture for Smc1-HD were lysed in high salt buffer (25 mM Tris-HCl pH 7.5, 2000 mM NaCl, 1 mM DTT and 10 mM imidazole) supplemented with 1 mM PSMF and 0.1 mg/ml lysozyme and sonicated followed by centrifugation. The supernatant was loaded onto a HisTrap FF-Column (Cytiva, Marlborough, MA), washed first with 10 CV high salt buffer and later with 10 CV low salt buffer (25 mM Tris-HCl pH 7.5, 100 mM NaCl, 1 mM DTT and 10 mM imidazole). The elution was performed in low salt buffer containing 200 mM imidazole, fractions were collected and loaded onto a HiTrap Heparin HP column (Cytiva). The column was washed with 10 CV of low salt buffer without imidazole and eluted in four steps of at least 5 CV with increasing proportions of the high salt buffer without imidazole (10%, 30%, 50% and 100%). Samples of the input, flow-through, wash and the different elution peaks were further analyzed by SDS-PAGE. Fractions containing the protein of interest were concentrated for mono- and dimeric SMC-HDs and loaded onto a size-exclusion column (Superdex 200 Increase 10/300 GL, Cytiva) equilibrated in (25 mM NaCl pH 7.5, 500 mM NaCl and 2 mM $MgCl_2$). Constructs, which tend to oligomerize (e.g. Smc1-HD-link-cScc1-link-SpyCatcher003), were not concentrated, but directly loaded onto a larger column (HiLoad Superdex

200, Cytiva). Peaks were evaluated by SDS-PAGE. Fractions containing the protein of interest were flash frozen and stored at -80°C.

For the fluorescently labeled Smc3-HD construct the size-exclusion peak fractions were concentrated (30 μM) and attached to the CoA-coupled dye LD-555 (Lumidyne, New York, NY, custom synthesis, 1 mM) in a 1:2 ratio with additional 10 mM $MgCl_2$ and 30 μM Sfp1.

The crosslinking procedure using BMOE was previously described (12, 35). For samples containing ATP the proteins were incubated on ice for 15 minutes with 1 mM fresh ATP.

Mass Photometry

Mass photometry measurements were performed on a Refeyn OneMP mass photometer (Oxford, UK). For measurement, borosilicate cover slips were cleaned by sonication in isopropanol and drying with nitrogen. Silicone gaskets were placed on the cover slip to create wells. The samples were diluted 1:20 to a typical final concentration of 50 nM in measuring buffer (25 mM Tris-HCl pH 7.5, 100 mM NaCl, 1 mM DTT and 2 mM $MgCl_2$) in the sample well and proteins binding to the glass surface were recorded for 60s using the software AcquireMP (Refeyn). For data analysis, the software DiscoverMP (Refeyn) was utilized to convert measured contrasts into molecular masses (protein standard of known molecular mass was used for calibration).

ATPase activity assay

The ATPase activity assay was performed using malachite green. 100 nM or 200 nM of the SMC-HDs were incubated with ATP for 10, 30 or 60 minutes at 37°C followed by a 2 min incubation at room temperature with a 4-fold access of MGAM reagent (1:3 mixture of a 4.2% ammonium molybdate solution and a 0.045% malachite green solution with 0.04% Triton X-100). The reaction was stopped by the addition of citrate solution (34 g in 100 ml water, 20 μl on 200 μl reaction volume). Absorption was measured at 600 nm, blanked by a control without protein, which captures the auto-hydrolysis of ATP and amount of produced phosphate was calculated by using a standard curve made from K_2HPO_4.

III.3.4 Results

Purification of SMC-HDs (Fig. 19) following a standard purification protocol (12) resulted into precipitation of most of the proteins and a high degree of DNA contamination of the remaining fraction. To avoid insolubility issues and DNA contaminations the lysis for all SMC-HD constructs was performed hereafter in high salt (2 M NaCl), which is possible due to the single-peptide chain character of the constructs. Salt concentration was reduced to 100 mM NaCl during Ni-NTA purification, which allowed direct loading onto the heparin column.

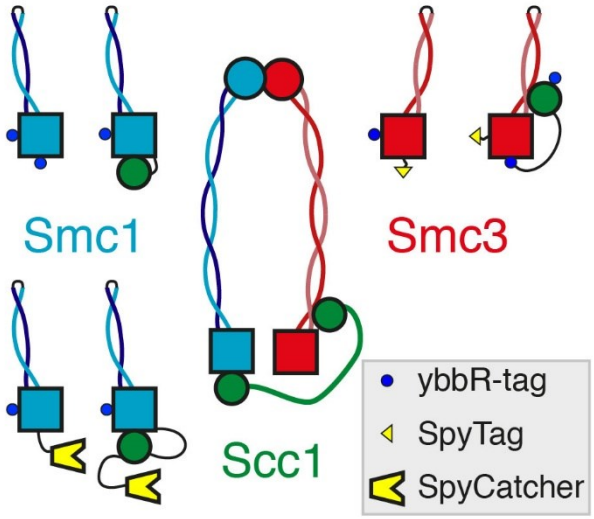

Figure 19: Construct design of SMC head domain (SMC-HD) constructs including terminal kleisin domains. Left: The Smc1-HD (light blue) constructs include ybbR-tags (dark blue sphere) or the SpyCatcher domain (yellow block). Fusions of Smc1-HD with the C-terminal Scc1 domain (cScc1, green) were equipped with one ybbR-tag and the SpyCatcher domain was added in another construct. Additionally, the linker length between the protein domains was increased to ensure sample integrity (e.g. Smc1-link-cScc1-link-SpyCatcher003). Right: The Smc3-HD constructs (red) include a ybbR-Tag and a SpyTag (yellow arrow), which can be covalently coupled to the SpyCatcher domain of the Smc1-HD constructs. This results in the formation of a single-peptide chain including a full SMC ATPase subcomplex. Lastly, the fusion of Smc3-HD with the N-terminal Scc1 includes ybbR-tags on positions, which allows the direct manipulation of the neck gate, as well as a SpyTag to introduce the Smc1-HD-cScc1 subcomplex for engagement.

On the heparin column, a linear gradient was not enough to separate observed protein peaks. Therefore, a stepwise purification strategy with prolonged washing steps was pursued (Fig. 19, see Methods). The Smc1-HD and Smc3-HD constructs behaved similarly during the purification, therefore only the Smc1-HD is shown in Figure 17 as a representative. The single SMC-HD constructs were found to elute at 10% buffer B (Peak 1, ~ 300 mM NaCl), whereas the different SMC-HD-Scc1 fusion constructs eluted additionally at a second peak at 30% buffer B (Peak 2, ~ 700 mM NaCl).

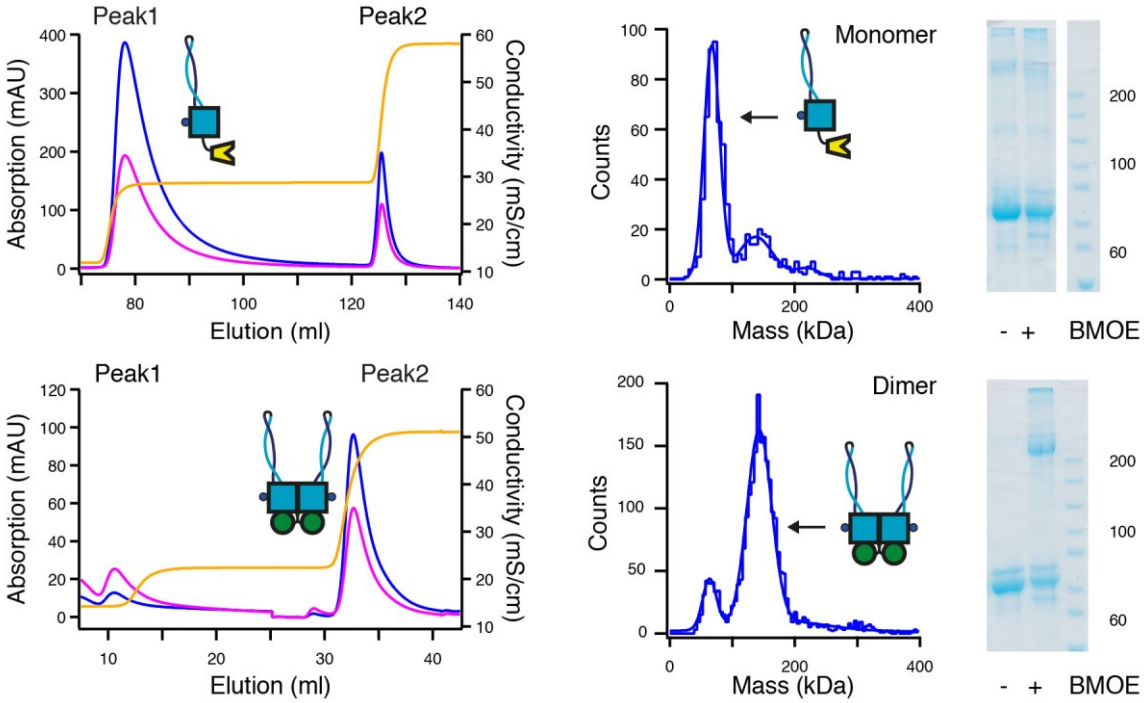

Figure 20: Homodimerization of Smc1 head domain by the C-terminal Scc1 domain. Top: The single Smc1 head domain (and the construct with the SpyCatcher003 domain) elutes from the heparin column (UV280: blue, UV260: lilac, conductivity: orange) mostly in Peak1 (10% buffer B). Mass photometry revealed that most of the proteins are monomeric at a concentration of 50 nM. BMOE-based crosslinking indicates that the small fraction of dimeric proteins is not due to a dimerization at the engagement interface. Bottom: The fusion construct of the Smc1 head domain and the C-terminal Scc1 domain elutes in Peak2 at high salt concentrations (30% buffer B). Mass photometry revealed that most of the proteins are dimeric at the same concentration. BMOE-based crosslinking indicates that these proteins dimerize at the engagement interface (black arrow). Bands at higher molecular sizes can be found in the lanes of both constructs, which could represent folded, thermophilic proteins, which even survived different denaturing procedures.

During size-exclusion chromatography, the proteins, which were found in Peak1, eluted mostly as monomers, whereas e.g. the Smc1-HD-cScc1 construct was found in Peak2 and eluted as homodimers (data not shown). Mass photometry confirmed that Smc1-HD and Smc3-HD constructs are monomeric, whereas the Smc1-HD-cScc1 construct forms dimers at the same concentration (Fig. 20, middle). This homodimerization could be further verified by crosslinking using cysteine point mutations located inside the engagement interface (Fig. 20, right), which was performed in the absence of ATP. This indicates that the Smc1-HD-cScc1 homodimer represents a “pre-engaged” apo state in contrast to the ATP bound engaged heterodimer. The homodimerization also agrees with the retarded elution from the heparin column in Peak2, as engaged heads possess an increased DNA binding surface leading to stronger interactions with the heparin column.

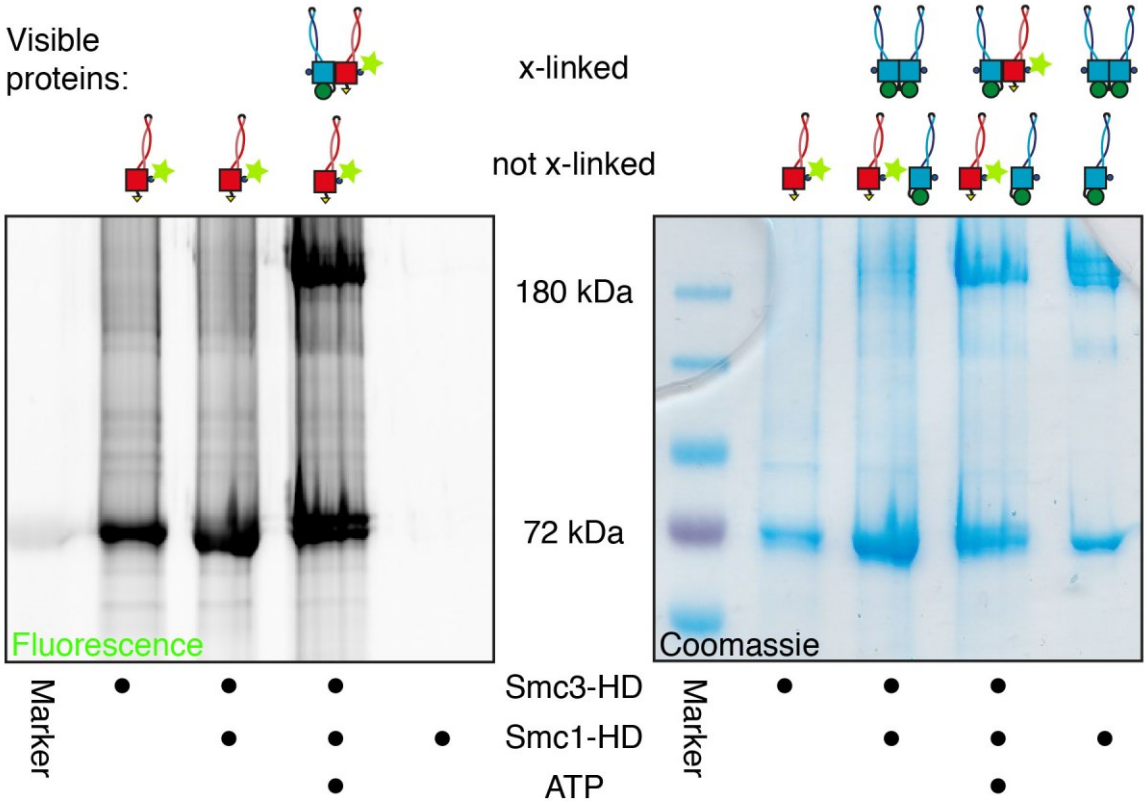

Figure 21: BMOE crosslinking of the heterodimer of the Smc3 head domain to the Smc1 head domain, which is bound to the C-terminal Scc1 domain. To evade the issue that the Smc1 homodimer and the Smc1/Smc3 heterodimer have nearly the same size and would be undistinguishable during SDS-PAGE after crosslinking, a fluorescently labeled Smc3 head domain (indicated by the green star in the cartoons) was used to distinguish the proteins. Left: The fluorescent image of the gel. Right: The Coomassie stain of the same gel. Top: Cartoons of the visible proteins, which can stay monomeric during the crosslinking (not x-linked) or a dimerization at the engagement interface can be fixed by the crosslink, which results into a clear mass shift of crosslinked proteins (x-linked). First lane, the Smc3 head domain, which is fluorescently tagged, did not show any additional bands due to the BMOE crosslinking. Second lane, a mixture of the Smc3 head domain and the Smc1 head domain with the C-terminal Scc1 domain does not form heterodimers in the absence of ATP. A crosslinked band is only visible in the Coomassie stain and represents the homodimers of the Smc1 head domain bound to the C-terminal Scc1 domain. Third lane, the heterodimer of the Smc3 head domain with the Smc1 head domain bound to the C-terminal Scc1 domain is formed in the presence of ATP. The fluorescent signal of the crosslinked complex clearly indicates that the fluorescently tagged Smc3 head domain was incorporated in the crosslinked band. Forth lane, the Smc1 head domain bound to the C-terminal Scc1 domain forms a crosslinked homodimer, which shows no fluorescence.

