PORTLAND
PRESS

Research Article

SMCHD1's ubiquitin-like domain is required for N-terminal dimerization and chromatin localization

Alexandra D. Gurzau[1,2], Christopher R. Horne[1,2], Yee-Foong Mok[3], Megan Iminitoff[1,2], Tracy A. Willson[1,2], Samuel N. Young[1,2], Marnie E. Blewitt[1,2] and James M. Murphy[1,2]

[1]The Walter and Eliza Hall Institute of Medical Research, 1G Royal Parade, Parkville, Melbourne, Victoria 3052, Australia; [2]Department of Medical Biology, University of Melbourne, Melbourne, Victoria 3010, Australia; [3]Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Melbourne, Victoria 3010, Australia

Correspondence: Marnie E. Blewitt (blewitt@wehi.edu.au) or James M. Murphy (jamesm@wehi.edu.au)

OPEN ACCESS

Structural maintenance of chromosomes flexible hinge domain-containing 1 (SMCHD1) is an epigenetic regulator that mediates gene expression silencing at targeted sites across the genome. Our current understanding of SMCHD1's molecular mechanism, and how substitutions within SMCHD1 lead to the diseases, facioscapulohumeral muscular dystrophy (FSHD) and Bosma arhinia microphthalmia syndrome (BAMS), are only emerging. Recent structural studies of its two component domains — the N-terminal ATPase and C-terminal SMC hinge — suggest that dimerization of each domain plays a central role in SMCHD1 function. Here, using biophysical techniques, we demonstrate that the SMCHD1 ATPase undergoes dimerization in a process that is dependent on both the N-terminal UBL (Ubiquitin-like) domain and ATP binding. We show that neither the dimerization event, nor the presence of a C-terminal extension past the transducer domain, affect SMCHD1's *in vitro* catalytic activity as the rate of ATP turnover remains comparable to the monomeric protein. We further examined the functional importance of the N-terminal UBL domain in cells, revealing that its targeted deletion disrupts the localization of full-length SMCHD1 to chromatin. These findings implicate UBL-mediated SMCHD1 dimerization as a crucial step for chromatin interaction, and thereby for promoting SMCHD1-mediated gene silencing.

Introduction

In 2008, SMCHD1 was identified as an epigenetic regulator essential for the maintenance of X-chromosome inactivation [1]. Subsequently, SMCHD1 has been attributed roles in the transcriptional silencing of clustered autosomal genes important in normal development and disease [2–8]. Yet exactly how SMCHD1 interacts with chromatin to facilitate gene expression regulation remains unknown, and thus remains of outstanding interest.

Full-length human SMCHD1 is a large, 2005-amino acid multidomain protein that exerts ATPase activity through its N-terminal GHKL (Gyrase B, Hsp90, histidine kinase and MutL) domain [9–11], and interacts with nucleic acids via its C-terminal SMC (structural maintenance of chromosomes) hinge domain [12–14]. These two functional domains are separated by a linker region that remains largely uncharacterized, yet constitutes more than half of the full-length protein as it comprises approximately 1200 amino acids. In canonical SMC proteins, the C-terminal hinge domain acts as a primary dimerization site that mediates heterodimeric SMC complex formation, establishing functional complexes such as cohesin and condensin [15–17]. Further interactions with non-SMC subunits at their N-terminus, where an ABC-type ATPase domain resides, leads to the formation of a closed ring structure that has been proposed to topologically entrap or encircle DNA [18,19]. Recently, we described the X-ray crystal structure of SMCHD1's hinge domain, which revealed the domain to assemble into an unusual donut shape via homodimerization [13]. Coupled with site-directed

mutagenesis studies, the hinge domain crystal structure enabled us to identify two positively charged surface patches, rather than the central pore, as mediators of nucleic acid interactions, in addition to the domain's role in dimerizing SMCHD1.

SMCHD1 is considered a member of the GHKL superfamily of ATPases owing to its N-terminal catalytic domain. The GHKL superfamily encompasses members such as the highly studied molecular chaperone heat shock protein 90 (Hsp90), the microrchidia (MORC) family of proteins, DNA topoisomerases and DNA mismatch repair proteins of the MutL family, among others [20]. While GHKL family members take part in diverse functions, they possess a common structure as they each comprise an α/β sandwich where three α-helices form a layer parallel to a four β-stranded antiparallel sheet, also known as a Bergerat ATP binding fold which is essential for ATPase function [20]. This structure contains various conserved residues, including a glutamic acid (Glu) in Motif I that activates the water molecule for ATP hydrolysis, and an asparagine (Asn) that is responsible for coordinating a magnesium ion (Mg^{2+}) to the active site [20]. An interesting feature common to several GHKL members is the ability to undergo ATP-induced homodimerization. Their catalytic cycle is often described as a molecular clamp mechanism, whereby interaction with ATP triggers dimerization at the GHKL N-terminus and capture of substrate (protein or nucleic acids), followed by ATP hydrolysis which leads to opening of the dimer and substrate release [20,21].

SMCHD1's GHKL ATPase has been of particular interest due to the identification of disease-related variants that are frequently located within this region of the protein. Variations in the human *SMCHD1* gene have been associated with two debilitating conditions: facioscapulohumeral muscular dystrophy (FSHD) [5] and Bosma arhinia microphthalmia syndrome (BAMS) [22,23]. While FSHD-associated variants in *SMCHD1* span the entire gene, all reported BAMS-associated variants are located exclusively within the ATPase region of SMCHD1, indicating that an altered catalytic activity may underlie the disease pathogenesis. Indeed, we previously showed that several BAMS-associated variants exhibit an enhanced catalytic activity in the context of the SMCHD1 ATPase protein, while FSHD-related variants almost exclusively display a reduced ATP hydrolysis activity [24]. However, the overall role of ATP hydrolysis in SMCHD1's function as an epigenetic regulator, and therefore the mechanisms by which pathogenic variants influence SMCHD1 function, remain poorly understood.

The first three-dimensional structure of a human SMCHD1 ATPase (residues 25–580) was recently reported, revealing the presence of an α/β sandwich arrangement that constitutes its active site—a conserved feature across members of the GHKL superfamily [10]. Furthermore, the crystal structure revealed a novel ubiquitin-like (UBL) domain located at the N-terminus that was proposed to mediate dimerization by undergoing a domain-swapping event via an N-terminal β-strand. Important to note is that the presented structure was obtained from a catalytically inactive point variant, E147A, which is unable to hydrolyze ATP. Upon examining the SMCHD1 ATPase protein under native PAGE conditions, Pedersen *et al.* [10] showed that a proportion of the E147A variant migrated as a dimer, yet surprisingly the wild-type counterpart or other disease-associated variants remained largely monomeric. This finding prompted us to explore the conformation of the wild-type human SMCHD1 ATPase and determine whether N-terminal dimerization is a native property of the protein, rather than a point variant-specific behavior.

Using both sedimentation velocity analytical centrifugation (AUC) and small-angle X-ray scattering (SAXS) analyses, we revealed that the wild-type SMCHD1 ATPase is able to undergo dimerization in a manner dependent on both its UBL domain and the ligand, ATP. While we observed that addition of the non-hydrolysable ATP analog, AMPPNP, promoted self-association of the SMCHD1 UBL construct, approximately half the protein population remained monomeric across varying protein concentrations, consistent with the existence of a monomer : dimer equilibrium in solution. Immunofluorescence microscopy demonstrated that deletion of the UBL domain from full-length SMCHD1 abrogates its chromatin localization in cells, suggesting a requirement for SMCHD1 dimerization via the ATPase domain in recruitment to target genes. Interestingly, SMCHD1's *in vitro* ATPase activity was unaffected by the presence or absence of the UBL domain, or extending the C-terminal sequence beyond the transducer domain, indicating that dimerization does not impact the rate of ATP hydrolysis. Because ATPase activity has been implicated in dissociation of other GHKL ATPases, and how SMCHD1's ATPase activity compares with other family members was unknown, we compared SMCHD1's catalytic activity to other GHKL family exemplars, revealing comparable ATP turnover rates. These studies provide insights into the connection between dimerization and the ATPase activity of SMCHD1, and highlight the importance of N-terminal dimerization in the recruitment of the full-length protein to chromatin where it can promote gene silencing.

Experimental

Protein expression

The DNA sequence encoding human SMCHD1 residues 25–580, 111–580, 111–702 or 25–702, as well as the N-terminal region of human MORC2 (residues 1–603) [25], were cloned into pFastBac Htb (Invitrogen), for expression using the Bac-to-Bac system (Invitrogen) following generation of bacmids in *E. coli* DH10MultiBac (ATG Biosynthetics), using established procedures [26]. These constructs encode an N-terminal, TEV protease-cleavable His$_6$ tag for affinity purification. Proteins were expressed and purified from *Sf*21 insect cells as described previously [24]. Transfections were performed as described [26], where lipid complexes of 1 μg of bacmid DNA was prepared by mixing with CellFectin II, before application in a total volume of 1 ml Insect-XPRESS protein-free medium with L-glutamine (Lonza) to 0.9×10^6 *Sf*21 cells adhered to the well of a 6-well plate. After 5 h static incubation at 27°C, the media was replaced with 2 ml Insect-XPRESS and the supernatant containing P1 virus harvested after 4 days static incubation at 27°C. Typically, *Sf*21 cells were maintained in Insect-XPRESS media, in suspension at 27°C, under shaking conditions at 130 rpm. P2 virus was generated by addition of 1 ml P1 virus to 100 ml *Sf*21 cells at a density of 1.5×10^6 cells/ml and incubated at 27°C, under shaking conditions for 4 days at 130 rpm. For protein expression, 0.5 l cell cultures were grown in 2.8 l Fernbach flasks, shaking at 90 rpm, 27°C to a density of $3.0–3.5 \times 10^6$ cells/ml before infection with an empirically defined ratio of P2 virus for an optimal protein yield. Cell pellets were harvested 48 h following infection, by centrifugation at 500 *g* for 5 min at room temperature, and pellets were snap-frozen in liquid nitrogen for storage at −80°C.

Full-length human Hsp90α [27] was sub-cloned into pPROEX HTb, whereas the N-terminal domain of human MLH1 (residues 1–340) was cloned into pET28-MHL vector (Addgene plasmid #26096). For both constructs, proteins were expressed in *E. coli* BL21-Codon Plus (DE3)-RIL cells. Briefly, transformed cells were cultured in Super broth at 37°C under shaking conditions at 200 rpm, until an OD$_{600}$ ∼ 0.6–0.8 was reached. The temperature was then reduced to 18°C and protein expression was induced with the addition of 0.5 mM IPTG overnight.

Recombinant protein purification

*Sf*21 insect cell or bacterial cell pellets were resuspended in lysis buffer (0.5 M NaCl, 20 mM Tris–HCl pH 8.0, 20% (v/v) glycerol, 5 mM imidazole pH 8.0, 0.5 mM TCEP), supplemented with 1 mM PMSF and 1X cOmplete EDTA-free protease inhibitor (Roche). Cells were disrupted by sonication while maintaining the lysate at 4°C, and insoluble material was removed by centrifugation at 45 000 *g* for 30 min at 4°C. Supernatants were subjected to Ni-chromatography (cOmplete His-Tag purification resin, Roche) and following extensive washing, eluted in lysis buffer containing 250 mM imidazole pH 8.0. Following cleavage of the His-tag with TEV protease, human SMCHD1 and Hsp90 proteins were buffer-exchanged into Buffer A (50 mM NaCl, 25 mM HEPES pH 7.5, 0.5 mM TCEP, 10% (v/v) glycerol) and loaded onto a MonoQ 5/50 GL column (GE Healthcare) pre-equilibrated with Buffer A, and eluted via a 0–100% gradient of Buffer B (500 mM NaCl, 25 mM HEPES pH 7.5, 0.5 mM TCEP, 10% (v/v) glycerol) over 40 column volumes for protein elution. SMCHD1-containing fractions of interest eluted at ∼250 mM NaCl and Hsp90-containing fractions eluted at ∼600 mM NaCl. These were pooled and concentrated, and further purified by Superdex-200 10/300 GL size exclusion chromatography (Cytiva) with elution in 100 mM NaCl, 20 mM HEPES pH 7.5, 0.5 mM TCEP. For the human MORC2 protein, ion exchange chromatography was omitted and size exclusion chromatography (SEC) was performed following Ni-chromatography; for human MLH1, a subtractive Ni-chromatography step was included prior to SEC to remove the His-tagged TEV protease. Briefly, the cleaved protein was diluted in lysis buffer and incubated with Ni-NTA resin (cOmplete His-Tag, Roche) for 1 h at 4°C, on rollers. The resin was washed with increasing imidazole concentrations of up to 35 mM imidazole, where the protein of interest remained in the unbound fractions and the TEV protease bound to the resin. Unbound fractions were pooled, concentrated and further purified by SEC. Protein purity was evaluated by reducing SDS–PAGE with Stain-Free visualization (Biorad) and fractions of interest were pooled, snap-frozen in liquid nitrogen and stored at −80°C until required.

