Supplementary Data

Structure and regulatory role of the C-terminal winged helix domain of the archaeal minichromosome maintenance complex

Christoph Wiedemann[1], Anna Szambowska[2,4], Sabine Häfner[1], Oliver Ohlenschläger[1], Karl-Heinz Gührs[3], and Matthias Görlach[1,*]

[1]*Research Group Biomolecular NMR Spectroscopy,*

[2]*Research Group Biochemistry,*

[3]*Protein laboratory, Leibniz Institute for Age Research, Fritz Lipmann Institute, Beutenbergstr. 11, D-07745 Jena, Germany;*

[4]*Laboratory of Molecular Biology IBB PAS, affiliated with University of Gdansk, Wita Stwosza 59, Gdansk, Poland*

[*]*Corresponding author: mago@fli-leibniz.de*

Table S 1: Primers used in this study

Primer	Sequence 5′-3′	Construct[a]
CW72	gtc ctc gag cta gac ttt ttt gta aca ttc	rev *Sso* MCM C-term. domain, XhoI site
CW74	gtc ctc gag tca gac tat ctt aag gta tcc	rev *Mth* MCM C-term. domain, XhoI site
CW81	agc agc ggc ctg gtg ccg cgc gaa aac ctg tat ttt cag ggc gta tta gat gag	fwd thrombine to TEV site
CW82	gcg cgg cac cag gcc gct gct gtg atg gtg atg	rev thrombine to TEV site
CW83	agt gta gga gtt gat atg taa tga aag tgg aaa a	fwd 2 stop codons distal to *Sso* MCM M604
CW84	cat atc aac tcc tac act ctc tag gaa aag	rev 2 stop codons distal to *Sso* MCM M604
CW85	aca ata atg act ggt aaa taa tga cct aaa agc gct	fwd 2 stop codons distal to *Sso* MCM K618
CW86	ttt acc agt cat tat tgt atc tat atc	rev 2 stop codons distal to *Sso* MCM K618
CW89	gaa aca cct gat agc cca gat agc gat act cca aga caa tta g	fwd *Sso* MCM I555_I557delinsDSD
CW90	tgg gct atc agg tgt ttc tga gct ttt ctt cct	rev *Sso* MCM I555_I557delinsDSD
CW91	gtc cca tgg aaa ttc cta gta aac aga ttg ac	fwd *Sso* MCM, NcoI site
CW106	agt gta gga gtt gat atg gcg gcc gcg gcc gcg gaa agt gga aaa	fwd *Sso* MCM M604_E605ins(A)5, rev CW84
CW107	agt gta gga gtt gat atg gat ata gat aca ata	fwd *Sso* MCM E605_I609del, rev CW84
CW108	aca ata atg act ggt aaa gcg gcc gcg gcc gcg cct aaa agc gct	fwd *Sso* MCM K618_P619ins(A)5, rev CW86
CW109	gga aaa ata gat ata gat aca cct aaa agc gct	fwd *Sso* MCM I614_K618del
CW110	tgt atc tat atc tat ttt tcc act ttc cat atc aac	rev *Sso* MCM I614_K618del
CW111	gga aaa ata gat ata gat aca gcg gcc gcg gcc gcg ata atg act ggt	fwd *Sso* MCM T613_I614ins(A)5, rev CW110
CW112	agt gta gga gtt gat atg gat ata gat aca cct	fwd *Sso* MCM E605_I609del, I614_K618del, rev CW84
CW118	gag tgt agg agt tga tat ggc gag tgg agc aag tgg agc gag tgg tgc gag tgg cgg taa acc taa aag cg	fwd E605_T616delins(ASG)4, rev CW84
CW119	gct cgg tac ccg ggg atc ctc tag a	fwd double strand DNA for EMSA, rev CW120[b]
CW120	tct aga gga tcc ccg ggt acc gag c	rev double strand DNA for EMSA, fwd CW119[b]
CW121	ctt gca tgc ctg cag gtc gac tct aga gga tcc ccg ggt acc gag c	fwd 3′ overhang DNA for EMSA, rev CW122[b]
CW122	tcc tct aga gtc gac ctg cag gca tgc aag	rev 3′ overhang DNA for EMSA, fwd CW121[b]
CW123	gct cgg tac ccg ggg atc ctc tag agt cga	rev 5′ overhang DNA for EMSA, fwd CW121[b]
CW124	tct acc tgg acg acc ggg ttt ttt ttt ttt ttt ttt ttg ggc cag cag gtc cat ca	fwd bubble DNA for EMSA, rev CW125[b]
CW125	tga tgg acc tgc tgg ccc ttt ttt ttt ttt ttt ttt ttc ccg gtc gtc cag gta ga	rev bubble DNA for EMSA, fwd CW124[b]
CW126	gct cgg tac ccg ggg atc ctc tag att ttt ttt ttt ttt ttt ttt	fwd fork DNA for EMSA, rev CW127[b]
CW127	ttt ttt ttt ttt ttt ttt ttt cta gag gat ccc cgg gta ccg agc	rev fork DNA for EMSA, fwd CW126[b]
CW134	agt gta gga gtt gat atg gaa acc gga aaa ata gat ata gat aaa gtt gaa ggt cgt acc cct aaa agc gc	fwd substitute *Sso* MCM E605_K618 by *Mth* MCM E583_T596, rev CW84

Table S 1: Primers used in this study

Primer	Sequence 5′-3′	Construct[a]
CW143	agt gta gga gtt gat atg gag aca ggc aag ata gat ata gac aag	fwd substitute *Sso* E605_V686 by *Mth* MCM E583_V666, rev CW74
CW144	gcc tgt ctc cat atc aac tcc tac act ctc tag gaa aag	rev substitute *Sso* E605_V686 by *Mth* MCM E583_V666, fwd CW91
CW145	agt gta gga gtt gat atg gaa acc gga aaa ata gat ata gat	fwd *Sso* MCM S606T, rev CW84
CW146	ata gat ata gat aca ata ccg act ggt aaa	fwd *Sso* MCM M615P, rev CW149
CW147	ata gat ata gat aca ata ggc act ggt aaa	fwd *Sso* MCM M615G, rev CW149
CW148	ata gat ata gat aca ata gcg act ggt aaa	fwd *Sso* MCM M615A, rev CW149
CW149	tat tgt atc tat atc tat ttt tcc act ttc	rev mutation of *Sso* MCM M615
CW150	agt gga aaa ata gat ata gat ccg ata ccg act ggt ccg cct aaa agc gc	fwd *Sso* MCM T613P, M615P and K618P
CW151	agt gga aaa ata gat ata gat ggc ata ggc act ggt ggc cct aaa agc gc	fwd *Sso* MCM T613G, M615G and K618G
CW152	agt gga aaa ata gat ata gat gcg ata gcg act ggt gcg cct aaa agc gc	fwd *Sso* MCM T613A, M615A and K618A
CW153	atc tat atc tat ttt tcc act ttc cat atc aac	rev mutation of *Sso* MCM T613 M615 and K618
CW156	ggt gat ccc ggt act gcc gcg tca caa atg	fwd *Sso* MCM K346A Walker A-mutation
CW157	ggc agt acc ggg atc acc tat tat cag tat gtg	rev *Sso* MCM K346A Walker A-mutation
CW158	cca ata ttg ata act cca gcg caa tta gag	fwd *Sso* MCM R560A Sensor II-mutation
CW159	tgg agt tat caa tat tgg gct atc agg tgt ttc	rev *Sso* MCM R560A Sensor II-mutation
Y-shpF	$(t)_{44}$ gct cgt gca gac gtc gag gtg agg acg agc tcc tcg tga cca cg	fwd Y-shapped DNA for helicase assay, rev Y-shpR[c]
Y-shpR	cgt ggt cac gag gag ctc gtc ctc acc tcg acg tct gca cga gc$(t)_{44}$	rev Y-shapped DNA for helicase assay, fwd Y-shpF[c]

[a] nomenclature for the description of sequence variations as recommended [1]

[b] DNA oligo sequences for EMSA as published in [2]

[c] DNA oligo sequences for helicase assay as published in [3]

Biophysical characterisation

Purified C-terminal domain of *Sso* MCM represents a folded and thermostable monomer

To obtain initial information about secondary structure elements of the purified C-terminal domain of *Sso* MCM far-UV CD spectra were recorded. The same initial biophysical characterization is provided for the *Mth* MCM C-terminal domain (Supplementary Figure S1, S2, S4-S5 and Table S2-S4). The spectra display two minima of ellipticity at 208 and 222 nm and one maximum at 192 nm (Supplementary Figure S1), respectively, that are typical for a high α-helical content (Supplementary Table S2). The secondary structure composition was calculated using CAPITO [4] and CDPro [5]. The heat induced protein unfolding as monitored by CD spectroscopy revealed that the isolated *Sso* MCM C-terminal domain is highly thermostable (Supplementary Figure S2 and Table S3).

Circular dichroism analysis

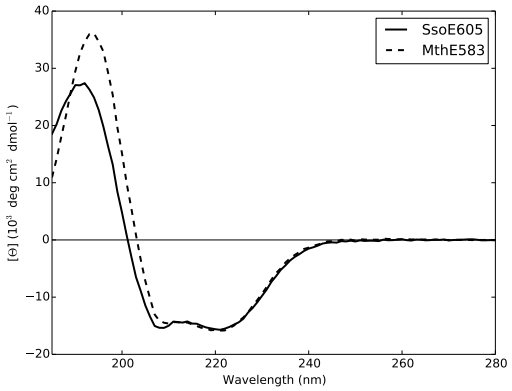

Fig. S 1: Far UV-CD spectra of *S. solfataricus* (solid line) and *M. thermautotrophicus* (dashed line) MCM C-terminal domains at 4℃. Circular dichroism (CD) spectra were collected on a JASCO J-710 CD spectropolarimeter at 4℃ in a 1 mm quartz cuvette to evaluate the proper fold of the proteins and to estimate the secondary structure content. The instrument was calibrated with D-10-camphorsulphonic acid. Purified proteins were exchanged into pure water using NAP-5 columns (GE Healthcare). The protein concentration was in the 1–10 μM range and verified spectrophotometrically at 280 nm with extinction coefficients calculated using ProtParam (http://web.expasy.org/protparam/). Each CD spectrum represents the average of 10 accumulated scans at 10 nm/min with a 1 nm slit width and a time constant of 4 s for a nominal resolution of 0.7 nm. Data were collected between 185 and 260 nm. No further zeroing was applied after background subtraction.

Table S 2: The secondary structure content of *Mth* and *Sso* MCM C-terminal domains derived from CD spectra

	Mth			*Sso*		
	helical	β-strand	irregular	helical	β-strand	irregular
CAPITO	46–57	6–20	30–34	34–44	12–26	39–54
SELCON3[a]	52	9	38	47	9	45
CONTIN[a]	47	13	40	46	8	46
CDSSTR[a]	53	11	37	51	9	40
DSSP[b]	54	15	31	53	17	30

[a] included in the software CDPro [5]
[b] secondary structure determination based on the pdb entries [6] (*Mth* PDB id: 2MA3, *Sso* PDB id: 2M45)

Table S 3: Heat induced unfolding

Construct	Denaturation temperature
Sso MCM	$\geq$ 85℃
Sso MCMΔWH	$\geq$ 85℃
Sso MCM WH	n. o.[a]
Mth MCM WH	80℃

[a] not observable in the temperature range up to 90℃

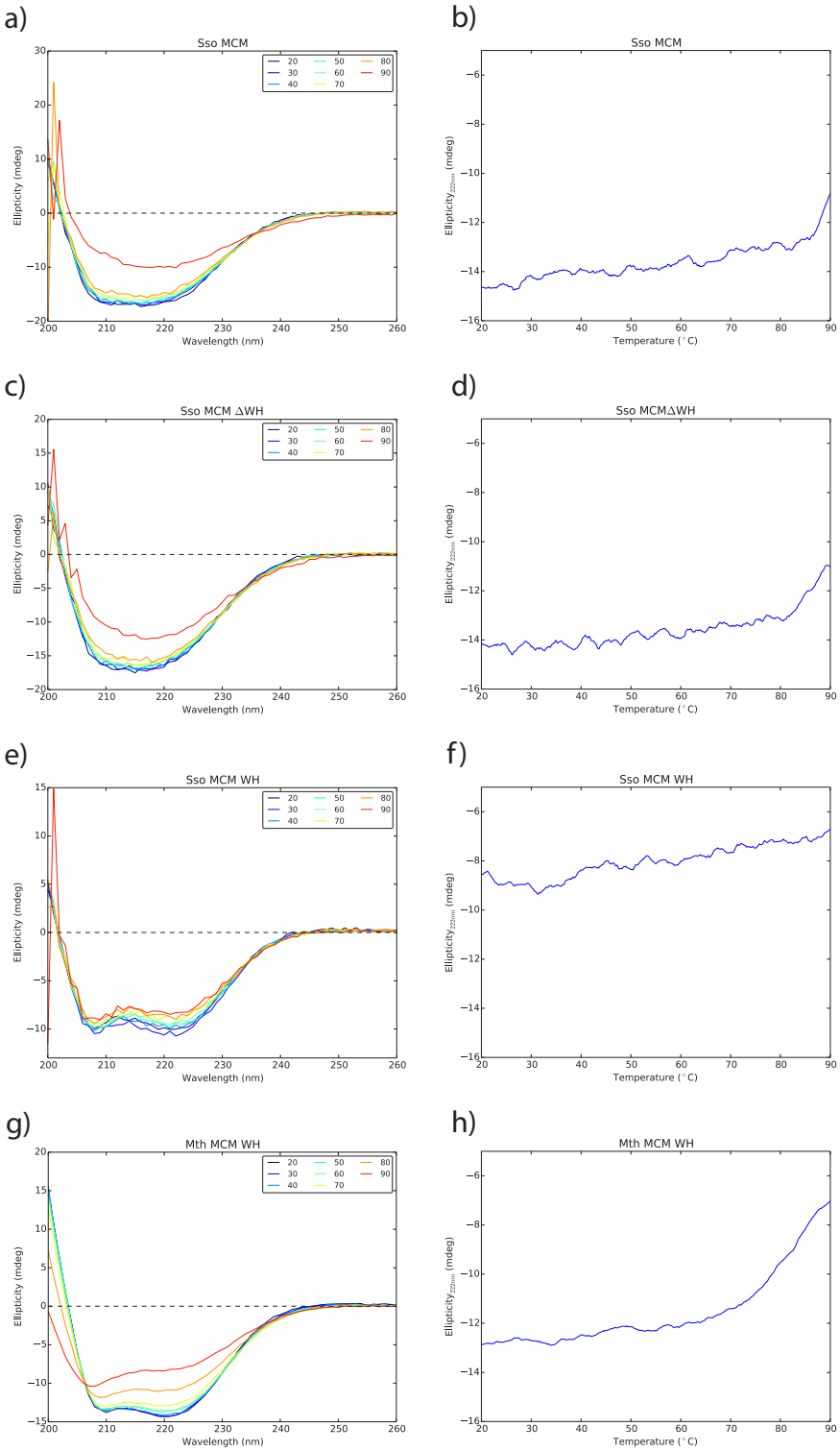

Fig. S 2: Heat denaturation as observed by CD spectroscopy of *Sso* MCM (a, b), *Sso* MCMΔWH (c, d), the *Sso* MCM C-terminal domain (WH) (e, f) and the *Mth* MCM C-terminal domain (WH) (g, h), respectively. At 20, 30, 40, 50, 60, 70, 80 and 90°C CD spectra were collected between 200 nm and 260 nm (left column). Each CD spectrum represents the average of 5 accumulated scans at 50 nm/min with a 1 nm slit width and a time constant of 1 s for a nominal resolution of 0.83 nm. Ellipticity was observed at 222 nm at temperatures ranging between 20–90°C (right column). Data points were collected in 0.2°C steps with a response time of 1 sec and a bandwidth of 1 nm. Temperature slope was 1°C/ min. Denaturation temperature was calculated with Spectra Manager™ software (JASCO).