Since both, Smc3-HD and Smc1-HD, possess cysteine point mutations (see Methods) in the engagement interface, their engagement during heterodimerization can be verified by crosslinking using the zero-length crosslinker BMOE, which will result into a significant mass shift during SDS-PAGE (37). Unfortunately, the Smc3-HD-Smc1-HD-cScc1 heterodimer and the Smc1-HD-cScc1 homodimer have the same size, which would impede the interpretation of the crosslinking assay. Therefore, the Smc3-HD was first fluorescently labeled at its ybbR-tag by using a CoA-coupled dye and then further purified to discard the reaction components. The crosslinking was induced for the single proteins and the combination of both in the absence or presence of ATP (Fig. 21). It turned out that the heterodimer only forms in the presence of ATP, which suggests that the homodimeric configuration is repealed during ATP and Smc3-HD binding.

The Smc1-HD constructs tend to show bands in SDS-PAGE at high molecular masses (see Fig. 21 right), which might represent not-denatured proteins as this *C. thermophilum* protein could have an extreme stability or these bands represents aggregated or misfolded proteins due to the construct design. These bands hinder the evaluation of the crosslinking assay or further SpyTag/SpyCatcher coupling. To test if these bands are a result of the artificial construct design, the linker lengths between the domains were increased (e.g. Smc1-HD-link-cScc1-link-SpyCatcher003) to exclude steric clashes. This new construct shows the typical elution profile in two steps as Smc1-HD-cScc1 (Fig. 21), but increased levels of oligomerization were found in the second heparin peak by size-exclusion chromatography (Fig. 22), which could result from a domain swap between two single-peptide chains. This means that the cScc1 domain of one peptide chain binds to the Smc1-HD of a second peptide chain, which results in an unbound Smc1-HD of the first chain and an unbound cScc1 domain of the second chain. These unbound domains are docking sites for further assemblies. In addition, the engagement interface is still prone to dimerize other Smc1-HDs to these oligomeric structures. The existence of oligomers was further verified with mass photometry (Fig. 22). These constructs were accordingly not concentrated during the protein purification and monomeric species were used for further experiments. Although the increased linker length did not improved the purification process and the initial problem of the high molecular bands during SDS-PAGE, these constructs revealed aspects of cohesin oligomerization.

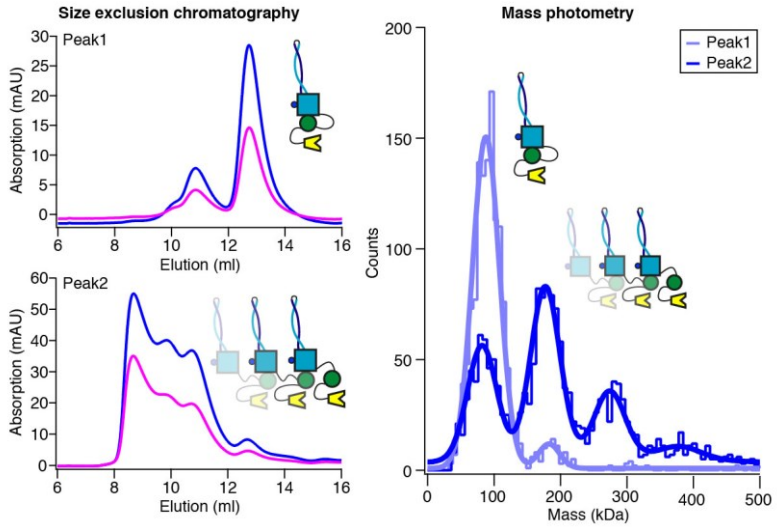

Figure 22: Oligomerization of the Smc1-HD-link-cScc1-link-SpyCatcher003 construct. Left: The size-exclusion chromatogram for Smc1-HD-link-cScc1-link-SpyCatcher003 (UV280: blue, UV260: lilac). Peak1 from the heparin column mostly contained monomeric proteins, whereas Peak2 from the heparin column mostly contained oligomeric proteins. Right: In mass photometry, the proportion of monomers for Peak1 (light blue) was close to 90 % at 200 nM, whereas Peak 2 (dark blue) included monomers (~ 30 %), dimers (~ 45 %), trimers (~ 20 %) and higher oligomers (~ 5 %) at 200 nM.

After successful purification of the two SMC-HD constructs, the next step included the covalent coupling of both constructs using the SpyTag-SpyCatcher system. For this, the SpyTag of the Smc3-HD interacts with the SpyCatcher at the Smc1-HD-cScc1 to form a non-engaged Smc3-HD-Smc1-HD-cScc1 heterodimer as a single-peptide chain construct, which can be pulled from both N-termini in the optical tweezers assay. The reaction between the SpyCatcher003 domain and the SpyTag003 is completed after only a few minutes at room temperature (157). The coupling of the HDs can be analyzed by SDS-PAGE and the gel clearly shows a newly created band (Fig. 23), which represents the non-engaged Smc3-HD-Smc1-HD-cScc1 single-peptide chain construct, which is a full SMC ATPase subcomplex.

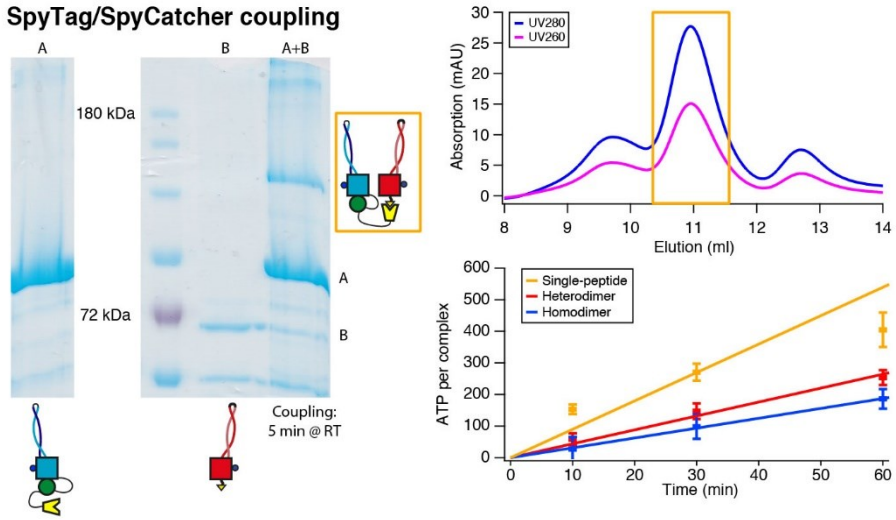

Figure 23: Generation of a full SMC ATPase subcomplex as a single-peptide chain. Left: The Smc1-HD-link-cScc1-link-SpyCatcher003 construct (A) can bind covalently to the Smc3-HD (B) via the SpyTag003. The reaction mix (A+B) contains a species, which runs at higher molecular masses and which is not present in either (A) or (B). Therefore, this species represents a single-peptide chain of the full SMC ATPase subcomplex. Right, top: Size-exclusion chromatogram of the reaction mix (UV280: blue, UV260: lilac), which separates the single-peptide chain ATPase subcomplex from single head domains (12 - 13 ml) and oligomeric structures (9 - 10 ml). Right, bottom: ATPase activity of the SMC head domains. The homodimer of the Smc1-HD-cScc1 construct without the SpyCatcher domain hydrolyzes ~ 3 ATP/min (blue). The addition of the Smc3-HD, which has no ATPase activity on its own, lead to the formation of Smc1/Smc3 heterodimers, which have an increased ATPase activity of

~ 4 ATP/min (red). Without the SpyCatcher domain these head domains cannot form a single-peptide chain during the measurement, which would distort the interpretation of the ATPase activity. The purified single-peptide chain ATPase subcomplex (orange) hydrolyzes ~ 9 ATP/min, which indicates that the limited dissolution of the head domains lead to an increase in ATP turnover.

The components of the reaction mix can be separated using size-exclusion chromatography and ATPase activity can be measured with the malachite green ATPase assay (Fig. 23, see Methods). The single-peptide ATPase construct hydrolyzes ~ 9 ATP/min, which is higher than the activity of the Smc1-HD-cScc1 homodimer (~ 3 ATP/min) or the Smc1/Smc3 heterodimer (~ 4 ATP/min). This suggest that the limited dissolution of the SMC-HDs by the SpyTag/SpyCatcher linkage leads to an increase in ATP turnover. Hence, it is likely that the stimulation of cohesin's ATPase activity is caused, at least partially, by the linkage of the HDs.

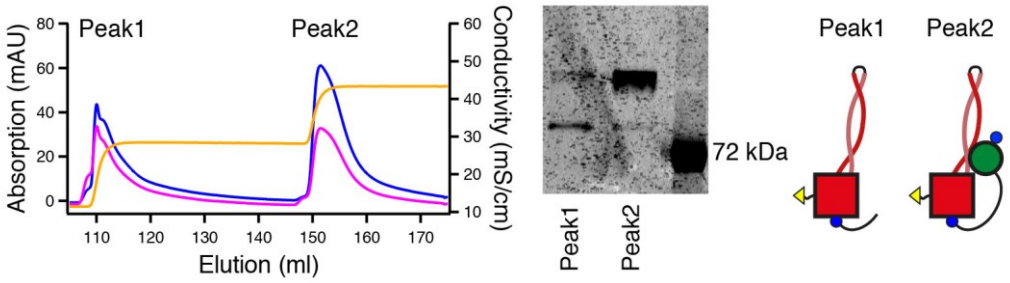

Figure 24: Purification of the neck gate construct, which includes the Smc3 head domain and the N-terminal Scc1 kleisin domain. Left: The protein elutes from the heparin column in two peaks as other SMC head domain constructs (UV280: blue, UV260: lilac, conductivity: orange). Middle: After fluorescently labeling of the fractions, it turned out that the protein can be digested at the linker between the head and the kleisin domain. The heparin column is able to separate digested and full-length proteins. Right: Cartoons of the digested protein, which can be mostly found in Peak1 and the full-length construct, which possesses two ybbR-tags to allow the manipulation of the neck gate in the optical tweezers system.

The neck gate construct formed by Smc3-HD linked to the N-terminal Scc1 domain (Smc3-HD-nScc1) was expressed in *E. coli* and purified similarly to the other SMC-HD constructs presented here. Also Smc3-HD-nScc1 was found in two peaks after the heparin column (Fig. 24). SDS-PAGE and fluorescently labeling identified the full-length Smc3-HD-nScc1 construct in the second peak, whereas the protein in the first peak was most likely digested. The separation of these two populations could be a result of the N-terminal Scc1 domain, which extends the DNA binding area of the Smc3-HD and therefore causes the retarded elution. In contrast to the Smc1-HD-cScc1 constructs, the size exclusion, crosslinking assay and mass photometry did not suggest any formation of dimers or oligomers for the Smc3-HD-nScc1 construct or for the Smc3-HD construct (data not shown). The neck gate construct was purified with sufficient purity to be used in future optical tweezers experiments.

III.3.5 Discussion

The purification procedure for SMC-HDs produced active ATPase subcomplexes as single-peptide chains, which reproduced the ATPase experiments of the individual domains (37, 39, 42). This supports further analysis on a single-molecular basis in the optical tweezers assay. Furthermore, the purified HD-constructs could mimic the properties of SMC complexes in the apo state. Cohesin complexes could rest in a dimeric state with low activity if ATP is unbound, like the homodimer of Smc1-HD-cScc1. These pre-engaged dimeric complexes could likely be involved in the initiation of loop extrusion. Additionally, cohesin complexes could form higher oligomers, like Smc1-HD-link-cScc1-link-SpyCatcher003, which would help in cohesin clustering. Furthermore, the single-peptide chain ATPase subcomplex can mimic the linkage of cohesin's heads by the loader complex.

Dimerization and loop extruding SMC complexes

Loop extruding cohesin and condensin complexes including all necessary subunits are monomeric complexes bound to DNA (6), whereby little is known about the initiation of loop extrusion (30). If these complexes dimerize at their HD, they can initially bind DNA at both hinge domains. Passive coiled-coil bending (59) could bring the DNA-bound hinge domains in close proximity to the HDs against a lower bending penalty as monomeric complexes (115). ATP binding could decouple the complexes by the resolution of the homodimer as found in Fig. 21. The transfer of DNA from one hinge domain to the engaged heads would lead to the dissociation of one complex, which ensures monomeric complexes capable of performing loop extrusion.

On the contrary, the loop extrusion activity of Smc5/6 complexes was found to strictly rely on their dimeric form (95), whereas the Nse5/6 subcomplex inhibits loop extrusion by preventing the dimerization of Smc5/6 complexes. This finding aligns with the homodimerization found by the Smc1-HD-cScc1 construct, which only forms dimers if the domain is bound to the C-terminal Scc1 domain (Fig. 20). In the case of Smc5/6, this could indicate that the Nse5/6 subcomplex competes for the binding site at the HD, which enables the formation of dimers. Further functional dimers have been reported for bacterial MukBEF (155) and JetABCD complexes (158), which suggests that dimerization is an evolutionary conserved feature of SMC complexes, which is lost in loop extruding cohesin and condensin complexes.

Oligomerization and cohesive cohesin

In our study, we identified the domain swap between neighboring subcomplexes by introducing an artificial linker (Fig. 22). However, this effect can occur naturally in SMC complexes because the kleisin is simultaneously bound at both heads via its terminal domains. This suggests that during late G1-phase newly expressed kleisin proteins (1) could bridge different Smc1/3 dimers resulting in an oligomerization as found for the purified fusion constructs even in the absence of ATP (Fig. 22). Since interphase chromosomes underlie loop extrusion of monomeric complexes (94, 159), oligomeric complexes must be resolved during loading. This could be achieved by the opening of the neck gate during head engagement followed by the closure of the neck gate by the association of the loader complex. The loader complex could further prefer the formation of monomeric SMC complexes by binding between the HDs.

