PORTLAND PRESS

BIOCHEMICAL JOURNAL
ACCEPTED MANUSCRIPT

The epigenetic regulator Smchd1 contains a functional GHKL-type ATPase domain

Kelan Chen, Renwick C.J. Dobson, Isabelle S. Lucet, Samuel N. Young , F. Grant Pearce , Marnie E. Blewitt and James M. Murphy

Structural maintenance of chromosomes flexible hinge domain containing 1 (Smchd1) is an epigenetic regulator that plays critical roles in gene regulation during development. Mutations in SMCHD1 were recently implicated in the pathogenesis of facioscapulohumeral muscular dystrophy (FSHD), although the mechanistic basis remains of outstanding interest. We have previously shown that Smchd1 associates with chromatin via its homodimeric C-terminal hinge domain, yet little is known about the function of the putative GHKL (gyrase, Hsp90, histidine kinase, MutL)-type ATPase domain at its N-terminus. To formally assess the structure and function of Smchd1's ATPase domain, we have generated recombinant proteins encompassing the predicted ATPase domain and the adjacent region. Here, we show that the Smchd1 Nterminal region exists as a monomer and adopts a conformation resembling that of monomeric full-length Hsp90 protein in solution, even though the two proteins share only ~8% overall sequence identity. Despite being monomeric, the N-terminal region of Smchd1 exhibits ATPase activity, which can be antagonised by the reaction product, ADP, or the Hsp90 inhibitor, radicicol, at a nanomolar concentration. Interestingly, introduction of an analogous mutation to that identified in SMCHD1 of an FSHD patient compromised protein stability, suggesting a possible molecular basis for loss of protein function and pathogenesis. Together, these results reveal important structure-function characteristics of Smchd1 that may underpin its mechanistic action at the chromatin level.

Cite as Biochemical Journal (2016) DOI: 10.1042/BCJ20160189

The epigenetic regulator Smchd1 contains a functional GHKL-type ATPase domain

Kelan Chen[*,†], Renwick C.J. Dobson[‡,§], Isabelle S. Lucet[*,†], Samuel N. Young[*], F. Grant Pearce[‡], Marnie E. Blewitt[*,†,1] and James M. Murphy[*,†,1,2]

[*]The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria 3052, Australia
[†]Department of Medical Biology, University of Melbourne, Parkville, Victoria 3050, Australia
[‡]Biomolecular Interactions Centre and School of Biological Sciences, University of Canterbury, Christchurch 8042, New Zealand
[§]Bio21 Institute, Department of Biochemistry and Molecular Biology, University of Melbourne, Parkville, Victoria 3010, Australia

[1] These authors share senior authorship
[2] To whom correspondence should be addressed: jamesm@wehi.edu.au

Short title Smchd1 contains a GHKL ATPase domain
Keywords GHKL ATPase, Smchd1, Hsp90

Summary statement There is a growing interest in the mechanistic and structural-functional properties of the epigenetic regulator Smchd1. Here we demonstrate that Smchd1's N-terminal region shares a similar overall conformation and ATP-binding site with Hsp90 and possesses discernable ATPase activity.

Abbreviations SMC, structural maintenance of chromosomes; Smchd1, SMC flexible hinge domain containing 1; GHKL, gyrase, Hsp90, histidine kinase, MutL; FSHD, facioscapulohumeral muscular dystrophy; Hsp90, heat shock protein 90; IEX, ion-exchange; SEC, size exclusion chromatography; SAXS, small angle X-ray scattering; PK-LDH, pyruvate kinase-lactate dehydrogenase; R_g, radius of gyration; *P(r)*, interatomic distance distribution.

Abstract

Structural maintenance of chromosomes flexible hinge domain containing 1 (Smchd1) is an epigenetic regulator that plays critical roles in gene regulation during development. Mutations in SMCHD1 were recently implicated in the pathogenesis of facioscapulohumeral muscular dystrophy (FSHD), although the mechanistic basis remains of outstanding interest. We have previously shown that Smchd1 associates with chromatin via its homodimeric C-terminal hinge domain, yet little is known about the function of the putative GHKL (gyrase, Hsp90, histidine kinase, MutL)-type ATPase domain at its N-terminus. To formally assess the structure and function of Smchd1's ATPase domain, we have generated recombinant proteins encompassing the predicted ATPase domain and the adjacent region. Here, we show that the Smchd1 N-terminal region exists as a monomer and adopts a conformation resembling that of monomeric full-length Hsp90 protein in solution, even though the two proteins share only ~8% overall sequence identity. Despite being monomeric, the N-terminal region of Smchd1 exhibits ATPase activity, which can be antagonised by the reaction product, ADP, or the Hsp90 inhibitor, radicicol, at a nanomolar concentration. Interestingly, introduction of an analogous mutation to that identified in SMCHD1 of an FSHD patient compromised protein stability, suggesting a possible molecular basis for loss of protein function and pathogenesis. Together, these results reveal important structure-function characteristics of Smchd1 that may underpin its mechanistic action at the chromatin level.

Introduction

Structural maintenance of chromosomes flexible hinge domain containing 1 (Smchd1) is a large nuclear protein that belongs to the SMC protein family [1, 2]. Originally identified as an epigenetic regulator with critical roles in X-inactivation [1, 2], Smchd1 has subsequently been shown to regulate expression of a number of autosomal genes including ones from the imprinted clusters and protocadherin clusters [3-7]. More recently, much interest has been generated in studying Smchd1 as its highly conserved human homolog, SMCHD1, is implicated in the pathogenesis of facioscapulohumeral muscular dystrophy (FSHD) [8-10].

Smchd1 contains two recognizable domains: a putative N-terminal ATPase domain and a C-terminal SMC hinge domain, with ~1,500 residues connecting the two entities (Figure 1). We have recently shown that the hinge domain mediates the dimerisation of two Smchd1 protomers and is critical for Smchd1's interaction with chromatin to elicit epigenetic control [7]. However, much less is known about the rest of the Smchd1 protein. The putative N-terminal ATPase domain of Smchd1 belongs to a functionally diverse protein family referred to as the GHKL (DNA gyrase, Heat shock protein 90, histidine kinase and DNA mismatch repair protein MutL) family [11]. Despite members of the GHKL protein family exhibiting divergent protein structures and biological functions, they all contain a similar ATP-binding pocket [12, 11]. Apart from the GHKL-ATPase domain in histidine kinases, which acts as a monomeric catalytic moiety[13], GHKL-ATPase domains tend to dimerise upon ATP-binding, as shown in Hsp90 [14, 15], DNA gyrase [16, 17] and MutL [18, 19]. Importantly, the ATPase activity of these proteins is coupled to their biological function, such as folding of client proteins by Hsp90 [15] and cutting and passaging of DNA in the case of DNA gyrase [17].

Several proteins that bear a GHKL-ATPase domain are emerging as new players in epigenetic regulation. DMS11/MORC6, itself a GHKL-ATPase and a partner of the plant orthologue of Smchd1, DMS3, was implicated in RNA-directed DNA methylation and heterochromatin condensation in *Arabidopsis* [20-22]. Another example is MORC1, which has been shown to participate in heterochromatin condensation in *Arabidopsis* [21] and establishment of DNA methylation at transposable elements in the male germline of mice [23]. These findings highlight a potential theme where the GHKL-ATPase domain is essential to bring about biological function of these proteins at the chromatin level.

The putative GHKL-ATPase present in Smchd1 contrasts the ABC-type ATPase domain constituted by the N- and C-terminal subunits of conventional SMC proteins (Figure 1A). For other canonical SMC proteins, ATP hydrolysis catalysed by the ABC-type ATPase 'head' domain is critical for regulating engagement and disassociation of two 'head' domains and energy-dependent conformational change of the SMC dimer, allowing entrapment of chromatin within the ring-shaped protein complex formed by SMC and non-SMC subunits [24-26]. The presence of the putative GHKL-ATPase domain in Smchd1 thus raises several interesting questions. One is whether Smchd1 might possess a distinct conformation unlike the conventional SMC proteins. On the other hand, when we initiated this study, it was not clear whether the putative GHKL-ATPase domain of Smchd1 possessed catalytic activity and whether such an activity would be of functional significance, like in the ABC-type ATPase in canonical SMC proteins. While recent studies have provided a glimpse into Smchd1's dimeric assembly and

propensity towards ATPase activity [27, 7, 28], detailed structure-function characterisation is still lacking.

To provide further insight into the structure and function of the putative ATPase domain of Smchd1, we generated recombinant protein encompassing the predicted GHKL-ATPase domain and its adjacent region, which, together, we term the N-terminal region. We provide evidence that the N-terminal region exists as a monomer in solution, both in the presence and absence of ATP, and is capable of hydrolysing ATP. We observed that Smchd1 ATPase activity is proportional to protein concentration, discounting the possibility that this activity arises from transient dimerisation, in contrast to Hsp90, which is known to function as a homodimer. However, as proposed for Hsp90, Smchd1 was subject to product (ADP) inhibition, which led to a 10-fold decrease in catalytic rate. Although there was only ~8% sequence identity between the N-terminal region of Smchd1 and full-length Hsp90, we demonstrated using small angle X-ray scattering (SAXS) that their overall conformations are analogous in solution. Additionally, we found that radicicol, one of three Hsp90 inhibitors tested in our study, is a potent antagonist of Smchd1 ATPase activity. These results reveal important structural features of the N-terminal region of Smchd1 and corroborate that Smchd1 is a functional ATPase. We have therefore provided a basis upon which to further mechanistically dissect the involvement of the Smchd1 ATPase domain in epigenetic control.

Experimental

Cloning, expression and purification of recombinant protein

cDNAs encoding the N-terminal region of mouse Smchd1 were PCR-amplified from a full-length cDNA clone and ligated into pFastBac Htb vector (Life Technologies). Site-directed mutagenesis was performed by overlap PCR, as before [7] and all insert sequences verified by Sanger sequencing (Micromon Facility, Monash University, VIC, Australia). Recombinant proteins were expressed and purified from *Sf*21 cells using the Bac-to-Bac system (Life Technologies) essentially as previously described with modifications [29]. Briefly, pFastBac Htb vectors were transformed into DH10MultiBac cells (ATG Biosynthetics) for generation of bacmids. *Sf*21 cells were maintained in Insect-XPRESS protein-free insect cell media with L-glutamine (Lonza). High-infectivity baculoviruses were produced by transfection of *Sf*21 cells with bacmids and subsequent amplification passages. *Sf*21 cells (500 ml) at a density of 3-4 x 10^6 were infected with high-infectivity baculovirus for protein expression.

