PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

The solution structure of full-length dodecameric MCM by SANS and molecular modeling

Susan Krueger,[1]* Jae-Ho Shin,[2] Joseph E. Curtis,[1] Kenneth A. Rubinson,[1,3] and Zvi Kelman[4,5]

[1] NIST Center for Neutron Research, National Institute of Standards and Technology, Gaithersburg, Maryland 20899

[2] Division of Applied Biology and Chemistry, College of Agriculture and Life Sciences, Kyungpook National University, Daegu, Republic of Korea

[3] Department of Biochemistry and Molecular Biology, Wright State University, Dayton, Ohio 45435

[4] Biomolecular Measurement Division, National Institute of Standards and Technology, Gaithersburg, Maryland 20899

[5] Institute for Bioscience and Biotechnology Research, Rockville, Maryland 20850

ABSTRACT

The solution structure of the full-length DNA helicase minichromosome maintenance protein from *Methanothermobacter thermautotrophicus* was determined by small-angle neutron scattering (SANS) data together with all-atom molecular modeling. The data were fit best with a dodecamer (dimer of hexamers). The 12 monomers were linked together by the B/C domains, and the adenosine triphosphatase (AAA+) catalytic regions were found to be freely movable in the full-length dodecamer both in the presence and absence of Mg^{2+} and 50-meric single-stranded DNA (ssDNA). In particular, the SANS data and molecular modeling indicate that all 12 AAA+ domains in the dodecamer lie approximately the same distance from the axis of the molecule, but the positions of the helix–turn–helix region at the C-terminus of each monomer differ. In addition, the A domain at the N-terminus of each monomer is tucked up next to the AAA+ domain for all 12 monomers of the dodecamer. Finally, binding of ssDNA does not lock the AAA+ domains in any specific position, which leaves them with the flexibility to move both for helicase function and for binding along the ssDNA.

Proteins 2014; 82:2364–2374.

Key words: DNA helicase; small-angle neutron scattering; minichromosome maintenance protein; Monte Carlo simulations; DNA replication.

INTRODUCTION

The minichromosome maintenance (MCM) proteins are the replicative helicases in archaea and eukarya responsible for the separation of the duplex DNA in front of the replication fork. In eukarya MCM is a family of six different but related polypeptides (Mcm2–7), but most archaeal species studied contain a single MCM homologue.[1–6] Although the genomes of several species contain multiple genes encoding for MCM, only one gene has been shown to be essential for cell viability.[7,8]

Similar to the eukaryotic MCM complex, the archaeal helicase was found to utilize energy from adenosine triphosphate (ATP) hydrolysis to translocate along one strand of the DNA and displace the complementary strand. The enzyme binds single-stranded DNA (ssDNA) and translocates in the 3′–5′ direction. The enzyme can also bind and translocate along duplex DNA. However, because the enzyme continues with the same 3′–5′ directionality as it progresses onto duplex DNA, it is generally accepted that it interacts with only one strand. Also, the MCM helicase can displace bound proteins from DNA and unwind DNA–RNA hybrids.[1–5]

Additional Supporting Information may be found in the online version of this article.

Abbreviations: HTH, helix–turn–helix; mtMCM, minichromosome maintenance protein from *Methanothermobacter thermautotrophicus*; NIST, National Institute of Standards and Technology; NCNR, NIST Center for Neutron Research; ssoMCM, minichromosome maintenance protein from *Sulfolobus solfataricus*; SANS, small-angle neutron scattering.

Grant sponsor: National Science Foundation (NSF); Grant number: DMR-0944772; Grant sponsor: NSF (to ZK); Grant number: MCB-0815646; Grant sponsor: CCP-SAS software developed through a joint EPSRC; Grant number: (EP/K039121/1); Grant sponsor: NSF; Grant number: CHE-1265821.

*Correspondence to: Susan Krueger, NIST Center for Neutron Research, National Institute of Standards and Technology, 100 Bureau Drive, Stop 6102, Gaithersburg, MD 20899. E-mail: susan.krueger@nist.gov

Received 27 February 2014; Revised 24 April 2014; Accepted 29 April 2014
Published online 9 May 2014 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/prot.24598

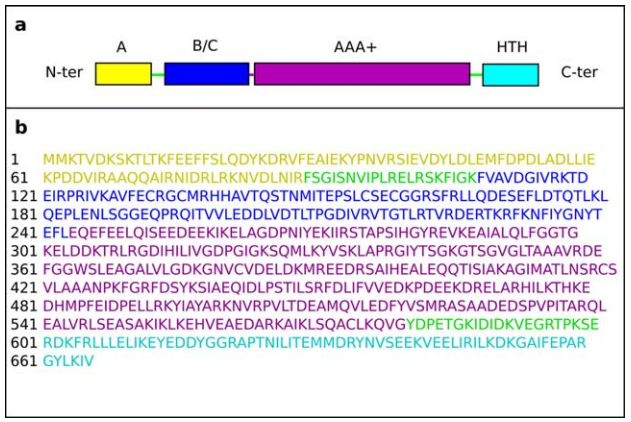

Figure 1

(**a**) Linear structure for full-length mtMCM, showing the N-terminal part that includes domain A, shown in yellow, and domain B/C shown in blue, the ATPase (AAA+) catalytic region, shown in purple, and the C-terminal HTH motif, shown in cyan. Unstructured loop regions between domains A and B/C and between domains AAA+ and HTH are shown in green. A smaller unstructured region between domains B/C and AAA+ is shown as part of domain AAA+ in purple. (**b**) Full-length mtMCM sequence with the domains and unstructured loop regions shown in the colors corresponding to those in (a).