Additionally, the domain swap could be involved in the maintenance of sister chromatid cohesion by enhancing the clustering of cohesive cohesin (156), which resists the forces generated by the spindle apparatus (85). The increased rupture forces of cohesin clusters were already reported using optical tweezers (154). Interestingly, the stimulation of the MukBF ATPase activity by the terminal kleisin domains (152) supports the notion that the kleisin domains does not necessarily need to originate from the same peptide chain to fulfill their functional role.

ATP hydrolysis and ATPase stimulation

The SMC-HD engagement regulates the ATPase activity of SMC complexes and leads to the formation of a DNA binding site (36, 37). However, the purpose of the energy gained by ATP hydrolysis is unknown. Molecular simulations have estimated that the dissociation energy of the Smc1/Smc3 heterodimer, as formed in Fig. 21, can consume up to 80% of the energy gained by ATP hydrolysis (160). This suggests that cohesin complexes and likely all the SMCs use the majority of the energy for the resolution of the tight DNA binding by the gripping state. The amount of free energy for further conformational changes of the SMC motor, like the O-shape/B-shape transition, is therefore very limited, which restricts loop extrusion modeling approaches (30, 63, 115). The energy stored in the engagement interface can be quantified by unfolding experiments of the single-peptide chain HDs presented in this study.

By covalently linking Smc1 and Smc3 heads, we created a single-peptide construct with increased ATPase activity compared to the isolated domains (Fig. 23), which did not reach the levels of full cohesin complexes stimulated by DNA and the loader by an order of magnitude (59, 154). Hence, cohesin's ATPase stimulation can be retraced partially to the physical linkage of the heads by the loader, which can bind to the Smc1-HD and the neck gate simultaneously (44–46). The linkage increases the ATPase activity most likely by reducing the diffusion of the heads and increasing the probability of head engagement (Fig. 23). Thereby, the loader adopts a bent conformation in the ATP bound gripping state with engaged heads (45), whereas in the apo-bridged state (ATP-free, disengaged) the corresponding subunit of condensin bridges the heads in a more relaxed conformation (60). This suggests that after ATP hydrolysis the bent loader conformation relaxes and drives the separation of the heads during disengagement (161). It is likely that our single-peptide chain construct coupled with the SpyTag/SpyCatcher system simulates the loader induced linkage in a comparable manner (Fig. 25). This would allow the separation of the linkage effect of the loader from further conformational changes induced by the loader, which propagate to the active sites to stimulate the ATP turnover.

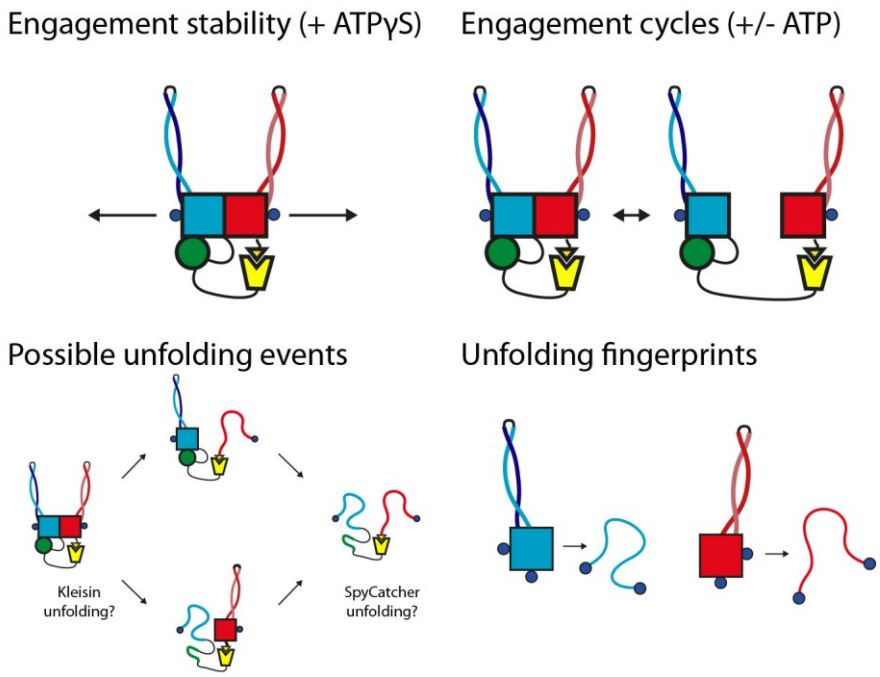

Figure 25: Unfolding of the SMC ATPase. Top left: The stability of the engagement interface of the SMC head domains can be measured under ATPγS conditions. Top right: Alternatively, the engagement and disengagement cycles under ATP consumption can be observed. Bottom left: During these experiments the response of the engagement interface should be clearly distinguishable from unfolding events of the individual components. Otherwise, the unfolding of the entire polypetide chain could identify the origin of an unfolding event and if the investigated tether was formed by a single subcomplex or by multiple tethers. Bottom right: Unfolding experiments of the individual head domains could give complementary unfolding fingerprints, which likely eases the identification of unfolding events of the heterodimer.

DNA clamping and DNA release at the neck gate

After understanding the conformational changes within the engaged HDs, their consequences for the opening or closure of the neck gate can be recapitulated using the Smc3-HD-nScc1 construct instead of the Smc3-HD construct (Fig. 26). Furthermore, future studies with the loader or with acetylated heads, Pds5 and Wpl1 could investigated DNA clamping (44–46) or DNA release (40, 41, 162) at the neck gate in more detail.

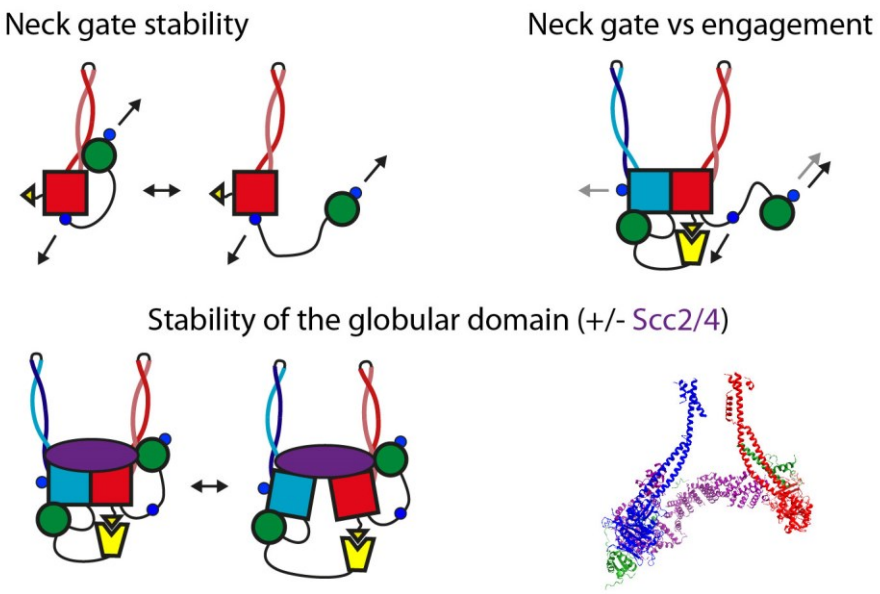

Figure 26: Unfolding of the SMC neck gate. Top left: The neck gate can be unfolded by pulling on attachment points after the head domain and after the kleisin domain, which will lead to the opening of the neck gate. Top right: In the literature, it was found that the head engagement upon ATP binding opens the neck gate. This process could be investigated in real time and different attachment points are conceivable (gray or black arrows). Bottom: Future experiments could reveal the stability of the globular domain and how the loader bridges the head domains, like in the apo-bridged conformation (PDB: 6YVV), or how the loader keeps the neck gate close. Alternatively, the mechanism behind unloading via the neck gate by Pds5-Wpl1 could be investigated.

To conclude, the successful purification of head domains as single-peptide chains, which can form active SMC ATPases, enables various single-molecule studies, which promise to investigate the SMC head engagement process, the conformational changes during ATP turnover and the neck gate regulation.

III.3.6 Author Contribution

Protein purification: MJ, JDL, JP, conducted experiments: MF, SZ; analyzed data: MF, SZ, JS; conceived of the study: JS; wrote paper: MF, SZ, JS.

III 3.7 Acknowledgments

MF acknowledges Daniel Panne (University of Leicester, UK) for providing the plasmids containing the SMC-HDs and the kleisin domains and Felix Metzner and Franziska Kunert (both Hopfner Lab, LMU) for helpful discussions about ATPase assays.

JS acknowledges support by the LMU Center for NanoScience CeNS. JS is supported by a DFG Emmy Noether grant STI672/2-1 and an ERC starting grant 758124.

IV Discussion

Cohesin and condensin were found to perform loop extrusion (LE) as monomers under ATP consumption (9, 94) and to perform the O-shape/B-shape transitions (8, 59). Both phenomenon inspired first iterations of the scrunching model (8), which explains LE as the consequence of these large-scale conformational changes enabled by the coiled-coil bending (30, 59). The ATP-cycle could control this movement (6, 30, 110), similar to the ATP-dependent stepping of cargo-carrying transporters (53). In contrast, my findings about the coiled-coil architecture in chapter III.2 and related research (59, 161) support that the O-shape/B-shape transition is not dependent on ATP but is rather a passive diffusion event. More and more findings shed light on single aspects of SMC complexes. However, there are still plenty of open questions in the SMC field: What is the mechanism behind loop extrusion? What is the sequence of events during loop extrusion? And is there a unified LE model for all the SMCs, which is frequently suggested (9, 30, 115)?

To discuss these questions in detail, the first part of this chapter focuses on the different variants of the cohesin complexes. Hence, I will examine what aspects make cohesin a loop extruding complex and which properties change during the transition between extruding and cohesive cohesin complexes. This is important to get a more comprehensive understanding of the interwoven properties of cohesin.

In a second part, I will discuss the latest models of LE based on our findings and related research. In a third part, I will discuss how the SMC complexes use their conserved features in a detailed way. This discussion will point out open questions about loop extrusion, which could be tackled with single-molecule studies of the SMC head domain constructs described in chapter III.3.

IV.1 Cohesive vs loop extruding cohesin

Latest research suggests that cohesin complexes can have two “operating modes”. Cohesin can either perform LE to spatially arrange the chromosomes (loop extruding cohesin) (6, 11, 94) or it can hold the sister chromatids together by entrapping the DNA molecules inside its ring structure (cohesive cohesin) (2, 21, 163). While loop extruding cohesin is bound to the loader complex, consumes ATP with a high turnover and is able to perform LE without the opening of any interface (94), cohesive cohesin is acetylated, bound to Pds5, inhibited in ATP hydrolysis and is able to entrap DNA inside its ring (78, 164). The transition between the two states must happen through DNA entrapment, which is regulated tightly.

The hinge domain enables topologically entrapment of DNA

Recently, the different cohesin functions could be separated by intense studies on the so-called DDD-mutant, which possesses charge neutralizing mutations of the inner channel of the hinge domain. This mutant was characterized first in yeast (165) and later in human cells (166), where it was shown to perform LE, but not to topologically entrap DNA inside its ring structure. These findings indicate that the hinge and especially its inner channel is utilized by cohesive cohesin complexes to allow DNA entrapment and sister chromatid cohesion (167). Artificially closing the hinge domain was already shown to abolish the establishment of sister chromatid cohesion (66). Although the mechanism behind hinge-mediated DNA entry is unclear, the evidence for this hypothesis are strong.

The hinge domain can adopt a conformation, where the inner channel is accessible by opening of an interface in the hinge domain (44). This suggests that the dimerized hinge domain (48, 165) represents a floodgate for DNA entry. If the hinge is opened at one interface to allow transient DNA binding in the inner channel (165), a reclosing of this interface and the simultaneous opening of the second interface would lead to the release of DNA inside the SMC lumen (44). This could explain how the hinge enables topologically loading of cohesin complexes by the DNA entry reaction (Fig. 27). DNA binding at the hinge was reported several times (8, 48, 59), including one study (35), which demonstrated that DNA binding leads to the opening of the coiled coils emerging from the hinge domain. This could likely be the result of a conformational change in the hinge domain, like the proposed opening of the floodgate-like interfaces.

Additionally, topological DNA entrapment depends on the presence of Scc2 and Scc3 in an *in vitro* reconstitution experiment (88) and occurred after elongated incubation times (41). Different interaction partners could likely stabilize a specific step in the hinge-mediated DNA entry reaction. It was shown that different Hawk proteins including Scc2, Pds5 and Scc3 can associate with cohesin's hinge domain (61, 115). This suggests that after Scc2 establishes initial DNA contacts (41), Scc3 could further regulate the hinge opening/reclosing transition (61). This would lead to the topologically entry of DNA in the SMC lumen and could be further stabilized by Pds5 (Fig. 27).