Purification of His_6-tagged protein was performed with Ni-NTA resin (Qiagen) or cOmplete His-tag purification resin (Roche) as previously described [7]. The tag was cleaved by overnight incubation with tobacco etch virus (TEV) protease at 4 °C. Cleaved protein was concentrated with a 30 kD molecular mass cutoff concentrator (Millipore) at 4 °C by centrifugation at 3300 x *g* and then diluted with ion exchange (IEX) Buffer A (25 mM Tris pH 7.5, 0.5 mM TCEP) and loaded onto MonoQ 5/50 GL column (GE Healthcare) pre-equilibrated with IEX Buffer A. Proteins were eluted with a gradual 0-100% gradient of IEX Buffer B (500 mM NaCl, 25 mM Tris pH 7.5, 0.5 mM TCEP) over 20-column volumes. Fractions containing the N-terminal region of Smchd1 were pooled and subjected to size exclusion chromatography (SEC) via a Superdex-200 10/300 GL column (GE Healthcare) in SEC Buffer (100 mM NaCl and 20 mM HEPES pH 7.5). Fractions containing the N-terminal region of Smchd1 were pooled,

concentrated to 10 mg/ml, and then aliquoted and snap-frozen in a liquid nitrogen bath for storage at -80 °C. All purification steps were performed at 4 °C. Protein was analysed by reducing SDS-PAGE or by Western Blot using an anti-SMCHD1 antibody (A302-871A; Bethyl), using established procedures [7].

Small-angle X-ray scattering (SAXS) data collection and analysis
SAXS data were collected at the Australian Synchrotron SAXS/WAXS beamline [30] using an in-line gel filtration chromatography setup as previously described [31, 28]. Essentially, 50 µl of 5 mg/ml recombinant protein samples were loaded onto an in-line Superdex 200 5/150 column (GE Healthcare) pre-equilibrated with 200 mM NaCl, 20 mM HEPES pH 7.5, 5% glycerol and eluted via a 1.5 mm glass capillary at 16 °C positioned in the X-ray beam. Scattering data were collected with a Pilatus detector (1 M, 170 mm X 170 mm) at 2-second intervals over the course of the elution and processed by the beamline control software, ScatterBrain. 2D intensity plots from the peak of the size exclusion chromatography elution corresponding to the protein sample were radially averaged, normalised to sample transmission, with the scattering profiles collected on buffer earlier in the chromatographic elution averaged for background subtraction.

Data analyses were performed with the ATSAS suite [32] essentially as previously described [31, 28]. Guinier analysis was performed using PRIMUS [33] to examine the scattering curve at very small scattering angles ($qRg \leq 1.3$). This analysis allows estimation of two parameters: the radius of gyration value (*Rg*), which is the square root of the average distance of each scattering atom from the particle centre, and zero angle intensity *I*(0), which is proportional to the square of the molecular weight for a given protein concentration. The linearity of the Guinier plot reflects that neither high molecular weight aggregates nor inter-particle interference contributes measurably to scattering, and thus the data was of high quality. The real-space interatomic distance distribution function *P*(r) and the maximum dimension of the scattering particle, D_{max} was computed by indirect Fourier transform using GNOM [32]. *Rg* and *I*(0) were also estimated using the *P*(r) function with data collected at all the angles. Low-resolution shape envelopes were reconstructed using the *ab initio* bead-modeling program DAMMIF [32]. Ten independent models were built from scattering profile of each sample and all models were compared using program DAMSEL [32], with the most probable models aligned using program DAMSUP [32], from which an averaged model was obtained using program DAMAVER [32]. The averaged model was adjusted to correspond to the experimentally determined excluded volume by running the program DAMFILT [32]. The final bead model was superimposed with crystal structures using the program SUPCOMB [32]. The theoretical scattering curves of the crystal structure of Hsp90 were calculated from atomic coordinates for comparison with the experimental data by running program CRYSOL, where consistency between the theoretical and experimental curves are inferred from the χ value [32].

Analytical ultracentrifugation
Analytical ultracentrifugation experiments were conducted in a Beckman model XL-I instrument. The protein was diluted to 0.5 mg/ml (7 □M) in buffer (100 mM NaCl, 20 mM HEPES, pH 7.5, 20 mM $MgCl_2$, and 0.1 mM ATPγS (Sigma-Aldrich) where applicable), loaded into double sector quartz cells, and mounted in a Beckman 8-hole An-50 Ti rotor. Solvent density (1.00523 g/ml) and viscosity (0.01169 cp) were computed using the amino acid composition and the program SEDNTERP [34]. Partial specific volume of the protein was estimated to be 0.73257

ml/g, from the protein sequence using the program SEDNTERP [34]. For the sedimentation velocity experiments, 380 µl of sample and 400 µl of reference buffer were centrifuged at a rotor speed of 50,000 rpm. Data was collected at a single wavelength (285 nm) in continuous mode for 120 scans using a time interval of 0 s and a step-size of 0.003 cm without averaging. The experiments were conducted at 14 °C to counter any protein instability over the several hours it took to collect the data. Sedimentation velocity data at multiple time points were fitted to a continuous sedimentation-coefficient model and continuous mass model [35-37], using the program SEDFIT [36].

Thermal denaturation assay

The wild-type and E147A ATPase domains (111-702aa) were examined in a thermal denaturation assay, using the previously described method for the thermal shift assay [29, 7]. Briefly, proteins were diluted in 150 mM NaCl, 20 mM Tris (pH 8.0), and 1 mM DTT to a final concentration of 5 µM and assayed in a total reaction volume of 25 µL in the presence SYPRO Orange (Molecular Probes), whose fluorescence was detected at 530 nm. The temperature was raised by 1 °C per minute from 25 °C to 95 °C, with fluorescent readings taken at each interval.

ATPase assay

The competitive fluorescence polarisation ATPase assay based on the detection of ADP was performed with the Transcreener ADP2 fluorescence polarisation assay kit (BellBrook Labs) essentially as previously described [38] with some modifications. Briefly, 10 µl reactions in duplicates or triplicates were set up in 384-well (low volume, black) plates, containing 7 µl reaction buffer (50 mM HEPES pH 7.5, with or without 4 mM $MgCl_2$, and 2 mM EGTA), 1 µl of recombinant Smchd1 (111-702 aa) at concentrations ranging from 0.1-0.4 µM or SEC buffer control, 1 µl of inhibitor or solvent control and 1 µl of 10 µM ATP substrate or nuclease-free water control. Reactions were incubated at 25 °C for one or two hours in the dark before adding 10 µl of detection mix (1X Stop & Detection Buffer B, 4 nM ADP AlexaFluor 633 Tracer, 128 µg/ml ADP2 antibody) for a further hour incubation. Fluorescence polarisation readings (mP) were measured with an Envision plate reader fitted with excitation filter 620/40 nm, emission filters 688/45 nm (s and p channels) and D658/fp688 dual mirror (PerkinElmer Life Sciences). Readings from a no antibody (free tracer) control were set as 20 mP as the normalisation baseline of the assay for all reactions. The amount of ADP present in each reaction was estimated by using a 12-point standard curve following manufacturer's instructions. Hsp90 inhibitors 17-AAG and NVP-BEP800 (both from Selleckchem) were dissolved in DMSO and Radicicol (Sigma-Aldrich) was dissolved in 70% ethanol. We also examined ATPase activity using two further assays. ADP-Glo (Promega) luminescence assays were used to measure ADP production over 1-2 hours for 0.1-0.4 µM ATPase in a final volume of 25 µl, according to manufacturer's instructions. Additionally, we monitored ADP generation using a conventional pyruvate kinase-lactate dehydrogenase coupled assay [15], where ADP regeneration to ATP is coupled with NADH oxidation, which was monitored by absorbance at 340 nm. Assays were performed using 0.35-10.8 µM Smchd1 ATPase, in a 250 µL reaction volume at 25°C, with spinach Rubisco activase used as a positive control (k_{cat} 30-60 µM), as before [39]. Although reaction rates were constant for the course of the experiment until NADH was consumed, including for 0.35 µM Smchd1 ATPase for 100 minutes, we have presented rate data for the period between 100-200 sec for consistency.

Results

Expression and purification of the N-terminal region of Smchd1

Initially, we generated several constructs corresponding to the predicted ATPase domain of Smchd1 as defined by homology search, corresponding to amino acids (aa) 111-365. However, attempts at protein expression and purification with these constructs in both bacteria and insect cells were unsuccessful; the expressed protein eluted in the void-volume during size exclusion chromatography (SEC), indicative of protein aggregation (details not shown).

While the ATPase domain alone of several members of the GHKL protein family could fold into an independent entity and could be successfully purified, it was noticed that, in Hsp90, the ATPase domain and its adjacent region corresponding to the middle-domain are intimately connected by a linker region (Figure 1B) [40]. We therefore hypothesized that including regions downstream of residue 365 of Smchd1 might confer better protein behaviour and stability.

To define domain boundaries of such a longer construct, we performed sequence analysis to predict the secondary structure of Smchd1 and generated a multiple sequence alignment of mouse Smchd1 and its chicken and frog orthologues with other GHKL-ATPase sequences, including Hsp90 from three evolutionarily distinct species, *Arabidopsis* DMS11 and its potential orthologues (Figure 1B). Despite only limited sequence homology between these proteins, we envisaged that recombinant constructs of Smchd1 comparable to the full-length DMS11 and Hsp90, with C-terminal boundaries at approximately 702-789 aa, would be suitable candidates for recombinant protein production (Figure 1B). Therefore we generated several constructs with 111 aa as the N-terminus and C-terminal truncation at potential boundaries of secondary structure elements, corresponding to 702 aa, 713 aa, 725 aa, 737 aa, 756 aa, 771 aa and 789 aa (Figure 1B).

We expressed these constructs as N-terminally His_6-tagged recombinant proteins using the baculovirus insect cell system. Among tested constructs, we found that 111-702 aa, 111-713 aa, 111-725 aa and 111-737 aa could be readily obtained following Ni^{2+} and size-exclusion chromatography based purification procedure. As exemplified in Figure 2A, both 111-702 aa and 111-725 aa are eluted as single peaks from size-exclusion chromatography, with elution volumes consistent with molecular weights of approximately 72 and 81 kilodaltons (kD), respectively, and obtained at high purity (Figure 2B). Analytical ultracentrifugation studies of Smchd1 111-702 confirmed that the protein exists as a single species in solution (at 7 μM) in the absence or presence of ATPγS. The species sedimented at 4.58 S, with no high molecular weight aggregated species detected (Figure 2C). By fitting the data to a continuous mass-distribution model, we confirmed the single species has a mass of ~67.1-67.3 kD (Figure 2D), which is in agreement with the predicted molecular weight of the monomer (68.4 kD). The hydrodynamic properties of 111-702 aa of Smchd1 are summarised in the Supplementary Table.

Although recombinant proteins with C-terminal boundaries beyond 737 aa gave comparable yields of soluble protein, a higher proportion of protein eluted in the void volume during SEC under the same purification conditions, suggesting a predisposition toward aggregation (data not shown). Thus, the 111-702 aa and 111-725 aa constructs are likely well-behaved entities that

encapsulate the ATPase domain and its adjacent region, which we hereinafter refer to as the N-terminal region of Smchd1. In parallel, we produced a putative catalytically dead form of the 111-702 aa protein, by substituting alanine for the conserved glutamic acid residue within Motif I of the GHKL-ATP binding pocket critical for the enzymatic reaction (E147A mutant) [14, 15]. The mutant protein was expressed and purified in the same manner as for the wild-type protein, yielding a recombinant protein with a comparable size-exclusion chromatography profile and purity as the wild-type protein (Figure 2B). We performed thermal denaturation assays and confirmed that the E147A mutant exhibits a comparable thermal stability to that of the wild-type protein as indicated by their melting temperature of approximately 38 and 42°C, respectively (Figure 2E).