Biochemical and biophysical studies suggest that the archaeal MCM proteins form dodecameric structures in solution—a head-to-head dimer of hexameric rings.[9–11] Dodecameric structures were also suggested by the three dimensional structure of the N-terminal part of the enzyme.[12–14] However, reconstructions from electron microscopy images show that the protein can also form other structures including hexamers, heptamers, octamers, filaments, and open rings.[1,2,5]

Although the ability of MCM to form both hexamers and dodecamers was first reported with the archaeal enzymes, it was later shown that the eukaryotic MCM[15] and the bacterial replicative helicase DnaB[16] could form similar structures. It is thought that these replicative helicases are assembled as dodecameric rings at the origins of replication. It is not clear, however, if the two hexamers associate during DNA synthesis or if the two paired hexamers separate and move away from each other[17] to carry out its function.

As illustrated in Figure 1(a), the archaeal MCM proteins can be divided into three main parts: the N-terminal region consisting of the A (yellow) and B/C (blue) domains, an ATPase (AAA+) catalytic region (purple) comprised of two domains in the center of the molecule, and the C-terminal region, which has been suggested to form a helix–turn–helix (HTH) domain[1,5,12] (cyan). Unstructured loop regions[12] between domains A and B/C and between domains AAA+ and HTH are shown in green. A smaller unstructured region between domains B/C and AAA+ is shown as part of the purple AAA+ domain. The domains defined here represent parts of the protein that are stable in isolation.

Both the zinc finger and the β-hairpin in domain B/C have been shown to participate in DNA binding.[12,18] The unstructured loop connecting the A domain to the B/C domain contains 19-residues. Consequently, the A domain can show major movement relative to the B/C domains.[14,19] Biochemical and mutational analyses suggest that the flexibility of domain A is required for it to play a role in regulating helicase activity.[20,21] Mutational analysis suggests that the B/C domain also drives the multimer (dodecamer) formation.[21–24] In addition, the B/C domain contains a loop that was shown to play a role in communication between the N-terminal DNA binding site and the AAA+ region.[25,26]

Neither the crystal nor solution structures of the dodecameric MCM have been reported. However, a 4.35Å-resolution crystal structure of the near-full-length monomer of the MCM protein from *Sulfolobus solfataricus* (ssoMCM) that is missing 6 N-terminal and 85 C-terminal residues has been determined.[27] This structure shows that the AAA+ catalytic domain is folded into two distinct domains; together, the two domains contain all the conserved motifs found in other members of the AAA+ family of ATPases.[28] The C-terminal part of ssoMCM was suggested to fold into an HTH domain,[29] although it was not seen clearly in the X-ray structure of the monomer.[27] Biochemical studies suggested that the HTH might play a regulatory role, since its removal results in increased helicase activity *in vitro*.[30]

Given the lack of structural information for full-length dodecameric MCM, the solution structures of the MCM protein from *Methanothermobacter thermautotrophicus* (mtMCM) in the presence and absence of DNA were determined using small-angle neutron scattering (SANS). The SANS data described below show that the MCM protein forms dodecamers in solution, consistent with the results from other biophysical methods. For structure modeling purposes, a mtMCM dodecamer molecule was built from the X-ray crystal structure coordinates of the N-terminal mtMCM monomer[12] and from the coordinates of the near-full-length monomer of ssoMCM,[27] as described in Materials and Methods, using the full-length mtMCM sequence shown in Figure 1(b), with colors corresponding to the domains and the unstructured regions shown in Figure 1(a). Ensembles of energetically relevant all-atom model structures were generated from this starting model structure using a Monte Carlo method that treats each of the unstructured loops of the 12 monomers independently.[14,31]

Comparing the SANS curves calculated for the ensemble of model structures with the experimental data revealed that large conformational changes can occur at each unstructured loop region that is shown in Figure 1. In accordance with previous SANS[14] and EM studies,[19] the calculated SANS curves show great sensitivity to the

positions of the A domains. The movement of domain A (the domain push mechanism) was suggested to play an important role in the regulation of MCM function.[20,30] In the present study, the data also suggest a large movement of the AAA+ catalytic domain occurs relative to the rest of the molecule. In addition, binding of ssDNA to the dodecamer does not lock the AAA+ domains in any position, which leaves them with the flexibility to move both for helicase function and for binding along the ssDNA.