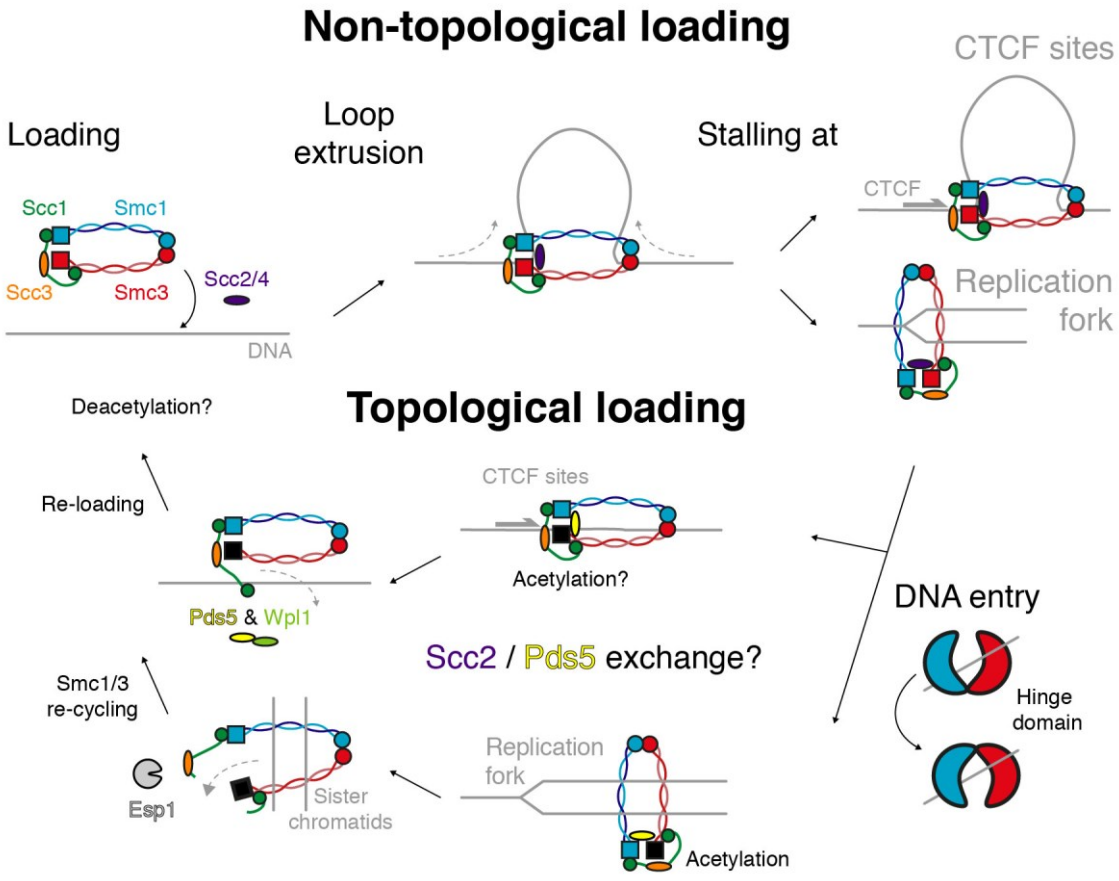

Figure 27: Topological vs non-topological loading. After loading of tetrameric cohesin complexes (Smc1, Smc3, Scc1 and Scc3) by the loader complex Scc2/4, cohesin is non-topologically bound to the DNA and able to perform loop extrusion. If

loop extrusion stalls at CTCF sites or at encounters with replication forks, cohesin can be topologically loaded by the DNA entry reaction mediated by the hinge domain. Alongside this reaction, it is likely that Scc2/4 is exchanged with Pds5 by the acetylation of conserved residues in the head domain. Topologically bound cohesin can be unloaded via Pds5-Wpl1 after unbinding from CTCF sites. Deacetylation of cohesin allows re-loading on DNA if Scc2/4 re-binds to cohesin. Otherwise, cohesin is protected during sister chromatid cohesion by the acetylation of cohesin and further binding partners (cohesive cohesin) from Wpl1-mediated unloading. Cohesive cohesin is released from DNA by separase, which is followed by deacetylation and re-cycling of the Smc1/3 dimers (see Figure 6).

How can loop extruding cohesins be converted to topologically bound cohesins?

The DDD-mutant additionally interacted with human CTCF in a reduced manner (166). Hence, it was speculated that the DNA-binding hinge conformation is relevant for encounters of loop extruding cohesin complexes (including Scc2 and Scc3) with CTCF. CTCF was already found to interact with Scc1 and Scc3 (121, 168) and therefore an interaction with the hinge domain via Scc3 is likely (61, 115). Since the DDD-mutant is deficient in topologically loading on DNA (166), the impairment of CTCF interactions with the hinge domain suggests a conversion of loop extruding cohesin complexes to topologically bound complexes at CTCF sites. This hypothesis benefits from the longer residence time of cohesin complexes at CTCF sites (92, 169), which matches the requirements with the hinge-mediated DNA entry reaction (41). Interestingly, artificial cleavage of the kleisin subunit was shown to release more cohesin complexes from CTCF sites than for other regions in the genome (170). Kleisin cleavage destroys the ring integrity of cohesin complexes and lead to the release of topologically bound complexes from DNA. This further suggests that cohesins can be topologically bound at CTCF sites, whereas cohesins at other genomic sites are more likely bound non-topologically.

During LE, cohesin complexes are bound non-topologically to chromosomal DNA (113), which includes DNA clamping in the gripping state (44–46). The DNA binding through the gripping state is accordingly not dependent on the kleisin integrity (41). This implies that after Scc2-mediated loading to the chromosome (Fig. 27), LE is performed until either the loader dissociates (6), which would likely be followed by the Wpl1-dependent release reaction (91, 102), or LE stalls at a CTCF site (99, 168), which could be followed by topologically loading of cohesin (see above). Since Pds5 binding and Scc2 binding are mutually exclusive (78) and Pds5 is found to be required for the CTCF boundary function (162), a dissociation of Scc2 during stalling or during the DNA entry reaction can be suggested. A similar exchange of Hawk proteins was already demonstrated for condensin (60). Furthermore, LE can be stalled at replication forks (105), where cohesin complexes are also found to be converted to cohesive cohesins, which entrap DNA topologically (171). These aspects strengthen a correlation between topologically loading and LE stalling.

Unloading of cohesin complexes

Analogous to the kleisin cleavage (170), the Pds5-Wpl1-mediated unloading pathway (Fig. 7) releases cohesin from chromosomes by the opening of the neck gate, which also destroys the ring integrity (172).

This scenario is plausible for topologically bound cohesin complexes, whereas loop extruding cohesins, which are bound to Scc2 but not to Pds5 (21, 78), should not be affected by encounters with Wpl1. Therefore, topological loading could be a prerequisite for the Wpl1-dependent interphase release pathway (87), which was not demonstrated yet.

Although CTCF was shown to protect cohesin from Wpl1 (121, 173), cohesins can be released from CTCF sites at the latest after 25 minutes (169). This suggests that cohesin first performs LE at CTCF sites (99, 174) until likely the hinge DNA entry reaction converts cohesin to topologically bound complexes while the Scc2/Pds5 exchange happens. Ultimately, the unbinding of Pds5-bound cohesin from CTCF sites could favor the Pds5-Wpl1 unloading reaction by the recruitment of Wpl1 (Fig. 27).

During S-phase topological bound cohesin is acetylated (80, 164), which stabilizes Pds5 binding to cohesins (83) and likely explains the prolonged protection of Wpl1 (173) in combination with further binding partners (175, 176). Cohesion is maintained by this cohesive cohesin variant until cleavage of the kleisin destroys the ring integrity by separase (1). Recently, the recruitment of the acetyl-transferase to CTCF sites was described (173), which results in the non-canonical acetylation of cohesins at CTCF sites (81). It was further speculated that this activity could be conserved (82) and counteracted by the deacetylation at CTCF sites (81) (Fig. 22). Therefore, the Scc2/Pds5 exchange could be driven by the SMC head acetylation, which preferably recruits Pds5 to cohesin. Future studies will shed light on the nature of topological cohesin binding, acetylation and the contrary roles of Pds5, which either is involved in the release of cohesin from chromosomes or in the maintenance of sister chromatid cohesion. With the SMC-HD constructs introduced in chapter III.3 the influence of these regulatory subunits on the neck gate could be addressed.

To summarize: There is growing evidence that cohesin complexes have two operating modes (cohesive vs loop extruding cohesin), which are linked to the topological or non-topological DNA binding mode (21). After loading, cohesin performs LE until stalling occurs at CTCF sites or replication forks (99, 105). This is likely followed by the DNA entry reaction at the hinge domain (41), which topologically load cohesin to the DNA. Furthermore, a Scc2/Pds5 exchange could be a consequence of the (non-)canonical acetylation of cohesin, which can convert loop extruding cohesin (Scc2-bound) to cohesive cohesin (Pds5-bound) by the different affinities of the Hawk proteins to the acetylated cohesin complex (81).

IV.2 Models of loop extrusion

In chapter I.5 we have been introduced to the scrunching model (8, 30, 59), the Brownian ratchet model (115) and the DNA segment capture model (63, 112, 116) for LE. These models have their individual strength, but rely mainly on the data proposed in the respective paper, whereas differences and contradictions are underestimated or ignored. Hence, the models are partially mutually exclusive, which could be due to undisclosed aspects of LE, like the recently reported asymmetry of cohesin-mediated LE (6). Such missing links would condense all modeling approaches to one conserved SMC LE mechanism.

In this chapter, I will discuss the latest models of LE one after the other with the goal to find answers for the central open questions: How is LE performed? Is there also a conserved mechanistic model? And if that is not the case, why do the SMC complexes have a conserved architecture?

The scrunching model

In the scrunching model (Fig. 9), the DNA is anchored at a safety-belt and the growing DNA loop is fed in by a Hinge-Hawk clamp, which can be brought in close proximity to the globular domain by the coiled-coil bending dynamics. The step size of LE can be convincingly explained by the elbow-based bending transitions of the coiled coils, which brings a DNA bound Hinge-Hawk clamp in close proximity to the globular domain (6). The coiled-coil bending is postulated to occur during ATP binding to a preferred site (59), which likely explains directionality and asymmetry of LE. Furthermore, the DNA transfer from the Hinge-Hawk clamp to the globular domain is performed after the engagement of the head domains, which could bind the DNA with a higher affinity than the Hinge-Hawk clamp. The anchoring of the loop at the safety-belt, which is a static DNA binding pore formed by the kleisin and a Hawk protein, ensures the growth of a loop. Next, the hinge domain without the DNA is released by a swing-out movement, which resets the LE cycle for further DNA capture (8). Otherwise, the shown cartoons in Fig. 9 for the scrunching model are partially derived from findings of cohesin (Hinge-Hawk clamp) (59, 115) and partially from condensin (safety-belt at a peripheral position from the complex). This raises the questions: Is there a safety-belt for cohesin? And is there a Hinge-Hawk clamp for condensin? Furthermore, this model is largely based on the elbow bending dynamics and could not explain LE based on the dimerization of Smc5/6 complexes (95).

The Hinge-Hawk clamp was not found for condensin. In addition, the position of condensin's hinge domain is more distal to the globular domain in the bent conformation than the hinge of cohesin (8, 61), which is speculated to impede interactions of a hypothetical condensin Hinge-Hawk clamp with the globular domain (177). This still allows the possibility that condensin's hinge domain can bind DNA by itself (8) and that the scrunching model based on the safety-belt could describe condensin-mediated LE. Otherwise, clear evidence for a safety-belt-like anchoring of cohesin, have not been found, even in the presence of CTCF (94, 99). These considerations and the novel Smc5/6 LE mechanism (95) suggest that a conserved working model for the eukaryotic loop extruding complexes cannot be supported. Hence, it is likely that the different SMC complexes utilize their conserved SMC features (ATP-dependent head engagement and ATP turnover, neck gate regulation, elongated coiled coils dimerized at a hinge domain, multiple DNA binding sites and regulatory subunits) in unique and different ways.

The Brownian ratchet model

The Brownian ratchet model shares the dependency on a DNA transfer in combination with coiled-coil bending, but does not rely on a safety-belt-like anchoring. Here, the Hinge-Hawk clamp is formed with Scc3 instead of Scc2 (115). Scc2 stays stably bound at the globular domain in either the gripping state (ATP bound) or in the apo-bridged (ATP free) state of the complex, which is supported by recent optical tweezers study on the complex (161). Furthermore, the Brownian ratchet model supports the magnetic tweezers experiment of EQ-condensin, where the coiled-coil bending was found to take place only once and during ATP binding (110). If ATP is bound and the head domains engage, a novel interface for the binding of the Hinge-Hawk clamp could be created as sparsely resolved for the gripping state (45) and CTCF-bound state of cohesin (178). This novel interface could freeze the diffusive coiled-coil bending dynamics (12, 59) in the bent conformation by the binding of the Hinge-Hawk clamp at the globular domain to allow further LE steps. Hence, this binding event represents an ATP-binding dependent Brownian ratchet and is enabled by the coiled-coil bending dynamics. Therefore, the Brownian ratchet relies on the segmented architecture of SMC coiled coils found in chapter III.2, which is in contrast to some scrunching models (8, 30) , where the coiled-coil bending dynamics is related to the ATPase cycle.

Side notes on elbow-based loop extrusion models

Although the Brownian ratchet model seems convincing so far, there are experimental results that contradict the proposed version of this model. First, the prohibited LE activity, if the kleisin unit is cleaved. In the model, the unstructured linkers of the kleisin protein spans from the globular domain to the Hinge-Hawk clamp and back again. The cleavage at one position would not necessarily lose the connection of the clamp to the globular domain because the Hawk subunits are postulated to bind in a static way to their kleisin binding site. Secondly, an opening of the neck gate during LE and a slipping state, where the Hinge-Hawk clamp scans along the DNA, were proposed (115). These ideas contradict the observations of LE by the covalently linked complex (94) and the obstacle bypass (113). Therefore, the Brownian ratchet model accounts for most observation, but need some extension. The Brownian ratchet model could benefit from the unbinding of the Hinge-Hawk clamp from the DNA and the directed coiled-coil bending dynamics to a preferred site as postulated in the scrunching model (59), which would add the directionality and the obstacle bypass to the model. Therefore, a smart combination of the scrunching model and the Brownian ratchet model could likely explain cohesin-mediated LE in more detail, but further experiments are needed to discriminate between the manifold possibilities of SMC activity.

In the future, LE experiments with altered SMC motor properties would be of high interest. For example, a shortening of the kleisin protein, which preserves the binding sites of the Hawks, could result in a smaller LE step size. This could validate the Brownian ratchet model. My unfolding experiment of truncated coiled coils from chapter III.2 revealed that the truncated coiled coils had a similar fold like the wild-type coiled coils, which motivates *in vitro* LE reconstitution experiments of truncated complexes. Additionally, I have found that alterations of the elbow showed misfolded configurations and impaired cohesin function *in vivo*. LE reconstitution experiments with SMC complexes, which possess truncated or altered coiled coils, could validate the relationship between the elbow bending dynamics and the LE

step size and could additionally reveal why the lengths of the SMC coiled coil are conserved. Furthermore, the binding site of the Hinge-Hawk clamp to the globular domain can be identified with the SMC head domain constructs introduced in chapter III.3. The stability of this interface, which represents the gripping state, could be tested with optical tweezers experiments.

Alternative loop extrusion models

Smc5/6 was shown to perform LE exclusively in dimerized form, which is not reported for the other SMC complexes. This, and the fact that Smc5/6 does not possess an elbow feature or a safety-belt anchoring, suggest that Smc5/6 complexes use a different LE mechanism. A model for LE of dimerized Smc5/6 complexes is not published yet, but it is reasonable to believe that the combination of two individually translocating Smc5/6 could allow this process (95), which raises questions about the translocation of these Smc5/6 complexes.

The DNA segment capture model postulated that the hinge-bound DNA is transported by a dynamic pumping movement of the coiled coils towards the globular domain. This leads to a capturing of the DNA in different pores of the complex in an ATP-independent manner (63, 112, 116). Without a safety-belt-like anchoring, this process could likely explain the translocation of monomeric Smc5/6 complexes. Further experiments are needed to shed light on the translocation and LE of Smc5/6 complexes, whereas the elbow-based modeling approaches could benefit from the DNA segment capture model by the identification of the different pores (112), which could help in the DNA anchoring or DNA transfer. Of high interest would be an experiment, in which the different SMC complexes are crosslinked at one (or only a few) position in the coiled coils to stop the postulated DNA transfer along the coiled-coils. This would reveal if the DNA transfer is happening outside or inside the coiled coils.