Thermostability of *Sso* MCM C-terminal domain assessed by [^{1}H,^{15}N]-HSQC spectra

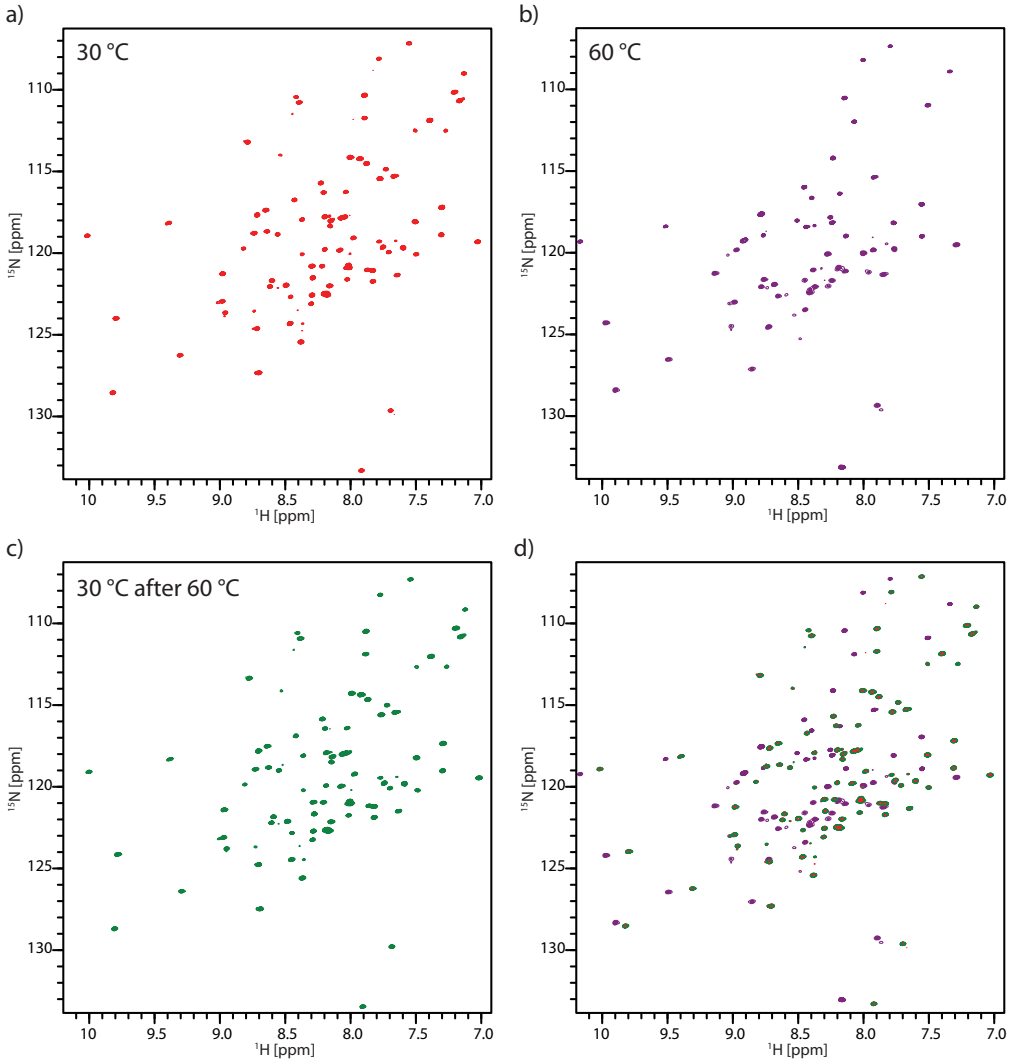

Fig. S 3: Temperature sensitivity of the *Sso* MCM C-terminal domain. [^{1}H,^{15}N]-HSQC spectra of an *Sso* MCM C-terminal domain sample (250 μM) were recorded at 600 MHz with spectral widths of 6893 Hz sampled over 1024 complex points in the ω_2 (^{1}H) dimension, and 1944 Hz over 256 complex points in the ω_1 (^{15}N) dimension with 4 scans per increment in the indirect dimension. The ^{1}H and ^{15}N radio frequency carriers were set at 4.69 ppm and 116 ppm, respectively. Recycle time was 1 s. a) and b) [^{1}H,^{15}N]-HSQC spectra of *Sso* MCM C-terminal domain recorded at 30°C and 60°C. c) [^{1}H,^{15}N]-HSQC spectra of *Sso* MCM C-terminal domain recorded at 30°C after cooling down from 60°C to 30°C. d) Superpostion of the spectra shown in a-c.
The observed [^{1}H,^{15}N]-HSQC fingerprint spectra resulted in a highly similar signal pattern for both temperatures suggesting that the overall fold is not affected at higher temperatures.

Light scattering

To gain access to hydrodynamic parameters of *Sso* MCM C-terminal domain we performed dynamic and static light scattering analysis. The hydrodynamic radius derived from dynamic light scattering is 1.55±0.12 nm (Supplementary Figure S4 and Table S4). This value is in line with the calculated one (R_h: 1.56–1.72 nm) as obtained for globular spherical proteins on the basis of their molecular weight by applying the Stokes' law [7]. The molecular weight estimated from R_h (9.3 kD) indicates that the purified, recombinant C-terminal domain of *Sso* MCM is monomeric in solution. Additionally, results obtained from static light scattering analysis showed

that the *Sso* MCM C-terminal domain has a molecular mass of about 10.5 kD, which is consistent with the dynamic light scattering results and a calculated molecular weight of 9.5 kD (Supplementary Figure S5).

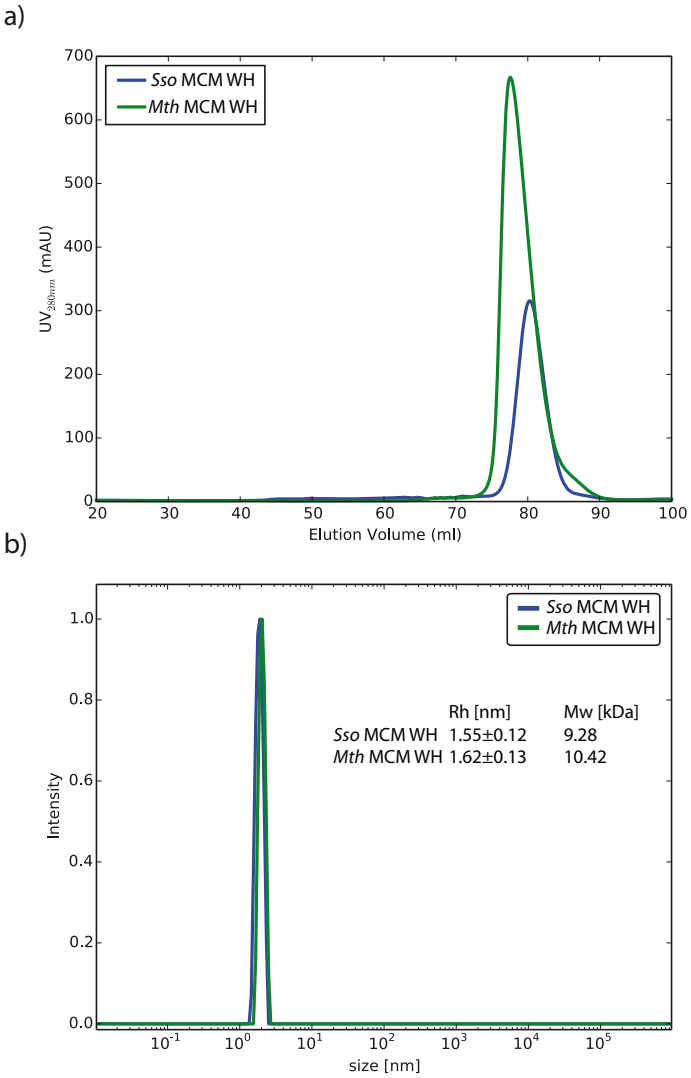

Fig. S 4: Size exclusion chromatography (a) and dynamic light scattering (b) of purified *Sso* (blue) and *Mth* MCM C-terminal domain (WH) (green), respectively.
a) Size exclusion chromatography was performed on a HighLoad 16/60 SD 75 column (GE Healthcare) equilibrated in 10 mM Na-Phosphate pH 6.2 and 150 mM NaCl. Flow rate was 1.5 ml/min.
b) Dynamic light scattering (DLS) on purified *Mth* and *Sso* MCM WH was performed on a Viscotek 802 DLS equipped with a 50 mW fiber coupled 830 nm diode laser. Fifty transients of 5 s each were recoreded at 20°C in a 50 μl volume. Protein concentrations were in the range of 50 μM in 10 mM Na-Phosphate, 150 mM NaCl, pH 6.2. Mass weighted distribution of hydrodynamic radii and a model of globular proteins were utilized for converting hydrodynamic radii to molecular weights. Mass weighted scattering intensity distribution from dynamic light scattering is plotted. OmniSIZE 3.0 software (Viscotek) was used for acquisition and analysis of DLS data.

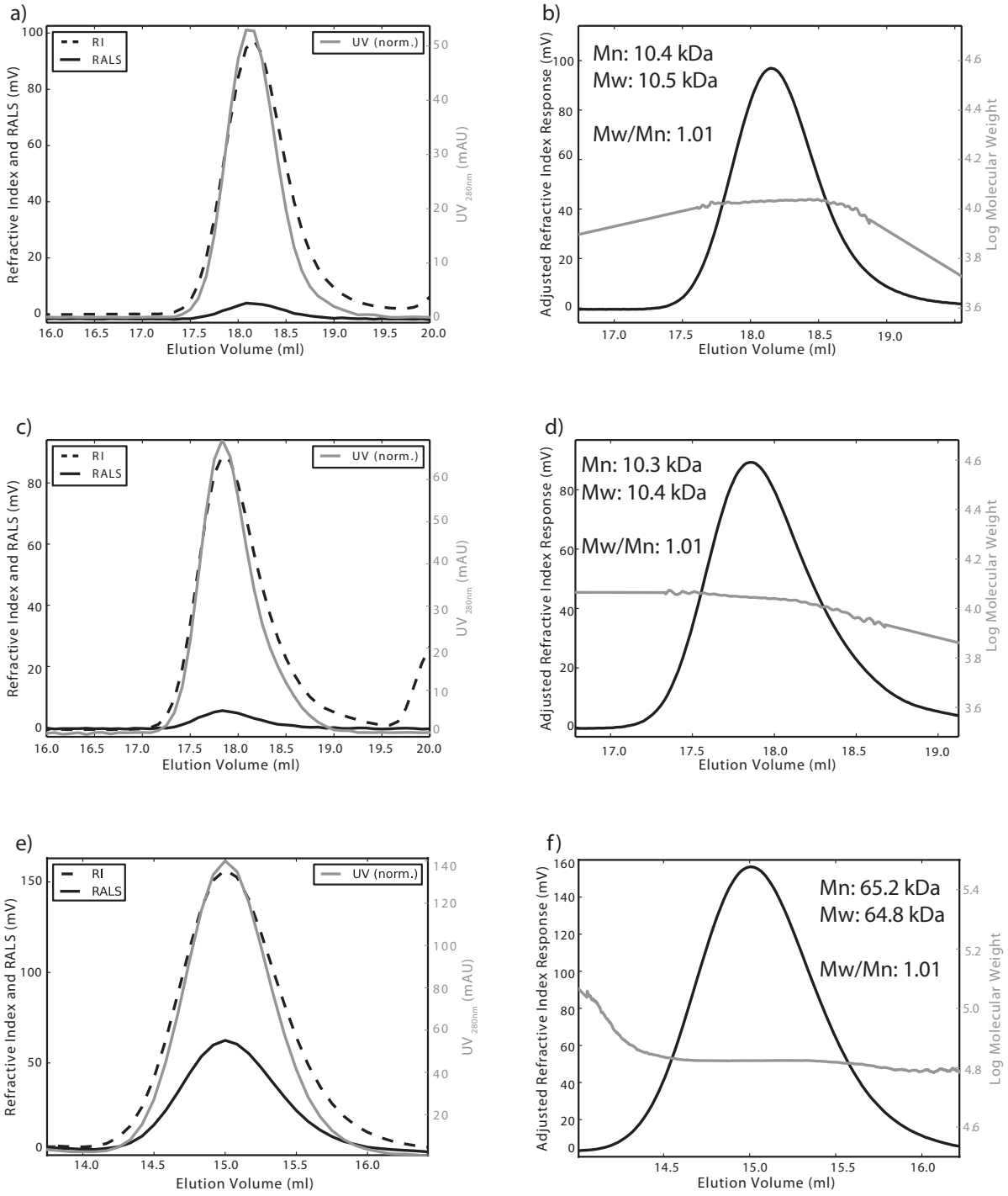

Fig. S 5: Static light scattering of *Sso* (a, b) and *Mth* (c, d) MCM C-terminal WH domain. Static light scattering (SLS) on purified *Mth* and *Sso* MCM WH were performed using a Superdex 200 10/300 GL column (GE Healthcare), equilibrated with the same buffer as for DLS, and 270 Dual detectors (Viscotek) coupled to an Äkta Explorer system (GE Healthcare). 100 μl of purified *Mth* and *Sso* MCM WH were applied to the system with a flow rate of 0.5 ml/min. Protein concentration was in the range of 1 mg/ml. Bovine serum albumin standard (Thermo scientific) was used for calibration and the protein concentration was 2 mg/ml in a 0.9% aquaeous NaCl solution. The resulting data were evaluated with the OmniSEC software. Bovine serum albumin (e, f) was used for calibration. Refractive index (RI), right angle light scattering (RALS) and UV signal at 280 nm is shown on in a, c, and e. Adjusted refractive index and calculated molecular weight of the eluted protein are shown on in b, d, and f.