The overall conformation of the N-terminal region of Smchd1 is similar to that of full-length Hsp90

To gain insight into the structure of the N-terminal region of Smchd1 in solution, we performed small-angle X-ray scattering (SAXS) analysis with recombinant protein of 111-702 aa and 111-725 aa. The background-subtracted scattering profiles of both proteins are shown in Figure 3. The Guinier plots of the two proteins are both linear, indicating the samples are monodisperse, free of aggregation or inter-particle interference and the data collected are of high quality (Figure 3). We next calculated the radius of gyration values (R_g), indicating the overall size of the proteins, using both Guinier approximation and computation of real-space interatomic distance distribution function P(r). The calculated *Rg* values are approximately 32 and 35 Å for 111-702 aa and 111-725 aa, respectively (Figure 3, Table 1). The estimated maximum diameters (D_{max}) of 111-702 aa and 111-725 aa, are 105 and 115 Å, respectively, indicating that the latter is more elongated than 111-702 aa, consistent with the included C-terminal extension (Figure 3, Table 1).

To infer the overall structure of the N-terminal region, we reconstructed the three-dimensional bead models of the proteins from the distance distribution function P(r) using DAMMIF. A consensus low-resolution envelope structure was obtained by averaging and filtering ten independent *ab initio* bead models using DAMAVER and DAMFILT and is shown in Figure 4. We found that the resulting models for both 111-702 aa and 111-725 aa adopt an elongated structure that grossly resembles the crystal structure of full-length Hsp90 [41] (Figure 4A and B, Hsp90 superimposed). In light of this observation, we calculated the theoretical scattering curve based on the crystal structure of Hsp90 to fit with the experimental scattering data of 111-702 aa and 111-725 aa using the program CRYSOL. As shown in Figure 4C and D, the computed theoretical scattering curve of Hsp90 fit the experimental data collected on 111-702 aa and 111-725 aa with excellent agreement, with χ values of 0.66 and 0.56, respectively. These results suggest that the overall structure of the N-terminal region of Smchd1 is highly similar to that of Hsp90, albeit that the sequence similarity between the matched regions of the two proteins is very low with only 8% identity (Figure 1B).

The N-terminal region of Smchd1 is a catalytically active ATPase in its monomeric state

We next examined if the N-terminal region of Smchd1 is capable of hydrolysing ATP using Smchd1 (111-702 aa) protein preparations. To eliminate the possibility that trace contaminants in the protein preparation sample might account for any observed catalytic activity, we tested the

putative catalytically inactive form of the N-terminal region containing the E147A mutation, as described above, as a negative control. To measure the ATPase activity, we made use of a fluorescence polarisation based Transcreener assay that has been used successfully to determine the ATPase activity of Hsp90 and several other proteins [38]. We found that the 111-702 aa protein is capable of converting ATP to ADP in a time, concentration and Mg^{2+}-dependent manner, with a calculated k_{cat} of 0.039±0.006 min^{-1}, whereas there was negligible ATP hydrolysis by the E147A mutant (Figure 5A). We further verified our results using a luminescent ADP-Glo assay and obtained comparable results (Supplementary Figure 1). The ATPase activity of the wild-type protein measured at 25°C thus indicates that the N-terminal region of Smchd1 represents a stable, folded and catalytically active protein. Given that the E147A mutant exhibits a similar thermal stability (Figure 2E), we attributed the loss of activity of the E147A mutant to substitution of a key catalytic residue, rather than protein misfolding.

We noticed that the measured ATPase activity of N-terminal region of Smchd1 is approximately 10-fold lower than the reported ATPase activity of Hsp90, which is within the range of 0.1-1.2 min^{-1} [15, 42]. To formally compare the enzymatic activity of Smchd1 and Hsp90, we adopted a pyruvate kinase-lactate dehydrogenase (PK-LDH) coupled assay that has been extensively used to study the *in vitro* ATPase activity of Hsp90. Unexpectedly, we observed a higher turnover number of the N-terminal region of Smchd1, with a k_{cat} of approximately 0.6 min^{-1} that is independent of protein concentration within the range of 0.35-10.8 □M (Figure 5B). A key difference between the PK-LDH coupled assay and the fluorescence polarisation and luminescence based assays is that the ATPase reaction product, ADP, is regenerated to ATP in the coupled assay, but ADP persists in fluorescence polarisation and luminescence based assays. Consequently, we reasoned that Smchd1 ATPase domain is sensitive to inhibition by its product, ADP, a phenomenon that has been described for Hsp90 [14, 43]. As a result, ADP may attenuate Smchd1's ATPase activity measured by the endpoint mode assays as described above. In contrast, ADP is regenerated into ATP in the coupled assay, diminishing its capacity to inhibit Smchd1 and resulting an elevated turnover number. Notably, these data indicate that Smchd1 ATPase activity is proportional to protein concentration, consistent with the idea that it functions as a monomer. This conclusion is supported by our SAXS and AUC analyses, which indicate that Smchd1 ATPase domain is monomeric at concentrations up to 7 μM in solution, unlike Hsp90.

Given the structural similarities between the N-terminal region of Smchd1 and Hsp90, we next explored whether Hsp90 inhibitors could antagonize the ATPase activity of Smchd1. We found that two Hsp90 inhibitors from different chemical classes, the geldanamycin analogue, 17-AAG, and NVP-BEP800, only imposed modest reduction of the ATPase activity of Smchd1 at concentrations up to 10 μM (Figure 5C). By contrast, the Hsp90 inhibitor, radicicol, is a potent inhibitor of Smchd1's ATPase activity, specifically, when incubated with 0.2 μM of 111-702 aa in the presence of 10 μM ATP, the half-maximal inhibitory concentration (IC_{50}) value of radicicol was estimated to be approximately 144 nM (Figure 5C). The susceptibility of Smchd1's ATPase domain to radicicol inhibition suggests that the Smchd1 active site might share some basic structural features with Hsp90. However, other classes of Hsp90 inhibitors that also target the Hsp90 ATP-binding pocket exerted comparatively little inhibition upon Smchd1 ATPase activity, which is unsurprising in light of the divergent Smchd1 and Hsp90 sequences.

A FSHD2 mutation in the N-terminal region may cause loss-of-function of SMCHD1

We next attempted to characterise a mutant form of N-terminal region of Smchd1 with a single amino acid substitution, a glycine-to-arginine substitution at residue 425 (G425R), equivalent to the change caused by a missense *SMCHD1* mutation in a FSHD2 patient [44]. Intriguingly, while recombinant protein harboring this mutation could be expressed using the baculovirus insect cell system, the yield was significantly diminished and the mutation appears to alter the isoelectric properties of the recombinant protein relative to the wild-type counterpart, as judged by its elution profile from ion-exchange chromatography (Supplementary Figure 2). While further studies are required to examine whether the G425R mutation reduces the abundance of SMCHD1 protein *in cellula*, these results are the first indication that the G>R transition may potentially compromise the function of SMCHD1 protein and underlie the pathogenicity.

Discussion

Significant efforts have recently been devoted into unravelling the molecular mechanisms by which Smchd1 mediates epigenetic regulation due to its implication in disease. To better understand one of Smchd1 protein's two component domains, the putative ATPase domain, we have generated recombinant proteins, performed SAXS analysis to examine its structure in solution and *in vitro* assays to characterise its ATPase activity.

In contrast to canonical SMC proteins that possess bipartite ABC-type ATPase domains, Smchd1 contains a putative GHKL-ATPase domain, which is a member of a functionally diverse protein superfamily. Although initial attempts to generate a recombinant protein representing the predicted ATPase domain were unsuccessful, we empirically delineated C-terminal boundaries downstream of the ATPase domain on the basis of its weak sequence homology to Hsp90 and obtained high-quality recombinant proteins encompassing ~600 residues of the N-terminal region of Smchd1. Thus, it is possible that the region adjacent the ATPase domain might be intimately connected to the ATPase domain and that the intact N-terminal entity exhibits improved overall stability and solubility compared with the truncated domain in isolation.

Upon determining the solution structure of the N-terminal region, specifically, for 111-702 aa and 111-725 aa of Smchd1, we found that both constructs exist as monomers with an elongated topology similar to that of the full-length Hsp90 monomer, despite the lack of apparent overall sequence conservation. It is worth noting some structural homology beyond the ATPase domain has previously been recognized among several GHKL-ATPase containing proteins [45]. In contrast, this region in Smchd1 was recently proposed to encode a domain with weak homology to a Bromo Adjacent Homology (BAH) domain [27]. However, considering the context of this region downstream of a GHKL-type ATPase and a molecular envelope that resembles the Hsp90 structure, we consider it more likely that domains resembling Hsp90's middle and C-terminal regions will be present. Based on their sequences, however, the secondary structure composition of Smchd1 and Hsp90 of the regions C-terminal to the ATPase domains are likely to be rather divergent. Given that the middle-domain and C-terminal domain of Hsp90 are important for regulating the ATPase activity of Hsp90 [46-48], it would be interesting to explore in the future whether other structural components within the N-terminal region of Smchd1 are involved in modulating its enzymatic activity.

In this study, we have sought to examine the ATPase activity of the N-terminal region of Smchd1 using three independent assays. Using a coupled ATPase assay, we found that while the N-terminal region of Smchd1 only exhibited weak ATPase activity, the turnover number is in fact comparable to that of yeast Hsp90. It is also notable that our enzymatic characterisation was performed at 25°C, rather than the 37°C typically used for Hsp90 studies, which may contribute to a relative underestimation of catalytic activity in our assays.

Because we have demonstrated that the N-terminal region of Smchd1 exists as a monomer in solution in the presence or absence of ATP, the oligomeric state of an active GHKL-ATPase domain of Smchd1 contrasts with that of other members of the GHKL-ATPase family, such as Hsp90 and MutL, which are known to undergo ATP-binding induced dimerisation. However, Smchd1's monomeric nature is not without precedent amongst the GHKL-ATPase protein family. For example, the GHKL-ATPase domain in histidine kinases [49, 50] and Pms2 [51], a subunit of human MutL heterodimer, are enzymatically active in their monomeric state. Furthermore, a recent study indicates that the GHKL-ATPase domain of a prokaryotic MutL homodimer binds to ATP without dimerisation [52]. Additionally, it has been shown that Hsp90 lacking the C-terminal dimerization domain exhibits approximately one-tenth of the ATPase activity of the full-length protein [43, 53, 42] and dimerisation of the ATPase domain of Hsp90 synergistically stimulates its catalytic activity [53, 54, 42]. While we cannot strictly rule out the possibility of transient dimerization of the N-terminal region upon ATP-binding and any induced ATPase activities upon dimerisation, our prior work has demonstrated that Smchd1 homodimerises via the C-terminal hinge domain (residues 1652-1965 aa) [7, 28] (Figure 6). Consequently, it remains to be determined whether the ATPase domains of two Smchd1 protomers in the hinge domain-mediated homodimer are in spatial proximity in cells and whether the ATPase domain of Smchd1 exhibits enhanced catalytic activity in the context of dimeric full-length protein. Further investigations of the relationship between the oligomeric state and the ATPase activity will thus provide further structure-function insights into the GHKL-ATPase domain.