IV.3 Creative usage of conserved SMC features

After discussing the individual LE models, it is still hard to judge, which model fits best and what sequence of events is truly performed during LE. Therefore, a closer look at the conserved SMC features, like done in chapter III.2 and III.3, promise to provide detailed insights for SMCs and SMC-like complexes in general.

Multiple DNA binding sites

The Brownian ratchet model assumes two binding sites for the LE mechanism (115), where the DNA loop growths between the static globular domain and the moving Hinge-Hawk clamp. Assuming only two binding sites, the unbinding of one site, like the Hinge-Hawk clamp, would release the extruded DNA loop from the complex. This implies that the Hinge-Hawk clamp must slip along the DNA, but as discussed, this does not agree with the observation of the obstacle bypass (113). Hence, it is likely that condensin and cohesin have at least three DNA binding sites, which are responsible for LE, like in the described scrunching model (8, 59, 110). Consequently, the DNA must be anchored within a pore likely formed by the kleisin subunit (safety-belt compartment) and transiently bound at the globular domain and the Hinge-Hawk clamp. While the Hinge-Hawk clamp transfers DNA to the globular domain, the initial bound DNA at the globular DNA gets repressed. The loop is formed between the anchoring at the

kleisin pore and the globular domain, while the hinge can bind a distal DNA position to start the next cycle of LE.

How is the ATPase cycle involved during loop extrusion?

The canonical ATPase cycle consists of an apo (ATP-free) state, an ATP-bound state, and an ADP state, where the cleaved phosphate could be still bound or released (1). For SMCs, the apo state is well described and seems to be inconsistent with DNA binding at the globular domain or inside the coiled coils (Fig. 4). The coiled-coil bending dynamics are likely independent from the ATPase cycle, while the Hinge-Hawk clamp was speculated to get trapped at the globular domain during ATP binding (110). The ATP-bound state represents the DNA-bound state at the engaged heads (44–46), which is only transiently populated. The lifetime of the engaged heads configuration is not known, but can be investigated with single-molecule optical tweezers experiments in dependence on ATP (Fig. 25). The engagement could be resolved more frequently during loader and DNA binding because of the increased ATPase activity (164). Little is known about the next stages of the cycle, which are speculated to release somehow the DNA from the globular domain by the disengagement of the head domains. DNA release could happen during phosphate cleavage, phosphate release or ADP release.

In chapter III.3 we have been introduced to SMC head domain constructs, which are able to engage and consume ATP as individual entities. With these constructs the cascade of conformational changes, which modulates the ATP-binding pocket from low to high activity, and also the influence of ADP can be investigated on a single-molecule level. The found engagement of Smc1/Smc3 heterodimers was able to resolve homodimers of the Smc1 head domains, which formed in the apo-state (Fig. 21). This likely ensures the formation of monomeric SMC complexes capable of performing LE. Additionally, the dissociation energy of the Smc1/Smc3 heterodimer and further the lifetime of the engaged conformation in the presence of ATP can be addressed with optical tweezers experiments (Fig. 25). This could reveal how long the SMCs “wait” for DNA binding at the globular domain, which then accelerates further LE steps or ATPase steps. With this information a lower limit for the duration of the entire LE cycle can be estimated. Otherwise, the loader is speculated to significantly drive the disengagement after ATP hydrolysis by the relaxation of its stressed conformation in the gripping state (161) and ensures a high ATP turnover rate (164) by the bridging of the head domains in the apo-bridged state (60, 161). Further experiments are needed to shed light on the disengagement process, the roles of the ADP-states and the physical linkage of the head domains during LE.

Why are elongated CC conserved?

Another conserved property of the SMC family is the length of the elongated coiled coils. In chapter III.2 we have learned that the coiled coils have a spacing purpose, but the relatively high sequence conservation indicates an additional role in propagating conformational changes. For cohesin and condensin, this conformational changes are enabled by the bending of the elbow. This O-shape/B-shape transition is likely enabled by the modular architecture of the coiled coils and therefore a consequence of the flexible linkage of the coiled-coil segments flanking the elbow. Speculations of possible helix sliding, which could fuel this transition in analogy to cargo-carrying motors, cannot be supported for cohesin. As we have learned that SMCs use their features in creative ways, this could still be possible

for Smc5/6 or other SMC-like complexes, which do not have an elbow and therefore could use their coiled coils for different mechanisms.

We have shown that alterations of the elbow destabilized the coiled-coil fold and led to mitotic defects *in vivo*, which are a consequence of the distorted cohesin complexes (chapter III.2). These observations established the elbow as a folding guide, which ensures proper coiled-coil alignment and which further suggested that elongated coiled coils in general could need discontinuities to avoid misfolded conformations. Interestingly, the length of the coiled coils and the persistence length of DNA are similar. This enables step sizes, which are larger than the complex diameter and therefore ensures the obstacle bypass by the capture of a distal DNA segment, which is important in the nuclear context to bypass DNA-bound proteins, like nucleosomes. Truncated complexes should have difficulties to bypass obstacles, because the chance of finding a DNA segment, which is not along the track, is reduced and the bending penalty for loop initialization is increased for complexes with shorter SMC arms. These considerations indicate that a length conservation of cohesin and condensin complexes ensures the correct DNA hand-over and that bacterial complexes or Smc5/6, which might follow the DNA segment model, are likely restricted by the DNA stiffness, which dictates the bending penalty of loop initialization.

Clustering

In addition to LE of monomeric complexes, a modeling approach, which is based on clustering of SMC (179), could reproduce the LE experiments in a comprehensive way. This idea contradicts the common opinion about LE, although clusters of SMC have been reported in the past (156, 180, 181). I have found that dimerization in the apo state can occur via the engagement interface of Smc1 head domains (Fig. 20) and that a domain swap between different complexes can lead to an oligomerization of cohesin complexes, like observed for the Smc1-HD-link-cScc1 constructs (Fig. 22). SMC or SMC-like complexes can further dimerize or oligomerize using their kleisin domains (43, 158) or their coiled coils (58), which provides many possibilities for an SMC cluster formation. The clusters of cohesin were found to be specific interactions of hinge and head domains (156), which are cell cycle regulated and peak during S-phase, whereas the coiled coils and the elbow were found to be not in close proximity to the sites of clustering (61, 156). Therefore, clustering occurs in the context of cohesive cohesin complexes and can be favored by acetylation and Pds5 binding. This suggests that there could be two types of LE mechanisms, one for monomeric complexes (see above) and one for clustered SMCs, where individual complexes bind to DNA and the clustering of the SMCs would bridge the DNA resulting in a DNA loop with SMCs at its basis. Since Pds5 and acetylation can be found at cohesins at CTCF-sites (81, 173), clustering could not only be involved during sister chromatid cohesion, but also help to maintain chromosomal loops and TAD boundaries (179, 180). Future experiments will show if functional clustering is possible for condensin or Smc5/6 and if SMC complexes can perform LE in the suggested way or if clustering or dimerization effects are responsible for the observed phenomena.

Conclusion

Once too simple models could be rejected (10, 182), a variety of experimental results lead to the proposal of contradicting loop extruding models (Fig. 9), which were derived from different SMC complexes. Since Smc5/6 was reported to perform loop extrusion by dimerization and therefore in a

different way than cohesin and condensin, the possibility that the different SMC complexes use their conserved features in different and maybe unique ways grows, which hampers the LE modeling by its diversity. Additionally, it cannot be excluded that different LE reaction cycles can coexist, like the different Hinge-Hawk clamps (Scc2-Hinge or Scc3-Hinge). Future experiments, which tackle the LE properties of the SMC complexes, will shed more light on the mechanism behind LE. Besides this high aspiration, the investigation of individual SMC domains on a single-molecular basis promise to reveal properties of all SMCs in more detail. The unfolding studies of the SMC coiled-coil domain have revealed that the elbow is a folding guide for proper coiled-coil alignment and that the bending dynamics are rather based on passive fluctuations than on helix sliding coupled to the ATPase cycle. The unfolding of the SMC head domains promise to give more insights about the role of engagement and disengagement for LE, while aspects of cohesin clustering could already be rebuilt with these constructs. By adding Hawk subunits to the setup, it will be investigated if they bind to the engaged Smc1/Smc3 heterodimer or the neck gate and how this influences SMC activity. This could reveal how DNA release (Pds5-Wpl1) or DNA entrapment (Scc2) is performed mechanistically. Furthermore, the planned unfolding experiments can benefit from the miscalibration procedure from chapter III.1.

V References

1. Yatskevich, S., J. Rhodes, and K. Nasmyth. 2019. Organization of Chromosomal DNA by SMC Complexes. *Annu. Rev. Genet.* 53:445–482.

2. Guacci, V., D. Koshland, and A. Strunnikov. 1997. A Direct Link between Sister Chromatid Cohesion and Chromosome Condensation Revealed through the Analysis of MCD1 in S. cerevisiae. *Cell*. 91:47–57.

3. Aragón, L. 2018. The Smc5/6 Complex: New and Old Functions of the Enigmatic Long-Distance Relative. *Annu. Rev. Genet.* 52:89–107.

4. Nasmyth, K. 2001. Disseminating the genome: joining, resolving, and separating sister chromatids during mitosis and meiosis. *Annu Rev Genet*. 35:673–745.

5. Henrikus, S.S., and A. Costa. 2021. Towards a Structural Mechanism for Sister Chromatid Cohesion Establishment at the Eukaryotic Replication Fork. *Biology (Basel)*. 10:466.

6. Kim, E., R. Barth, and C. Dekker. 2023. Looping the Genome with SMC Complexes. *Annu. Rev. Biochem.* 92:annurev-biochem-032620-110506.

7. Eeftens, J.M., A.J. Katan, M. Kschonsak, M. Hassler, L. de Wilde, E.M. Dief, C.H. Haering, and C. Dekker. 2016. Condensin Smc2-Smc4 Dimers Are Flexible and Dynamic. *Cell Reports*. 14:1813–1818.

8. Ryu, J.-K., A.J. Katan, E.O. van der Sluis, T. Wisse, R. de Groot, C.H. Haering, and C. Dekker. 2020. The condensin holocomplex cycles dynamically between open and collapsed states. *Nat Struct Mol Biol*. 27:1134–1141.

9. Ganji, M., I.A. Shaltiel, S. Bisht, E. Kim, A. Kalichava, C.H. Haering, and C. Dekker. 2018. Real-time imaging of DNA loop extrusion by condensin. *Science*. 360:102–105.

10. Fudenberg, G., N. Abdennur, M. Imakaev, A. Goloborodko, and L.A. Mirny. 2017. Emerging Evidence of Chromosome Folding by Loop Extrusion. *Cold Spring Harb Symp Quant Biol*. 82:45–55.

11. Lieberman-Aiden, E., N.L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B.R. Lajoie, P.J. Sabo, M.O. Dorschner, R. Sandstrom, B. Bernstein, M.A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L.A. Mirny, E.S. Lander, and J. Dekker. 2009. Comprehensive mapping of long range interactions reveals folding principles of the human genome. *Science*. 326:289–293.

12. Freitag, M., S. Jaklin, F. Padovani, E. Radzichevici, S. Zernia, K.M. Schmoller, and J. Stigler. 2022. Single-molecule experiments reveal the elbow as an essential folding guide in SMC coiled-coil arms. *Biophys J*. 121:4702–4713.

13. Freitag, M., D. Kamp, M. Synakewicz, and J. Stigler. 2021. Identification and correction of miscalibration artifacts based on force noise for optical tweezers experiments. *J. Chem. Phys.* 155:175101.

14. Hu, B., T. Itoh, A. Mishra, Y. Katoh, K.-L. Chan, W. Upcher, C. Godlee, M.B. Roig, K. Shirahige, and K. Nasmyth. 2011. ATP Hydrolysis Is Required for Relocating Cohesin from Sites Occupied by Its Scc2/4 Loading Complex. *Current Biology*. 21:12–24.

15. Cutter, A.R., and J.J. Hayes. 2015. A brief review of nucleosome structure. *FEBS Lett*. 589:2914–2922.

16. Ghoneim, M., H.A. Fuchs, and C.A. Musselman. 2021. Histone Tail Conformations: A Fuzzy Affair with DNA. *Trends in Biochemical Sciences*. 46:564–578.

17. Bell, O., A. Burton, C. Dean, S.M. Gasser, and M.-E. Torres-Padilla. 2023. Heterochromatin definition and function. *Nat Rev Mol Cell Biol*. 1–4.

18. Morrison, O., and J. Thakur. 2021. Molecular Complexes at Euchromatin, Heterochromatin and Centromeric Chromatin. *Int J Mol Sci*. 22:6922.

19. Oberbeckmann, E., V. Niebauer, S. Watanabe, L. Farnung, M. Moldt, A. Schmid, P. Cramer, C.L. Peterson, S. Eustermann, K.-P. Hopfner, and P. Korber. 2021. Ruler elements in chromatin remodelers set nucleosome array spacing and phasing. *Nat Commun*. 12:3232.

20. Xie, L., and Z. Liu. 2021. Single-cell imaging of genome organization and dynamics. *Molecular Systems Biology*. 17:e9653.

21. Davidson, I.F., and J.-M. Peters. 2021. Genome folding through loop extrusion by SMC complexes. *Nat Rev Mol Cell Biol*. 22:445–464.

22. Mirny, L., and J. Dekker. 2022. Mechanisms of Chromosome Folding and Nuclear Organization: Their Interplay and Open Questions. *Cold Spring Harb Perspect Biol*. 14:a040147.

23. Branco, M.R., and A. Pombo. 2006. Intermingling of Chromosome Territories in Interphase Suggests Role in Translocations and Transcription-Dependent Associations. *PLOS Biology*. 4:e138.

24. Gibcus, J.H., K. Samejima, A. Goloborodko, I. Samejima, N. Naumova, J. Nuebler, M.T. Kanemaki, L. Xie, J.R. Paulson, W.C. Earnshaw, L.A. Mirny, and J. Dekker. 2018. A pathway for mitotic chromosome formation. *Science*. 359:eaao6135.

25. Karpinska, M.A., and A.M. Oudelaar. 2023. The role of loop extrusion in enhancer-mediated gene activation. *Current Opinion in Genetics & Development*. 79:102022.

26. McArthur, E., and J.A. Capra. 2021. Topologically associating domain boundaries that are stable across diverse cell types are evolutionarily constrained and enriched for heritability. *The American Journal of Human Genetics*. 108:269–283.