ATPase assays

Fluorescence polarization ATPase assays were performed as outlined in Chen K *et al.* [9]. Ten microliter reactions were set up in triplicates in 384-well low flange, black, flat-bottom plates (Corning) containing 7 μl

reaction buffer (50 mM HEPES pH 7.5, 4 mM MgCl$_2$, 2 mM EGTA), 1 µl recombinant protein at concentrations ranging from 0.1 to 0.6 µM or SEC buffer control, 1 µl nuclease-free water and 1.25–10 µM ATP substrate. Reactions were incubated at 20°C for 1 h in the dark. Reactions were stopped by the addition of 10 µl detection mix (1× Detection buffer, 4 nM ADP Alexa Fluor 633 Tracer, 128 µg/ml ADP^2 antibody) and incubated for another hour in the dark. Fluorescence polarization readings (mP) were measured using an Envision plate reader (PerkinElmer Life Sciences) fitted with excitation filter 620/40 nm, emission filters 688/45 nm (s and p channels) and D658/fp688 dual mirror. Readings from a free tracer (no antibody) control were set as 20 mP as the normalization baseline of the assay for all reactions. The amount of ADP produced by each reaction was estimated by a 12-point standard curve, as outlined in the manufacturer's protocol. Data were plotted and analyzed in GraphPad Prism. A second type of ATPase assay was performed using the ADP-Glo Kinase Assay kit (Promega). Each reaction was performed in a total of 5 µl, consisting of a final 3.125, 6.25, 12.5 or 25 µM SMCHD1 protein (residues 25–702 or 111–702), either 20, 50 or 100 µM ATP, and reaction buffer (50 mM HEPES pH 7.5, 4 mM MgCl$_2$, 2 mM EGTA). Reactions were incubated at 20°C for 1 h, followed by the addition of 5 µl ADP-Glo reagent to terminate the reaction by depleting ATP, and a further incubation for 40 min at 20°C. Ten microliter of Kinase Detection reagent was then added to each reaction and incubated for a further 30 min at 20°C. Luminescence was measured using the FLUOstar Omega microplate reader (BMG Labtech). Data were plotted and analyzed in GraphPad Prism.

Analytical ultracentrifugation

Sedimentation velocity experiments were performed with a XL-I analytical ultracentrifuge (Beckman Coulter) using double sector quartz cells and epon center-pieces in an An-50 Ti 8-hole rotor. Data were obtained at 50 000 rpm using 350 µl protein at 1.0 mg/ml concentration for the initial experiment, and varying concentrations (0.25, 0.5, 1.0, 1.5 and 2.0 mg/ml) for the following experiment. The protein samples were diluted in 100 mM NaCl, 20 mM HEPES pH 7.5, 0.5 mM TCEP, which was used as the reference buffer. A total of 100 scans were collected at 20°C using radial absorbance scans at 290 nm and a step size of 0.003 cm. Data were analyzed using SEDFIT [28] and Ultrascan [29]. Sedimentation data were fitted to a continuous size distribution [c(s)], and fit data were plotted using GUSSI [30] and analyses are reported in Table 1. The buffer density, buffer viscosity and an estimate of the partial specific volume of the protein based on the amino acid sequence were determined using SEDNTERP. Data were also subjected to two-dimensional spectrum analyses, and van Holde-Weischet analyses in UltraScan 4.0 [29,31].

Small-angle x-ray scattering (SAXS) data collection and analysis

SAXS data were collected at the Australian Synchrotron SAXS/WAXS beamline using the co-flow, in-line SEC setup [32], as previously described [9,24]. Data collection and analysis statistics are shown in Supplementary Table S1. Fifty microliter of 5 mg/ml recombinant protein was resolved by injection on to an in-line Superdex-200 Increase 5/150 GL column (Cytiva) in 200 mM NaCl, 20 mM HEPES pH 7.5, 5% (v/v) glycerol, 0.5 mM TCEP (Pierce) and eluted in the path in the beam via a quartz capillary. For samples with added ligand, protein was incubated with 1 mM AMPPNP/Mg^{2+} on ice for 30 min prior to the commencement of the experiment, and the buffer used for these samples was as described above for the apo samples but with the addition of 1 mM ATP/ Mg^{2+}. One second exposures of scattering data were collected with a PILATUS3 × 2 M detector, radially averaged, and scattering data from the apex of the SEC peak were background subtracted using data collected for buffer-only shots earlier in the data collection using Scatterbrain software (Stephen

Table 1. Summary of AUC analysis of SMCHD1 ATPase (residues 25–702) in the presence of AMPPNP/Mg^{2+}

[SMCHD1] mg/ml	Weight average sedimentation coefficient	Monomer (%)	Dimer (%)
0.25	4.62	88.23	8.53
0.5	4.69	76.84	11.94
1	5.02	61.48	26.22
1.5	5.23	53.42	35.30
2	5.28	57.64	38.00

Mudie, Australian Synchrotron). Data analyses were performed using the ATSAS suite [33] as described previously [24]. Guinier analyses were performed using PRIMUS [34] to examine scatter at very low q ($qR_g \leq 1.3$) to estimate radius of gyration, R_g, and zero angle intensity ($I(0)$), with linearity indicating the absence of both high molecular mass aggregates and interparticle interference. The real space interatomic distance distribution function, $P(r)$, and the maximum dimension of the particle, D_{MAX}, were computed by indirect Fourier transform using GNOM [35], which also allowed estimation of R_g and $I(0)$. The atomic co-ordinates from a SMCHD1 (residues 25–702) monomer or dimer model (Figure 1c) were used to obtain theoretical scattering curves for comparison with experimental data, using the program CRYSOL [33].

CRISPR-cas9-mediated SMCHD1-knockout in HEK293 cells

We designed guide RNAs (gRNAs) targeting the first exon of SMCHD1 (forward 5′CAAACAAGTACACCG TCCTG; reverse 5′ GGGGAGCGCTCGGACTACGC). These were cloned into the lentiGuide-puro vector (Addgene #52963), whereas Cas9 was delivered in a blasticidin resistant lentivirus generated from a pLentiGuide-BlastR vector. To produce lentivirus, HEK293T cells were cultured in DMEM supplemented with 10% (v/v) fetal bovine serum (FBS), penicillin (100 U ml^{-1}), streptomycin (100 µg ml^{-1}) at 37°C in a humidified atmosphere with 5% (v/v) CO_2. Cells plated on a 10 cm plate at 80% confluency were transfected with a lentiviral packaging cocktail (14 µg of psPAX vector, 5.6 µg of pVSV-G vector, 8.4 µg of LentiGuide-BlastR vector or LentiGuide-puro vector, 144 µL of 0.5 M $CaCl_2$, 1.2 ml HBS and 1 ml of DNAse-free water), which was vortexed for 10 s and incubated for 10 min at room temperature prior to adding onto cells. The cells were incubated overnight and media was changed the next day. One day later, the supernatant was harvested and filtered. HEK293 cells were first transduced with the cas9 lentivirus and selected for with blasticidin, followed by the gRNA lentiviral transduction and selection with puromycin. For either, lentiviral supernatant was prepared 1 : 10 in media containing 4 µg/ml polybrene (Sigma–Aldrich) and added to HEK293 cells at 50% confluence. After 24 h, the media was changed and 5 µg/ml puromycin (Sigma–Aldrich) or 5 µg/ml blasticidin S (InvivoGen) were added for selection of transduced cells. Clonal cell lines were generated via single cell isolation, and verified by Illumina Next Generation sequencing (FWD: 5′GTGACCTATGAAC TCAGGAGTCtcgcgtacctgacacacaca, REV 5′ CTGAGACTTGCACATCGCAGCcgctgtcttttctccttttc), as described previously [36].

Deletion of the UBL domain (Δ1–110) in mouse full-length SMCHD1 was accomplished using PCR-mediated cloning before introduction into a pcDNA3 vector. For transfection of SMCHD1-knockout HEK293 cells with constructs containing full-length SMCHD1 or the UBL-deletion variant, $\sim 2 \times 10^4$ clonal SMCHD1-KO HEK293 cells were seeded in a 12-well plate on a 13 mm coverslip (Marienfield Superior). 24 h following plating, cells at $\sim$80% confluency were transfected with 1.2 µg of the corresponding construct using calcium phosphate-mediated transfection. Immunofluorescence was performed 24 h post-transfection.

Immunofluorescence microscopy

Immunofluorescence microscopy studies were performed as described previously [37]. Briefly, cells were washed in PBS and fixed in 3% (w/v) paraformaldehyde made in PBS for 10 min at room temperature. Cells were washed three times with PBS for 5 min each, then permeabilized on ice with 0.5% (v/v) Triton X-100 in PBS, followed by three washes in PBS for 5 min each. Cells were blocked in 1% (w/v) bovine serum albumin (BSA) (Life Technologies) for 15 min, followed by an overnight incubation in a dark and humid chamber at 4°C with a primary anti-SMCHD1 antibody (in-house clone 2B8; available from Millipore under catalog number MABS2292) diluted 1 : 100 in 1% (w/v) BSA. Cells were washed three times in PBS for 5 min each and incubated for 40 min at room temperature in a dark and humid chamber with a secondary anti-rat-568 antibody for SMCHD1 (Life Technologies, A-11077) diluted 1 : 500 in 1% (w/v) BSA. Cells were washed three times in PBS for 5 min each and stained with DAPI for 10 min at room temperature, followed by another two PBS washes. Coverslips were mounted in Vectashield H1000 mounting medium (Vector Laboratories). Cells were visualized using the Zeiss LSM 880 NLO microscope at 63× magnification and z-stacks were acquired. Images were analyzed using the open source ImageJ distribution package, FIJI.