It was also interesting to note that our findings suggest that the GHKL-ATPase domain of Smchd1 is subject to ADP inhibition, a scenario that has been reported for yeast Hsp90, which has approximately 3- to 5-fold higher binding affinity to ADP than ATP [14, 43]. On the other hand, addition of ADP does not affect ATP hydrolysis by bacterial MutL, despite the ATP-binding pocket being well conserved between MutL and Hsp90 [19]. Therefore, it will be of immense interest to examine the exact catalytic mechanism of ATP hydrolysis by Smchd1 in future.

We have also evaluated the potency of several small molecule Hsp90 inhibitors on Smchd1's ATPase activity. Remarkably, we found that radicicol could efficiently antagonize the hydrolysis of ATP by Smchd1 with an IC_{50} value lying within the nanomolar range, which is essentially comparable to the IC_{50} value of radicicol on yeast Hsp90 [55]. We noticed that a very recent study showed that the effect of radicicol on Smchd1's ATPase activity is only modest, with the IC_{50} value in the millimolar range [27]. The discrepancy between our findings and results from Brideau *et al.* is potentially due to the differences in the recombinant constructs and biochemical assays employed. The Brideau *et al.* study examined only the core ATPase domain expressed as

a maltose-binding protein fusion in *E. coli*, raising the possibility that either the absence of the region C-terminal to the core ATPase domain or the choice of host may lead to protein instability or incomplete folding and consequent reduced susceptibility to radicicol inhibition. Our data support the idea that radicicol could bind to the ATP-binding pocket in Smchd1 in a similar way as it binds to Hsp90 [55]. Future efforts toward a high-resolution structure of Smchd1's ATPase domain in complex with radicicol will be required to reveal the exact details at atomic level. Furthermore, whether inhibition of Smchd1's ATPase activity by radicicol observed *in vitro* here has any consequences *in cellula* remains to be elucidated.

While the exact biological roles played by the GHKL-ATPase domain remains to be determined, we envisage that hydrolysis of ATP by the ATPase domain could potentially fuel an energy-dependent conformational change of Smchd1 protein necessary for its engagement with chromatin or active manipulation of chromatin (Figure 6). Importantly, several *SMCHD1* mutations in FSHD patients map to the N-terminal region (Figure 6) [8-10, 44], highlighting the likely importance of the N-terminal region to Smchd1's function. As a first attempt, we introduced a single amino acid substitution, G425R, corresponding to a mutation observed in the N-terminal region of SMCHD1 in a FSHD2 patient. We found that this mutation potentially destabilises the Smchd1 ATPase domain, which may provide an explanation for its association with FSHD2 in this patient. Future efforts will be devoted to studying recombinant proteins harbouring other mutations found in FSHD patients. By examining protein stability, oligomeric state, protein conformation and enzymatic activity of the mutant proteins, it is possible to deduce whether these mutations might cause structural or enzymatic defects at the protein level. As discussed above, although the likelihood and necessity of ATPase domain dimerisation remain to be determined, it is possible that mutations identified in FSHD patients might disrupt inter-ATPase domain interactions within a SMCHD1 homodimer or ATP hydrolysis-competent conformation. Thus, our study not only provides a much-needed avenue to assess the potential structural or enzymatic defects caused by those mutations at the protein level but also provides a basis to dissect the mechanistic details of ATP hydrolysis catalysed by Smchd1. Furthermore, while the fluorescence polarisation and luminescence based ATPase assay used in this study may underestimate the ATPase activity in comparison to the highly sensitive coupled ATPase assay, we foresee that the experimental approaches herein could provide a basis for high-throughput screening to identify small molecule allosteric activators of SMCHD1 activity. Such compounds will provide a starting point for developing much-needed agents to therapeutically combat FSHD.

Acknowledgements
We thank the staff of the Australian Synchrotron SAXS/WAXS beamline for their assistance with data collection.

Funding
This study was supported by Project grant 1045936 (to MEB and JMM) from the National Health and Medical Research Council (NHMRC) and fellowship support from the Australian Research Council (to MEB and JMM), the NHMRC to MEB (1110206) and JMM (1105754) and Cancer Council Victoria (to KC). Additional support was provided by the Victorian State Government Operational Infrastructure Support, NHMRC IRIISS grant (9000220), the Australian Cancer Research Foundation, the Ministry of Business, Innovation and Employment (contract UOCX1208 to RCJD), the New Zealand Royal Society Marsden Fund (contract UOC1013 to RCJD), and the US Army Research Laboratory and US Army Research Office (contract W911NF-11-1-0481 to RCJD).

Author contributions
KC designed, performed and analysed experiments; RCJD, ISL and FGP performed experiments and analysed data; SNY performed experiments; MEB contributed to experimental design and data analysis; JMM contributed to experimental design, data collection and data analysis. The manuscript was written mostly by KC with contributions from the other authors.

References

1 Blewitt, M. E., Vickaryous, N. K., Hemley, S. J., Ashe, A., Bruxner, T. J., Preis, J. I., Arkell, R. and Whitelaw, E. (2005) An N-ethyl-N-nitrosourea screen for genes involved in variegation in the mouse. Proceedings of the National Academy of Sciences of the United States of America. **102**, 7629-7634

2 Blewitt, M. E., Gendrel, A. V., Pang, Z., Sparrow, D. B., Whitelaw, N., Craig, J. M., Apedaile, A., Hilton, D. J., Dunwoodie, S. L., Brockdorff, N., Kay, G. F. and Whitelaw, E. (2008) SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. Nature genetics. **40**, 663-669

3 Gendrel, A. V., Tang, Y. A., Suzuki, M., Godwin, J., Nesterova, T. B., Greally, J. M., Heard, E. and Brockdorff, N. (2013) Epigenetic functions of smchd1 repress gene clusters on the inactive X chromosome and on autosomes. Molecular and cellular biology. **33**, 3150-3165

4 Leong, H. S., Chen, K., Hu, Y., Lee, S., Corbin, J., Pakusch, M., Murphy, J. M., Majewski, I. J., Smyth, G. K., Alexander, W. S., Hilton, D. J. and Blewitt, M. E. (2013) Epigenetic regulator Smchd1 functions as a tumor suppressor. Cancer research. **73**, 1591-1599

5 Mould, A. W., Pang, Z., Pakusch, M., Tonks, I. D., Stark, M., Carrie, D., Mukhopadhyay, P., Seidel, A., Ellis, J. J., Deakin, J., Wakefield, M. J., Krause, L., Blewitt, M. E. and Kay, G. F. (2013) Smchd1 regulates a subset of autosomal genes subject to monoallelic expression in addition to being critical for X inactivation. Epigenetics & chromatin. **6**, 19

6 Massah, S., Hollebakken, R., Labrecque, M. P., Kolybaba, A. M., Beischlag, T. V. and Prefontaine, G. G. (2014) Epigenetic characterization of the growth hormone gene identifies SmcHD1 as a regulator of autosomal gene clusters. PloS one. **9**, e97535

7 Chen, K., Hu, J., Moore, D. L., Liu, R., Kessans, S. A., Breslin, K., Lucet, I. S., Keniry, A., Leong, H. S., Parish, C. L., Hilton, D. J., Lemmers, R. J., van der Maarel, S. M., Czabotar, P. E., Dobson, R. C., Ritchie, M. E., Kay, G. F., Murphy, J. M. and Blewitt, M. E. (2015) Genome-wide binding and mechanistic analyses of Smchd1-mediated epigenetic regulation. Proceedings of the National Academy of Sciences of the United States of America. **112**, E3535-3544

8 Lemmers, R. J., Tawil, R., Petek, L. M., Balog, J., Block, G. J., Santen, G. W., Amell, A. M., van der Vliet, P. J., Almomani, R., Straasheijm, K. R., Krom, Y. D., Klooster, R., Sun, Y., den Dunnen, J. T., Helmer, Q., Donlin-Smith, C. M., Padberg, G. W., van Engelen, B. G., de Greef, J. C., Aartsma-Rus, A. M., Frants, R. R., de Visser, M., Desnuelle, C., Sacconi, S., Filippova, G. N., Bakker, B., Bamshad, M. J., Tapscott, S. J., Miller, D. G. and van der Maarel, S. M. (2012) Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. Nature genetics. **44**, 1370-1374

9 Sacconi, S., Lemmers, R. J., Balog, J., van der Vliet, P. J., Lahaut, P., van Nieuwenhuizen, M. P., Straasheijm, K. R., Debipersad, R. D., Vos-Versteeg, M., Salviati, L., Casarin, A., Pegoraro, E., Tawil, R., Bakker, E., Tapscott, S. J., Desnuelle, C. and van der Maarel, S. M. (2013) The FSHD2 gene SMCHD1 is a modifier of disease severity in families affected by FSHD1. American journal of human genetics. **93**, 744-751

10 Larsen, M., Rost, S., El Hajj, N., Ferbert, A., Deschauer, M., Walter, M. C., Schoser, B., Tacik, P., Kress, W. and Muller, C. R. (2015) Diagnostic approach for FSHD revisited: SMCHD1 mutations cause FSHD2 and act as modifiers of disease severity in FSHD1. European journal of human genetics : EJHG. **23**, 808-816

11 Dutta, R. and Inouye, M. (2000) GHKL, an emergent ATPase/kinase superfamily. Trends in biochemical sciences. **25**, 24-28