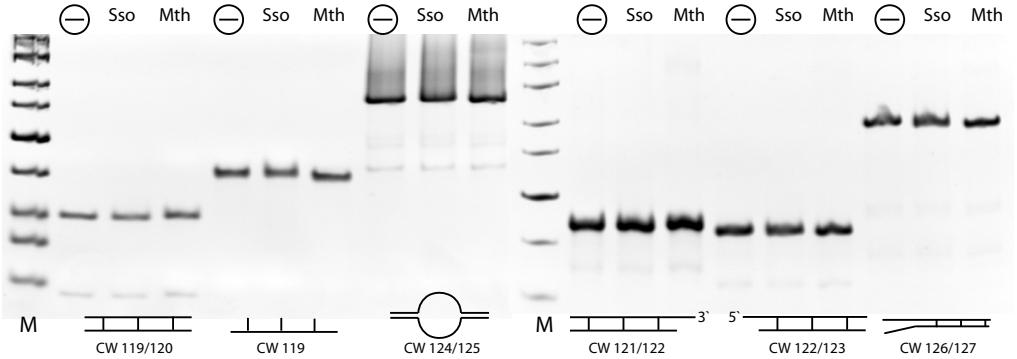

Fig. S 6: Electrophoretic mobility shift assay. DNA (10 μM) of different topologies as indicated below the gels was incubated at room temperature for 30 min with a 100-fold molar excess of isolated archaeal MCM C-terminal domains of *Sso* or *Mth*, respectively. Samples were loaded onto a native 12.5% polyacrylamide gel. Gels were run at 20 mA for 2 h in 40 mM Tris-acetat, pH 7.2 at 4°C. Different incubation temperatures, buffer compositions and gel running temperatures were also tested with comparable outcome (not shown).

NMR relaxation parameters

{^{1}H,^{15}N} relaxation parameters of selectively ^{15}N-labeled *Sso* and *Mth* MCM C-terminal domains were acquired using two-dimensional, ^{1}H-detected heteronuclear NMR experiments by recording inversion recovery [^{1}H,^{15}N]-HSQC for T_1 measurements, CPMG [8] spin echo [^{1}H,^{15}N]-HSQC for T_2 measurements and heteronuclear steady-state NOE spectra, respectively. T_1 values were measured in a series of spectra with relaxation delays of 0.003, 0.05, 0.1, 0.15, 0.2, 0.35, 0.5, 0.7, 1.2, 2 and 4 s. T_2 measurements were taken with relaxation delays of 0.016, 0.033, 0.048, 0.065, 0.081, 0.114, 0.146 and 0.179 s. The relaxation delay for T_1 and T_2 measurements were 5 s and 3 s, respectively. To allow NOE evolution, ^{1}H-^{15}N steady-state NOE values were measured with two different data sets, one collected without initial proton presaturation and a second with an initial proton presaturation period of 3 s. Protein concentrations were in the 1 mM range.

Estimates of the rotational correlation time, τ_c, of *Sso* and *Mth* MCM C-terminal domains were obtained from the ratios of ^{15}N T_1 and T_2 values, using a modified equation as reported in [9]. The Debye-Stokes-Einstein equation was used to calculate the hydrodynamic radius from τ_c [10]. The hydrodynamic properties of *Sso* and *Mth* MCM C-terminal domains were also predicted from their NMR-derived structure ensembles using the software HYDROPRO [11].

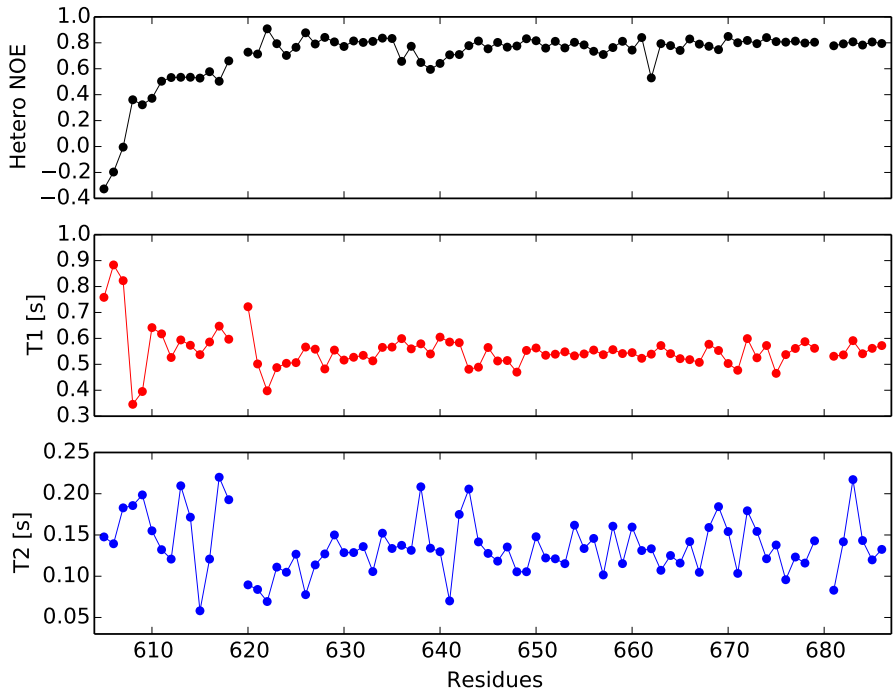

Fig. S 7: {^{1}H,^{15}N} Hetero-NOE, relaxation times T_1 and T_2 for the *Sso* MCM C-terminal domain at 17.6 T plotted as a function of the residue number. T_1 and T_2 spectra were recorded with spectral widths of 8710 Hz sampled over 1024 complex points in the ω_2 (^{1}H) dimension, and 2280 Hz over 94 complex points in the ω_1 (^{15}N) dimension with 8 scans for each increment in the indirect dimension. Heteronuclear steady-state NOE spectra were acquired with a spectral width of 8710 Hz over 1024 complex points in the ω_2 (^{1}H) dimension and 2280 Hz over 68 complex points in the ω_1 (^{15}N) dimension with 32 scans for each increment in the indirect dimension. The ^{1}H and ^{15}N radio frequency carriers were set at 4.69 ppm and 116 ppm, respectively. Temperature was at 30℃.

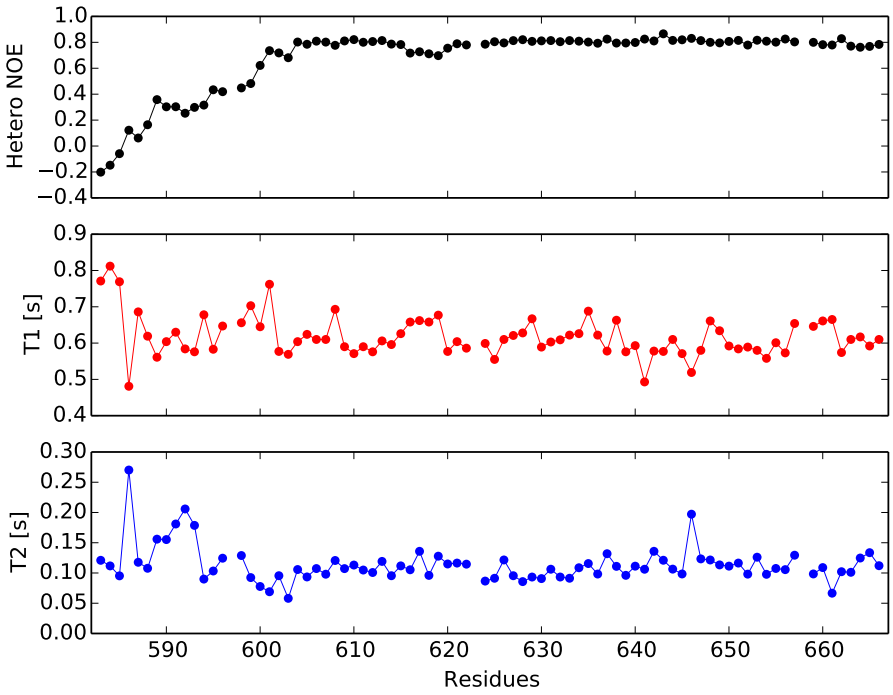

Fig. S 8: $\{^1H,^{15}N\}$ Hetero-NOE, relaxation times T_1 and T_2 for the *Mth* MCM C-terminal domain at 17.6 T plotted as a function of the residue number. T_1 and T_2 spectra were recorded with spectral widths of 7653 Hz sampled over 1024 complex points in the ω_2 (^1H) dimension, and 2204 Hz over 128 complex points in the ω_1 (^{15}N) dimension with 4 scans for each increment in the indirect dimension. Heteronuclear steady-state NOE spectra were acquired with a spectral width of 7653 Hz over 1024 complex points in the ω_2 (^1H) dimension and 2204 Hz over 128 complex points in the ω_1 (^{15}N) dimension with 8 scans for each increment in the indirect dimension. The ^1H and ^{15}N radio frequency carriers were set at 4.69 ppm and 119 ppm, respectively. Temperature was at 30°C.

Table S 4: Comparison of hydrodynamic parameters of *Mth* and *Sso* MCM C-terminal domains

	Mth				*Sso*			
	DLS[a]	SLS[b]	HYDROPRO[c]	NMR[d]	DLS	SLS	HYDROPRO	NMR
τ_c[e][ns]	4.4±1.1	–	7.7±1.04	5.2±1.4	3.9±0.9	–	5.2±0.4	4.4±1.2
R_h[f][nm]	1.62±0.13	–	2.04±0.1	1.73±0.15	1.55±0.12	–	1.79±0.04	1.64±0.15
M_W[g][kD]	10.5	10.4	17.6–22.9	9.6–13.8	9.3	10.5	10.9–14.9	7.9–11.3
M_W[h][kD]			10.1				9.5	

[a] Dynamic light scattering
[b] Static light scattering
[c] calculation with HYDROPRO software [11] from NMR-derived structure ensembles, the average value is indicated
[d] τ_c was obtained from the T_1/T_2 ratio [9] measured on a 750 MHz spectrometer at 30°C for *Mth* and *Sso* MCM-WH
[e] rotational correlation time
[f] hydrodynamic radius
[g] for molecular weight calculation a hydration radius (0.16–0.32 nm) was subtracted from the protein hydrodynamic radius [7]
[h] theoretical molecular weight based on amino acid composition

For DLS, SLS, HYDROPRO and NMR methods, τ_c, R_h and M_W, respectively, are mean values ± standard deviations. The errors are obtained by applying the error propagation law. The underlined values indicate those determined from the experimental data. The nonunderlined values are determined using the Stokes-Einstein and the Debye-Stokes-Einstein equation [10].

Table S 5: NMR and refinement statistics for the *Mth* MCM C-terminal domain

Total distance restraints[a]	2494
NOE based distance restraints	
intra-residual (|i-j|=0)	291
sequential (|i-j| = 1)	652
medium-range (2≤|i-j|≤4)	812
long-range (|i-j|≥5)	701
Restrained H-bonds	38
Dihedral angle restraint	182
Constraint violations	
Distance constraints (Å)	0.007±0.001
Max. distance constraint violation (Å)	0.20±0.08
Dihedral angle constraints (°)	0.21±0.14
Max. dihedral angle violation (°)	3.22±2.25
CYANA target function ($Å^2$)	0.84±0.23
AMBER energies (kcal mol^{-1})	-1437.1±62.3
Deviations from idealized geometry	
Bond lengths (Å)	0.005
Bond angles (°)	1.450
mean global RMSD (Å)	
Total[b](backbone)	6.37±1.95
Total[b](heavy atoms)	6.72±1.87
Ordered[c](backbone)	0.92±0.28
Ordered[c](heavy atoms)	1.82±0.31
Ramachandran statistics (%)[d]	
Most favored	85
Additionally allowed	14
Generously allowed	1
Disallowed	0

[a] upper & lower distance restraints
[b] *Mth* MCM WH: residues E583-V666
[c] *Mth* MCM WH: residues K598-V666
[d] PROCHECK classification [12]
Expression vector derived residues were not included in the structural statistics.