Immunoblot

Samples were resolved by standard reducing SDS–PAGE analysis on a 4–12% Bis-Tris gel (Thermo Fisher Scientific) in MES buffer and transferred to a PVDF membrane (Osmonics, GE Healthcare) by wet transfer at 100 V for 1 h in transfer buffer (25 mM Tris, 192 mM glycine, 20% v/v methanol). Membranes were blocked
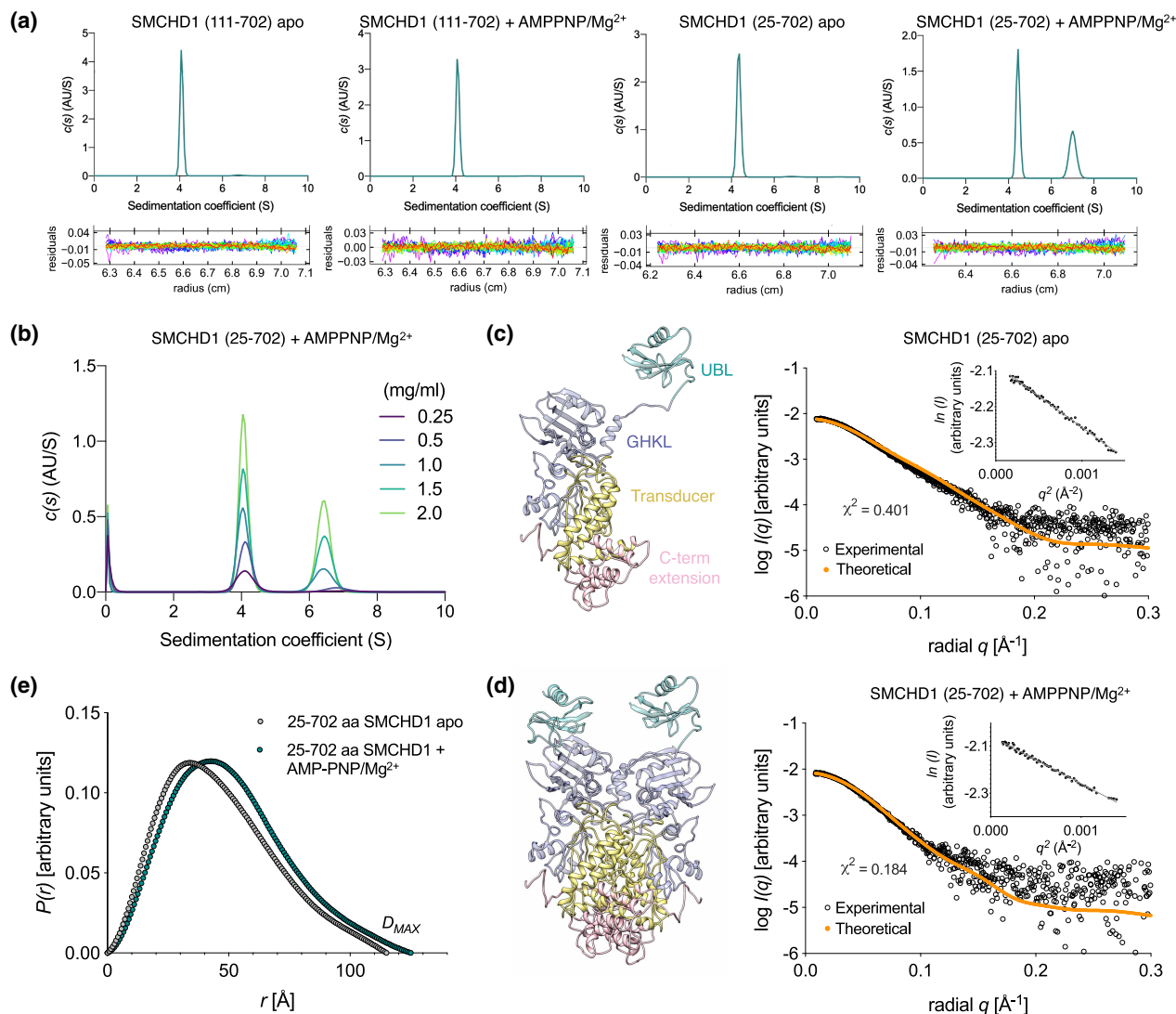
Figure 1. Wild-type SMCHD1 ATPase dimerizes in solution when in the presence of the UBL domain and AMPPNP/Mg^{2+}.
(**a**) Sedimentation velocity data obtained at protein concentrations of 1.0 mg/ml were fitted to a continuous sedimentation coefficient [c(s)] distribution model, with residuals for the best fit of the data shown below each panel. Protein samples were either in an apo condition, denoting the absence of ligand, or in the presence of 1 mM AMPPNP and Mg^{2+}. (**b**) Sedimentation velocity analysis was performed as in (**a**), but only for the SMCHD1 (residues 25–702) construct in the presence of 1 mM AMPPNP and Mg^{2+} and at varying protein concentrations, as indicated in the legend. (**c,d**; left) Three-dimensional models of the monomeric (**c**) and dimeric (**d**) UBL-containing SMCHD1 ATPase (residues 25–702) were obtained by merging a previously developed model of the 111–702 residue SMCHD1 ATPase based on the high-resolution structure of the GHKL protein Hsp90 (PDB:1CG9) [9], and the SMCHD1 ATPase (residues 25–580) crystal structure (PDB:6MW7) [10]. Cartoons were drawn using PyMOL. (**c,d**; right) SAXS scattering profiles for UBL-containing SMCHD1 (residues 25–702) under apo conditions (**c**) or in the presence of AMPPNP/Mg^{2+} (**d**), displaying the log intensity of scattered X-rays, $I(q)$, as a function of momentum transfer, q, in Å^{-1}. Experimental scatter is shown in black, whereas the theoretical scatter, which was obtained via CRYSOL from the models on the left is shown in orange, with the quality of fit estimated by a chi-squared (χ^2) value. Guinier plots are shown as insets, where linearity indicates the presence of monodisperse particles, free from any measurable amounts of aggregation or interparticle interference. (**e**) Pair-distribution functions, $P(r)$, for UBL-containing SMCHD1 (residues 25–702) apo (gray) and SMCHD1 with AMPPNP/Mg^{2+} (turquoise), obtained via Fourier transformations of the scattering intensity using the software GNOM. The maximum dimension (D_{MAX}) values for each occur where the curve bisects the x-axis. Data shown in (**a–e**) are representative of two independent experiments.

with a 5% (v/v) skim milk powder in 0.1% (v/v) Tween-20/PBS for 1 h at room temperature. Primary antibody was added to the membrane in 5 ml blocking buffer and incubated overnight at 4°C in a capped tube, on rollers. Membranes were washed for 30 min at room temperature with 0.1% (v/v) Tween-20/PBS, followed by incubation with HRP-conjugated secondary antibody for 1 h at room temperature, which was diluted in 5 ml blocking buffer. The 30-min washing step was repeated once more, and antibody binding was visualized using the Luminata ECL system (Millipore) following the manufacturer's instructions on a ChemiDoc Touch Imaging System (Bio-Rad).

Results
Dimerization of the SMCHD1 ATPase is dependent on the ubiquitin-like (UBL) domain and ATP binding

The recently published structure of the SMCHD1 ATPase (residues 25–580) harbored the catalytically inactive point mutant, E147A, for which preferential dimerization was observed over the wild-type counterpart [10]. This led us to examine whether the wild-type SMCHD1 ATPase similarly adopts a dimeric conformation. To do so, we investigated an extended SMCHD1 ATPase construct that encompasses residues 25–702, extending past the transducer domain at the C-terminus while incorporating the newly identified UBL domain (residues 25–110). We compared this to a SMCHD1 construct that lacks the UBL domain (ΔUBL; residues 111–702; Supplementary Figure S1a,b), which we previously demonstrated to occur as a monomeric species in solution [9,24] by performing sedimentation velocity AUC experiments. When the sedimentation data were fitted with a continuous sedimentation coefficient [c(s)] distribution model, we observed only a single species for the ΔUBL SMCHD1 ATPase (residues 111–702), both in the absence and presence of a non-hydrolysable analog of ATP, AMPPNP, and a cofactor for the hydrolysis reaction, Mg^{2+}, with sedimentation coefficients of 4.09S and 4.10S, respectively (Figure 1a). The estimated molecular mass values for the single species were 65.1 kDa and 65.3 kDa for the apo and AMPPNP conditions, respectively, consistent with a calculated monomeric mass of 68.0 kDa for the ΔUBL SMCHD1 ATPase construct (residues 111–702). For the UBL-containing SMCHD1 (residues 25–702), we detected a monomeric species in the absence of ligand with a sedimentation coefficient of 4.45S and an estimated molecular mass of 73.8 kDa. However, in the presence of AMPPNP/Mg^{2+}, we detected both a monomeric population in addition to a higher molecular mass species that corresponds to 7.01S and a molecular mass of 126.0 kDa, where the calculated monomeric mass for this construct is 78.0 kDa (Figure 1a). These data are consistent with the wild-type SMCHD1 ATPase undergoing a dimerization event that is dependent on both the presence of the N-terminal UBL domain and an ATP-mimetic ligand.

To gain further insight into the self-association properties of the UBL-containing SMCHD1 ATPase (residues 25–702), we repeated the sedimentation velocity AUC studies in the presence of AMPPNP/Mg^{2+}, this time at varying protein concentrations of 0.25, 0.5, 1.0, 1.5 and 2.0 mg/ml (Figure 1b). Under these experimental conditions, we observe a shift towards an increasing dimer population with a corresponding rise in protein concentration, which indicates the presence of a concentration-dependent self-association between monomer and dimer. Furthermore, this is demonstrated both by the proportion of each oligomeric species and an increase in the weight average sedimentation coefficient, calculated by integrating the area under the monomer and dimeric species within the c(s) distribution (Table 1). The sedimentation velocity data were also analyzed using the van Holde–Weischet method [31], which is a model-independent analysis that directly assesses the shape of the sedimenting boundary to provide an estimate of the proportion of each species in solution. These analyses support the idea that increased protein concentration favors the prevalence of the dimer over the monomeric form (Supplementary Figure S1c). We also assessed whether the dimeric SMCHD1 population remained associated in solution, as opposed to reverting back to a monomer-dimer ratio. Here, we pooled the fractions from size exclusion chromatography containing SMCHD1 ATPase dimer (Supplementary Figure S2a) and subjected the protein to a second round of size exclusion chromatography. Interestingly, the reloaded protein eluted as a single peak that corresponded to the dimeric protein, indicating that the SMCHD1 ATPase homodimer is preserved in the presence of AMPPNP/Mg^{2+} (Supplementary Figure S2b).

We further examined the conformation of wild-type SMCHD1 UBL-containing ATPase protein (residues 25–702) using SAXS. SMCHD1 ATPase was eluted from a Superdex-200 size exclusion column into the capillary in the path of the SAXS beamline at the Australian Synchrotron and images were collected every 1 s. The capillary was equipped with a 'co-flow' setup in which buffer was delivered in a laminar manner most proximal to the capillary to prevent capillary fouling/deposition if the protein were to undergo radiation damage [32,38].

The Guinier analysis for the protein under either apo or ligand conditions produced linear plots (Figure 1c–d; insets), which is consistent with a single monodisperse species in solution and an absence of interparticle interference, enabling more detailed analyses of the samples. We previously established the radius of gyration (R_g) of the wild-type, ΔUBL SMCHD1 ATPase (residues 111–702) as ∼32 Å, with a maximum particle dimension (D_{MAX}) of 105 Å [9], and further confirmed these parameters in this study (Supplementary Figure S3a,b; Supplementary Table S1). These values correspond to a monomeric configuration, additionally validated here by our sedimentation velocity AUC studies (Figure 1a). To establish the conformational states of the unliganded and ligand-bound SMCHD1 UBL-containing ATPase (residues 25–702), we computed the theoretical scattering curves for either a monomeric or dimeric UBL-containing SMCHD1 ATPase (residues 25–702) and compared these with experimentally obtained data (Figure 1c,d). These models are in agreement with the experimental data, as evidenced by χ^2 values of 0.401 for the fit of the monomeric SMCHD1 ATPase model to the monomeric scatter pattern, and 1.191 for the AMPPNP-bound SMCHD1 (Supplementary Figure S3c). The former is an excellent fit to the data; the reference value for this fit is 0.25, because the data reduction software, Scatterbrain, records 2 standard errors rather than the more conventional 1 standard error. Based on the elution volume from size exclusion chromatography, we concluded that the AMPPNP-bound UBL-containing SMCHD1 ATPase was the dimer form. By comparison, the dimer model fitted the AMPPNP-bound UBL-containing SMCHD1 scatter with a χ^2 value of 0.184 (Figure 1d). Interestingly, the dimer model also fit the monomeric (*sans* AMPPNP) UBL-containing SMCHD1 ATPase very well, with $\chi^2 = 0.180$ (Supplementary Figure S3d). For the experimental data of the UBL-containing SMCHD1 ATPase (-AMPPNP) we observed a R_g of 36 Å and a D_{MAX} of 115 Å, whereas in the presence of AMPPNP, the indicated R_g was 39 Å and the D_{MAX} value was 125 Å (Figure 1e; Supplementary Table S1). The similarity between the parameters calculated for the monomer and dimer forms was surprising, but illustrates that the general topology and envelope of the monomeric and dimeric forms are remarkably similar. Importantly, molecular mass estimates from our scattering data using SAXSMoW [39] confirmed that, despite the similarity of their fits to the monomeric SMCHD1 ATPase scattering data, a molecular mass estimate of 95.7 kDa (cf. calculated molecular mass of 78 kDa) was determined from the monomer (−AMPPNP) scattering data, whereas the estimated molecular mass from the scattering of UBL-containing SMCHD1 (+AMPPNP) was 139.7 kDa, consistent with a dimeric species (Supplementary Table S1). Taken together, these results are consistent with a dimeric arrangement of the SMCHD1 ATPase in the presence of AMPPNP/Mg^{2+}.