12 Bergerat, A., de Massy, B., Gadelle, D., Varoutas, P. C., Nicolas, A. and Forterre, P. (1997) An atypical topoisomerase II from Archaea with implications for meiotic recombination. Nature. **386**, 414-417
13 Dutta, R., Qin, L. and Inouye, M. (1999) Histidine kinases: diversity of domain organization. Molecular microbiology. **34**, 633-640
14 Prodromou, C., Roe, S. M., Piper, P. W. and Pearl, L. H. (1997) A molecular clamp in the crystal structure of the N-terminal domain of the yeast Hsp90 chaperone. Nature structural biology. **4**, 477-482
15 Panaretou, B., Prodromou, C., Roe, S. M., O'Brien, R., Ladbury, J. E., Piper, P. W. and Pearl, L. H. (1998) ATP binding and hydrolysis are essential to the function of the Hsp90 molecular chaperone in vivo. The EMBO journal. **17**, 4829-4836
16 Berger, J. M., Gamblin, S. J., Harrison, S. C. and Wang, J. C. (1996) Structure and mechanism of DNA topoisomerase II. Nature. **379**, 225-232
17 Brino, L., Urzhumtsev, A., Mousli, M., Bronner, C., Mitschler, A., Oudet, P. and Moras, D. (2000) Dimerization of Escherichia coli DNA-gyrase B provides a structural mechanism for activating the ATPase catalytic center. The Journal of biological chemistry. **275**, 9468-9475
18 Ban, C. and Yang, W. (1998) Crystal structure and ATPase activity of MutL: implications for DNA repair and mutagenesis. Cell. **95**, 541-552
19 Ban, C., Junop, M. and Yang, W. (1999) Transformation of MutL by ATP binding and hydrolysis: a switch in DNA mismatch repair. Cell. **97**, 85-97
20 Lorkovic, Z. J., Naumann, U., Matzke, A. J. and Matzke, M. (2012) Involvement of a GHKL ATPase in RNA-directed DNA methylation in Arabidopsis thaliana. Current biology : CB. **22**, 933-938
21 Moissiard, G., Cokus, S. J., Cary, J., Feng, S., Billi, A. C., Stroud, H., Husmann, D., Zhan, Y., Lajoie, B. R., McCord, R. P., Hale, C. J., Feng, W., Michaels, S. D., Frand, A. R., Pellegrini, M., Dekker, J., Kim, J. K. and Jacobsen, S. E. (2012) MORC family ATPases required for heterochromatin condensation and gene silencing. Science. **336**, 1448-1451
22 Brabbs, T. R., He, Z., Hogg, K., Kamenski, A., Li, Y., Paszkiewicz, K. H., Moore, K. A., O'Toole, P., Graham, I. A. and Jones, L. (2013) The stochastic silencing phenotype of Arabidopsis morc6 mutants reveals a role in efficient RNA-directed DNA methylation. The Plant journal : for cell and molecular biology. **75**, 836-846
23 Pastor, W. A., Stroud, H., Nee, K., Liu, W., Pezic, D., Manakov, S., Lee, S. A., Moissiard, G., Zamudio, N., Bourc'his, D., Aravin, A. A., Clark, A. T. and Jacobsen, S. E. (2014) MORC1 represses transposable elements in the mouse male germline. Nature communications. **5**, 5795
24 Nasmyth, K. and Haering, C. H. (2009) Cohesin: its roles and mechanisms. Annual review of genetics. **43**, 525-558
25 Hirano, T. (2016) Condensin-Based Chromosome Organization from Bacteria to Vertebrates. Cell. **164**, 847-857
26 Verver, D. E., Hwang, G. H., Jordan, P. W. and Hamer, G. (2016) Resolving complex chromosome structures during meiosis: versatile deployment of Smc5/6. Chromosoma. **125**, 15-27
27 Brideau, N. J., Coker, H., Gendrel, A. V., Siebert, C. A., Bezstarosti, K., Demmers, J., Poot, R. A., Nesterova, T. B. and Brockdorff, N. (2015) Independent Mechanisms Target SMCHD1 to Trimethylated Histone H3 Lysine 9-Modified Chromatin and the Inactive X Chromosome. Molecular and cellular biology. **35**, 4053-4068

28 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. The Biochemical journal. **473**, 733-742
29 Murphy, J. M., Zhang, Q., Young, S. N., Reese, M. L., Bailey, F. P., Eyers, P. A., Ungureanu, D., Hammaren, H., Silvennoinen, O., Varghese, L. N., Chen, K., Tripaydonis, A., Jura, N., Fukuda, K., Qin, J., Nimchuk, Z., Mudgett, M. B., Elowe, S., Gee, C. L., Liu, L., Daly, R. J., Manning, G., Babon, J. J. and Lucet, I. S. (2014) A robust methodology to subclassify pseudokinases based on their nucleotide-binding properties. The Biochemical journal. **457**, 323-334
30 Kirby, N. M., Mudie, S. T., Hawley, A. M., Cookson, D. J., Mertens, H. D. T., Cowieson, N. and Samardzic-Boban, V. (2013) A low-background-intensity focusing small-angle X-ray scattering undulator beamline. J Appl Crystallogr. **46**, 1670-1680
31 Varghese, L. N., Ungureanu, D., Liau, N. P., Young, S. N., Laktyushin, A., Hammaren, H., Lucet, I. S., Nicola, N. A., Silvennoinen, O., Babon, J. J. and Murphy, J. M. (2014) Mechanistic insights into activation and SOCS3-mediated inhibition of myeloproliferative neoplasm-associated JAK2 mutants from biochemical and structural analyses. The Biochemical journal. **458**, 395-405
32 Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A. G., Gajda, M., Gorba, C., Mertens, H. D., Konarev, P. V. and Svergun, D. I. (2012) New developments in the program package for small-angle scattering data analysis. J Appl Crystallogr. **45**, 342-350
33 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
34 Laue, T. M., Shah, B. D., Ridgeway, T. M. and Pelletier, S. L. (1992) Computer-aided interpretation of analytical sedimentation data for proteins. The Royal Society of Chemistry, Cambridge
35 Perugini, M. A., Schuck, P. and Howlett, G. J. (2000) Self-association of human apolipoprotein E3 and E4 in the presence and absence of phospholipid. The Journal of biological chemistry. **275**, 36758-36765
36 Schuck, P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. Biophysical journal. **78**, 1606-1619
37 Schuck, P., Perugini, M. A., Gonzales, N. R., Howlett, G. J. and Schubert, D. (2002) Size-distribution analysis of proteins by analytical ultracentrifugation: strategies and application to model systems. Biophysical journal. **82**, 1096-1111
38 Rowlands, M., McAndrew, C., Prodromou, C., Pearl, L., Kalusa, A., Jones, K., Workman, P. and Aherne, W. (2010) Detection of the ATPase activity of the molecular chaperones Hsp90 and Hsp72 using the TranscreenerTM ADP assay kit. Journal of biomolecular screening. **15**, 279-286
39 Keown, J. R. and Pearce, F. G. (2014) Characterization of spinach ribulose-1,5-bisphosphate carboxylase/oxygenase activase isoforms reveals hexameric assemblies with increased thermal stability. The Biochemical journal. **464**, 413-423
40 Huai, Q., Wang, H., Liu, Y., Kim, H. Y., Toft, D. and Ke, H. (2005) Structures of the N-terminal and middle domains of E. coli Hsp90 and conformation changes upon ADP binding. Structure. **13**, 579-590

41 Ali, M. M., Roe, S. M., Vaughan, C. K., Meyer, P., Panaretou, B., Piper, P. W., Prodromou, C. and Pearl, L. H. (2006) Crystal structure of an Hsp90-nucleotide-p23/Sba1 closed chaperone complex. Nature. **440**, 1013-1017
42 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
43 Weikl, T., Muschler, P., Richter, K., Veit, T., Reinstein, J. and Buchner, J. (2000) C-terminal regions of Hsp90 are important for trapping the nucleotide during the ATPase cycle. Journal of molecular biology. **303**, 583-592
44 Lemmers, R. J., Goeman, J. J., van der Vliet, P. J., van Nieuwenhuizen, M. P., Balog, J., Vos-Versteeg, M., Camano, P., Ramos Arroyo, M. A., Jerico, I., Rogers, M. T., Miller, D. G., Upadhyaya, M., Verschuuren, J. J., Lopez de Munain Arregui, A., van Engelen, B. G., Padberg, G. W., Sacconi, S., Tawil, R., Tapscott, S. J., Bakker, B. and van der Maarel, S. M. (2015) Inter-individual differences in CpG methylation at D4Z4 correlate with clinical variability in FSHD1 and FSHD2. Human molecular genetics. **24**, 659-669
45 Meyer, P., Prodromou, C., Hu, B., Vaughan, C., Roe, S. M., Panaretou, B., Piper, P. W. and Pearl, L. H. (2003) Structural and functional analysis of the middle segment of hsp90: implications for ATP hydrolysis and client protein and cochaperone interactions. Molecular cell. **11**, 647-658
46 Harris, S. F., Shiau, A. K. and Agard, D. A. (2004) The crystal structure of the carboxy-terminal dimerization domain of htpG, the Escherichia coli Hsp90, reveals a potential substrate binding site. Structure. **12**, 1087-1097
47 Hawle, P., Siepmann, M., Harst, A., Siderius, M., Reusch, H. P. and Obermann, W. M. (2006) The middle domain of Hsp90 acts as a discriminator between different types of client proteins. Molecular and cellular biology. **26**, 8385-8395
48 Retzlaff, M., Stahl, M., Eberl, H. C., Lagleder, S., Beck, J., Kessler, H. and Buchner, J. (2009) Hsp90 is regulated by a switch point in the C-terminal domain. EMBO reports. **10**, 1147-1153
49 Wang, C., Sang, J., Wang, J., Su, M., Downey, J. S., Wu, Q., Wang, S., Cai, Y., Xu, X., Wu, J., Senadheera, D. B., Cvitkovitch, D. G., Chen, L., Goodman, S. D. and Han, A. (2013) Mechanistic insights revealed by the crystal structure of a histidine kinase with signal transducer and sensor domains. PLoS biology. **11**, e1001493
50 Rivera-Cancel, G., Ko, W. H., Tomchick, D. R., Correa, F. and Gardner, K. H. (2014) Full-length structure of a monomeric histidine kinase reveals basis for sensory regulation. Proceedings of the National Academy of Sciences of the United States of America. **111**, 17839-17844
51 Guarne, A., Junop, M. S. and Yang, W. (2001) Structure and function of the N-terminal 40 kDa fragment of human PMS2: a monomeric GHL ATPase. The EMBO journal. **20**, 5521-5531
52 Iino, H., Hikima, T., Nishida, Y., Yamamoto, M., Kuramitsu, S. and Fukui, K. (2015) Small-angle X-ray scattering analysis reveals the ATP-bound monomeric state of the ATPase domain from the homodimeric MutL endonuclease, a GHKL phosphotransferase superfamily protein. Extremophiles : life under extreme conditions
53 Richter, K., Muschler, P., Hainzl, O. and Buchner, J. (2001) Coordinated ATP hydrolysis by the Hsp90 dimer. The Journal of biological chemistry. **276**, 33689-33696

54 Vaughan, C. K., Piper, P. W., Pearl, L. H. and Prodromou, C. (2009) A common conformationally coupled ATPase mechanism for yeast and human cytoplasmic HSP90s. FEBS J. **276**, 199-209
55 Roe, S. M., Prodromou, C., O'Brien, R., Ladbury, J. E., Piper, P. W. and Pearl, L. H. (1999) Structural basis for inhibition of the Hsp90 molecular chaperone by the antitumor antibiotics radicicol and geldanamycin. Journal of medicinal chemistry. **42**, 260-266
56 Biegert, A., Mayer, C., Remmert, M., Soding, J. and Lupas, A. N. (2006) The MPI Bioinformatics Toolkit for protein sequence analysis. Nucleic acids research. **34**, W335-339
57 Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic acids research. **22**, 4673-4680
58 Robert, X. and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. Nucleic acids research. **42**, W320-324
59 Svergun, D. I. (1992) Determination of the Regularization Parameter in Indirect-Transform Methods Using Perceptual Criteria. J Appl Crystallogr. **25**, 495-503
60 Kozin, M. B. and Svergun, D. I. (2001) Automated matching of high- and low-resolution structural models. J Appl Crystallogr. **34**, 33-41
61 Svergun, D., 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 Crystallogr. **28**, 768-773

Figure legends

Figure 1. Schematic diagrams of domain architecture and sequence alignments.
(A) Domain architecture of Smchd1, Hsp90 and canonical SMC proteins with their component domains shown schematically. SMC protein features an ABC-type ATPase (head), dimerises via the hinge domain and forms a ring-shaped complex with non-SMC subunits. Smchd1 contains a predicted GHKL-type ATPase (111-365 aa) with limited homology to other members of GHKL-ATPase protein family, such as Hsp90. The N-terminal region of Smchd1 (111-702 aa) defined by this study is shaded in magenta.
(B) Protein sequence of *Mus musculus* (*Mm*) Smchd1 (111-790 aa) and orthologous sequences in *Gallus gallus* (*Gg*) and *Xenopus tropicalis* (*Xt*) aligned to full-length *Saccharomyces cerevisiae* (*Sc*) Hsp90, *Mm* Hsp90aa1, *Caenorhabditis elegans* (*Cs*) Hsp90, *Arabidopsis thaliana* (*At*) DMS11 (101-663 aa), *Capsella rubella* (*Cr*) Uncharacterised protein (UniProt ID: R0IMK0, 103-655 aa) and *Eutrema salsugineum* (*Es*) Uncharacterised protein (UniProt ID: V4MSW7, 97-674 aa). Identical residues are coloured in red and similar residues are shaded in yellow. Secondary structures deduced from crystal structure of Hsp90 (PDB: 2CG9) and predicted from the primary sequence of Smchd1 using Quick2D [56] are shown on the top and bottom, respectively. The conserved GHKL ATPase motifs I-IV are shaded in blue. Positions of recombinant construct boundaries and positions of mutated residues are indicated with the vertical arrows in red. Image was generated using CLUSTAL W [57] and ESPript3.0 [58].