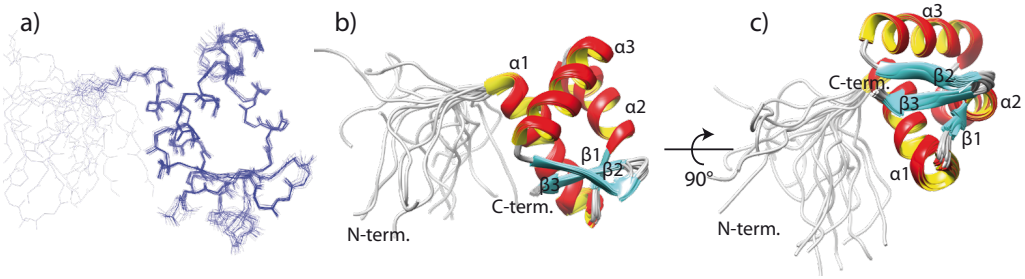

Fig. S 9: NMR solution structure of the *M. thermautotrophicus* MCM C-terminal domain. a) Superimposition of backbone traces for the 20 structures with the lowest energy after OPAL refinement. b and c) Ribbon representation of the structural ensemble with labeled secondary structure elements. Here, the less ordered residues 579–591 of the *M. thermautotrophicus* MCM C-terminal WH domain are omitted for clarity. The *Mth* MCM C-terminal domain was expressed and purified as reported [13]. The ^{1}H, ^{13}C and ^{15}N backbone and side chain chemical shift assignments for the *M. thermautotrophicus* MCM C- terminal WH domain are available from the BioMagResBank (accession numbers: 19187) [13]. The structural coordinates of the *M. thermautotrophicus* MCM C-terminal domain have been deposited to protein data bank (PDB: 2MA3). α1-3: α helix 1-3; β1-3: β strand 1-3

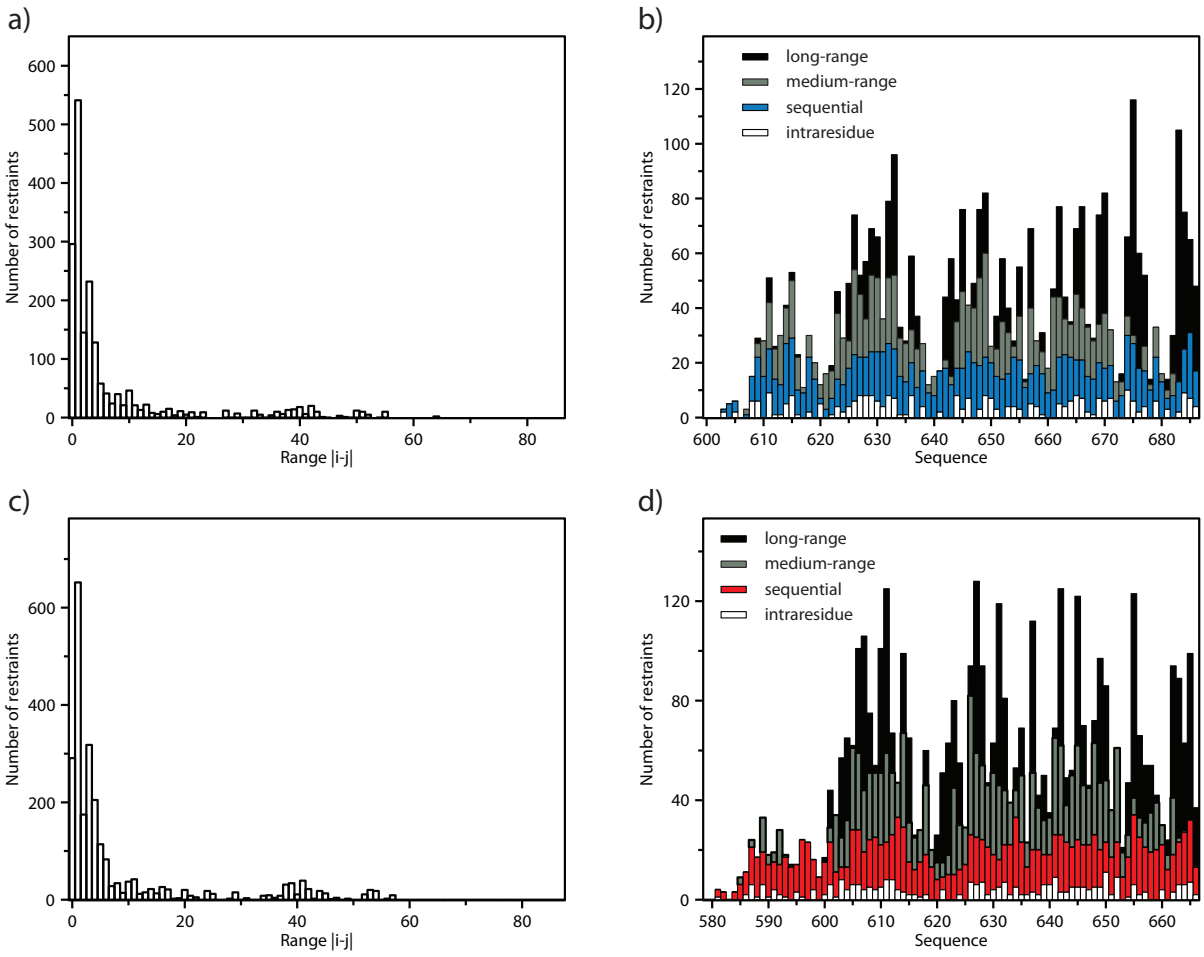

Fig. S 10: Distribution and number of NOEs of the *Sso* (a, b) and *Mth* (c, d) MCM C-terminal domain. a) and c) The panels show the number of NOEs as a function of residue |i-j| range. NOE constraints were classified into 4 range categories according to the sequential distance of involved residues (intraresidue k=i, sequential k=i±1, medium-range k=i±2-4, long-range k=i±n n>4). b) and d) The number of NOE constraints used in the structure calculation as a function of residue number.

Table S 6: 20 top ranked homologues of *Sso* MCM C-terminal domain[a] returned by the DALI web server [14]

No	PDB[b]	Z-score[c]	rmsd[d]	lali[e]	nres[f]	%id[g]	Description
1	2m45-A	13.8	0.5	67	87	100	Minichromosome maintenance protein MCM
2	2ma3-A	10.9	1.1	64	88	20	Minichromosome maintenance protein MCM
3	4o6j-A	10.4	2.0	62	213	13	Iron-dependent transcription repressor related protein
4	2jt1-A	9.9	1.7	63	71	19	PefI protein
5	1dpu-A	9.9	2.3	64	69	14	Replication protein a (rpa32) c-terminal domain
6	4chu-B	9.9	1.7	63	127	17	HTH-type transcriptional regulator iscr
7	2heo-A	9.8	1.4	58	59	14	Z-DNA binding protein 1
8	4kmf-A	9.8	1.6	60	62	17	INF-inducible and ds-dep. eIF-2 kinase
9	3mwm-A	9.7	2.0	64	127	8	Putative metal uptake regulation protein
10	2xrn-B	9.7	1.5	61	241	7	HTH-type transcriptional regulator ttgv
11	4ija-A	9.6	1.6	59	363	17	XylR protein
12	2x4h-C	9.6	1.8	61	131	16	Hypothetical protein sso2273
13	1lva-A	9.5	1.8	62	258	13	Selenocysteine-specific elongation factor
14	1okr-B	9.5	1.5	61	122	23	Methicillin resistance regulatory protein MecI
15	4lb5-B	9.4	1.5	58	63	24	Protein kinase containing Z-DNA binding domains
16	1ldj-A	8.9	1.8	63	725	13	Cullin homolog 1
17	3dqv-D	8.8	1.8	63	378	11	NEDD8
18	2qww-A	8.8	2.4	60	146	17	Transcriptional regulator, MARR family
19	2mh2-A	8.7	1.9	60	64	18	Homologous-pairing protein 2 homolog
20	2g9w-A	8.6	1.7	62	119	11	Conserved hypothetical protein

[a] the mean structure of the region K620-V686 was submitted
[b] Protein Data Bank entry code and chain identifier
[c] normalized score
[d] root-mean-square deviation of C^{α} atoms in the least-squares superimposition of the structurally equivalent C^{α} atoms
[e] number of structurally equivalent residues
[f] total number of amino acids in the hit protein
[g] percentage of identical amino acids over structurally equivalent residues

Table S 7: 20 top ranked homologues of *Mth* MCM C-terminal domain[a] returned by the DALI web server [14]

No	PDB[b]	Z-score[c]	rmsd[d]	lali[e]	nres[f]	%id[g]	Description
1	2ma3-A	14.4	0.4	68	88	100	Minichromosome maintenance protein MCM
2	2m45-A	10.5	2.7	67	87	19	Minichromosome maintenance protein MCM
3	4o6j-A	9.8	2.1	62	213	18	Iron-dependent transcription repressor related protein
4	2jt1-A	9.4	2.3	64	71	13	PefI protein
5	1dpu-A	9.2	1.8	61	69	16	Replication protein A (rpa32) C-terminal domain
6	4ija-A	9.2	1.6	57	363	16	XylR protein
7	1hst-A	9.1	2.4	64	74	14	Histone H5
8	3cta-A	9.0	1.7	60	187	23	Riboflavin kinase
9	2mh2-A	8.9	1.8	61	64	11	Homologous-pairing protein 2 homolog
10	2heo-A	8.8	1.7	58	59	16	Z-DNA binding protein 1
11	1tbx-A	8.8	1.9	63	94	14	SSV1 F-93
12	4chu-B	8.8	3.0	63	127	13	HTH-type transcriptional regulator ISCR
13	1ldj-A	8.7	1.7	61	725	11	Cullin homolog 1
14	1p6r-A	8.7	1.5	61	82	18	Penicillinase repressor
15	3mwm-A	8.7	2.1	62	127	13	Putative metal uptake regulation protein
16	2xrn-B	8.6	2.4	60	241	8	HTH-type transcriptional regulator ttgv
17	4kmf-A	8.5	1.8	57	62	9	INF-inducible and ds-dep. eIF-2 kinase
18	3bz6-A	8.4	2.1	66	168	14	UPF0502 protein PSPTO_2686
19	1ust-A	8.4	2.3	64	92	14	Histone H1
20	3eyy-A	8.3	1.8	61	133	15	Putative iron uptake regulatory protein

[a] the mean structure of the region K598-V666 was submitted
[b] Protein Data Bank entry code and chain identifier
[c] normalized score
[d] root-mean-square deviation of C^{α} atoms in the least-squares superimposition of the structurally equivalent C^{α} atoms
[e] number of structurally equivalent residues
[f] total number of amino acids in the hit protein
[g] percentage of identical amino acids over structurally equivalent residues

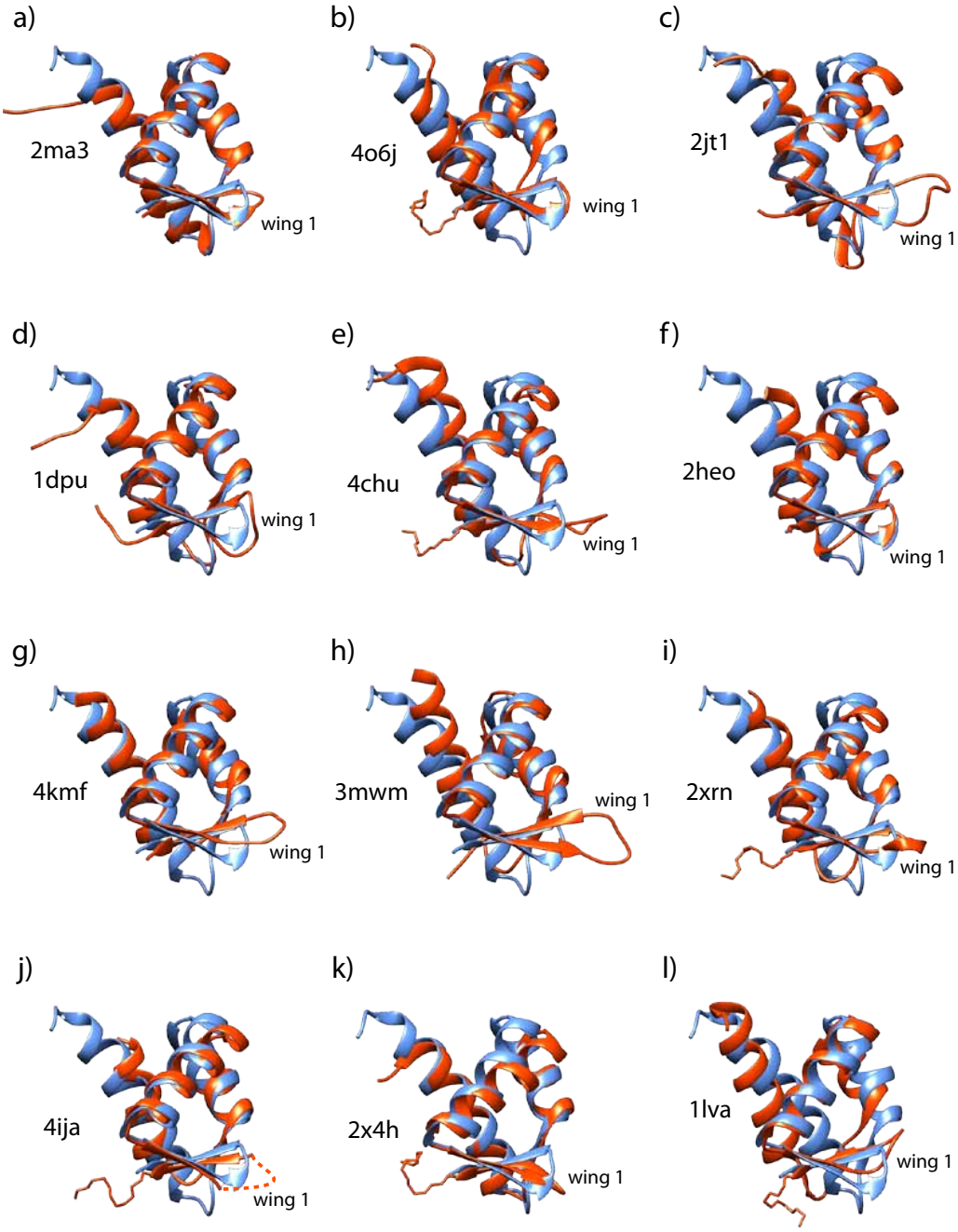

Fig. S 11: Superposition of the *Sso* MCM C-terminal truncated WH domain (blue) with homologous structures (red) found in a DALI search (see Tab. S6). Superpositions for the 12 structures exhibiting the highest Z score are shown.
Note that except for 2ma3 (*Mth* WH) and 2heo (mouse Z-DNA binding protein 1), the shown WH domains have a more extended wing 1 element and/or a C-terminal extension.