SMCHD1's UBL domain is required for the cellular localization of the full-length protein

To assess the functional role of the UBL domain in the context of the full-length protein, we designed a mouse SMCHD1 construct that excludes the N-terminal region where the UBL domain resides (Δ1–110). We transiently transfected constructs encoding either wild-type full-length mouse SMCHD1, the UBL-deleted SMCHD1 (Δ1–110) or the E147A SMCHD1 point variant that represents an established catalytically inactive mutant, into CRISPR-Cas9-edited, SMCHD1-knockout 293 cells (SMCHD1-KO 293). This is an ideal cellular system to study SMCHD1 function because mouse SMCHD1 complements the endogenous human SMCHD1 function and the exogenous mouse SMCHD1 constructs are not susceptible to editing by the human targeting gRNA sequences. We validated the SMCHD1-KO efficiency by both immunoblot (Figure 2a) as well as immunofluorescence (Figure 2b). Wild-type 293 cells illustrate the native nuclear localization pattern of SMCHD1, which is denoted by two bright nuclear foci per cell that correspond to SMCHD1's localization to the two inactive X-chromosomes present in tetraploid 293 cells (Figure 2b). Transfection of wild-type full-length SMCHD1 into the SMCHD1-KO 293 cells resulted in the formation of multiple nuclear foci, a localization pattern that differs from the wild-type control cells where two bright nuclear foci are observed in each cell, most likely as a result of the overexpression of the construct. Nonetheless, deletion of the UBL domain from the full-length protein (in a construct encoding residues 111–2007) compromised nuclear foci formation, resulting in a diffuse SMCHD1 staining pattern that does not resemble wild-type SMCHD1-transfected cells (Figure 2b). The E147A catalytically inactive SMCHD1 point variant exhibited a similar nuclear localization pattern, equally unable to form nuclear foci. These findings suggest that both ATPase domain dimerization mediated by SMCHD1's UBL domain, as well as ATPase activity, are required for proper chromatin localization and focus formation by the full-length protein.
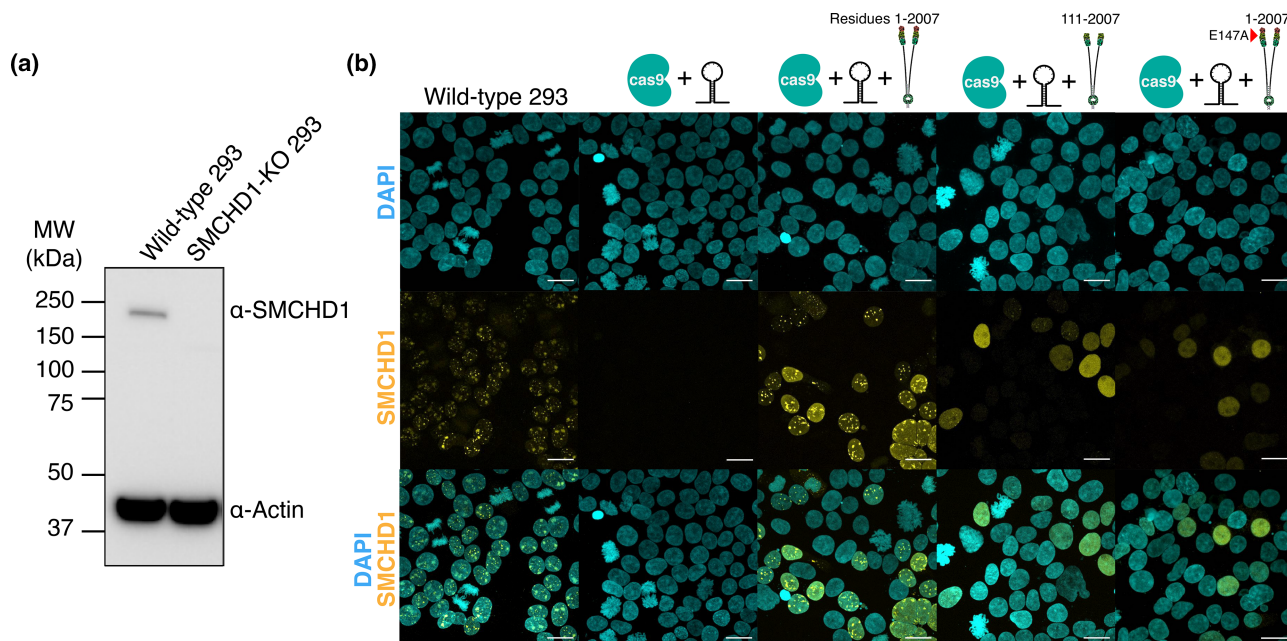
PORTLAND
PRESS

Figure 2. SMCHD1's N-terminal UBL domain is required for its localization to chromatin.
(**a**) Immunoblot of wild-type and CRISPR-Cas9-induced SMCHD1-KO 293 cells, blotting for SMCHD1 and actin as a loading control. Molecular mass markers are indicated on the left, in kilodaltons. (**b**) Immunofluorescence in (left to right) wild-type 293 cells, CRISPR-Cas9-induced SMCHD1-KO 293 cells, SMCHD1-KO 293 cells transfected with full-length wild-type mouse SMCHD1, SMCHD1-KO 293 cells transfected with a ΔUBL SMCHD1 (111–2007), or with full-length wild-type mouse SMCHD1 with an E147A mutation. DAPI staining is depicted in cyan and SMCHD1 is shown in yellow, with merged images shown below as indicated. Maximum intensity projection images are shown as representative of n > 50 nuclei positive for SMCHD1 overexpression per sample, and data shown are representative of two independent experiments. Scale bars, 20 μm.

The UBL domain or a C-terminal extension do not alter the catalytic activity of the SMCHD1 ATPase

Previously, we characterized a SMCHD1 construct that encompasses residues 111–702, extending past the transducer domain. This C-terminal extension (residues 580–702) contains five reported FSHD2-associated missense mutations in *SMCHD1*: W596G, V615D, P622L, V641L and P690S [14]. We formerly investigated one of these, P690S, and showed that it exhibits a decreased catalytic activity *in vitro* [24]. We attempted to study an additional *SMCHD1* variant, V615D, but were unable to produce sufficient recombinant protein due to a low expression yield, suggesting the Val to Asp substitution is likely destabilizing (data not shown). Overall, these data indicated that the C-terminal extension downstream of the transducer domain of SMCHD1 likely holds an important functional role. We, therefore, set out to examine the extended 25–702 amino acid wild-type SMCHD1 ATPase protein, which incorporates both the newly identified UBL domain and the C-terminal extension (Figure 3a), to determine its catalytic activity *in vitro* in comparison with SMCHD1 constructs encompassing residues 25–580, 111–580 and 111–702. While at lower protein concentrations, SMCHD1 constructs encompassing residues 25–580 and 111–580 exert a higher ATP turnover than the 111–702 SMCHD1 counterpart, similar trends were observed at higher protein concentrations across all three constructs (Figure 3b). The presence of the UBL domain within the extended SMCHD1 construct (residues 25–702) similarly shows no increase in catalytic activity compared with the ΔUBL SMCHD1 ATPase (residues 111–702) counterpart (Figure 3c,d). Similarly, when assayed at concentrations at which we observed the construct to undergo dimerization, the UBL-containing SMCHD1 ATPase (25–702) did not exhibit increased catalytic activity, but rather a slight decline (Figure 3e,f), raising the prospect that the ATP turnover might be conformationally or sterically impacted in the dimeric form. These findings suggest that the dimerization event does not enhance ATP turnover, inferring the absence of a synergistic behavior upon dimerization of the SMCHD1 ATPase. Our data are consistent with a previous report, where the presence of the UBL domain on the
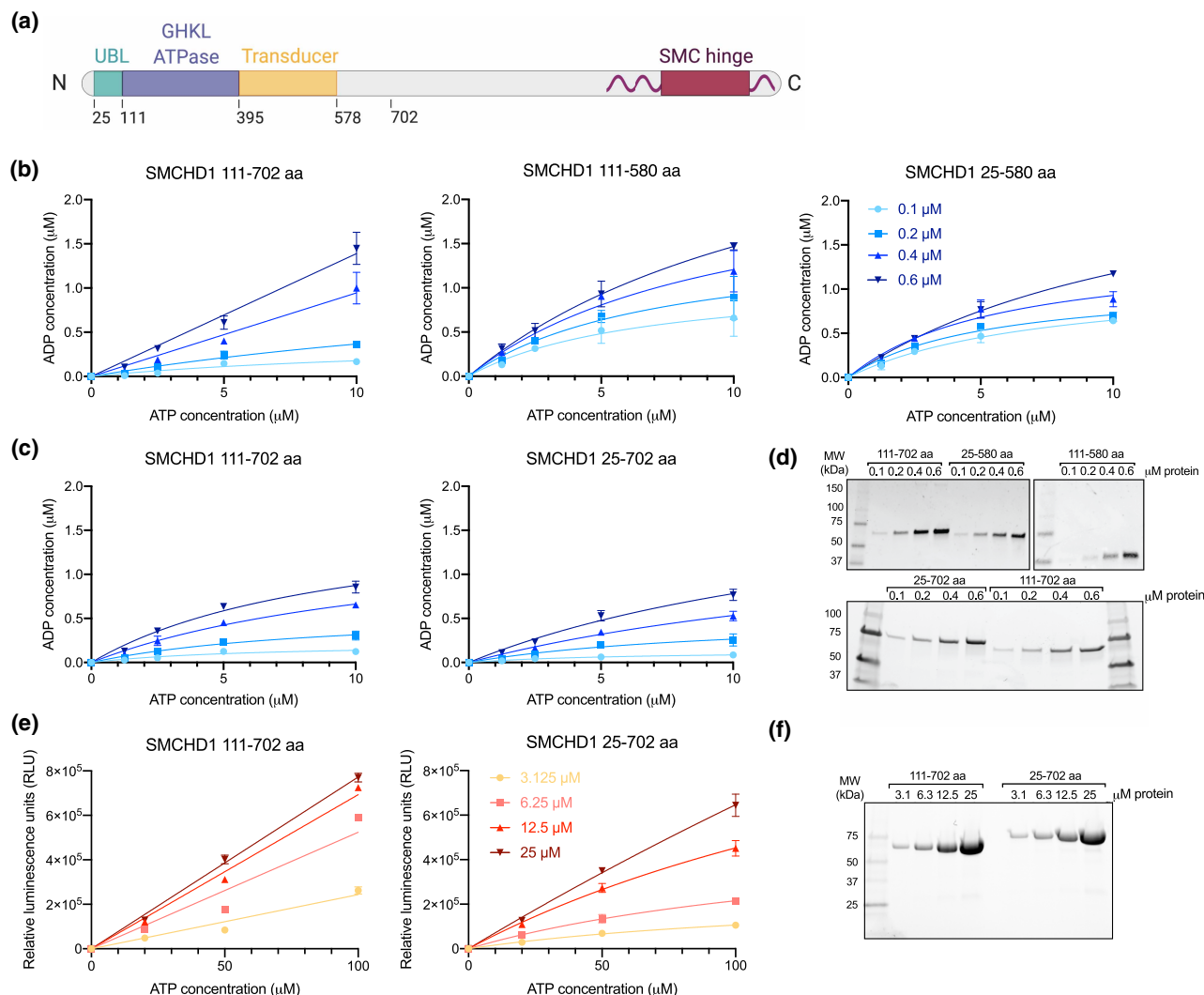
Figure 3. SMCHD1's UBL domain or C-terminal extension do not alter its catalytic activity *in vitro*.
(**a**) Schematic depiction of the architecture of full-length SMCHD1, with the amino acid residue numbers corresponding to the domain boundaries shown below. (**b,c**) *In vitro* fluorescence polarization ATPase assays of the indicated SMCHD1 constructs, where the *x*-axis denotes the ATP concentrations used: 1.25, 2.5, 5.0 and 10 µM and the *y*-axis indicates amount of ADP produced (µM). Protein concentrations (0.1, 0.2, 0.4 and 0.6 µM) are shown in different shades of blue as indicated in the legend. Each measurement was performed in triplicate for all constructs in parallel; mean ± standard deviation is shown for these triplicates. Each chart is representative of data obtained in duplicate independent replicates. (**d**) Reducing stain-free SDS–PAGE analyses of prepared SMCHD1 protein dilutions used in the corresponding assays. Molecular mass marker position is shown on the left, in kilodaltons. (**e**) ADP-Glo assay of ΔUBL SMCHD1 ATPase (residues 111–702) and UBL-containing SMCHD1 ATPase (residues 25–702) at higher protein concentrations of 3.125, 6.25, 12.5 and 25 µM (depicted in shades of red), The *y*-axis represents the relative luminescence units (RLU) produced by the reactions, whereas the *x*-axis denotes the final ATP concentrations present in the reactions (20, 50 and 100 µM). Data shown are from technical triplicates and are representative of two independent experiments. (**f**) Reducing stain-free SDS–PAGE gel of the prepared SMCHD1 protein dilutions used in the ADP-Glo assay, with molecular mass marker position depicted on the left in kilodaltons.