27. Rao, S.S.P., S.-C. Huang, B. Glenn St Hilaire, J.M. Engreitz, E.M. Perez, K.-R. Kieffer-Kwon, A.L. Sanborn, S.E. Johnstone, G.D. Bascom, I.D. Bochkov, X. Huang, M.S. Shamim, J. Shin, D. Turner, Z. Ye, A.D. Omer, J.T. Robinson, T. Schlick, B.E. Bernstein, R. Casellas, E.S. Lander, and E.L. Aiden. 2017. Cohesin Loss Eliminates All Loop Domains. *Cell*. 171:305-320.e24.

28. Strick, T.R., T. Kawaguchi, and T. Hirano. 2004. Real-Time Detection of Single-Molecule DNA Compaction by Condensin I. *Current Biology*. 14:874–880.

29. Eeftens, J.M., S. Bisht, J. Kerssemakers, M. Kschonsak, C.H. Haering, and C. Dekker. 2017. Real-time detection of condensin-driven DNA compaction reveals a multistep binding mechanism. *The EMBO Journal*. 36:3448–3457.

30. Takaki, R., A. Dey, G. Shi, and D. Thirumalai. 2021. Theory and simulations of condensin mediated loop extrusion in DNA. *Nat Commun*. 12:5865.

31. Bürmann, F., B.-G. Lee, T. Than, L. Sinn, F.J. O'Reilly, S. Yatskevich, J. Rappsilber, B. Hu, K. Nasmyth, and J. Löwe. 2019. A folded conformation of MukBEF and cohesin. *Nat Struct Mol Biol*. 26:227–236.

32. Choudhary, K., and M. Kupiec. 2023. The cohesin complex of yeasts: sister chromatid cohesion and beyond. *FEMS Microbiology Reviews*. 47:fuac045.

33. Potapova, T., and G. Gorbsky. 2017. The Consequences of Chromosome Segregation Errors in Mitosis and Meiosis. *Biology*. 6:12.

34. Piché, J., P.P. Van Vliet, M. Pucéat, and G. Andelfinger. 2019. The expanding phenotypes of cohesinopathies: one ring to rule them all! *Cell Cycle*. 18:2828–2848.

35. Soh, Y.-M., F. Bürmann, H.-C. Shin, T. Oda, K.S. Jin, C.P. Toseland, C. Kim, H. Lee, S.J. Kim, M.-S. Kong, M.-L. Durand-Diebold, Y.-G. Kim, H.M. Kim, N.K. Lee, M. Sato, B.-H. Oh, and S. Gruber. 2015. Molecular Basis for SMC Rod Formation and Its Dissolution upon DNA Binding. *Molecular Cell*. 57:290–303.

36. Käshammer, L., J.-H. Saathoff, K. Lammens, F. Gut, J. Bartho, A. Alt, B. Kessler, and K.-P. Hopfner. 2019. Mechanism of DNA End Sensing and Processing by the Mre11-Rad50 Complex. *Molecular Cell*. 76:382-394.e6.

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

38. Hassler, M., I.A. Shaltiel, M. Kschonsak, B. Simon, F. Merkel, L. Thärichen, H.J. Bailey, J. Macošek, S. Bravo, J. Metz, J. Hennig, and C.H. Haering. 2019. Structural Basis of an Asymmetric Condensin ATPase Cycle. *Molecular Cell*. 74:1175-1188.e9.

39. Arumugam, P., T. Nishino, C.H. Haering, S. Gruber, and K. Nasmyth. 2006. Cohesin's ATPase activity is stimulated by the C-terminal Winged-Helix domain of its kleisin subunit. *Curr Biol*. 16:1998–2008.

40. Gligoris, T.G., J.C. Scheinost, F. Bürmann, N. Petela, K.-L. Chan, P. Uluocak, F. Beckouët, S. Gruber, K. Nasmyth, and J. Löwe. 2014. Closing the cohesin ring: Structure and function of its Smc3-kleisin interface. *Science*.

41. Collier, J.E., and K.A. Nasmyth. 2022. DNA passes through cohesin's hinge as well as its Smc3–kleisin interface. *eLife*. 11:e80310.

42. Haering, C.H., D. Schoffnegger, T. Nishino, W. Helmhart, K. Nasmyth, and J. Löwe. 2004. Structure and Stability of Cohesin's Smc1-Kleisin Interaction. *Molecular Cell*. 15:951–964.

43. Bürmann, F., L.F.H. Funke, J.W. Chin, and J. Löwe. 2021. Cryo-EM structure of MukBEF reveals DNA loop entrapment at chromosomal unloading sites. *Molecular Cell*. 81:4891-4906.e8.

44. Shi, Z., H. Gao, X. Bai, and H. Yu. 2020. Cryo-EM structure of the human cohesin-NIPBL-DNA complex. *Science*.

45. Higashi, T.L., P. Eickhoff, J.S. Sousa, J. Locke, A. Nans, H.R. Flynn, A.P. Snijders, G. Papageorgiou, N. O'Reilly, Z.A. Chen, F.J. O'Reilly, J. Rappsilber, A. Costa, and F. Uhlmann. 2020. A Structure-Based Mechanism for DNA Entry into the Cohesin Ring. *Molecular Cell*. 79:917-933.e9.

46. Collier, J.E., B.-G. Lee, M.B. Roig, S. Yatskevich, N.J. Petela, J. Metson, M. Voulgaris, A. Gonzalez Llamazares, J. Löwe, and K.A. Nasmyth. 2020. Transport of DNA within cohesin involves clamping on top of engaged heads by Scc2 and entrapment within the ring by Scc3. *eLife*. 9:e59560.

47. Yu, Y., S. Li, Z. Ser, H. Kuang, T. Than, D. Guan, X. Zhao, and D.J. Patel. 2022. Cryo-EM structure of DNA-bound Smc5/6 reveals DNA clamping enabled by multi-subunit conformational changes. *Proc Natl Acad Sci U S A*. 119:e2202799119.

48. Griese, J.J., and K.-P. Hopfner. 2011. Structure and DNA-binding activity of the Pyrococcus furiosus SMC protein hinge domain. *Proteins: Structure, Function, and Bioinformatics*. 79:558–568.

49. Vazquez Nunez, R., Y. Polyhach, Y.-M. Soh, G. Jeschke, and S. Gruber. 2021. Gradual opening of Smc arms in prokaryotic condensin. *Cell Reports*. 35:109051.

50. Truebestein, L., and T.A. Leonard. 2016. Coiled-coils: The long and short of it. *BioEssays*. 38:903–916.

51. Kim, J.-S., A. DeGiovanni, J. Jancarik, P.D. Adams, H. Yokota, R. Kim, and S.-H. Kim. 2005. Crystal structure of DNA sequence specificity subunit of a type I restriction-modification enzyme and its functional implications. *Proc Natl Acad Sci U S A*. 102:3248–3253.

52. Herrmann, H., S.V. Strelkov, P. Burkhard, and U. Aebi. 2009. Intermediate filaments: primary determinants of cell architecture and plasticity. *J Clin Invest*. 119:1772–1783.

53. Rao, L., F. Berger, M.P. Nicholas, and A. Gennerich. 2019. Molecular mechanism of cytoplasmic dynein tension sensing. *Nat Commun*. 10:3332.

54. Lee, H., and H.N. and J.-K. Ryu. 2021. Structure-function relationships of SMC protein complexes for DNA loop extrusion. 9:1–13.

55. Anderson, D.E., A. Losada, H.P. Erickson, and T. Hirano. 2002. Condensin and cohesin display different arm conformations with characteristic hinge angles. *Journal of Cell Biology*. 156:419–424.

56. Lammens, A., A. Schele, and K.-P. Hopfner. 2004. Structural biochemistry of ATP-driven dimerization and DNA-stimulated activation of SMC ATPases. *Curr Biol*. 14:1778–1782.

57. de Jager, M., J. van Noort, D.C. van Gent, C. Dekker, R. Kanaar, and C. Wyman. 2001. Human Rad50/Mre11 Is a Flexible Complex that Can Tether DNA Ends. *Molecular Cell*. 8:1129–1135.

58. Rotheneder, M., K. Stakyte, E. van de Logt, J.D. Bartho, K. Lammens, Y. Fan, A. Alt, B. Kessler, C. Jung, W.P. Roos, B. Steigenberger, and K.-P. Hopfner. 2023. Cryo-EM structure of the Mre11-Rad50-Nbs1 complex reveals the molecular mechanism of scaffolding functions. *Mol Cell*. 83:167-185.e9.

59. Bauer, B.W., I.F. Davidson, D. Canena, G. Wutz, W. Tang, G. Litos, S. Horn, P. Hinterdorfer, and J.-M. Peters. 2021. Cohesin mediates DNA loop extrusion by a “swing and clamp” mechanism. *Cell*. 184:5448-5464.e22.

60. Lee, B.-G., F. Merkel, M. Allegretti, M. Hassler, C. Cawood, L. Lecomte, F.J. O’Reilly, L.R. Sinn, P. Gutierrez-Escribano, M. Kschonsak, S. Bravo, T. Nakane, J. Rappsilber, L. Aragon, M. Beck, J. Löwe, and C.H. Haering. 2020. Cryo-EM structures of holo condensin reveal a subunit flip-flop mechanism. *Nat Struct Mol Biol*. 27:743–751.

61. Petela, N.J., A. Gonzalez Llamazares, S. Dixon, B. Hu, B.-G. Lee, J. Metson, H. Seo, A. Ferrer-Harding, M. Voulgaris, T. Gligoris, J. Collier, B.-H. Oh, J. Löwe, and K.A. Nasmyth. 2021. Folding of cohesin’s coiled coil is important for Scc2/4-induced association with chromosomes. *eLife*. 10:e67268.

62. Hallett, S.T., I. Campbell Harry, P. Schellenberger, L. Zhou, N.B. Cronin, J. Baxter, T.J. Etheridge, J.M. Murray, and A.W. Oliver. 2022. Cryo-EM structure of the Smc5/6 holo-complex. *Nucleic Acids Research*. 50:9505–9520.

63. Nomidis, S.K., E. Carlon, S. Gruber, and J.F. Marko. 2022. DNA tension-modulated translocation and loop extrusion by SMC complexes revealed by molecular dynamics simulations. *Nucleic Acids Research*. gkac268.

64. Diebold-Durand, M.-L., H. Lee, L.B. Ruiz Avila, H. Noh, H.-C. Shin, H. Im, F.P. Bock, F. Bürmann, A. Durand, A. Basfeld, S. Ham, J. Basquin, B.-H. Oh, and S. Gruber. 2017. Structure of Full-Length SMC and Rearrangements Required for Chromosome Organization. *Molecular Cell*. 67:334-347.e5.

65. Waldman, V.M., T.H. Stanage, A. Mims, I.S. Norden, and M.G. Oakley. 2015. Structural mapping of the coiled-coil domain of a bacterial condensin and comparative analyses across all domains of life suggest conserved features of SMC proteins. *Proteins: Structure, Function, and Bioinformatics*. 83:1027–1045.

66. Gruber, M., J. Söding, and A.N. Lupas. 2006. Comparative analysis of coiled-coil prediction methods. *Journal of Structural Biology*. 155:140–145.

67. Ludwiczak, J., A. Winski, K. Szczepaniak, V. Alva, and S. Dunin-Horkawicz. 2019. DeepCoil—a fast and accurate prediction of coiled-coil domains in protein sequences. *Bioinformatics*. 35:2790–2795.

68. McDonnell, A.V., T. Jiang, A.E. Keating, and B. Berger. 2006. Paircoil2: improved prediction of coiled coils from sequence. *Bioinformatics*. 22:356–358.

69. Gabler, F., S.-Z. Nam, S. Till, M. Mirdita, M. Steinegger, J. Söding, A.N. Lupas, and V. Alva. 2020. Protein Sequence Analysis Using the MPI Bioinformatics Toolkit. *Current Protocols in Bioinformatics*. 72:e108.

70. Zimmermann, L., A. Stephens, S.-Z. Nam, D. Rau, J. Kübler, M. Lozajic, F. Gabler, J. Söding, A.N. Lupas, and V. Alva. 2018. A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *Journal of Molecular Biology*. 430:2237–2243.

71. Lupas, A.N., and J. Bassler. 2017. Coiled Coils – A Model System for the 21st Century. *Trends in Biochemical Sciences*. 42:130–140.

72. Hartmann, M.D., C.T. Mendler, J. Bassler, I. Karamichali, O. Ridderbusch, A.N. Lupas, and B. Hernandez Alvarez. 2016. α/β coiled coils. *eLife*. 5:e11861.

73. O'Shea, E.K., J.D. Klemm, P.S. Kim, and T. Alber. 1991. X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science*. 254:539–544.

74. Jumper, J., R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S.A.A. Kohl, A.J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A.W. Senior, K. Kavukcuoglu, P. Kohli, and D. Hassabis. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature*. 596:583–589.

75. Varadi, M., S. Anyango, M. Deshpande, S. Nair, C. Natassia, G. Yordanova, D. Yuan, O. Stroe, G. Wood, A. Laydon, A. Žídek, T. Green, K. Tunyasuvunakool, S. Petersen, J. Jumper, E. Clancy, R. Green, A. Vora, M. Lutfi, M. Figurnov, A. Cowie, N. Hobbs, P. Kohli, G. Kleywegt, E. Birney, D. Hassabis, and S. Velankar. 2022. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research*. 50:D439–D444.

76. Lin, Z., H. Akin, R. Rao, B. Hie, Z. Zhu, W. Lu, N. Smetanin, R. Verkuil, O. Kabeli, Y. Shmueli, A. Dos Santos Costa, M. Fazel-Zarandi, T. Sercu, S. Candido, and A. Rives. 2023. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*. 379:1123–1130.

77. Pathania, A., W. Liu, A. Matityahu, J. Irudayaraj, and I. Onn. 2021. Chromosome loading of cohesin depends on conserved residues in Scc3. *Curr Genet*. 67:447–459.

78. Petela, N.J., T.G. Gligoris, J. Metson, B.-G. Lee, M. Voulgaris, B. Hu, S. Kikuchi, C. Chapard, W. Chen, E. Rajendra, M. Srinivisan, H. Yu, J. Löwe, and K.A. Nasmyth. 2018. Scc2 Is a Potent Activator of Cohesin's ATPase that Promotes Loading by Binding Scc1 without Pds5. *Molecular Cell*. 70:1134-1148.e7.

79. Moldovan, G.-L., B. Pfander, and S. Jentsch. 2006. PCNA controls establishment of sister chromatid cohesion during S phase. *Mol Cell*. 23:723–732.

80. Beckouët, F., B. Hu, M.B. Roig, T. Sutani, M. Komata, P. Uluocak, V.L. Katis, K. Shirahige, and K. Nasmyth. 2010. An Smc3 acetylation cycle is essential for establishment of sister chromatid cohesion. *Mol Cell*. 39:689–699.