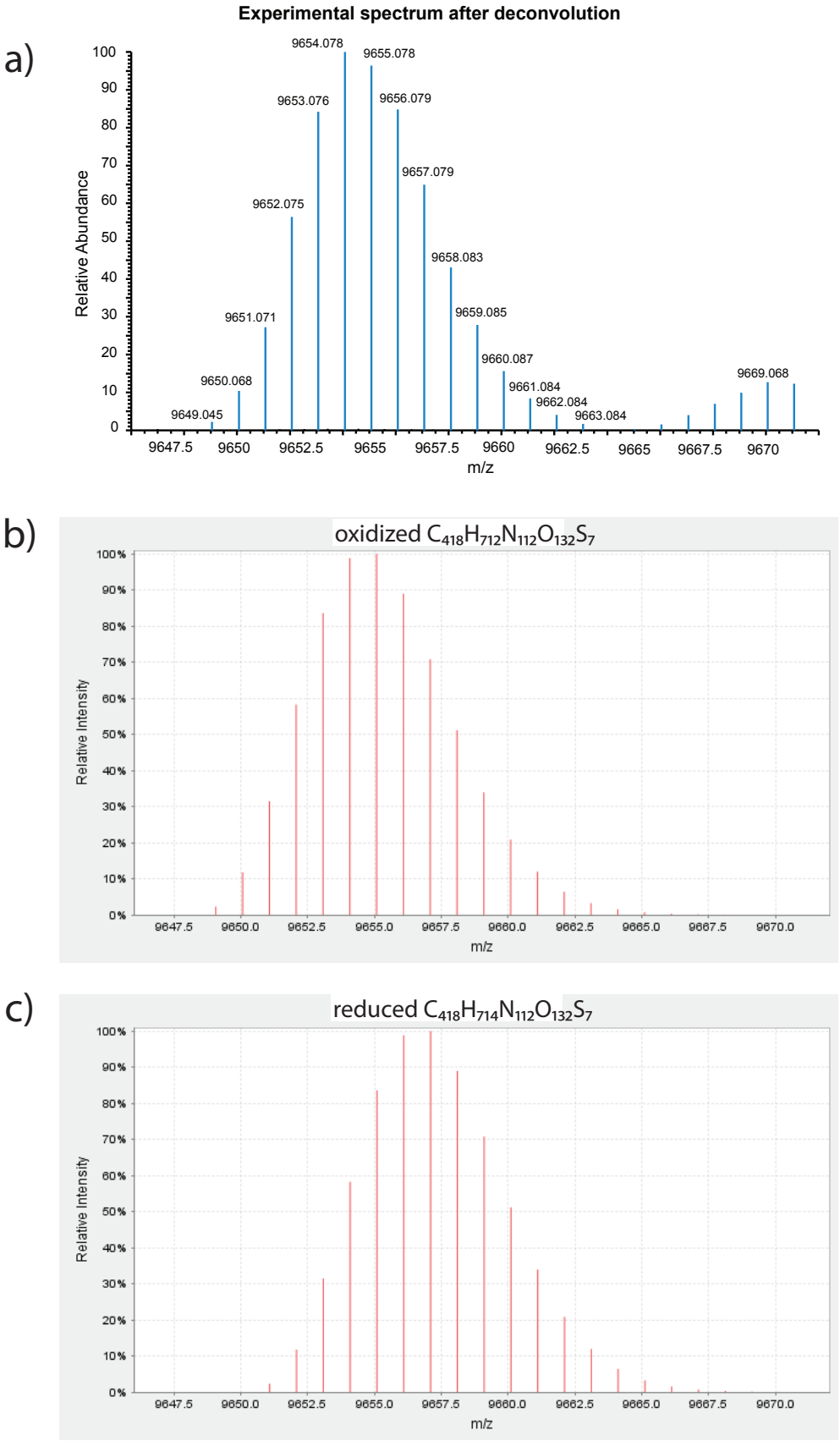

Fig. S 12: Mass spectrometry of the *Sso* MCM C-terminal WH domain. Isotope distribution of the *Sso* MCM C-terminal WH domain as experimentally determined by ESI mass spectrometry is shown in a). The experimentally determined isotope distribution was compared to b) predicted isotope distributions for the WH domain with disulfide bridge (oxidised) and c) with free cysteine residues (reduced), respectively. The predicted distributions were calculated using the freeware program "Isotope Pattern Calculator" (Pacific Northwest National Laboratory, http://omics.pnl.gov/software/ScalaBLAST.php).

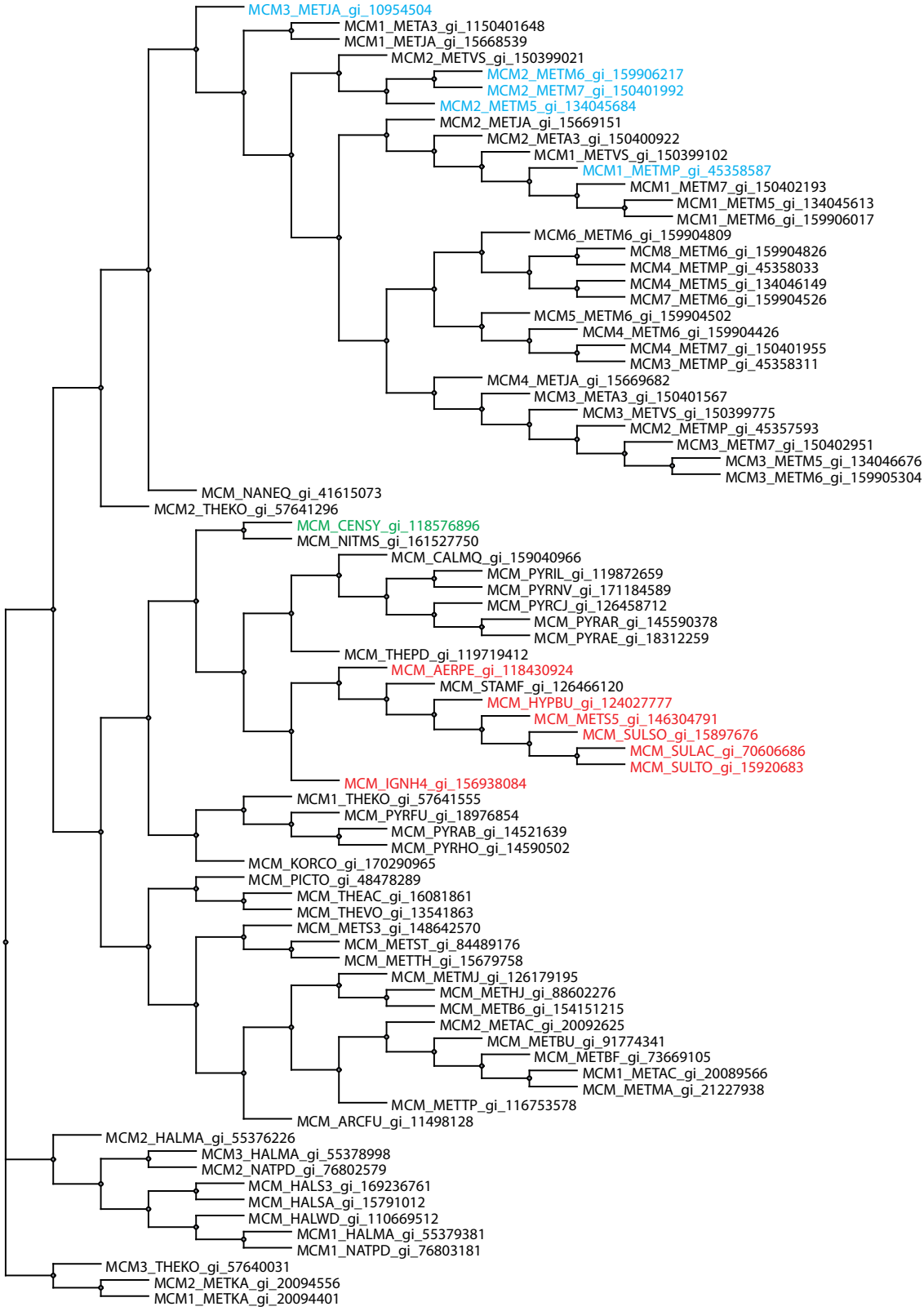

Fig. S 13: Phylogentic tree of of archaeal MCMs. The selection of species was reported [15]. The analysis was performed with MUSCLE [16, 17] under predefined settings on the EMBL-EBI web server [18]. Organism identification is based on the mnemonic UniProtKB entry code. For further information the NCBI protein sequence accession number (gi) is given. Protein sequences were cured by removing known intein sequences [19] where applicable. MCMs with two C-terminal cysteines are highlighted in red, those carrying a single cysteine ∼20-30 residues proximal of the C-terminus in cyan, and the MCM with one cysteine at the very C-terminus in green, respectively.

Fig. S 14: Amino acid sequence alignment of archaeal MCM C-terminal domains for the species shown in Figure S12. The analysis was performed with MUSCLE [16, 17] using the default settings of the EMBL-EBI web server [18] and manually refined using Unipro UGENE [20] . Identification of organisms is based on the mnemonic UniProtKB entry code and the NCBI protein sequence accession number (gi) is given. Cysteine residues are highlighted in purple. Species with MCM C-terminal domains harbouring two cysteines are boxed in red, in blue for a single cysteine ∼20-30 residues proximal of the C-terminus and green for one cysteine at the very MCM C-terminus, respectively.

ATPase activity of *Sso* MCM derivatives

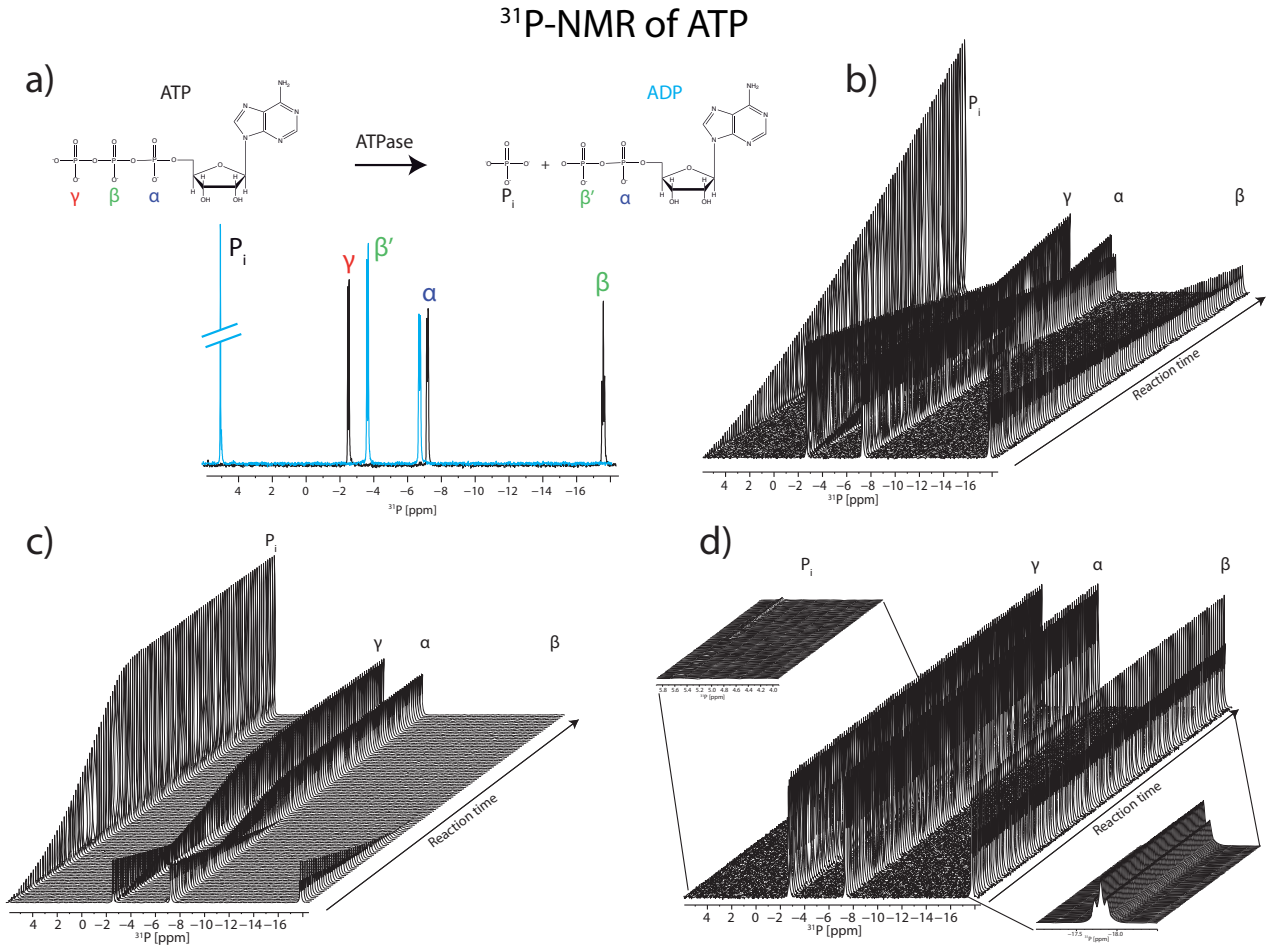

Fig. S 15: a) Hydrolysis of ATP into ADP and free inorganic phosphate. a, top) Schematic reaction scheme of ATP hydrolysis by an ATPase yields ADP and free inorganic phosphate. The α-, β- and γ-phosphates of ATP and the released free inorganic phosphate (P_i) and the α- and β'-phosphates of ADP are indicated. a, bottom) [^{31}P]-NMR spectra of ATP (black) before and after (blue) complete hydrolysis into ADP and free inorganic phosphate. b-d) Time series of [^{31}P]-NMR spectra indicating the progress of ATP hydrolysis catalysed by (b) *Sso* MCM, (c) *Sso* MCMΔWH (E605_V686del, C-terminal domain) and (d) buffered ATP diluted into ATPase reaction buffer (see Material and Methods). Expanded regions in d) to show the time-course for free inorganic phosphate (d, top left) and β-phosphate of ATP (d, bottom right), respectively.