SMCHD1 ATPase domain (residues 25–580) did not impact the ATP hydrolysis activity compared with a construct lacking the UBL (residues 111–580) [10]. Taken together, these data demonstrate that the *in vitro* catalytic activity of SMCHD1 remains unaltered in the presence of either the N-terminal UBL domain or the C-terminal extension downstream of the transducer domain.

SMCHD1's ATPase activity is comparable to that of related GHKL-type proteins

Initial structural analyses of the ΔUBL SMCHD1 ATPase (residues 111–702) using SAXS revealed a gross conformational similarity to full-length Hsp90 [9], and biochemical studies demonstrated a susceptibility to a well-established Hsp90 inhibitor, Radicicol [9,11]. Thus, while it is evident that SMCHD1 exhibits topological similarities to other GHKL family ATPases, including the capacity to undergo ATP-induced N-terminal dimerization, it has remained unclear whether there are intrinsic differences between SMCHD1 and other GHKL family members' catalytic activities in the absence of direct comparisons. We addressed this knowledge gap by directly comparing the catalytic activities of UBL-containing SMCHD1 ATPase (residues 25–702) with *H. sapiens* full-length Hsp90α, the MutL homolog 1 (MLH1) ATPase domain (residues 1–340) and the extended MORC2 ATPase (residues 1–603) *in vitro* using our established endpoint fluorescence polarization assay (Figure 4a–d). Upon assessing their activity, we found that full-length Hsp90 exhibits a comparable catalytic rate to SMCHD1, with an estimated turnover rate (k_{cat}) of ∼0.015 μM ADP/min/μM protein, akin to 0.018 μM ADP/min/μM protein obtained for SMCHD1 (Figure 4a,b). The monomeric ATPase domain of MLH1 displayed a slightly higher turnover rate of 0.025 μM ADP/min/μM protein (Figure 4c), whereas the MORC2 dimer exhibited a k_{cat} of 0.047 μM ADP/min/μM protein in this assay (Figure 4d). Overall, these results suggest
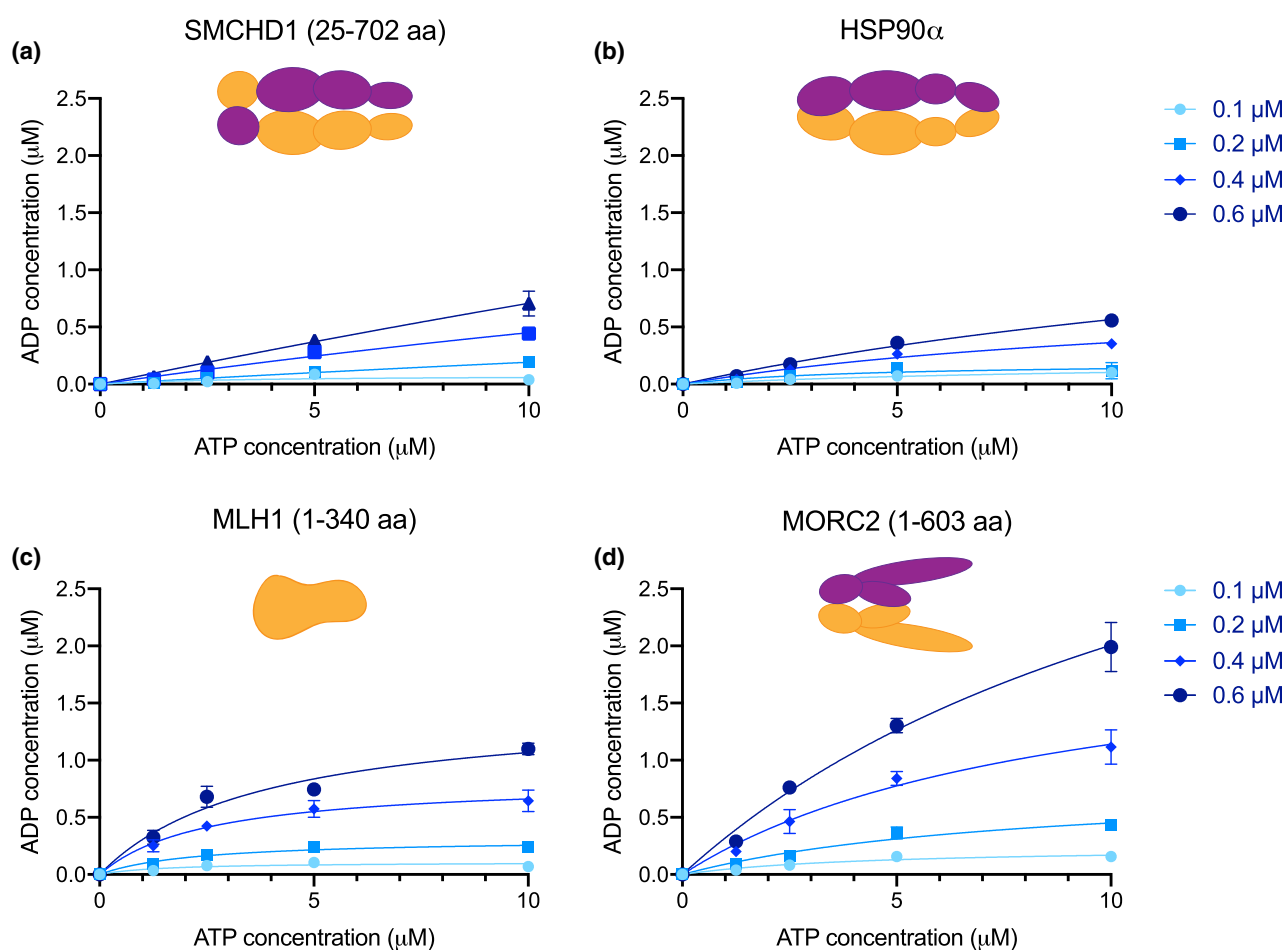

Figure 4. The catalytic rate of SMCHD1's ATPase is comparable to that of other GHKL-type proteins.
(**a–d**) *In vitro* fluorescence polarization ATPase assays of *Homo sapiens* (**a**) SMCHD1 (residues 25–702), (**b**) full-length Hsp90α (isoform 2), (**c**) MLH1 N-terminal domain (residues 1–340) and (**d**) MORC2 N-terminal region (residues 1–603). The *x*-axis denotes the ATP concentrations used: 1.25, 2.5, 5.0 and 10 μM and the *y*-axis indicates amount of ADP produced (μM). Protein concentrations (0.1, 0.2, 0.4 and 0.6 μM) are shown in different shades of blue as indicated in the legend on the right. Each measurement was performed in triplicate for all constructs in parallel; mean ± standard deviation is shown for these triplicates. Each chart is representative of data obtained in duplicate independent replicates.

that the catalytic rate of SMCHD1's ATPase is comparable to that of other GHKL-type proteins, but most similar to full-length Hsp90.

Discussion

The growing evidence for the role of SMCHD1 variants in human disease has led to increased interest and, consequently, many recent advances in our current understanding of SMCHD1's molecular structure and function. Both component domains of SMCHD1 — the N-terminal ATPase and the C-terminal hinge — can independently dimerize. While the hinge domain is constitutively dimeric [12,13], the ATPase can interchange between monomeric and dimeric forms [10], although the underlying basis was incompletely understood. In this study, we explored the dimerization properties of the wild-type SMCHD1 ATPase and demonstrated that, similarly to the recently reported E147A SMCHD1 mutant [10], the wild-type protein is able to dimerize via a mechanism reliant on the N-terminal UBL domain and ATP binding. GHKL-type proteins have been commonly described to function via a molecular clamp mechanism where ATP-dependent N-terminal dimerization dictates the opening and closing of the dimer and is directly coupled to the catalytic cycle [20,21]. SMCHD1 may likewise behave as a molecular clamp around chromatin, transitioning between the open and closed states to engage and disengage from target sites. The propensity of the SMCHD1 ATPase to undergo dimerization in the presence of ATP raises the possibility of an intrinsic behavior that may regulate SMCHD1's residency time and consequently gene silencing function at chromatin targets. ATP binding is known to trigger intermolecular interactions that further result in N-terminal dimerization among the GHKL ATPase family [20], a behavior that likely also occurs in SMCHD1, leading to UBL domain swapping and the dimerization event that we observe. While the presence of a UBL domain has not yet been reported in other members of the GHKL superfamily, the N-terminal β-strap that precedes the UBL domain in SMCHD1 and co-ordinates dimerization via a domain-swapping event is also found among other GHKL members, for example in the molecular chaperone Hsp90 [40,41]. Despite the primary dimerization interface of Hsp90 being situated within the C-terminal domain of the protein, N-terminal dimerization via the β-strap was suggested to promote efficient ATP hydrolysis [42].

The E147A SMCHD1 mutant from which the first three-dimensional structure of a SMCHD1 ATPase (residues 25–580) was solved, was reported to elute as a mixture of monomer and dimer from SEC in the presence of ATP, from which only the dimer peak was selected for crystallization [10]. Here, we observed a similar phenomenon for the extended wild-type SMCHD1 ATPase (residues 25–702) in the presence of non-hydrolysable AMPPNP, whereby monomer and dimer peaks occurred in both SEC and sedimentation velocity AUC analyses in solution. Our data support the existence of a proportion of the UBL-containing SMCHD1 ATPase in a conformation that is equipped to undergo a monomer to dimer transition in the presence of ATP (or analogs), which could plausibly be reversed upon ATP hydrolysis. It remains of outstanding interest whether, in the context of full-length SMCHD1, which harbors a C-terminal SMC hinge domain that confers constitutive dimerization upon SMCHD1, whether the ATPase domain would be poised for stoichiometric dimerization owing to the inherent high local concentration of protomers.