Figure 2. Expression and purification of the N-terminal region of Smchd1.
(A) Size exclusion chromatogram with absorbance at 280 nm (y-axis, arbitrary units) over retention time (x-axis, minutes) for recombinant Smchd1 protein corresponding to 111-702 aa (red line), 111-725 aa (purple line) and 111-702 aa with the E147A substitution (black line). Elution time of molecular weight (MW) standards are indicated as vertical lines with MW labelled. Linear regression plot of Log_{10}MW (y-axis, arbitrary units) over retention time (x-axis, minutes) was constructed based on the standards as in A. The line equation was constructed to solve the y based on x to estimate the molecular weight of 111-702 aa (red dot) and 111-725 aa (purple dot).
(B) Coomassie-stained 4-12% (w/v) reducing SDS-PAGE gel showing fractions collected during SEC as in A for 111-702 aa (left) and 111-725 aa (middle). Both the wild-type and E147A mutant of 111-702 aa were purified at high-purity as judged by Coomassie-staining (right). Red asterisk indicates the position of the recombinant protein migrated on gel, reflecting the MW of the denatured protein.
(C) Sedimentation velocity demonstrates that 111-702 aa is a single species in solution. Top panels: residuals for the continuous size, *c(s),* distribution best fits plotted as a function of radial position (cm) from the axis of rotation for 111-702 aa of Smchd1 at 7 μM in the absence of ATPγS (top) and the presence of ATPγS (0.1 mM) (bottom). Bottom panel: Continuous size, *c(s)*, distribution is plotted as a function of $s_{20,w}$ for 111-702 aa of Smchd1 at 7 μM in the absence of ATPγS (solid line) and the presence of ATPγS (0.1 mM) (dashed line). This analysis confirms that 111-702 aa (at 7 μM) is a single species in solution, even in the presence of ATPγS (0.1 mM). Analyses was performed using the program SEDFIT [36] at a resolution of 200, with s_{min} = 0.5, s_{max} = 9 and at a confidence level (F-ratio) = 0.95. Statistics for the nonlinear least squares best-fits were: in the absence of ATPγS, r.m.s.d. = 0.008, runs test-Z = 0.4; in the presence of ATPγS, r.m.s.d. = 0.008, runs test-Z = 3.9.

(D) Sedimentation velocity demonstrates that 111-702 aa is a monomer in solution. Top panels: residuals for the continuous mass, *c(M),*distribution best fits, plotted as a function of radial position (cm) from the axis of rotation for 111-702 aa of Smchd1 at 7 μM in the absence of ATPγS (top) and the presence of ATPγS (0.1 mM) (bottom). Bottom panel: Continuous mass, *c(M)*, distribution is plotted as a function of mass for 111-702 aa of Smchd1 at 7 μM in the absence of ATPγS (solid line) and the presence of ATPγS (0.1 mM) (dashed line). This analyse confirms that 111-702 aa of Smchd1 (at 7 μM) is a monomer in solution, even in the presence of ATPγS (0.1 mM): the mass calculated from its sequence is 68.4207 kD, which agrees well with the mass determined by sedimentation velocity experiments: 67.1 and 67.3 kD. Analyses was performed using the program SEDFIT [36] at a resolution of 200, with $mass_{min}$ = 5 kD, $mass_{max}$ = 210 kD and at a confidence level (F-ratio) = 0.95. Statistics for the nonlinear least squares best-fits were: in the absence of ATPγS, r.m.s.d. = 0.008, runs test-Z = 1; in the presence of ATPγS, r.m.s.d. = 0.008, runs test-Z = 3.8.
(E) Thermostability of the WT (blue) and E147A mutant (grey) of Smchd1 (111-702 aa), respectively, monitored by the fluorescence intensity of dye SYPRO orange (y-axis, 530 nm) over increasing temperature (x-axis, °C). The plot is representative of two independent experiments.

Figure 3. Small-angle X-ray scattering (SAXS) analysis of 111-702 aa and 111-725 aa of Smchd1.
(A, D) Scattering intensity profile for 111-702 aa (A) and 111-725 aa (D). The raw SAXS data are shown as black circles representing mean intensity $I(q)$ as a function of momentum transfer q in $Å^{-1}$.
(B, E) Guinier plot of 111-702 aa (B) and 111-725 aa (E) for $qRg \leq 1.3$. Linearity indicates that neither high molecular weight aggregates nor inter-particle interference contribute measurably to scattering. The radius of gyration and initial scattering intensity $I(0)$ were approximated using the Guinier equation with PRIMUS [33], giving values as indicated in Table 1.
(C) and (F) Fourier transformation of the scattering intensity as in (A) and (D), respectively, yields the pair-wise inter-atomic distance distribution function, $P(r)$, calculated using GNOM [59]. The Rg and maximum particle dimension D_{max} calculated from the $P(r)$ analysis are as indicated in Table 1.

Figure 4. Bead models of 111-702 aa and 111-725 aa of Smchd1 adopts an elongated structure similar to the crystal structure of Hsp90.
(A) and (B) *Ab initio* bead models of 111-702 aa (A) and 111-725 aa (B), respectively, represented as a grey surface, superimposed with crystal structure of *Saccharomyces cerevisiae* Hsp90 (PDB: 2CG9), represented using program SUPCOMB [60]. Residues 2-597 aa (A) and 2-628 aa (B) of Hsp90 that are comparable to 111-702 aa and 111-725 aa of Smchd1 (based on sequence alignment in Figure 1), respectively, were included in the superimposition. The GHKL-ATPase domain, middle-domain and C-terminal domain of Hsp90 are coloured as in Figure 1.
(C) and (D) The theoretical scattering profile (red line) calculated from the crystal structure in (A) and (B) was fitted to the experimental scattering data of 111-702 aa and 111-725 aa of Smchd1 using CRYSOL [61], with χ values of 0.66 and 0.56, respectively.

Figure 5. The N-terminal region of Smchd1 is a functional ATPase.
(A) Amount of ADP produced (μM, y-axis) by the WT (left) and E147A mutant (right) of Smchd1 (111-702 aa) at protein concentrations indicated on the x-axis (μM) after incubating with 10 μM of ATP. Results from one-hour reactions and two-hour reactions are shown in blue and red, respectively. Results from two-hour reactions in the absence of magnesium ions are shown in black. Each data point is the mean ± standard deviation of triplicate technical replicates. Plots are representative of two independent experiments using different batches of the same protein preparation, as measured by the fluorescence polarisation assay.
(B) Specific ATPase activity (μM ADP · min^{-1}) measured at a range of different protein concentrations (μM, x-axis) of Smchd1 (111-702 aa) using the PK-LDH coupled ATPase assay. Each point represents a single experiment. ATPase activity was proportional to protein concentration, with *kcat* ~0.6 μM ADP/min/μM protein.
(C) Amount of ADP produced (μM, y-axis) by 0.2 μM of Smchd1 (111-702 aa) after one hour incubation with 10 μM of ATP in the presence of 0, 0.25, 0.5 and 10 μM (x-axis) of Hsp90 inhibitor radicicol (red), 17-AAG (blue) and NVP-BEP800 (black) (left). Does responsive curve for radicicol at various concentrations (nM, x-axis) was obtained by fitting the data with four-parameter logistic curve, yielding an estimated IC_{50} of 144.6 nM (right). Each data point is the mean ± standard deviation of triplicate technical replicates as measured using the fluorescence polarisation assay.

Figure 6. Schematic diagram for the N-terminal region in the context of dimeric full-length Smchd1.
The N-terminal region of Smchd1 is connected to the SMC hinge domain at C-terminus that mediates dimerization. Several single amino acid substitutions of human SMCHD1 protein identified in FSHD patients are outside the notional GHKL-ATPase but mapped to the N-terminal region defined by this study.

Table 1. Data collection and scattering parameters for small-angle X-ray analysis.

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)	2	
Protein concentration	~5 mg/ml sample injected via in-line size exclusion chromatography	
Temperature (°C)	16	
Structural parameters		
Protein sample	111-702 aa	111-725 aa
I(0)(cm^{-1}) [from Guinier]	0.0164 ± 0.0001	0.01539 ± 0.00003
Rg (Å) [from Guinier]	31.6 ± 0.4	34.8 ± 0.9
I(0)(cm^{-1}) [from *P(r)*]	0.0165 ± 0.0001	0.01538 ± 0.00002
Rg (Å) [from *P(r)*]	32.3 ± 0.2	35.3 ± 0.5
D_{max} (Å)	105	115
Software		
Primary data reduction	ScatterBrain (Australia Synchrotron)	
Data processing	PRIMUS, GNOM	
Ab initio modeling	DAMSEL, DAMSUP, DAMAVER, DAMMIF, DAMFILT, SUPCOMB	
Computation of model intensities	CRYSOL	
3D graphic representations	MacPyMOL	