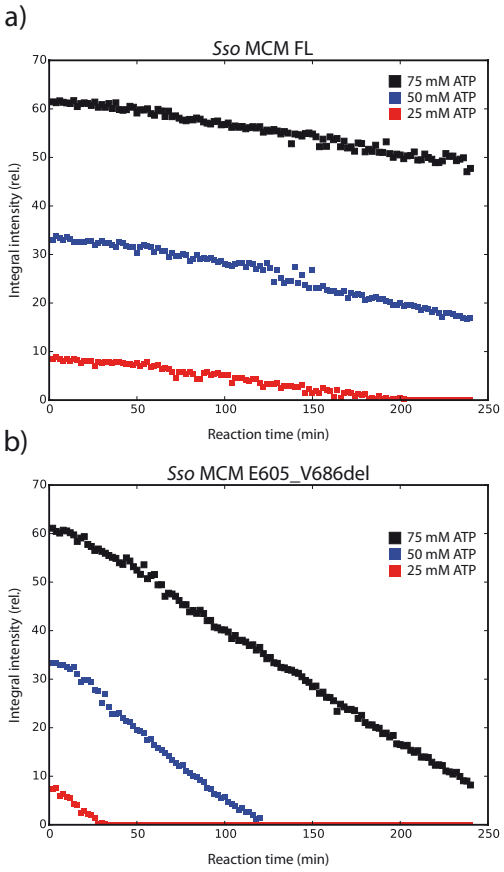

Fig. S 16: ATP hydrolysis by *Sso* MCM constructs under various ATP substrate concentrations. ATP hydrolysis was assayed by [^{31}P]-NMR spectroscopy. Time series of [^{31}P]-NMR spectra were recorded at 60°C. ^{31}P chemical shifts at 60°C were referenced relative to 75 % H_3PO_4 at 0 ppm. Spectra were processed using Bruker software TOPSPIN V2.1. The integrated peak areas of the free phosphate (red dots, 25 mM ATP; blue, 50 mM ATP; black, 75 mM ATP) are plotted as a function of the reaction time.
Increasing the ATP concentration to 75 mM or decreasing it to 25 mM, respectively, did not significantly change the rate of ATP hydrolysis as judged from the slope of the decrease of the signal intensity integral over time: under all three ATP concentrations the rate for a) the full length wild type MCM was 3.4 s^{-1} and 7.2 s^{-1} for b) the MCM construct lacking the C-terminal WH domain, respectively. This indicates that the assays shown in Fig. S15 and S17 were conducted under saturating substrate conditions.

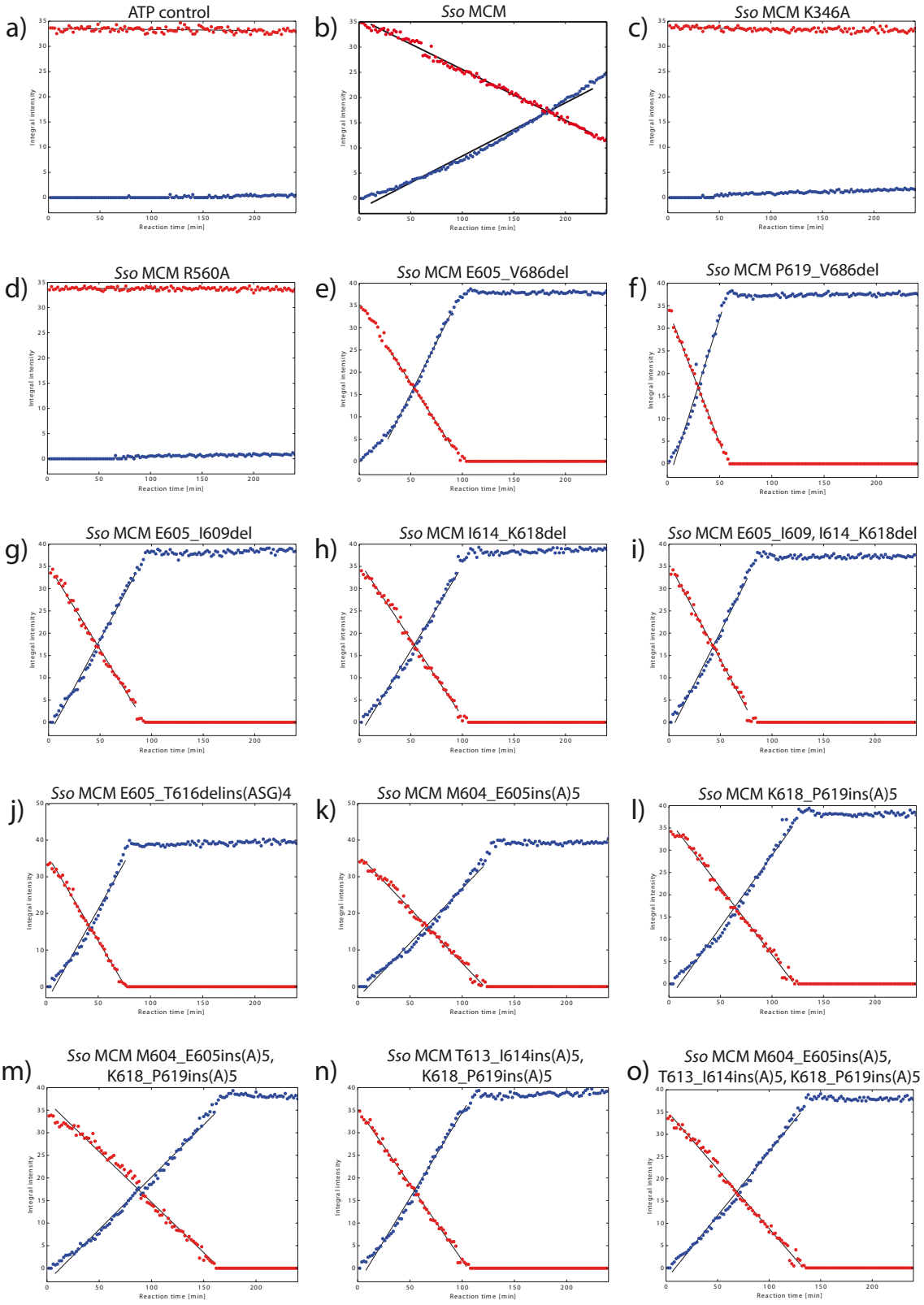

Fig. S 17: ATP hydrolysis by *Sso* MCM constructs. Time series of [^{31}P]-NMR spectra were recorded at 60°C. ^{31}P chemical shifts at 60°C were referenced relative to 75 % H_3PO_4 at 0 ppm. Spectra were processed using Bruker software TOPSPIN V2.1. The integrated peak areas of the free phosphate (red dots) and the β-phosphate of ATP (blue dots) are plotted as a function of the reaction time. Results of the regression analysis for increase of free inorganic phosphate and decrease of β-phosphate of ATP given as black line.

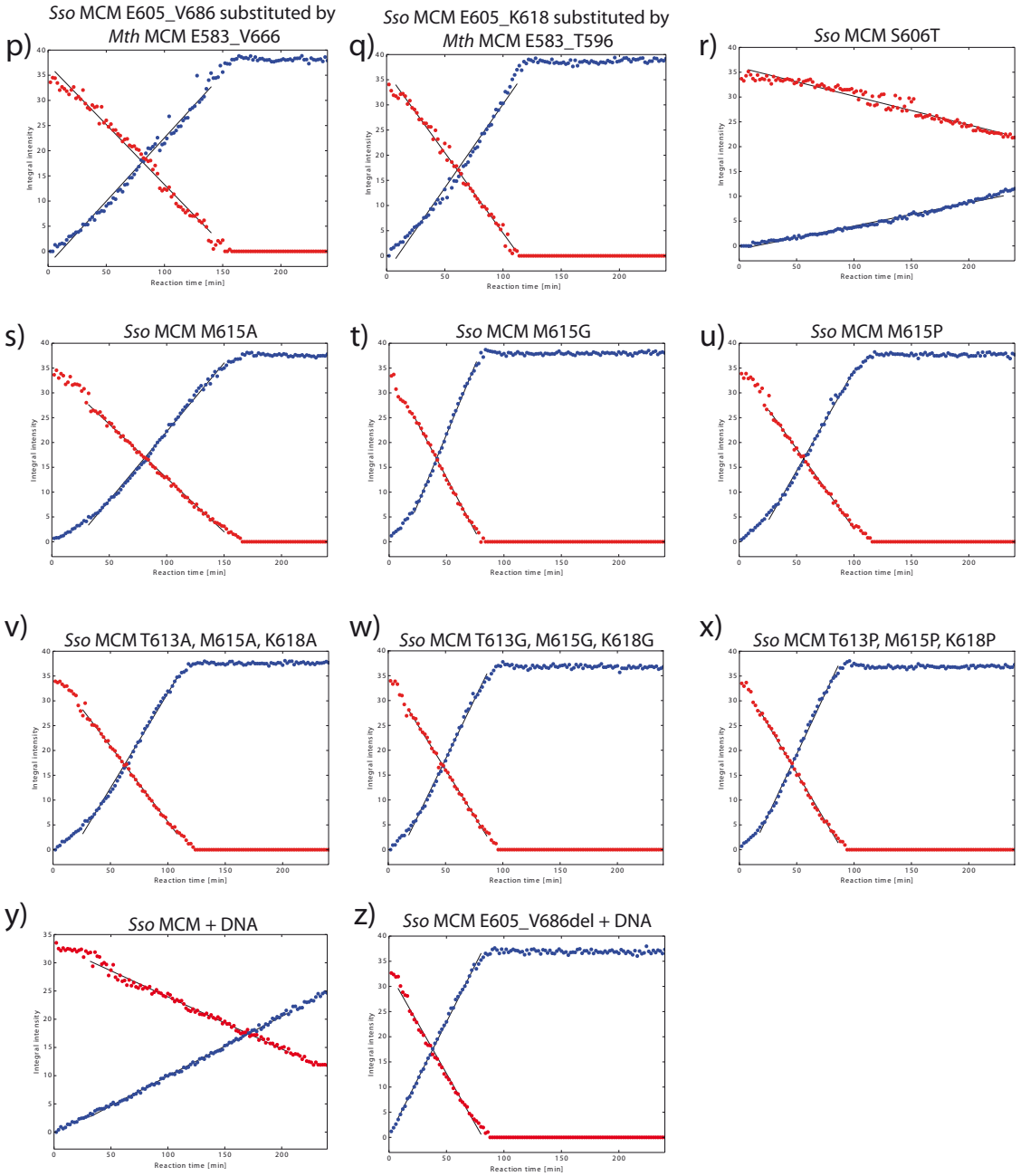

Fig. S 17: ATP hydrolysis by *Sso* MCM constructs continued.

Table S 8: ATP hydrolysis of the *Sso* MCM constructs under saturating substrate concentrations

Sso MCM construct[a]	free phosphate[b] [mM/ min]	ATP β-phosphate[b] [mM/ min]	k_{cat} [min^{-1}]
wild type (full length)	0.11	-0.10	2±0.6[c]
K346A	0.02	-0.01	0.2
R560A	0.01	0.00	0.0
E605_V686del	0.46	-0.36	7.2±0.5[c]
P619_V686del	0.71	-0.56	11.2
E605_I609del	0.42	-0.38	7.6
I614_K618del	0.38	-0.34	6.8
E605_I609del, I614_K618del	0.46	-0.43	8.6
E605_T616delins(ASG)4	0.50	-0.48	9.6
M604_E605ins(A)5	0.29	-0.29	5.8
K618_P619ins(A)5	0.32	-0.30	6.0
M604_E605ins(A)5, K618_P619ins(A)5	0.23	-0.22	4.4
T613_I614ins(A)5, K618_P619ins(A)5	0.37	-0.33	6.6
M604_E605ins(A)5, T613_I614ins(A)5, K618_P619ins(A)5	0.28	-0.27	5.4
substitute E605_V686 by *Mth* MCM E583_V666	0.25	-0.24	4.8
substitute E605_K618 by *Mth* MCM E583_T596	0.33	-0.31	6.1
S606T	0.05	-0.06	1.2
M615A	0.28	-0.22	4.4
M615G	0.56	-0.44	8.8
M615P	0.41	-0.32	6.4
T613A, M615A, K618A	0.38	-0.30	6.0
T613G, M615G, K618G	0.48	-0.37	7.4
T613P, M615P, K618P	0.49	-0.39	7.8
wild type + DNA[d]	0.11	-0.09	1.8
E605_V686del + DNA[d]	0.45	-0.40	8.0

[a] nomenclature for the description of sequence variations as recommended [1]
[b] slope of the regression line
[c] standard deviation from three independent experiments
[d] assay in the presence of 3-fold molar excess of sonicated pUC19-derived DNA (fragment size in the range of 80-160 bp)