GHKL-type proteins are a family of weak ATPases, where the energy released upon ATP hydrolysis is thought to drive complex conformational rearrangements within the protein itself rather than being used as a motor function [43,44]. Here, we were able to provide a direct comparison of the catalytic activity exhibited by SMCHD1's ATPase and representative human GHKL family members – Hsp90, MLH1 and MORC2 – revealing that they all display slow turnover rates, with k_{cat} values ranging between 0.015 and 0.047 μM ADP/min/μM protein. Previously reported turnover rates for full-length Hsp90 were in the range of 0.1–1.2 ADP/min/μM protein [45,46], 0.023 ADP/min/μM protein for MLH1 (residues 1–343) [47], while for MORC2 (residues 1–603) the fitted k_{cat} was reported as 0.10 ADP/min/μM protein [25]. While some of the values we obtained are ~2-fold lower than previously described turnover rates for the respective proteins, it is important to note that we employed an end point assay where we monitored ADP production during a linear relationship to ATP usage, as opposed to an NADH-coupled system [9], for example, which quantitates ATP hydrolysis in a continuous mode. We also performed the ATPase assays at 20°C rather than the 37°C often used for enzymatic assays. Together, these two aspects may contribute to an underestimation of turnover rates in our experimental setup, but overall show excellent agreement with published values. Most importantly, these data validate SMCHD1 ATPase activity as comparable to others in the GHKL family. It is interesting that ATPase assays of SMCHD1 at concentrations that favor dimerization (in the presence of ATP) led to slightly suppressed, rather than the anticipated elevated, ATPase activity. These data suggest that dimerization might suppress processivity and/or thwart ATP or ADP dissociation from SMCHD1 dimers, although the precise mechanism remains of ongoing interest.

PORTLAND PRESS

The identification of an N-terminal UBL domain in SMCHD1 raised great interest in its functional role. Most commonly, UBL domains have been associated with the recruitment of proteins to the 26S proteasome, stimulating proteasomal degradation in a similar fashion to ubiquitin [48,49]. In the case of SMCHD1, we have shown that its UBL domain is necessary for N-terminal dimerization and localization of the full-length protein to target sites on chromatin. This mechanism differs from canonical SMC proteins, where ring closure at the N-terminus of SMC heterodimers is dependent on interactions with non-SMC subunits rather than via an ATP-dependent domain-swapping event [18,50]. Whether the UBL domain has a direct role in SMCHD1's recruitment to chromatin, or rather, that the ability to dimerize at the N-terminus is sufficient for SMCHD1's localization to target sites, remains to be investigated. Nonetheless, the aberrant nuclear localization we observed for the E147A SMCHD1 variant, which forms a stable dimer that is unable to hydrolyze ATP, indicates that SMCHD1 relies on the UBL-mediated monomer-dimer conformational cycling at the N-terminus for faithful chromatin association.

Our studies validate the idea that SMCHD1's ATPase activity is connected to its dimerization and localization to chromatin. The precise nature and dynamics of ATP hydrolysis serving as an 'off' switch in SMCHD1 conformational cycling await further structural analyses for resolution. In particular, it remains of enormous interest to understand how the two protomers assemble into dimers, the dispositions of the component domains, and organization of the uncharacterized intermediate domain that connects the ATPase and SMC hinge domains in the context of the full-length, >2000 amino acid protein. Despite SMCHD1's essential role in epigenetic regulation, the atomic structure of the full-length protein and the molecular mechanisms underlying its function in both a healthy and diseased state remain to be elucidated. We anticipate that a detailed understanding of full-length SMCHD1 structure will enable interpretation of the functional effects of various substitutions in *SMCHD1* found in patients suffering of FSHD and BAMS, and provide a foundation for development of therapeutic interventions.

Data availability
All data and reagents are available from the authors upon request.

Competing Interests
The authors declare that there are no competing interests associated with the manuscript.

Funding

Open Access Statement
Open access for this article was enabled by the participation of the University of Melbourne in an all-inclusive *Read & Publish* pilot with Portland Press and the Biochemical Society under a transformative agreement with CAUL.

CRediT Author Contribution
Alexandra D. Gurzau: Conceptualization, formal analysis, investigation, methodology, writing — original draft. **Christopher Horne**: Investigation, writing — review and editing. **Yee-Foong Mok**: Investigation, writing — review and editing. **Megan Iminitoff**: Investigation, writing — review and editing. **Tracy A. Willson**: Investigation, writing — review and editing. **Samuel N. Young**: Investigation, writing — review and editing. **Marnie E. Blewitt**: Conceptualization, supervision, funding acquisition, writing — review and editing. **James M. Murphy**: Conceptualization, supervision, funding acquisition, writing — review and editing.

Acknowledgements

Abbreviations

AUC, analytical ultracentrifugation; BAMS, Bosma arhinia microphthalmia syndrome; FSHD, facioscapulohumeral muscular dystrophy; gRNAs, guide RNAs; MORC, microrchidia; SAXS, small-angle X-ray scattering; SEC, size exclusion chromatography; SMCHD1, structural maintenance of chromosomes flexible hinge domain-containing 1; UBL, ubiquitin-like.

References

1 Blewitt, M.E., Gendrel, A.V., Pang, Z., Sparrow, D.B., Whitelaw, N., Craig, J.M. et al. (2008) SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. *Nat. Genet.* **40**, 663–669 https://doi.org/10.1038/ng.142

2 Mould, A.W., Pang, Z., Pakusch, M., Tonks, I.D., Stark, M., Carrie, D. et al. (2013) Smchd1 regulates a subset of autosomal genes subject to monoallelic expression in addition to being critical for X inactivation. *Epigenetics Chromatin* **6**, 19 https://doi.org/10.1186/1756-8935-6-19

3 Gendrel, A.V., Tang, Y.A., Suzuki, M., Godwin, J., Nesterova, T.B., Greally, J.M. et al. (2013) Epigenetic functions of smchd1 repress gene clusters on the inactive X chromosome and on autosomes. *Mol. Cell. Biol.* **33**, 3150–3165 https://doi.org/10.1128/MCB.00145-13

4 Chen, K., Hu, J., Moore, D.L., Liu, R., Kessans, S.A., Breslin, K. et al. (2015) Genome-wide binding and mechanistic analyses of Smchd1-mediated epigenetic regulation. *Proc. Natl Acad. Sci. U.S.A.* **112**, E3535–E3544 https://doi.org/10.1073/pnas.1504232112

5 Lemmers, R.J., Tawil, R., Petek, L.M., Balog, J., Block, G.J., Santen, G.W. et al. (2012) Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. *Nat. Genet.* **44**, 1370–1374 https://doi.org/10.1038/ng.2454

6 Jansz, N., Keniry, A., Trussart, M., Bildsoe, H., Beck, T., Tonks, I.D. et al. (2018) Smchd1 regulates long-range chromatin interactions on the inactive X chromosome and at Hox clusters. *Nat. Struct. Mol. Biol.* **25**, 766–777 https://doi.org/10.1038/s41594-018-0111-z

7 Wanigasuriya, I., Gouil, Q., Kinkel, S.A., Del Fierro, T., Beck, A., Roper, T. et al. (2020) Smchd1 is a maternal effect gene required for genomic imprinting. *eLife* **9**, e55529 https://doi.org/10.7554/eLife.55529

8 Jansz, N., Chen, K., Murphy, J.M. and Blewitt, M.E. (2017) The epigenetic regulator SMCHD1 in development and disease. *Trends Genet.* **33**, 233–243 https://doi.org/10.1016/j.tig.2017.01.007

9 Chen, K., Dobson, R.C., Lucet, I.S., Young, S.N., Pearce, F.G., Blewitt, M.E. et al. (2016) The epigenetic regulator Smchd1 contains a functional GHKL-type ATPase domain. *Biochem. J.* **473**, 1733–1744 https://doi.org/10.1042/BCJ20160189

10 Pedersen, L.C., Inoue, K., Kim, S., Perera, L. and Shaw, N.D. (2019) A ubiquitin-like domain is required for stabilizing the N-terminal ATPase module of human SMCHD1. *Commun. Biol.* **2**, 255 https://doi.org/10.1038/s42003-019-0499-y

11 Brideau, N.J., Coker, H., Gendrel, A.V., Siebert, C.A., Bezstarosti, K., Demmers, J. et al. (2015) Independent mechanisms target SMCHD1 to trimethylated histone H3 lysine 9-modified chromatin and the inactive X chromosome. *Mol. Cell. Biol.* **35**, 4053–4068 https://doi.org/10.1128/MCB.00432-15

12 Chen, K., Czabotar, P.E., Blewitt, M.E. and Murphy, J.M. (2016) The hinge domain of the epigenetic repressor Smchd1 adopts an unconventional homodimeric configuration. *Biochem. J.* **473**, 733–742 https://doi.org/10.1042/BJ20151049

13 Chen, K., Birkinshaw, R.W., Gurzau, A.D., Wanigasuriya, I., Wang, R., Iminitoff, M. et al. (2020) Crystal structure of the hinge domain of Smchd1 reveals its dimerization mode and nucleic acid-binding residues. *Sci. Signal.* **13**, eaaz5599 https://doi.org/10.1126/scisignal.aaz5599

14 Gurzau, A.D., Blewitt, M.E., Czabotar, P.E., Murphy, J.M. and Birkinshaw, R.W. (2020) Relating SMCHD1 structure to its function in epigenetic silencing. *Biochem. Soc. Trans.* **48**, 1751–1763 https://doi.org/10.1042/BST20200242

15 Hirano, M. and Hirano, T. (2002) Hinge-mediated dimerization of SMC protein is essential for its dynamic interaction with DNA. *EMBO J.* **21**, 5733–5744 https://doi.org/10.1093/emboj/cdf575

16 Haering, C.H., Lowe, J., Hochwagen, A. and Nasmyth, K. (2002) Molecular architecture of SMC proteins and the yeast cohesin complex. *Mol. Cell* **9**, 773–788 https://doi.org/10.1016/S1097-2765(02)00515-4

17 Uchiyama, S., Kawahara, K., Hosokawa, Y., Fukakusa, S., Oki, H., Nakamura, S. et al. (2015) Structural basis for dimer formation of human condensin structural maintenance of chromosome proteins and its implications for single-stranded DNA recognition. *J. Biol. Chem.* **290**, 29461–29477 https://doi.org/10.1074/jbc.M115.670794

18 Gligoris, T.G., Scheinost, J.C., Burmann, F., Petela, N., Chan, K.L., Uluocak, P. et al. (2014) Closing the cohesin ring: structure and function of its Smc3-kleisin interface. *Science* **346**, 963–967 https://doi.org/10.1126/science.1256917

19 Haering, C.H., Farcas, A.M., Arumugam, P., Metson, J. and Nasmyth, K. (2008) The cohesin ring concatenates sister DNA molecules. *Nature* **454**, 297–301 https://doi.org/10.1038/nature07098

20 Dutta, R. and Inouye, M. (2000) GHKL, an emergent ATPase/kinase superfamily. *Trends Biochem. Sci.* **25**, 24–28 https://doi.org/10.1016/S0968-0004(99)01503-0

21 Li, S., Yen, L., Pastor, W.A., Johnston, J.B., Du, J., Shew, C.J. et al. (2016) Mouse MORC3 is a GHKL ATPase that localizes to H3K4me3 marked chromatin. *Proc. Natl Acad. Sci. U.S.A.* **113**, E5108–E5116 https://doi.org/10.1073/pnas.1609709113