81. van Ruiten, M.S., D. van Gent, Á. Sedeño Cacciatore, A. Fauster, L. Willems, M.L. Hekkelman, L. Hoekman, M. Altelaar, J.H.I. Haarhuis, T.R. Brummelkamp, E. de Wit, and B.D. Rowland. 2022. The cohesin acetylation cycle controls chromatin loop length through a PDS5A brake mechanism. *Nat Struct Mol Biol*. 29:586–591.

82. Bastié, N., C. Chapard, L. Dauban, O. Gadal, F. Beckouët, and R. Koszul. 2022. Smc3 acetylation, Pds5 and Scc2 control the translocase activity that establishes cohesin-dependent chromatin loops. *Nat Struct Mol Biol*. 29:575–585.

83. Chan, K.-L., T. Gligoris, W. Upcher, Y. Kato, K. Shirahige, K. Nasmyth, and F. Beckouët. 2013. Pds5 promotes and protects cohesin acetylation. *Proc Natl Acad Sci U S A*. 110:13020–13025.

84. Chapard, C., R. Jones, T. van Oepen, J.C. Scheinost, and K. Nasmyth. 2019. Sister DNA Entrapment between Juxtaposed Smc Heads and Kleisin of the Cohesin Complex. *Molecular Cell*. 75:224-237.e5.

85. Ferraro-Gideon, J., R. Sheykhani, Q. Zhu, M.L. Duquette, M.W. Berns, and A. Forer. 2013. Measurements of forces produced by the mitotic spindle using optical tweezers. *Mol Biol Cell*. 24:1375–1386.

86. Borges, V., C. Lehane, L. Lopez-Serra, H. Flynn, M. Skehel, T. Rolef Ben-Shahar, and F. Uhlmann. 2010. Hos1 deacetylates Smc3 to close the cohesin acetylation cycle. *Mol Cell*. 39:677–688.

87. Chatterjee, A., S. Zakian, X.-W. Hu, and M.R. Singleton. 2013. Structural insights into the regulation of cohesion establishment by Wpl1. *EMBO J*. 32:677–687.

88. Murayama, Y., and F. Uhlmann. 2015. DNA Entry into and Exit out of the Cohesin Ring by an Interlocking Gate Mechanism. *Cell*. 163:1628–1640.

89. Gandhi, R., P.J. Gillespie, and T. Hirano. 2006. Human Wapl is a cohesin-binding protein that promotes sister-chromatid resolution in mitotic prophase. *Curr Biol*. 16:2406–2417.

90. Beckouët, F., M. Srinivasan, M.B. Roig, K.-L. Chan, J.C. Scheinost, P. Batty, B. Hu, N. Petela, T. Gligoris, A.C. Smith, L. Strmecki, B.D. Rowland, and K. Nasmyth. 2016. Releasing Activity Disengages Cohesin's Smc3/Scc1 Interface in a Process Blocked by Acetylation. *Molecular Cell*. 61:563–574.

91. Kueng, S., B. Hegemann, B.H. Peters, J.J. Lipp, A. Schleiffer, K. Mechtler, and J.-M. Peters. 2006. Wapl controls the dynamic association of cohesin with chromatin. *Cell*. 127:955–967.

92. Tachibana-Konwalski, K., J. Godwin, L. van der Weyden, L. Champion, N.R. Kudo, D.J. Adams, and K. Nasmyth. 2010. Rec8-containing cohesin maintains bivalents without turnover during the growing phase of mouse oocytes. *Genes Dev*. 24:2505–2516.

93. Goldstein, L.S.B. 1980. Mechanisms of chromosome orientation revealed by two meiotic mutants in Drosophila melanogaster. *Chromosoma*. 78:79–111.

94. Davidson, I.F., B. Bauer, D. Goetz, W. Tang, G. Wutz, and J.-M. Peters. 2019. DNA loop extrusion by human cohesin. *Science*. 366:1338–1345.

95. Pradhan, B., T. Kanno, M. Umeda Igarashi, M.S. Loke, M.D. Baaske, J.S.K. Wong, K. Jeppsson, C. Björkegren, and E. Kim. 2023. The Smc5/6 complex is a DNA loop-extruding motor. *Nature*. 616:843–848.

96. Nagano, T., Y. Lubling, T.J. Stevens, S. Schoenfelder, E. Yaffe, W. Dean, E.D. Laue, A. Tanay, and P. Fraser. 2013. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature*. 502:59–64.

97. Merkenschlager, M., and E.P. Nora. 2016. CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation. *Annu. Rev. Genom. Hum. Genet.* 17:17–43.

98. de Wit, E., E.S.M. Vos, S.J.B. Holwerda, C. Valdes-Quezada, M.J.A.M. Verstegen, H. Teunissen, E. Splinter, P.J. Wijchers, P.H.L. Krijger, and W. de Laat. 2015. CTCF Binding Polarity Determines Chromatin Looping. *Mol Cell*. 60:676–684.

99. Davidson, I.F., R. Barth, M. Zaczek, J. van der Torre, W. Tang, K. Nagasaka, R. Janissen, J. Kerssemakers, G. Wutz, C. Dekker, and J.-M. Peters. 2023. CTCF is a DNA-tension-dependent barrier to cohesin-mediated loop extrusion. *Nature*. 616:822–827.

100. Nora, E.P., A. Goloborodko, A.-L. Valton, J.H. Gibcus, A. Uebersohn, N. Abdennur, J. Dekker, L.A. Mirny, and B.G. Bruneau. 2017. Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell*. 169:930-944.e22.

101. Tedeschi, A., G. Wutz, S. Huet, M. Jaritz, A. Wuensche, E. Schirghuber, I.F. Davidson, W. Tang, D.A. Cisneros, V. Bhaskara, T. Nishiyama, A. Vaziri, A. Wutz, J. Ellenberg, and J.-M. Peters. 2013. Wapl is an essential regulator of chromatin structure and chromosome segregation. *Nature*. 501:564–568.

102. Haarhuis, J.H.I., R.H. van der Weide, V.A. Blomen, J.O. Yáñez-Cuna, M. Amendola, M.S. van Ruiten, P.H.L. Krijger, H. Teunissen, R.H. Medema, B. van Steensel, T.R. Brummelkamp, E. de Wit, and B.D. Rowland. 2017. The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. *Cell*. 169:693-707.e14.

103. Busslinger, G.A., R.R. Stocsits, P. van der Lelij, E. Axelsson, A. Tedeschi, N. Galjart, and J.-M. Peters. 2017. Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. *Nature*. 544:503–507.

104. Olan, I., A.J. Parry, S. Schoenfelder, M. Narita, Y. Ito, A.S.L. Chan, G.S.C. Slater, D. Bihary, M. Bando, K. Shirahige, H. Kimura, S.A. Samarajiwa, P. Fraser, and M. Narita. 2020. Transcription-dependent cohesin repositioning rewires chromatin loops in cellular senescence. *Nat Commun*. 11:6049.

105. Cameron, G., D. Gruszka, S. Xie, K.A. Nasmyth, M. Srinivasan, and H. Yardimci. 2022. Sister chromatid cohesion establishment during DNA replication termination. 2022.09.15.508094.

106. Peters, J.-M. 2021. How DNA loop extrusion mediated by cohesin enables V(D)J recombination. *Current Opinion in Cell Biology*. 70:75–83.

107. Arnould, C., V. Rocher, A.-L. Finoux, T. Clouaire, K. Li, F. Zhou, P. Caron, P.E. Mangeot, E.P. Ricci, R. Mourad, J.E. Haber, D. Noordermeer, and G. Legube. 2021. Loop extrusion as a mechanism for formation of DNA damage repair foci. *Nature*. 590:660–665.

108. Stigler, J., G.Ö. Çamdere, D.E. Koshland, and E.C. Greene. 2016. Single-Molecule Imaging Reveals a Collapsed Conformational State for DNA-Bound Cohesin. *Cell Reports*. 15:988–998.

109. Davidson, I.F., D. Goetz, M.P. Zaczek, M.I. Molodtsov, P.J. Huis in 't Veld, F. Weissmann, G. Litos, D.A. Cisneros, M. Ocampo-Hafalla, R. Ladurner, F. Uhlmann, A. Vaziri, and J. Peters. 2016. Rapid movement and transcriptional re-localization of human cohesin on DNA. *EMBO J*. 35:2671–2685.

110. Ryu, J.-K., S.-H. Rah, R. Janissen, J.W.J. Kerssemakers, A. Bonato, D. Michieletto, and C. Dekker. 2022. Condensin extrudes DNA loops in steps up to hundreds of base pairs that are generated by ATP binding events. *Nucleic Acids Research*. 50:820–832.

111. Chen, J., S.A. Darst, and D. Thirumalai. 2010. Promoter melting triggered by bacterial RNA polymerase occurs in three steps. *Proc Natl Acad Sci U S A*. 107:12523–12528.

112. Shaltiel, I.A., S. Datta, L. Lecomte, M. Hassler, M. Kschonsak, S. Bravo, C. Stober, J. Ormanns, S. Eustermann, and C.H. Haering. 2022. A hold-and-feed mechanism drives directional DNA loop extrusion by condensin. *Science*. 376:1087–1094.

113. Pradhan, B., R. Barth, E. Kim, I.F. Davidson, B. Bauer, T. van Laar, W. Yang, J.-K. Ryu, J. van der Torre, J.-M. Peters, and C. Dekker. 2022. SMC complexes can traverse physical roadblocks bigger than their ring size. *Cell Rep*. 41:111491.

114. Ivanov, D., and K. Nasmyth. 2005. A topological interaction between cohesin rings and a circular minichromosome. *Cell*. 122:849–860.

115. Higashi, T.L., G. Pobegalov, M. Tang, M.I. Molodtsov, and F. Uhlmann. 2021. A Brownian ratchet model for DNA loop extrusion by the cohesin complex. *Elife*. 10:e67530.

116. Marko, J.F., P. De Los Rios, A. Barducci, and S. Gruber. 2019. DNA-segment-capture model for loop extrusion by structural maintenance of chromosome (SMC) protein complexes. *Nucleic Acids Research*. 47:6956–6972.

117. Lee, B.-G., J. Rhodes, and J. Löwe. 2022. Clamping of DNA shuts the condensin neck gate. *Proc. Natl. Acad. Sci. U.S.A.* 119:e2120006119.

118. Kschonsak, M., F. Merkel, S. Bisht, J. Metz, V. Rybin, M. Hassler, and C.H. Haering. 2017. Structural Basis for a Safety-Belt Mechanism That Anchors Condensin to Chromosomes. *Cell*. 171:588-600.e24.

119. Terakawa, T., S. Bisht, J.M. Eeftens, C. Dekker, C.H. Haering, and E.C. Greene. 2017. The condensin complex is a mechanochemical motor that translocates along DNA. *Science*. 358:672–676.

120. Kong, M., E.E. Cutts, D. Pan, F. Beuron, T. Kaliyappan, C. Xue, E.P. Morris, A. Musacchio, A. Vannini, and E.C. Greene. 2020. Human Condensin I and II Drive Extensive ATP-Dependent Compaction of Nucleosome-Bound DNA. *Mol Cell*. 79:99-114.e9.

121. Li, Y., J.H.I. Haarhuis, Á.S. Cacciatore, R. Oldenkamp, M.S. van Ruiten, L. Willems, H. Teunissen, K.W. Muir, E. de Wit, B.D. Rowland, and D. Panne. 2020. The structural basis for cohesin-CTCF anchored loops. *Nature*. 578:472–476.

122. Essiambre, R.-J. 2021. Arthur Ashkin: Father of the optical tweezers. *Proc Natl Acad Sci U S A*. 118:e2026827118.

123. Ashkin, A. 1980. Applications of laser radiation pressure. *Science*. 210:1081–1088.

124. Svoboda, K., and S.M. Block. 1994. Biological applications of optical forces. *Annu Rev Biophys Biomol Struct*. 23:247–285.

125. Chu, S., J.E. Bjorkholm, A. Ashkin, and A. Cable. 1986. Experimental Observation of Optically Trapped Atoms. *Phys. Rev. Lett.* 57:314–317.

126. Cecconi, C., E.A. Shank, C. Bustamante, and S. Marqusee. 2005. Direct observation of the three-state folding of a single protein molecule. *Science*. 309:2057–2060.

127. Bauer, D., S. Meinhold, R.P. Jakob, J. Stigler, U. Merkel, T. Maier, M. Rief, and G. Žoldák. 2018. A folding nucleus and minimal ATP binding domain of Hsp70 identified by single-molecule force spectroscopy. *PNAS*. 115:4666–4671.

128. Stigler, J., F. Ziegler, A. Gieseke, J.C.M. Gebhardt, and M. Rief. 2011. The Complex Folding Network of Single Calmodulin Molecules. *Science*. 334:512–516.

129. Ashkin, A., J.M. Dziedzic, and T. Yamane. 1987. Optical trapping and manipulation of single cells using infrared laser beams. *Nature*. 330:769–771.

130. Ishizaka, S., C. Yamamoto, and H. Yamagishi. 2021. Liquid-Liquid Phase Separation of Single Optically Levitated Water-Ionic Liquid Droplets in Air. *J Phys Chem A*. 125:7716–7722.

131. Bustamante, C.J., Y.R. Chemla, S. Liu, and M.D. Wang. 2021. Optical tweezers in single-molecule biophysics. *Nat Rev Methods Primers*. 1:25.

132. Gebhardt, J.C.M., T. Bornschlögl, and M. Rief. 2010. Full distance-resolved folding energy landscape of one single protein molecule. *Proc Natl Acad Sci USA*. 107:2013–2018.

133. Bustamante, C., J.F. Marko, E.D. Siggia, and S. Smith. 1994. Entropic elasticity of lambda-phage DNA. *Science*. 265:1599–1600.

134. Dietz, H., and M. Rief. 2004. Exploring the energy landscape of GFP by single-molecule mechanical experiments. *Proceedings of the National Academy of Sciences*. 101:16192–16197.

135. Morfill, J., F. Kühner, K. Blank, R.A. Lugmaier, J. Sedlmair, and H.E. Gaub. 2007. B-S Transition in Short Oligonucleotides. *Biophys J*. 93:2400–2409.

136. Wang, M.D., H. Yin, R. Landick, J. Gelles, and S.M. Block. 1997. Stretching DNA with optical tweezers. *Biophysical Journal*. 72:1335–1346.

137. Stigler, J., and M. Rief. 2012. Hidden Markov Analysis of Trajectories in Single-Molecule Experiments and the Effects of Missed Events. *ChemPhysChem*. 13:1079–1086.