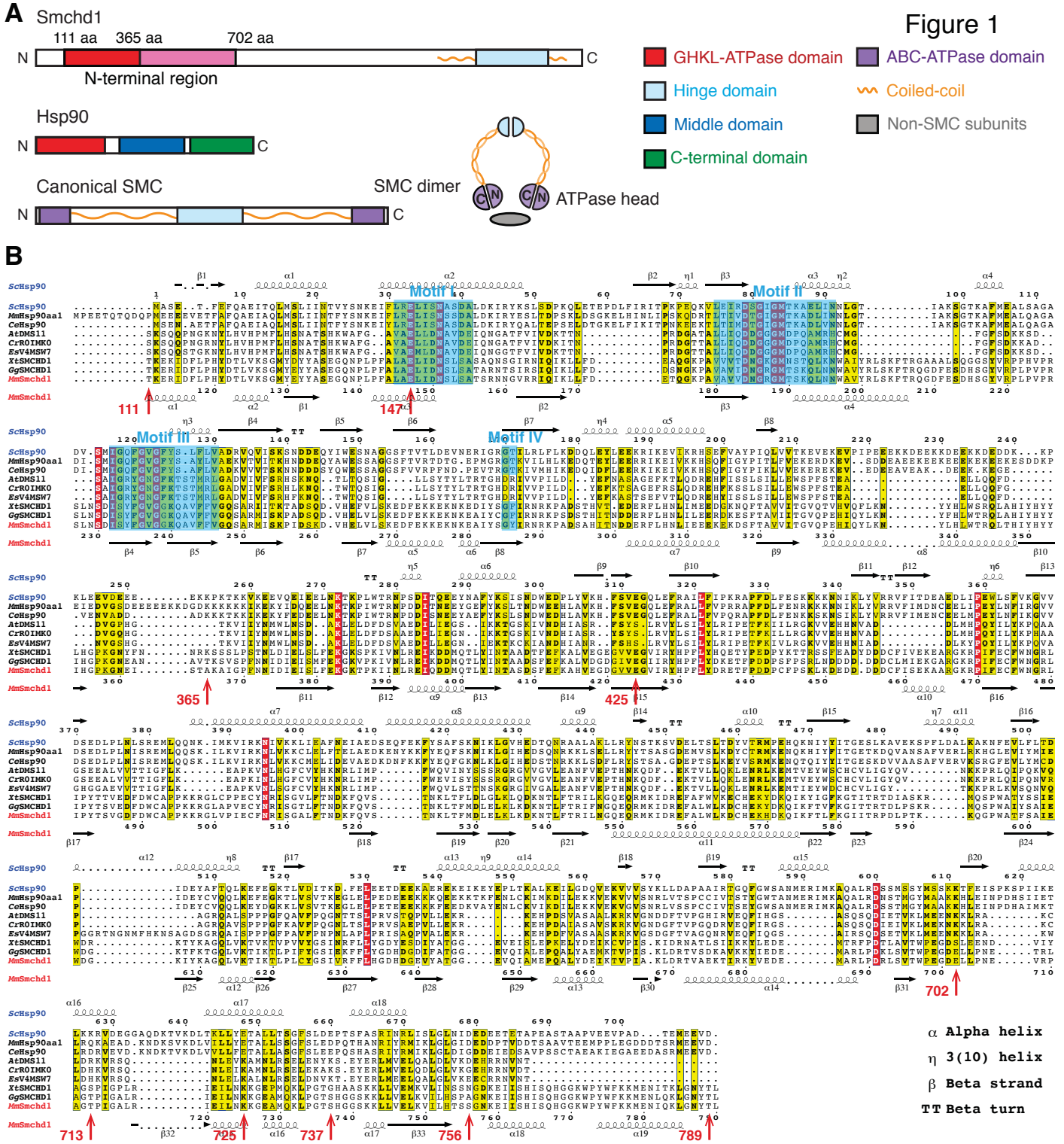
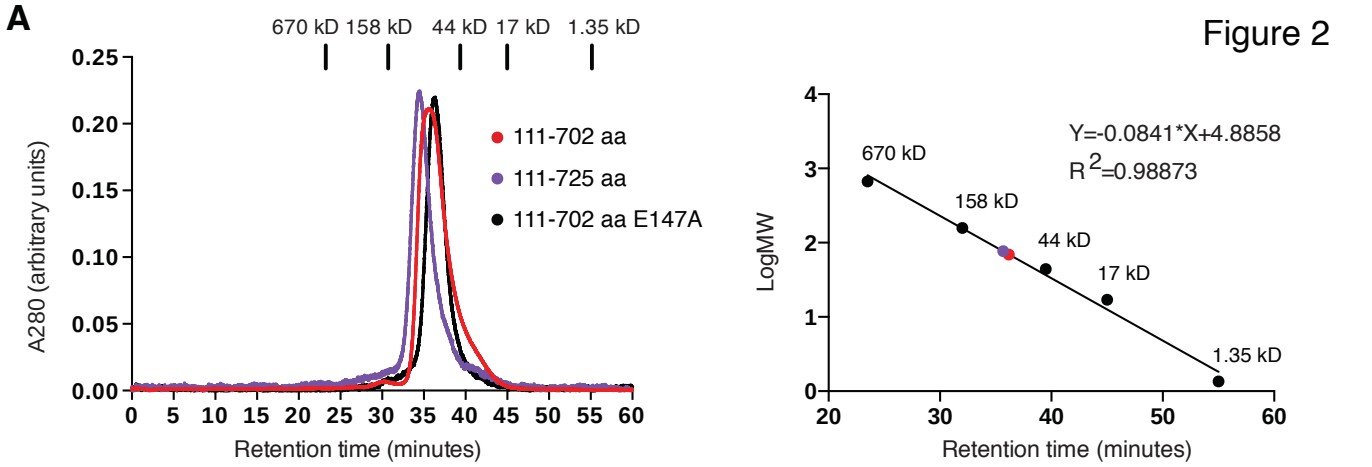
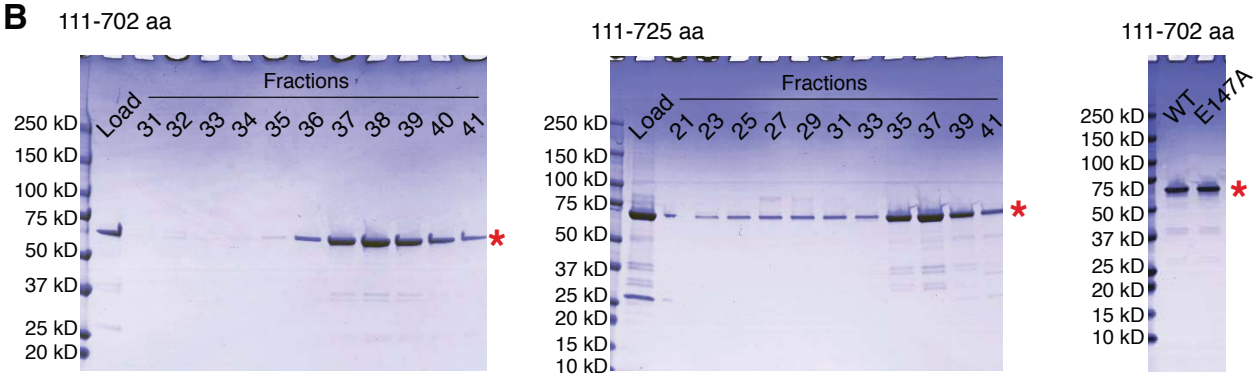
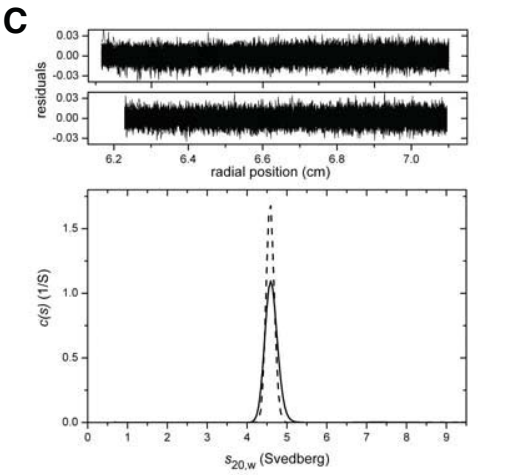

A
Smchd1
111 aa
365 aa
702 aa
N
C
N-terminal region
Hsp90
Canonical SMC
SMC dimer
ATPase head
GHKL-ATPase domain
ABC-ATPase domain
Hinge domain
Coiled-coil
Middle domain
Non-SMC subunits
C-terminal domain
B
Motif I
Motif II
Motif III
Motif IV
111
147
365
425
702
713
725
737
756
789
α Alpha helix
η 3(10) helix
β Beta strand
TT Beta turn

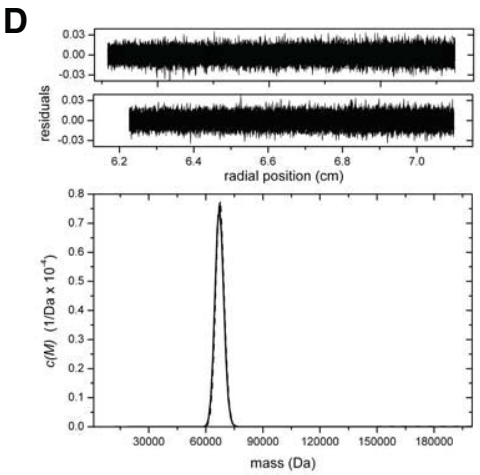
Figure 1

Figure 2


A
670 kD 158 kD 44 kD 17 kD 1.35 kD
A280 (arbitrary units)
Retention time (minutes)
111-702 aa
111-725 aa
111-702 aa E147A
LogMW
Y=-0.0841*X+4.8858
R2=0.98873
670 kD
158 kD
44 kD
17 kD
1.35 kD
Retention time (minutes)


B
111-702 aa
Fractions
Load
250 kD
150 kD
100 kD
75 kD
50 kD
37 kD
25 kD
20 kD
111-725 aa
Fractions
Load
250 kD
150 kD
100 kD
75 kD
50 kD
37 kD
25 kD
20 kD
15 kD
10 kD
111-702 aa
WT
E147A
250 kD
150 kD
100 kD
75 kD
50 kD
37 kD
25 kD
20 kD
15 kD
10 kD


C
residuals
radial position (cm)
c(s) (1/S)
s20,w (Svedberg)


D
residuals
radial position (cm)
c(M) (1/Da x 10-4)
mass (Da)

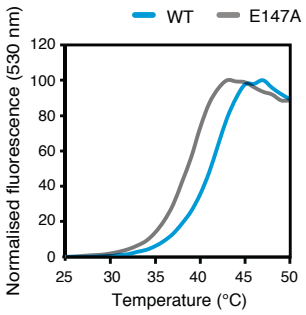

E
WT
E147A
Normalised fluorescence (530 nm)
Temperature (°C)