22 Gordon, C.T., Xue, S., Yigit, G., Filali, H., Chen, K., Rosin, N. et al. (2017) De novo mutations in SMCHD1 cause Bosma arhinia microphthalmia syndrome and abrogate nasal development. *Nat. Genet.* **9**, 249–255 https://doi.org/10.1038/ng.3765

23 Shaw, N. D., Brand, H., Kupchinsky, Z. A., Bengani, H., Plummer, L., Jones, T. I. et al. (2017) SMCHD1 mutations associated with a rare muscular dystrophy can also cause isolated arhinia and Bosma arhinia microphthalmia syndrome. *Nat. Genet.* **49**, 238–248 https://doi.org/10.1038/ng.3743

24 Gurzau, A.D., Chen, K., Xue, S., Dai, W., Lucet, I.S., Ly, T.T.N. et al. (2018) FSHD2- and BAMS-associated mutations confer opposing effects on SMCHD1 function. *J. Biol. Chem.* **293**, 9841–9853 https://doi.org/10.1074/jbc.RA118.003104

25 Douse, C.H., Bloor, S., Liu, Y., Shamin, M., Tchasovnikarova, I.A., Timms, R.T. et al. (2018) Neuropathic MORC2 mutations perturb GHKL ATPase dimerization dynamics and epigenetic silencing by multiple structural mechanisms. *Nat. Commun.* **9**, 651 https://doi.org/10.1038/s41467-018-03045-x

26 Babon, J.J. and Murphy, J.M. (2013) In vitro JAK kinase activity and inhibition assays. *Methods Mol. Biol.* **967**, 39–55 https://doi.org/10.1007/978-1-62703-242-1_3

27 Verba, K.A., Wang, R.Y., Arakawa, A., Liu, Y., Shirouzu, M., Yokoyama, S. et al. (2016) Atomic structure of Hsp90-Cdc37-Cdk4 reveals that Hsp90 traps and stabilizes an unfolded kinase. *Science* **352**, 1542–1547 https://doi.org/10.1126/science.aaf5023

28   Schuck, P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.* **78**, 1606–1619 https://doi.org/10.1016/S0006-3495(00)76713-0

29   Demeler, B. (2005) UltraScan - a comprehensive data analysis software package for analytical ultracentrifugation experiments. In *Analytical Ultracentrifugation: Techniques and Methods* (Scott, D. J., Harding, S. E. and Rowe, A. J., eds). pp. 210–230, The Royal Society of Chemistry, Cambridge, UK

30   Brautigam, C.A. (2015) Calculations and publication-quality illustrations for analytical ultracentrifugation data. *Methods Enzymol.* **562**, 109–133 https://doi.org/10.1016/bs.mie.2015.05.001

31   Demeler, B. and van Holde, K.E. (2004) Sedimentation velocity analysis of highly heterogeneous systems. *Anal. Biochem.* **335**, 279–288 https://doi.org/10.1016/j.ab.2004.08.039

32   Ryan, T.M., Trewhella, J., Murphy, J.M., Keown, J.R., Casey, L., Pearce, F.G. et al. (2018) An optimized SEC-SAXS system enabling high X-ray dose for rapid SAXS assessment with correlated UV measurements for biomolecular structure analysis. *J. Appl. Crystallogr.* **51**, 97–111 https://doi.org/10.1107/S1600576717017101

33   Petoukhov, M.V., Franke, D., Shkumatov, A.V., Tria, G., Kikhney, A.G., Gajda, M. et al. (2012) New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **45**, 342–350 https://doi.org/10.1107/S0021889812007662

34   Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J. and Svergun, D.I. (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277–1282 https://doi.org/10.1107/S0021889803012779

35   Svergun, D.I. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495–503 https://doi.org/10.1107/S0021889892001663

36   Kueh, A.J. and Herold, M.J. (2016) Using CRISPR/Cas9 technology for manipulating cell death regulators. *Methods Mol. Biol.* **1419**, 253–264 https://doi.org/10.1007/978-1-4939-3581-9_18

37   Chaumeil, J., Augui, S., Chow, J.C. and Heard, E. (2008) Combined immunofluorescence, RNA fluorescent in situ hybridization, and DNA fluorescent in situ hybridization to study chromatin changes, transcriptional activity, nuclear organization, and X-chromosome inactivation. *Methods Mol. Biol.* **463**, 297–308 https://doi.org/10.1007/978-1-59745-406-3_18

38   Kirby, N., Cowieson, N., Hawley, A.M., Mudie, S.T., McGillivray, D.J., Kusel, M. et al. (2016) Improved radiation dose efficiency in solution SAXS using a sheath flow sample environment. *Acta Crystallogr. D Struct. Biol.* **72**, 1254–1266 https://doi.org/10.1107/S2059798316017174

39   Piiadov, V., Ares de Araujo, E., Oliveira Neto, M., Craievich, A.F. and Polikarpov, I. (2019) SAXSMow 2.0: online calculator of the molecular weight of proteins in dilute solution from experimental SAXS data measured on a relative scale. *Protein Sci.* **28**, 454–463 https://doi.org/10.1002/pro.3528

40   Ali, M.M., Roe, S.M., Vaughan, C.K., Meyer, P., Panaretou, B., Piper, P.W. et al. (2006) Crystal structure of an Hsp90-nucleotide-p23/Sba1 closed chaperone complex. *Nature* **440**, 1013–1017 https://doi.org/10.1038/nature04716

41   Shiau, A.K., Harris, S.F., Southworth, D.R. and Agard, D.A. (2006) Structural analysis of *E. coli* hsp90 reveals dramatic nucleotide-dependent conformational rearrangements. *Cell* **127**, 329–340 https://doi.org/10.1016/j.cell.2006.09.027

42   Richter, K., Reinstein, J. and Buchner, J. (2002) N-terminal residues regulate the catalytic efficiency of the Hsp90 ATPase cycle. *J. Biol. Chem.* **277**, 44905–44910 https://doi.org/10.1074/jbc.M208457200

43   Corbett, K.D. and Berger, J.M. (2005) Structural dissection of ATP turnover in the prototypical GHL ATPase TopoVI. *Structure* **13**, 873–882 https://doi.org/10.1016/j.str.2005.03.013

44   Wei, H., Ruthenburg, A.J., Bechis, S.K. and Verdine, G.L. (2005) Nucleotide-dependent domain movement in the ATPase domain of a human type IIA DNA topoisomerase. *J. Biol. Chem.* **280**, 37041–37047 https://doi.org/10.1074/jbc.M506520200

45   Panaretou, B., Prodromou, C., Roe, S.M., O'Brien, R., Ladbury, J.E., Piper, P.W. et al. (1998) ATP binding and hydrolysis are essential to the function of the Hsp90 molecular chaperone in vivo. *EMBO J.* **17**, 4829–4836 https://doi.org/10.1093/emboj/17.16.4829

46   Cunningham, C.N., Southworth, D.R., Krukenberg, K.A. and Agard, D.A. (2012) The conserved arginine 380 of Hsp90 is not a catalytic residue, but stabilizes the closed conformation required for ATP hydrolysis. *Protein Sci.* **21**, 1162–1171 https://doi.org/10.1002/pro.2103

47   Hall, M.C., Shcherbakova, P.V. and Kunkel, T.A. (2002) Differential ATP binding and intrinsic ATP hydrolysis by amino-terminal domains of the yeast Mlh1 and Pms1 proteins. *J. Biol. Chem.* **277**, 3673–3679 https://doi.org/10.1074/jbc.M106120200

48   Collins, G.A. and Goldberg, A.L. (2020) Proteins containing ubiquitin-like (Ubl) domains not only bind to 26S proteasomes but also induce their activation. *Proc. Natl Acad. Sci. U.S.A.* **117**, 4664–4674 https://doi.org/10.1073/pnas.1915534117

49   Kim, H.T. and Goldberg, A.L. (2018) UBL domain of Usp14 and other proteins stimulates proteasome activities and protein degradation in cells. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E11642–E11650 https://doi.org/10.1073/pnas.1808731115

50   Burmann, F., Shin, H.C., Basquin, J., Soh, Y.M., Gimenez-Oya, V., Kim, Y.G. et al. (2013) An asymmetric SMC-kleisin bridge in prokaryotic condensin. *Nat. Struct. Mol. Biol.* **20**, 371–379 https://doi.org/10.1038/nsmb.2488
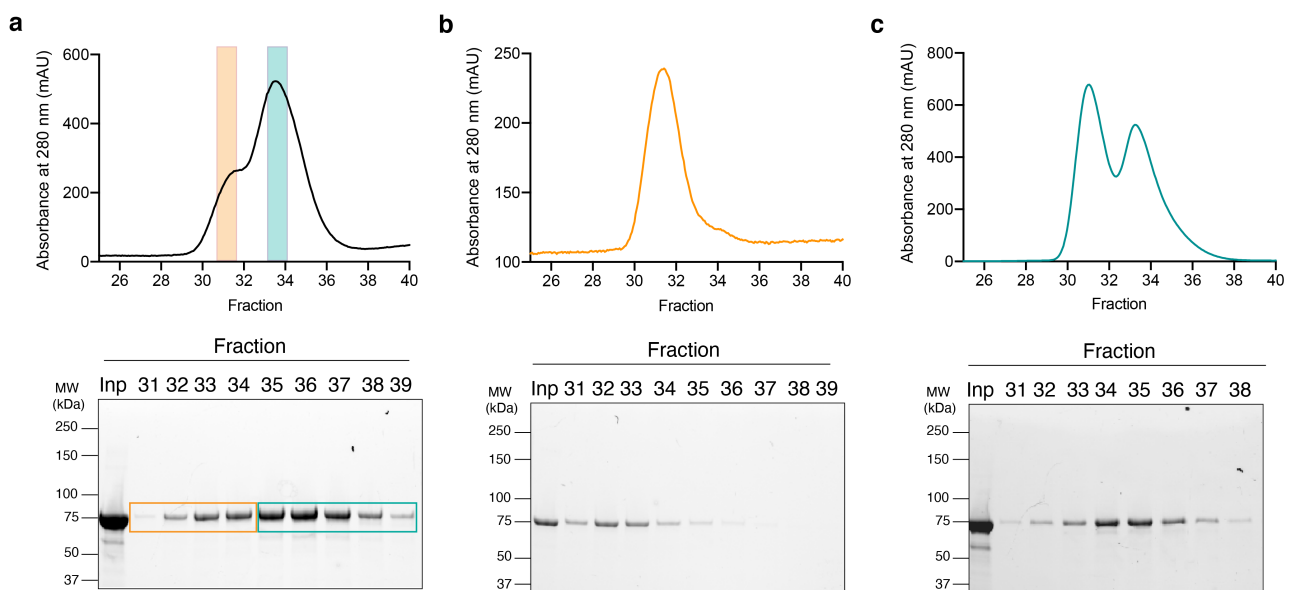
SUPPLEMENTARY FIGURES