138. Žoldák, G., J. Stigler, B. Pelz, H. Li, and M. Rief. 2013. Ultrafast folding kinetics and cooperativity of villin headpiece in single-molecule force spectroscopy. *Proceedings of the National Academy of Sciences*. 110:18156–18161.

139. Tolić-Nørrelykke, S.F., E. Schäffer, J. Howard, F.S. Pavone, F. Jülicher, and H. Flyvbjerg. 2006. Calibration of optical tweezers with positional detection in the back focal plane. *Review of Scientific Instruments*. 77:103101.

140. Berg-Sørensen, K., and H. Flyvbjerg. 2004. Power spectrum analysis for optical tweezers. *Review of Scientific Instruments*. 75:594–612.

141. von Hansen, Y., A. Mehlich, B. Pelz, M. Rief, and R.R. Netz. 2012. Auto- and cross-power spectral analysis of dual trap optical tweezer experiments using Bayesian inference. *Rev Sci Instrum*. 83:095116.

142. Cole, D., G. Young, A. Weigel, A. Sebesta, and P. Kukura. 2017. Label-Free Single-Molecule Imaging with Numerical-Aperture-Shaped Interferometric Scattering Microscopy. *ACS Photonics*. 4:211–216.

143. Young, G., N. Hundt, D. Cole, A. Fineberg, J. Andrecka, A. Tyler, A. Olerinyova, A. Ansari, E.G. Marklund, M.P. Collier, S.A. Chandler, O. Tkachenko, J. Allen, M. Crispin, N. Billington, Y. Takagi, J.R. Sellers, C. Eichmann, P. Selenko, L. Frey, R. Riek, M.R. Galpin, W.B. Struwe, J.L.P. Benesch, and P. Kukura. 2018. Quantitative mass imaging of single biological macromolecules. *Science*. 360:423–427.

144. Synakewicz, M., R.S. Eapen, A. Perez-Riba, P.J.E. Rowling, D. Bauer, A. Weißl, G. Fischer, M. Hyvönen, M. Rief, L.S. Itzhaki, and J. Stigler. 2022. Unraveling the Mechanics of a Repeat-Protein Nanospring: From Folding of Individual Repeats to Fluctuations of the Superhelix. *ACS Nano*. 16:3895–3905.

145. van der Sleen, L.M., and K.M. Tych. 2021. Bioconjugation Strategies for Connecting Proteins to DNA-Linkers for Single-Molecule Force-Based Experiments. *Nanomaterials (Basel)*. 11:2424.

146. Avellaneda, M.J., E.J. Koers, D.P. Minde, V. Sunderlikova, and S.J. Tans. 2020. Simultaneous sensing and imaging of individual biomolecular complexes enabled by modular DNA–protein coupling. *Commun Chem*. 3:20.

147. Swoboda, M., J. Henig, H.-M. Cheng, D. Brugger, D. Haltrich, N. Plumeré, and M. Schlierf. 2012. Enzymatic oxygen scavenging for photostability without pH drop in single-molecule experiments. *ACS Nano*. 6:6364–6369.

148. Soh, Y.-M., J. Basquin, and S. Gruber. 2021. A rod conformation of the Pyrococcus furiosus Rad50 coiled coil. *Proteins: Structure, Function, and Bioinformatics*. 89:251–255.

149. Green, N.S., E. Reisler, and K.N. Houk. 2001. Quantitative evaluation of the lengths of homobifunctional protein cross-linking reagents used as molecular rulers. *Protein Sci*. 10:1293–1304.

150. Wells, J.M., and S.A. McLuckey. 2005. Collision-induced dissociation (CID) of peptides and proteins. *Methods Enzymol*. 402:148–185.

151. Götze, M., J. Pettelkau, S. Schaks, K. Bosse, C.H. Ihling, F. Krauth, R. Fritzsche, U. Kühn, and A. Sinz. 2012. StavroX—A Software for Analyzing Crosslinked Products in Protein Interaction Studies. *J. Am. Soc. Mass Spectrom.* 23:76–87.

152. Zawadzka, K., P. Zawadzki, R. Baker, K.V. Rajasekar, F. Wagner, D.J. Sherratt, and L.K. Arciszewska. 2018. MukB ATPases are regulated independently by the N- and C-terminal domains of MukF kleisin. *eLife*. 7:e31522.

153. Ladurner, R., V. Bhaskara, P.J. Huis in 't Veld, I.F. Davidson, E. Kreidl, G. Petzold, and J.-M. Peters. 2014. Cohesin's ATPase activity couples cohesin loading onto DNA with Smc3 acetylation. *Curr Biol*. 24:2228–2237.

154. Gutierrez-Escribano, P., M.D. Newton, A. Llauró, J. Huber, L. Tanasie, J. Davy, I. Aly, R. Aramayo, A. Montoya, H. Kramer, J. Stigler, D.S. Rueda, and L. Aragon. 2019. A conserved ATP- and Scc2/4-dependent activity for cohesin in tethering DNA molecules. *Sci Adv*. 5:eaay6804.

155. Badrinarayanan, A., R. Reyes-Lamothe, S. Uphoff, M.C. Leake, and D.J. Sherratt. 2012. In Vivo Architecture and Action of Bacterial Structural Maintenance of Chromosome Proteins. *Science*. 338:528–531.

156. Xiang, S., and D. Koshland. 2021. Cohesin architecture and clustering in vivo. *eLife*. 10:e62243.

157. Keeble, A.H., P. Turkki, S. Stokes, I.N.A. Khairil Anuar, R. Rahikainen, V.P. Hytönen, and M. Howarth. 2019. Approaching infinite affinity through engineering of peptide–protein interaction. *Proc Natl Acad Sci U S A*. 116:26523–26533.

158. Liu, H.W., F. Roisné-Hamelin, B. Beckert, Y. Li, A. Myasnikov, and S. Gruber. 2022. DNA-measuring Wadjet SMC ATPases restrict smaller circular plasmids by DNA cleavage. *Mol Cell*. 82:4727-4740.e6.

159. Banigan, E.J., A.A. van den Berg, H.B. Brandão, J.F. Marko, and L.A. Mirny. 2020. Chromosome organization by one-sided and two-sided loop extrusion. *eLife*. 9:e53558.

160. Marcos-Alcalde, Í., J.I. Mendieta-Moreno, B. Puisac, M.C. Gil-Rodríguez, M. Hernández-Marcos, D. Soler-Polo, F.J. Ramos, J. Ortega, J. Pié, J. Mendieta, and P. Gómez-Puertas. 2017. Two-step ATP-driven opening of cohesin head. *Sci Rep*. 7:3266.

161. Pobegalov, G., L.-Y. Chu, J.-M. Peters, and M.I. Molodtsov. 2023. Single cohesin molecules generate force by two distinct mechanisms. *Nat Commun*. 14:3946.

162. Wutz, G., C. Várnai, K. Nagasaka, D.A. Cisneros, R.R. Stocsits, W. Tang, S. Schoenfelder, G. Jessberger, M. Muhar, M.J. Hossain, N. Walther, B. Koch, M. Kueblbeck, J. Ellenberg, J. Zuber, P. Fraser, and J.-M. Peters. 2017. Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *The EMBO Journal*. 36:3573–3599.

163. Haering, C.H., A.-M. Farcas, P. Arumugam, J. Metson, and K. Nasmyth. 2008. The cohesin ring concatenates sister DNA molecules. *Nature*. 454:297–301.

164. Boardman, K., S. Xiang, F. Chatterjee, U. Mbonu, V. Guacci, and D. Koshland. 2023. A model for Scc2p stimulation of cohesin's ATPase and its inhibition by acetylation of Smc3p. *Genes Dev*. 37:277–290.

165. Srinivasan, M., J.C. Scheinost, N.J. Petela, T.G. Gligoris, M. Wissler, S. Ogushi, J.E. Collier, M. Voulgaris, A. Kurze, K.-L. Chan, B. Hu, V. Costanzo, and K.A. Nasmyth. 2018. The Cohesin Ring Uses Its Hinge to Organize DNA Using Non-topological as well as Topological Mechanisms. *Cell*. 173:1508-1519.e18.

166. Nagasaka, K., I.F. Davidson, R.R. Stocsits, W. Tang, G. Wutz, P. Betty, M. Panarotto, G. Litos, A. Schleiffer, D.W. Gerlich, and J.-M. Peters. 2022. Cohesin mediates DNA loop extrusion and sister chromatid cohesion by distinct mechanisms. Cell Biology.

167. Kurze, A., K.A. Michie, S.E. Dixon, A. Mishra, T. Itoh, S. Khalid, L. Strmecki, K. Shirahige, C.H. Haering, J. Löwe, and K. Nasmyth. 2011. A positively charged channel within the Smc1/Smc3 hinge required for sister chromatid cohesion. *EMBO J*. 30:364–378.

168. Zhang, H., Z. Shi, E.J. Banigan, Y. Kim, H. Yu, X. Bai, and I.J. Finkelstein. 2022. CTCF and R-loops are boundaries of cohesin-mediated DNA looping. Molecular Biology.

169. Gerlich, D., B. Koch, F. Dupeux, J.-M. Peters, and J. Ellenberg. 2006. Live-cell imaging reveals a stable cohesin-chromatin interaction after but not before DNA replication. *Curr Biol*. 16:1571–1578.

170. Liu, Y., and J. Dekker. 2022. CTCF–CTCF loops and intra-TAD interactions show differential dependence on cohesin ring integrity. *Nat Cell Biol*. 24:1516–1527.

171. Minamino, M., C. Bouchoux, B. Canal, J.F.X. Diffley, and F. Uhlmann. 2023. A replication fork determinant for the establishment of sister chromatid cohesion. *Cell*. 186:837-849.e11.

172. Murayama, Y., and F. Uhlmann. 2015. DNA Entry into and Exit out of the Cohesin Ring by an Interlocking Gate Mechanism. *Cell*. 163:1628–1640.

173. Wutz, G., R. Ladurner, B.G. St Hilaire, R.R. Stocsits, K. Nagasaka, B. Pignard, A. Sanborn, W. Tang, C. Várnai, M.P. Ivanov, S. Schoenfelder, P. van der Lelij, X. Huang, G. Dürnberger, E. Roitinger, K. Mechtler, I.F. Davidson, P. Fraser, E. Lieberman-Aiden, and J.-M. Peters. 2020. ESCO1 and CTCF enable formation of long chromatin loops by protecting cohesinSTAG1 from WAPL. *Elife*. 9:e52091.

174. Sanborn, A.L., S.S.P. Rao, S.-C. Huang, N.C. Durand, M.H. Huntley, A.I. Jewett, I.D. Bochkov, D. Chinnappan, A. Cutkosky, J. Li, K.P. Geeting, A. Gnirke, A. Melnikov, D. McKenna, E.K. Stamenova, E.S. Lander, and E.L. Aiden. 2015. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc Natl Acad Sci U S A*. 112:E6456-6465.

175. García-Nieto, A., A. Patel, Y. Li, R. Oldenkamp, L. Feletto, J.J. Graham, L. Willems, K.W. Muir, D. Panne, and B.D. Rowland. 2023. Structural basis of centromeric cohesion protection. *Nat Struct Mol Biol*. 30:853–859.

176. Nishiyama, T., R. Ladurner, J. Schmitz, E. Kreidl, A. Schleiffer, V. Bhaskara, M. Bando, K. Shirahige, A.A. Hyman, K. Mechtler, and J.-M. Peters. 2010. Sororin mediates sister chromatid cohesion by antagonizing Wapl. *Cell*. 143:737–749.

177. Bürmann, F., and J. Löwe. 2023. Structural biology of SMC complexes across the tree of life. *Curr Opin Struct Biol*. 80:102598.

178. Zhang, H., Z. Shi, E.J. Banigan, Y. Kim, H. Yu, X.-C. Bai, and I.J. Finkelstein. 2023. CTCF and R-loops are boundaries of cohesin-mediated DNA looping. *Mol Cell*. S1097-2765(23)00520–8.

179. Ryu, J.-K., C. Bouchoux, H.W. Liu, E. Kim, M. Minamino, R. de Groot, A.J. Katan, A. Bonato, D. Marenduzzo, D. Michieletto, F. Uhlmann, and C. Dekker. 2021. Bridging-induced phase separation induced by cohesin SMC protein complexes. *Sci Adv*. 7:eabe5905.

180. Gu, B., C.J. Comerci, D.G. McCarthy, S. Saurabh, W.E. Moerner, and J. Wysocka. 2020. Opposing Effects of Cohesin and Transcription on CTCF Organization Revealed by Super-resolution Imaging. *Molecular Cell*. 80:699-711.e7.

181. Kentepozidou, E., S.J. Aitken, C. Feig, K. Stefflova, X. Ibarra-Soria, D.T. Odom, M. Roller, and P. Flicek. 2020. Clustered CTCF binding is an evolutionary mechanism to maintain topologically associating domains. *Genome Biol*. 21:5.

182. Nichols, M.H., and V.G. Corces. 2018. A tethered-inchworm model of SMC DNA translocation. *Nat Struct Mol Biol*. 25:906–910.

VI Acknowledgements

First, I would like to thank Prof. Dr. Johannes Stigler for being a fantastic supervisor, for helpful discussions, debates and ideas. I am thankful for the introduction to exciting cutting-edge techniques in the single-molecule field and to open research questions in the highly dynamic SMC field. I also appreciate the enormous trust he but in me after we had started the lab in the end of 2017.

I am also very grateful to Dr. Sarah Zernia, who supported me the most during the publication process of my paper and the making of this manuscript. She helped me to advance my writing and illustrating skills and had the best advices in any biochemical topics. Without her, I would still be stuck between misfolded states, like ΔEB at medium tension. I cannot imagine a better office partner.

Next, I want to thank the present and past members of the Stigler Crew, who helped me a lot during our time together in the lab and during our weekly meetings, where I gained more and more biochemical understanding of the things I did the previous week over the years. Also the exciting social events and retreats were highlights of my PhD time.

Further highlights were the seminars and retreats with my graduate school RTG 1721 in the beginning of my PhD time, which introduced me to a variety of helpful and kind researchers at the campus. Especially, the big help and resilience of the Hopfner lab, the Rief lab, the Gaub lab and Marie Synakewicz were key to kick-start the Stigler lab and to make my PhD thesis possible.

I want to thank Dr. Thomas Fröhlich, Dr. Franz Herzog and Dr. Daniel Bollschweiler for their support and the kind introduction into mass spectroscopy, crosslinking or mass photometry and for providing excellent core facilities.

Lastly, my biggest gratitude goes to my beloved wife, who endured my lab-related absences and gave me the opportunity to work efficiently at home during the pandemic, while the lab and also the child-care had to close for a long time span. I am also thankful for the persistent support of my parents and my family.

P.S.: Just pull on it! That’s always a great idea!