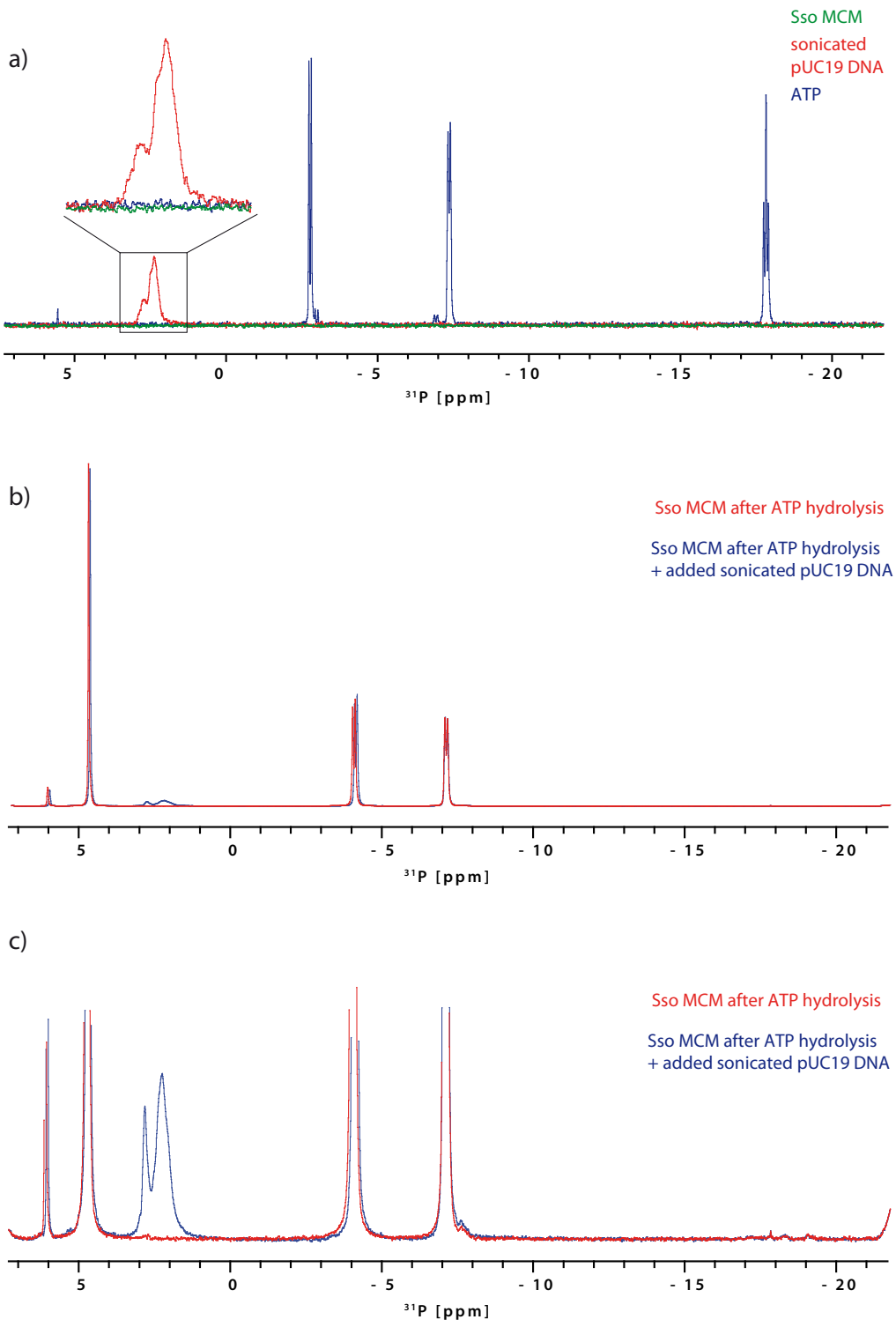

Fig. S 18: a) 1D [^{31}P]-NMR spectra of purified *Sso* MCM (green), sonicated pUC19-derived DNA with a fragment size in the range of 80-160 bp (red) and ATP buffered to pH 7.2 (blue). Spectra were recorded at 600 MHz with a spectral width of 7042 Hz sampled over 1228 complex points in the ω_1 (^{31}P) dimension. For both purified *Sso* MCM (green) and sonicated pUC19 derivate DNA (red) 60000 transients in the ^{31}P-dimension were sampled. The ATP spectra was recorded with 40960 transients. b) [^{31}P]-NMR spectra of *Sso* MCM after complete hydrolysis of ATP into ADP and free inorganic phosphate without DNA (red) and in the presence of DNA (blue). Spectral parameters same as in a) with 30000 transients in the ^{31}P-dimension. c) Zoomed spectral region of b). The ^{31}P radio frequency carrier was set at -10.5 ppm. Recycle time was 1 s and temperature was 60°C. ^{31}P chemical shifts at 60°C were referenced relative to 75 % H_3PO_4 at 0 ppm. Concentrations of *Sso* MCM protein, ATP and sonicated pUC19 derivate DNA were 50 μM, 50 mM and 150 μM, respectively.

DNA helicase activity of *Sso* MCM derivatives

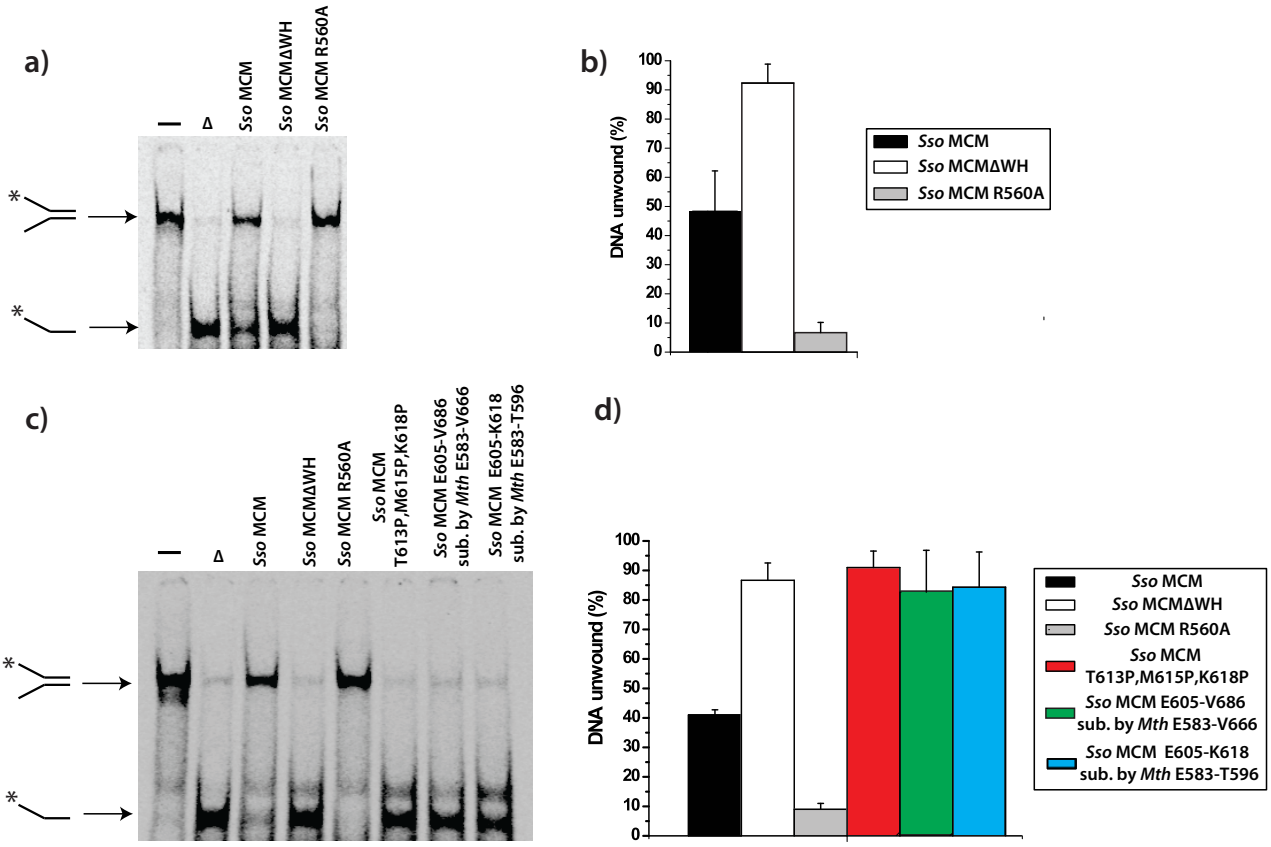

Fig. S 19: a, c) Helicase assays were performed with 2 nM 5′-labeled Y-shaped DNA substrate and 1 μM of indicated *Sso* MCM constructs. After incubation for 30 min at 70°C, the reaction mixtures were cooled to 0°C and 10 μl of helicase loading dye was added. The samples were electrophoresed through a 8% non-denaturing polyacrylamide gels in TBE buffer. Subsequently, gels were dried and exposed to a phosphoimaging screen. Gels were visualized using a phosphoimager (Typhoon Trio; GE Healthcare). Δ refers to boiled DNA substrate (no protein). b, d) The bar graphs represents the percentage of unwound DNA for individual *Sso* MCM mutants as calculated using the Image Quant software. The mean standard error of three independent experiments is indicated.

Docking approach

Paramagnetic relaxation enhancement

Proton relaxation rates (R_1) were obtained from a series of saturation-recovery [^{1}H,^{15}N]-HSQC experiments. Spectra were recorded at 30°C on a Bruker 750 MHz AvanceIII NMR system (5 mm triple resonance cryo-probe) with spectral widths of 8710 Hz sampled over 1024 complex points in the ω_2 (^{1}H) dimension, and 2280 Hz over 256 complex points in the ω_1 (^{15}N) dimension with 8 scans per increment in the indirect dimension. Data sets were acquired with recovery delays of 0.05, 0.1, 0.2, 0.5, 1, 1.5, 2, 4 and 6 s. The ^{1}H and ^{15}N radio frequency carriers were set at 4.69 ppm and 116 ppm, respectively. For determination of ^{1}H solvent PREs, a stock solution of 0.5 M Gd(DTPA-bis(methylamide)) (Omniscan™, GE Healthcare) was added stepwise to a final concentrations of 0, 2, 4, 6, 8 and 10 mM. Spectra were processed using NMRpipe [21] and analyzed with CCPNmr Analysis [22]. Peak intensities were fitted using PREdator [23] to

$$I = I_0(1 - e^{-R_1 t}).$$

I_0 is the intensity after an infinite recovery delay, t is the recovery time and R_1 is the longitudinal relaxation rate. PRE is defined as the increase in relaxation rate upon addition of a paramagnetic co-solvent. Here, PREs are represented by the slope of R_1 as function of the concentration of Gd(DTPA-bis(methylamide). For a detailed description of PRE calculation, see [24–26].

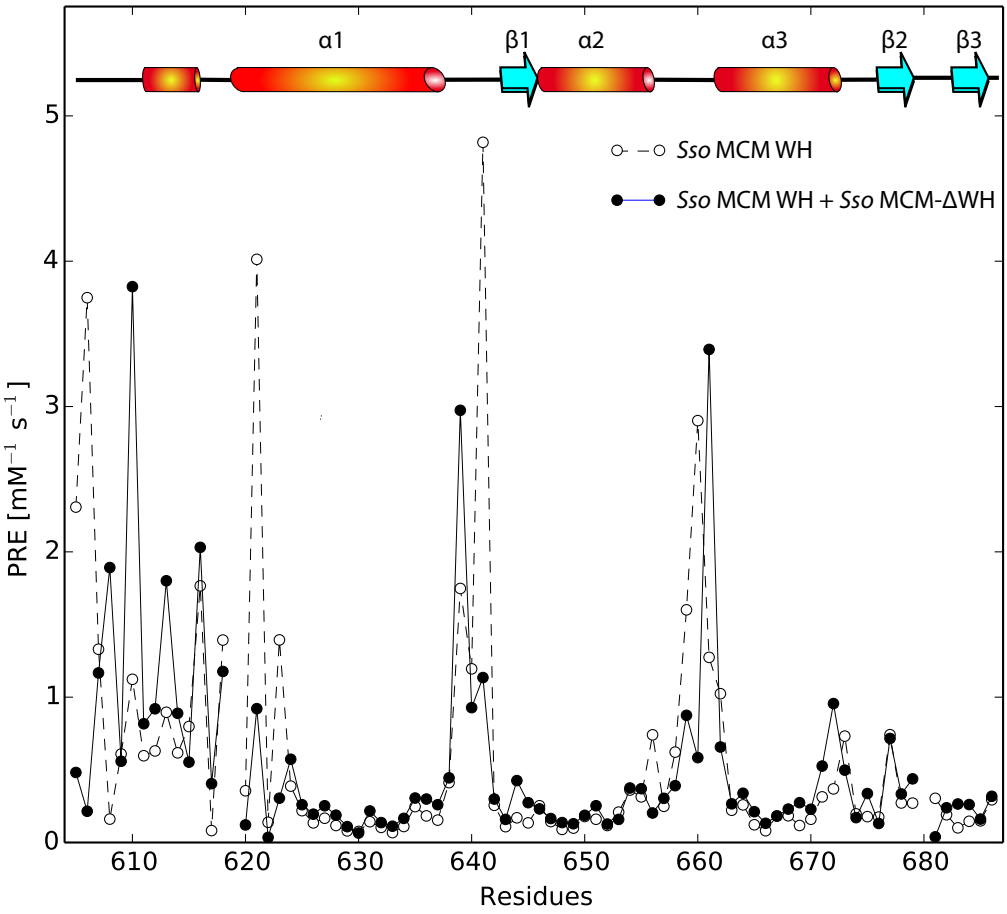

Fig. S 20: Paramagnetic relaxation enhancement (PRE) determined for the backbone amide protons of the *Sso* MCM C-terminal domain. PREs were measured for the *Sso* MCM C-terminal domain (100 μM) in absence (dashed line) and presence (solid line) of a 3-fold molar excess of *Sso* MCMΔWH. Secondary structure elements of the *Sso* MCM C-terminal domain NMR solution structure are indicated.