Figure 3

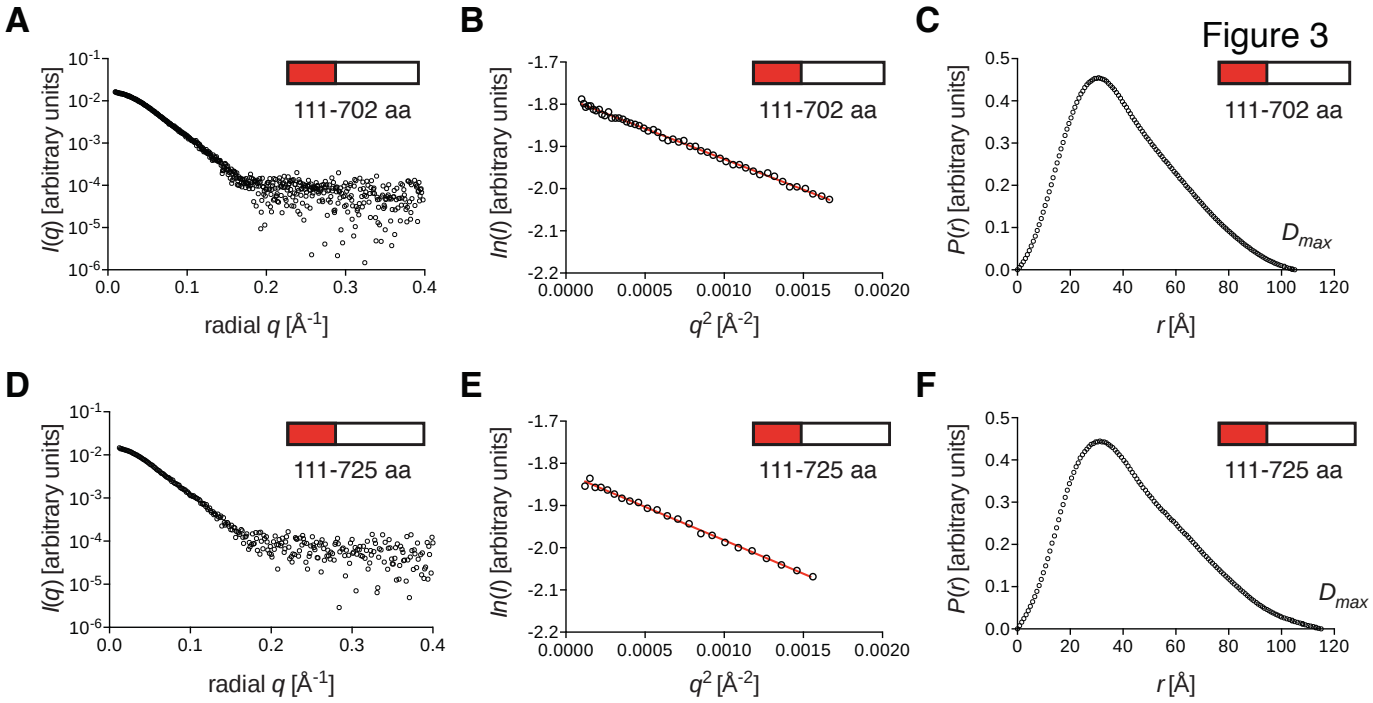

A
B
C
D
E
F
111-702 aa
111-725 aa
I(q) [arbitrary units]
radial q [Å-1]
ln(I) [arbitrary units]
q2 [Å-2]
P(r) [arbitrary units]
r [Å]
Dmax

Figure 4

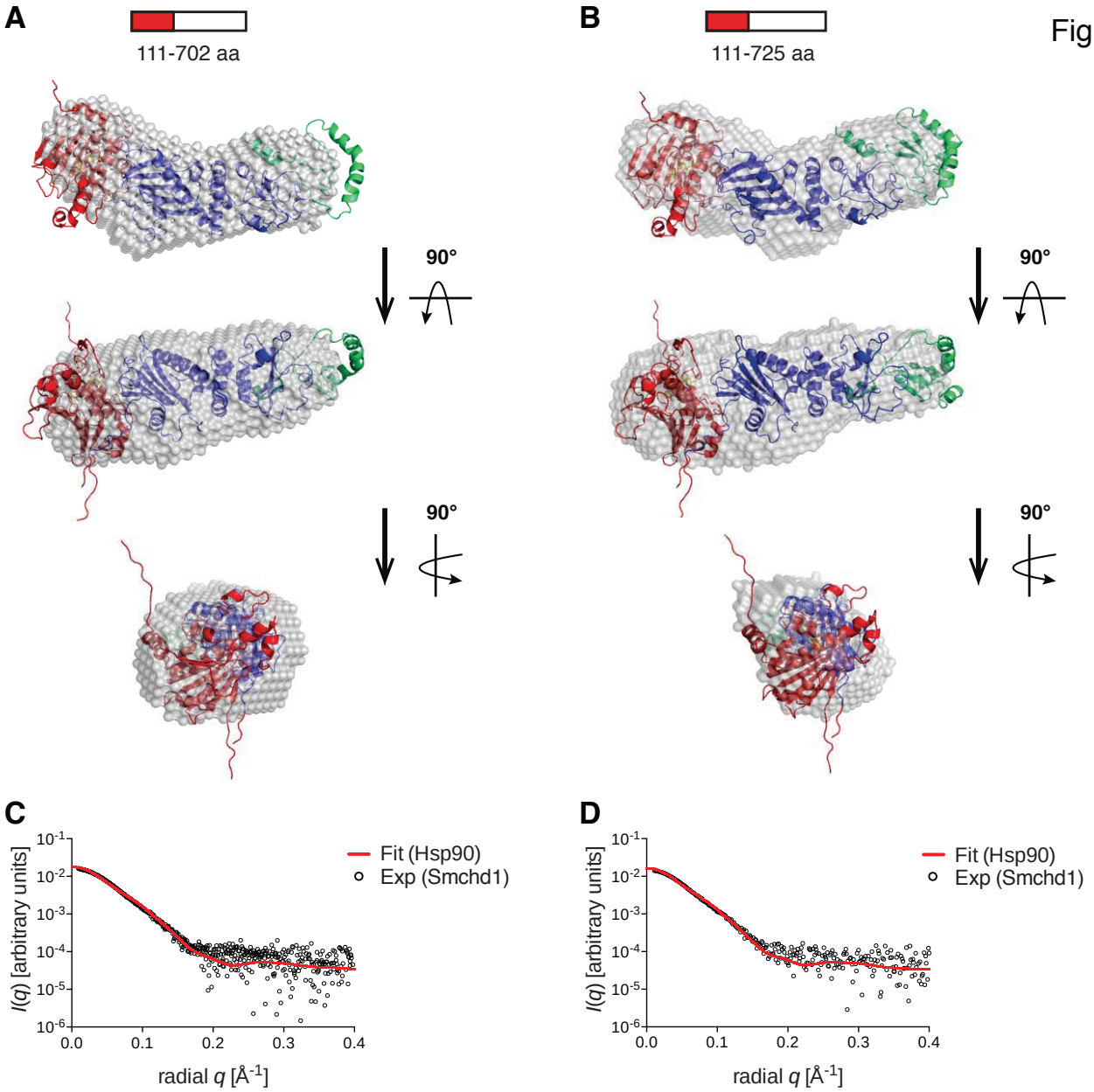

A
111-702 aa
90°
90°
B
111-725 aa
90°
90°
C
I(q) [arbitrary units]
radial q [Å-1]
Fit (Hsp90)
Exp (Smchd1)
D
I(q) [arbitrary units]
radial q [Å-1]
Fit (Hsp90)
Exp (Smchd1)

Figure 5

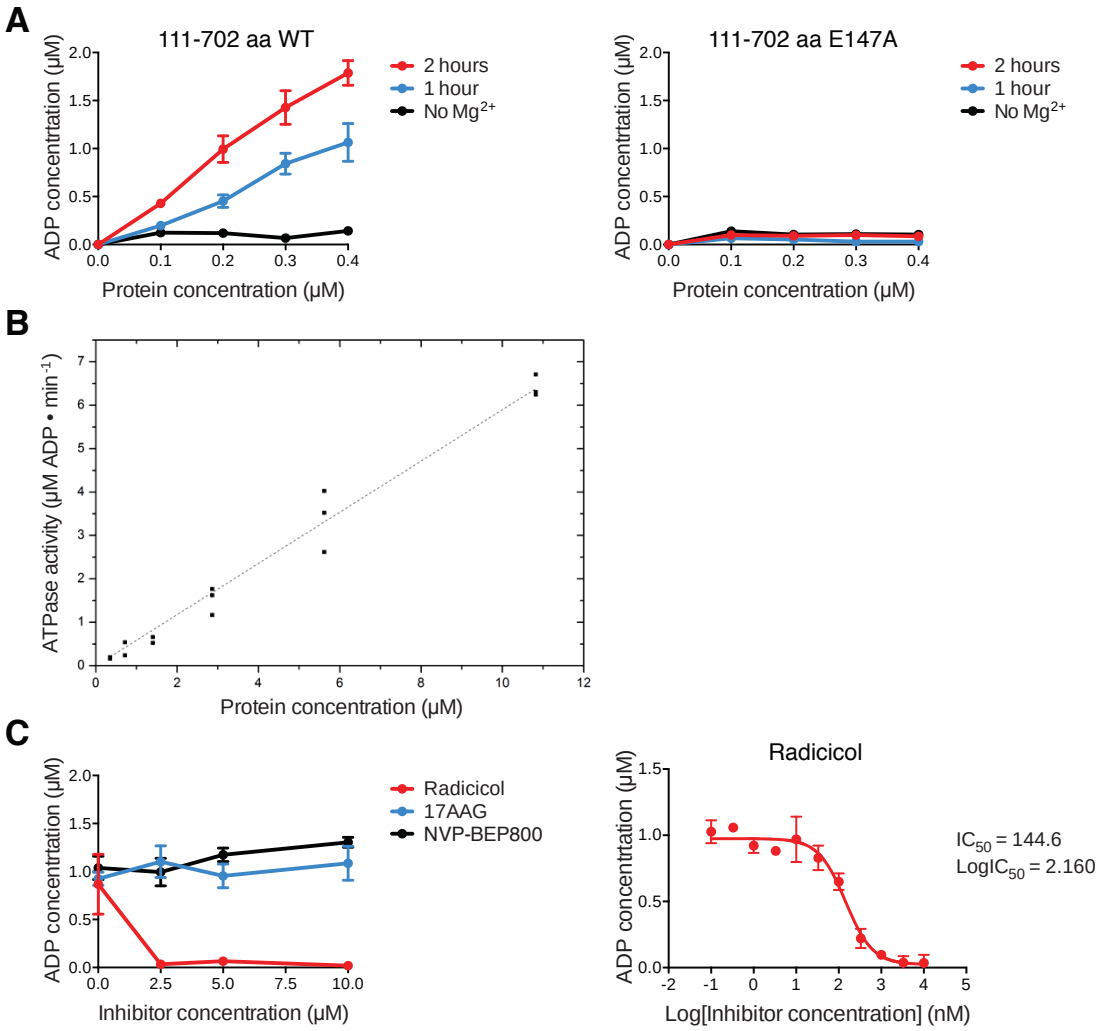
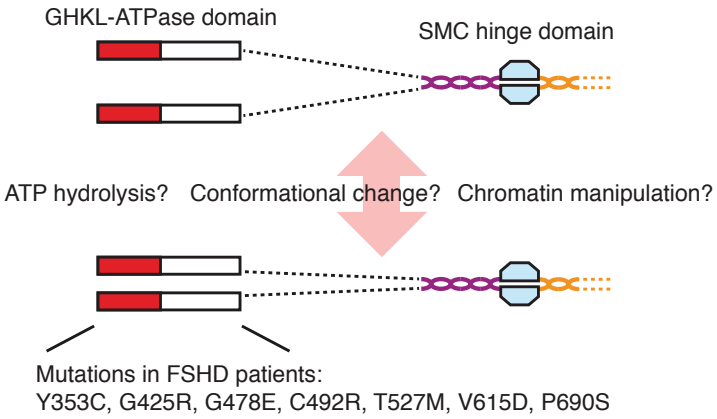

A
111-702 aa WT
ADP concentrtation (μM)
Protein concentration (μM)
2 hours
1 hour
No Mg2+
111-702 aa E147A
B
ATPase activity (μM ADP • min-1)
Protein concentration (μM)
C
Radicicol
17AAG
NVP-BEP800
Inhibitor concentration (μM)
IC50 = 144.6
LogIC50 = 2.160
Log[Inhibitor concentration] (nM)

Figure 6



Mutations in FSHD patients:
Y353C, G425R, G478E, C492R, T527M, V615D, P690S