Supplementary Figure 1. Supporting sedimentation velocity analytical ultracentrifugation (AUC) data. a-b. Raw sedimentation profiles of absorbance at 290 nm versus cell radius for initial sedimentation velocity analyses performed at 1.0 mg/mL protein, for the outlined samples described above each panel (**a**); and for additional sedimentation velocity analyses examining UBL-containing SMCHD1 ATPase (residues 25-702) in the presence of AMPPNP/Mg^{2+} at varying protein concentrations, as exemplified with a representative profile for the 0.5 mg/mL protein concentration sample (**b**). For both **a** and **b**, the bottom panels represent the residuals for the continuous size, c(s), distribution best fits plotted as a function of radial position (cm) from the axis of rotation. Analyses were performed using the program SEDFIT [1]. **c** van-Holde Weischet analysis [2] of UBL-containing SMCHD1 ATPase (residues 25-702) in the presence of AMPPNP/Mg^{2+} at varying protein concentrations.
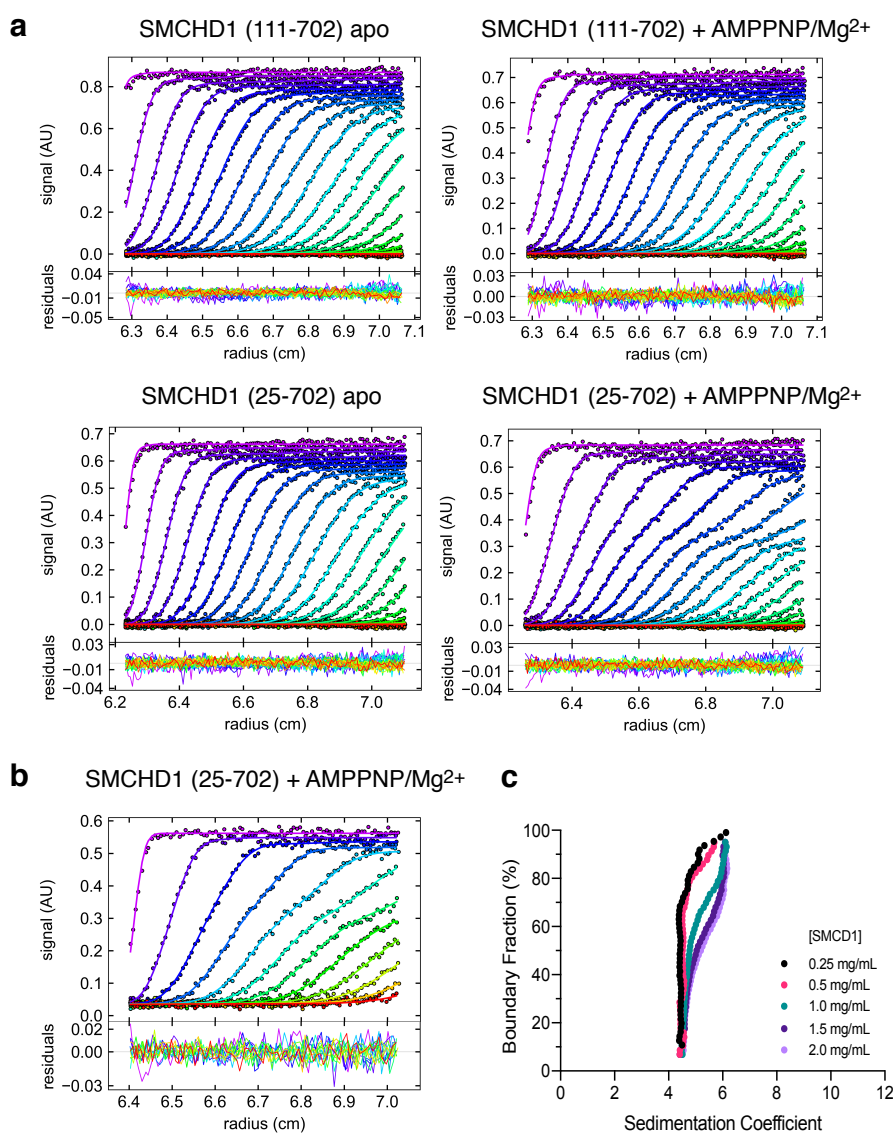
Supplementary Figure 2. Analytical size exclusion chromatography (SEC) of SMCHD1 (residues 25-702) reveals stable dimer formation in solution in the presence of AMPPNP/Mg^{2+}. a-c (Top panels) SEC profiles obtained using a Superdex-200 Increase 10/300 GL column for **a** SMCHD1 (residues 25-702) upon 30 min incubation on ice with 5 mM AMPPNP/Mg^{2+} showing **b** the re-injection of the pooled dimer peak from **a**, and **c** the re-injection of the pooled monomer peak from **a** after incubation with 5 mM AMPPNP/Mg^{2+}. The SMCHD1 (residues 25-702) dimer peak is highlighted in orange, and the monomeric peak is depicted in green. **a-c (Bottom panels)** Reducing Stain-Free SDS-PAGE analyses of selected fractions following SEC analyses of the corresponding samples above, and the injected sample (Inp) showing molecular weight marker position on the left, in kilodaltons (kDa).

Supplementary Figure 3. SAXS analyses of the ΔUBL SMCHD1 ATPase (residues 111-702) reveals a stable monomer in solution in the presence or absence of ligand, AMPPNP/Mg^{2+}.

a-b SAXS scattering profiles, with the log intensity of scattered X-rays $I(q)$, as a function of momentum transfer, q, in Å^{-1} for the ΔUBL SMCHD1 ATPase (residues 111-702) under apo conditions (**a**) and in the presence of AMPPNP/Mg^{2+} (**b**). The insets show Guinier plots, where linearity indicates the presence of monodisperse particles. **c-d** The corresponding real space pairwise-distribution functions, $P(r)$, for apo ΔUBL SMCHD1 ATPase (residues 111-702) (**c**) and in the presence of AMPPNP/Mg^{2+} (**d**), obtained via Fourier transformations of the scattering intensity using the software GNOM. The D_{MAX} (maximum dimension) values for each occur where the curve bisects the x-axis. (**e-f**) Three-dimensional models of the monomeric (**e**) and dimeric (**f**) SMCHD1 ATPase (residues 25-702) were obtained by merging a previously developed model of the ΔUBL SMCHD1 ATPase (residues 111-702) based on the high-resolution structure of the GHKL protein Hsp90 (PDB:1CG9) [3], and the SMCHD1 ATPase (residues 25-580) crystal structure (PDB:6MW7) [4]. Structure cartoons were generated using PyMOL. (**g-h**) SAXS scattering profiles for UBL-containing SMCHD1 ATPase (residues 25-702) in the presence of AMPPNP/Mg^{2+} (**g**) or in the absence of AMPPNP/Mg^{2+} (**h**), displaying the log intensity of scattered X-rays, $I(q)$, as a function of momentum transfer, q, in Å^{-1}. Experimental scatter is shown in black, whereas the theoretical scatter, which was obtained using CRYSOL from the models in **e** and **f** is shown in orange, with the quality of fit estimated by a chi-squared (χ^2) value.
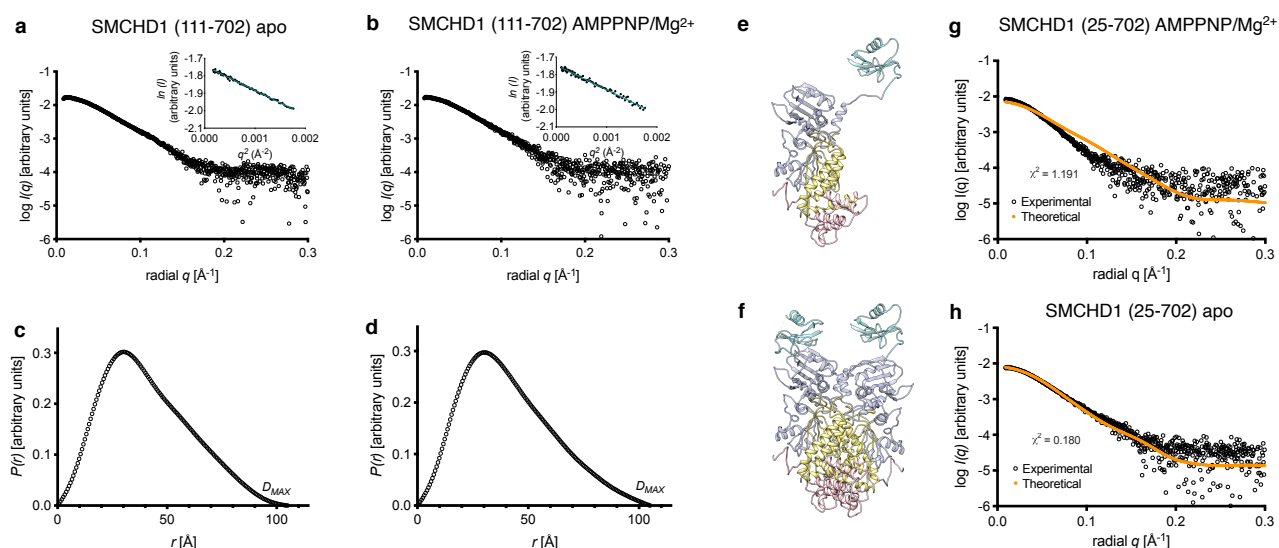
Supplementary Table 1. Data collection and scattering parameters for SAXS analyses

Data collection parameters

Instrument	Australian Synchrotron SAXS/WAXS beamline
Beam geometry	120 µm point source
Beam wavelength (Å)	1.033
q range (Å^{-1})	0.0114-0.4
Exposure time (seconds)	1
Protein concentration	~5 mg/ml sample injected via in-line size exclusion chromatography
Temperature (°C)	8

Structural parameters

Protein sample	SMCHD1 (111-702 apo)	SMCHD1 (111-702 AMPPNP/Mg^{2+})	SMCHD1 (25-702 apo)	SMCHD1 (25-702 AMPPNP/Mg^{2+})
$I(0)$(cm^{-1}) [from Guinier]	0.01806 ±0.00013	0.01791 ±0.00022	0.08244 ±0.00082	0.08785 ±0.00132
R_g (Å) [from Guinier]	31.2 ±0.344	31.1 ±0.630	34.6 ±0.546	37.6 ±0.878
$I(0)$(cm^{-1}) [from $P(r)$]	0.01825 ±0.00009	0.01812 ±0.00016	0.00837 ±0.00005	0.00894 ±0.00011
R_g (Å) [from $P(r)$]	32.17 ±0.206	32.33 ±0.383	36.10 ±0.244	39.20 ±0.515
D_{max} (Å) [from $P(r)$]	105	105	115	125
R_g (Å) [from *CRYSOL*]	29.3	29.4	34.4	36.3
D_{max} (Å) [from *CRYSOL*]	103	103	117	128
MW (kDa) [from SAXSMoW]	68.4 (68 kDa expected)	70.7 (68 kDa expected)	95.7 (78 kDa expected)	139.7 (156 kDa expected)

Software

Primary data reduction	ScatterBrain (Australia Synchrotron)
Data processing	PRIMUS [5], GNOM [6], CRYSOL [7], SAXSMoW [8]

SUPPLEMENTARY REFERENCES

1   Schuck, P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. Biophys J. **78**, 1606-1619

2   Demeler, B. and van Holde, K. E. (2004) Sedimentation velocity analysis of highly heterogeneous systems. Anal Biochem. **335**, 279-288

3   Chen, K., Dobson, R. C., Lucet, I. S., Young, S. N., Pearce, F. G., Blewitt, M. E. and Murphy, J. M. (2016) The epigenetic regulator Smchd1 contains a functional GHKL-type ATPase domain. Biochem J. **473**, 1733-1744

4   Pedersen, L. C., Inoue, K., Kim, S., Perera, L. and Shaw, N. D. (2019) A ubiquitin-like domain is required for stabilizing the N-terminal ATPase module of human SMCHD1. Commun Biol. **2**, 255

5   Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. and Svergun, D. I. (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. Journal of Applied Crystallography. **36**, 1277-1282

6   Svergun, D. I. (1992) Determination of the Regularization Parameter in Indirect-Transform Methods Using Perceptual Criteria. Journal of Applied Crystallography. **25**, 495-503

7   Svergun, D. I., Barberato, C. and Koch, M. H. J. (1995) CRYSOL - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. J Appl Cryst. **28**, 768-773

8   Piiadov, V., Ares de Araujo, E., Oliveira Neto, M., Craievich, A. F. and Polikarpov, I. (2019) SAXSMoW 2.0: Online calculator of the molecular weight of proteins in dilute solution from experimental SAXS data measured on a relative scale. Protein Sci. **28**, 454-463