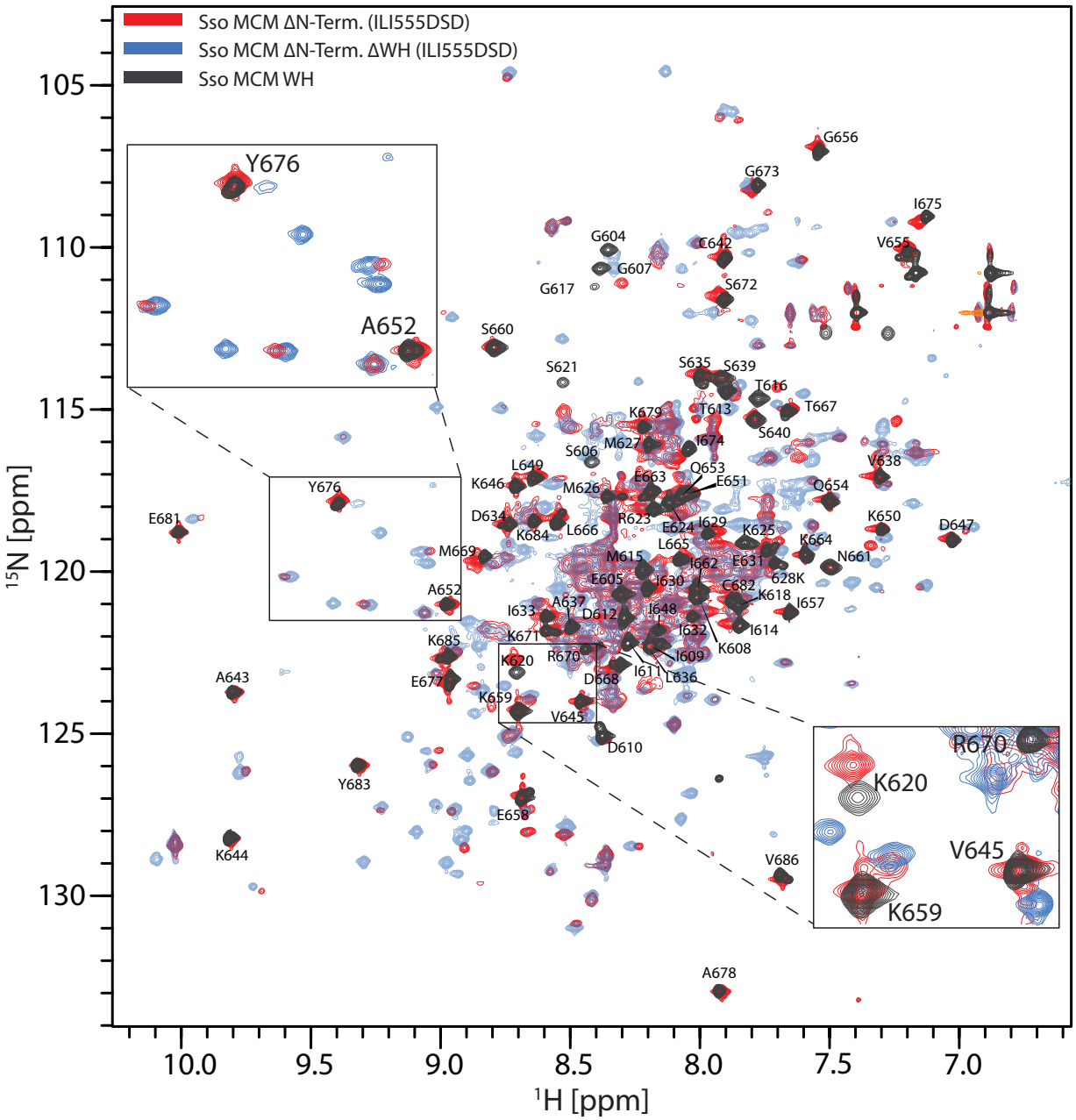

Fig. S 21: Superimposition of [^{1}H,^{15}N]-TROSY-HSQC [27, 28] spectra of the *Sso* MCM C-terminal domain (gray), *Sso* MCMΔN-term (ILI555DSD) (red) and *Sso* MCMΔN-termΔWH (ILI555DSD) (blue). Spectra were recorded at 750 MHz with spectral widths of 9014 Hz sampled over 900 complex points in the ω_2 (^{1}H) dimension, and 2432 Hz over 150 complex points in the ω_1 (^{15}N) dimension with 32 scans for each increment in the indirect dimension. The ^{1}H and ^{15}N radio frequency carriers were at 4.74 ppm and 119.0 ppm, respectively. Recycle time was 2.5 s and temperature was 30°C. Protein concentrations were in the 300 μM range.The assignment for the *Sso* MCM C-terminal domain is given.

Table S 9: Statistics of the HADDOCK docking run

HADDOCK score	$E_{electrostatic}$	$E_{van\text{-}der\text{-}Waals}$	Buried surface area	Cluster size best (vs. total)
0.1 ± 5.4	-621.4 ± 57.0	-26.8 ± 9.8	1762.1 ± 335.4	134 (191)

The program High Ambiguity Driven Protein-Protein Docking (HADDOCK) was used to generate a docked structure of *Sso* MCM including the C-terminal WH domain. Residues which are in close proximity to the potential domain interface were defined as active interacting residues (605, 606, 615, 616, 617, 620, 621, 623, 625, 626). Surrounding residues considered indirectly involved in or affected by the domain interaction were defined as passive interacting residues (673 and 674). Residue G601 of *Sso* MCM (PDB: 3F9V) was defined as active because it represents the most C-terminal residue of the AAA+ domain preceding the most N-terminal residue (V602) of the C-terminal WH domain. No further restrictions and structural assumptions were applied. HADDOCK clustered 191 structures into 9 clusters, which represent 95.5 % of the water-refined models generated by HADDOCK.

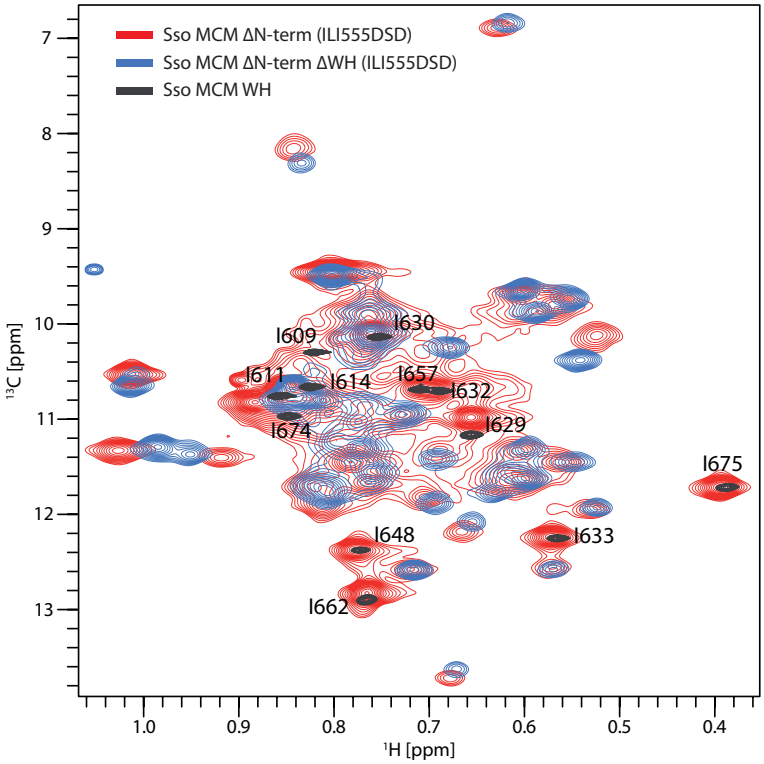

Fig. S 22: Superimposition of [^{1}H,^{13}C]-HSQC spectra of the *Sso* MCM C-terminal WH domain (black), *Sso* MCMΔN-term (ILI555DSD) (red) and *Sso* MCMΔN-termΔWH (ILI555DSD) (blue). In all samples the methyl groups of isoleucines are selectively labelled. For selective labeling of isoleucine methyl groups cells were grown in M9 media prepared in D_2O and supplemented with methyl-^{13}C, 3,3-D_2 α-ketobutyric acid (10 mg/100 ml) and ^{12}C-glucose as additional carbon source. The [^{1}H,^{13}C]-HSQC spectra were recorded at 750 MHz with spectral widths of 6009 Hz sampled over 600 complex points in the ω_2 (^{1}H) dimension, and 1886 Hz over 100 complex points in the ω_1 (^{13}C) dimension with 32 scans for each increment in the indirect dimension. The ^{1}H and ^{13}C radio frequency carriers were at 3.76 ppm and 10 ppm, respectively. Recycle time was 2.5 s. Protein concentrations were in the 300 μM range.

References Supplementary Data

[1] J T den Dunnen and S E Antonarakis. Nomenclature for the description of human sequence variations. *Human genetics*, 109(1):121–124, July 2001.

[2] Mariarita De Felice, Valentina Aria, Luca Esposito, Mariarosaria De Falco, Biagio Pucci, Mosè Rossi, and F Pisani. A novel DNA helicase with strand-annealing activity from the crenarchaeon Sulfolobus solfataricus. *Biochem. J*, 408:87–95, 2007.

[3] Matthew J Moreau, Adam T McGeoch, Alan R Lowe, Laura S Itzhaki, and Stephen D Bell. ATPase Site Architecture and Helicase Mechanism of an Archaeal MCM. *Molecular Cell*, 28(2):304–314, October 2007.

[4] Christoph Wiedemann, Peter Bellstedt, and Matthias Görlach. CAPITO—a web server-based analysis and plotting tool for circular dichroism data. *Bioinformatics (Oxford, England)*, 29(14):1750–1757, 2013.

[5] Narasimha Sreerama and Robert W Woody. Estimation of Protein Secondary Structure from Circular Dichroism Spectra: Comparison of CONTIN, SELCON, and CDSSTR Methods with an Expanded Reference Set. *Analytical biochemistry*, 287(2):252–260, December 2000.

[6] W Kabsch and C Sander. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22(12):2577–2637, December 1983.

[7] R M Venable and R W Pastor. Frictional models for stochastic simulations of proteins. *Biopolymers*, 27(6):1001–1014, June 1988.

[8] S Meiboom and D Gill. Modified Spin-Echo Method for Measuring Nuclear Relaxation Times. *Review of Scientific Instruments*, 29(8):688, 1958.

[9] L E Kay, D A Torchia, and A Bax. Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry*, 28(23):8972–8979, November 1989.

[10] Shenggen Yao, Jeffrey J Babon, and Raymond S Norton. Protein effective rotational correlation times from translational self-diffusion coefficients measured by PFG-NMR. *Biophysical Chemistry*, 136(2-3):145–151, August 2008.

[11] A Ortega, D Amorós, and J García de la Torre. Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. *Biophysical journal*, 101(4):892–898, August 2011.

[12] RomanA Laskowski, J AntoonC Rullmann, MalcolmW MacArthur, Robert Kaptein, and JanetM Thornton. AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *Journal of Biomolecular NMR*, 8(4):477–486, December 1996.

[13] Christoph Wiedemann, Oliver Ohlenschläger, Barbara Medagli, Silvia Onesti, and Matthias Görlach. (1)H, (15)N and (13)C chemical shift assignments for the winged helix domains of two archeal MCM C-termini. *Biomolecular NMR assignments*, 8(2):357–360, October 2014.

[14] Liisa Holm and Päivi Rosenström. Dali server: conservation mapping in 3D. *Nucleic Acids Research*, 38(Web Server issue):W545–9, July 2010.

[15] Nicholas Chia, Isaac Cann, and Gary J Olsen. Evolution of DNA replication protein complexes in eukaryotes and Archaea. *PLoS ONE*, 5(6):e10866, 2010.

[16] Robert C Edgar. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC bioinformatics*, 5:113, August 2004.

[17] R C Edgar. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5):1792–1797, March 2004.

[18] H McWilliam, W Li, M Uludag, S Squizzato, Y M Park, N Buso, A P Cowley, and R Lopez. Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Research*, 41(W1):W597–W600, June 2013.

[19] Francine B Perler. InBase: the intein database. *Nucleic Acids Research*, 30(1):383–384, 2002.

[20] Konstantin Okonechnikov, Olga Golosova, Mikhail Fursov, and UGENE team. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics (Oxford, England)*, 28(8):1166–1167, April 2012.

[21] F Delaglio, S Grzesiek, G W Vuister, G Zhu, J Pfeifer, and A Bax. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR*, 6(3):277–293, November 1995.

[22] Wim F Vranken, Wayne Boucher, Tim J Stevens, Rasmus H Fogh, Anne Pajon, Miguel Llinas, Eldon L Ulrich, John L Markley, John Ionides, and Ernest D Laue. The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins: Structure, Function, and Bioinformatics*, 59(4):687–696, June 2005.

[23] Christoph Wiedemann, Peter Bellstedt, and Matthias Görlach. PREdator: a python based GUI for data analysis, evaluation and fitting. *Source Code for Biology and Medicine*, 9(21):1–4, 2014.

[24] Michal Respondek, Tobias Madl, Christoph Göbl, Regina Golser, and Klaus Zangger. Mapping the Orientation of Helices in Micelle-Bound Peptides by Paramagnetic Relaxation Waves. *Journal of the American Chemical Society*, 129(16):5228–5234, April 2007.

[25] Tobias Madl, Wolfgang Bermel, and Klaus Zangger. Use of Relaxation Enhancements in a Paramagnetic Environment for the Structure Determination of Proteins Using NMR Spectroscopy. *Angewandte Chemie International Edition*, 48(44):8259–8262, September 2009.

[26] Tobias Madl, Thomas Güttler, Dirk Görlich, and Michael Sattler. Structural Analysis of Large Protein Complexes Using Solvent Paramagnetic Relaxation Enhancements. *Angewandte Chemie International Edition*, 50(17):3993–3997, March 2011.

[27] K Pervushin, R Riek, G Wider, and K Wüthrich. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proceedings of the National Academy of Sciences*, 94(23):12366–12371, November 1997.

[28] R Riek, G Wider, K Pervushin, and K Wüthrich. Polarization transfer by cross-correlated relaxation in solution NMR with very large molecules. *Proceedings of the National Academy of Sciences*, 96(9):4918–4923, April 1999.