MATERIALS AND METHODS*

Protein expression, purification, and sample preparation for SANS analysis

Full-length (residues 1–666) wild-type mtMCM was overexpressed in *Echerichia coli* BL21(DE3) pLysS cells (Novagen Biosciences, San Diego, CA) for 16 h in an autoinduction medium consisting of 1% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.33% (w/v) $(NH_4)_2SO_4$, 0.68% (w/v) KH_2PO_4, 0.71% (w/v) Na_2HPO_4, 0.05% (w/v) glucose, 0.2% (w/v) α-lactose, and 0.015% (w/v) $MgSO_4$. The proteins were purified on a Ni-NTA column (GE Healthcare, Piscataway, NJ) as previously described.[19] The purified proteins, eluted in D_2O, were dialyzed twice (10 mL into 100 mL) against D_2O-based SANS buffer consisting of 20 mmol L^{-1} (m*M*) Tris–HCl (pH 7.4), 100 m*M* NaCl, 0.5 m*M* EDTA, 2 m*M* dithiothreitol, and 5% (v/v) glycerol. Buffers for the mtMCM/DNA complex also contained 2 m*M* $MgCl_2$. Since the glycerol did not contain deuterium, the buffer was equivalent to a ≈ 95% (v/v) D_2O buffer without glycerol. The proteins were stored at 4 °C, and SANS measurements were started within 24 h after dialysis. The mtMCM dodecameric protein concentration was found to be 1.74 mg mL^{-1} (molecular mass ≈ 920 kDa, 1.9 μ*M*) by Bradford (Bio-Rad, Hercules, CA) assay with bovine serum albumin as the standard. Prior to the SANS measurements, the sample was clarified using centrifugation to remove filaments[19,23] that form at all concentrations. Thus, the final protein concentration is likely lower than that measured prior to centrifugation. To the sample containing mtMCM/DNA complex, 50-mer ssDNA (50 nt; 5′-GGGACGCGTCGGCCTGGCACGTCGGCCGCTGCGGCCAGGCACCCGATGGC-3′) in D_2O was added to a final concentration of 0.08 mg mL^{-1} (molecular mass ≈ 17.5 kDa, 4.6 μ*M*). This concentration represents greater than twofold molar excess in DNA over protein. The dissociation constant of ssDNA from mtMCM, K_d = 130 n*M*,[25] makes it essentially certain that at least one ssDNA molecule will be bound to each mtMCM molecule in solution. A 95% D_2O buffer sample containing 50-mer ssDNA at approximately 0.1 mg mL^{-1} was also prepared to confirm that free DNA in solution does not contribute significantly to the scattering, even if none of the DNA were bound to MCM in solution.

Two samples were prepared for SANS measurements: **1**, mtMCM without Mg^{2+} or DNA (–DNA) and **2**, mtMCM with Mg^{2+} and 50-mer ssDNA (+DNA).

SANS measurements

SANS measurements were performed on the 30-meter SANS instruments[32] at the National Institute of Standards and Technology (NIST) Center for Neutron Research (NCNR) in Gaithersburg, MD. The neutron wavelength, λ, was 5 Å, with a wavelength spread, $\Delta\lambda/\lambda$, of 0.15. Scattered neutrons were detected with a 64 cm × 64 cm two-dimensional position-sensitive detector with 128 × 128 pixels at a resolution of 0.5 cm $pixel^{-1}$. The data were reduced using the IGOR program with SANS macroroutines developed at the NCNR.[33] Raw counts were corrected for empty-cell counts, ambient room background counts, and nonuniform detector pixel response. Data were placed on an absolute scale relative to the concurrently measured incident beam flux. Finally, the data were radially averaged to produce scattered intensity, $I(q)$ versus q curves, where $q = 4\pi \sin(\theta)/\lambda$, and 2θ is the scattering angle. Sample-to-detector distances of 10.0 m and 2.0 m were used in order to cover the range 0.007 $Å^{-1} \leq q \leq 0.2$ $Å^{-1}$. The scattering intensities from the samples were then further corrected for buffer scattering and incoherent scattering from hydrogen in the samples.

Initial data analysis was performed using the Guinier approximation, $I(q) \approx I(0)\exp(-q^2R_g^2/3)$, on the low-$q$ portions of the data to obtain initial values for the radius of gyration, R_g, and the forward scattering intensity, $I(0)$, of the samples. This analysis is valid only in the region where $qR_g \lesssim 1$. Standard Kratky analysis was performed on each data set by plotting $I(q)\cdot q^2$ versus q to get a qualitative idea of the degree of protein folding.[34] Both the Guinier and Kratky analyses were performed using the NCNR IGOR SANS macroroutines.[33] Distance distribution functions, $P(r)$ versus r, were calculated using the program GNOM[35] using the full q-range of the data. A range of maximum distance values, D_{max}, was explored in order to determine its effect on the calculated $P(r)$ versus r curves.

Structure modeling

The structure for the mtMCM dodecamer molecule was built from the X-ray crystal structure coordinates of

*Certain commercial equipment, instruments, or materials are identified in this paper to foster understanding. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology nor does imply that the materials or equipment identified are necessarily the best available for the purpose.

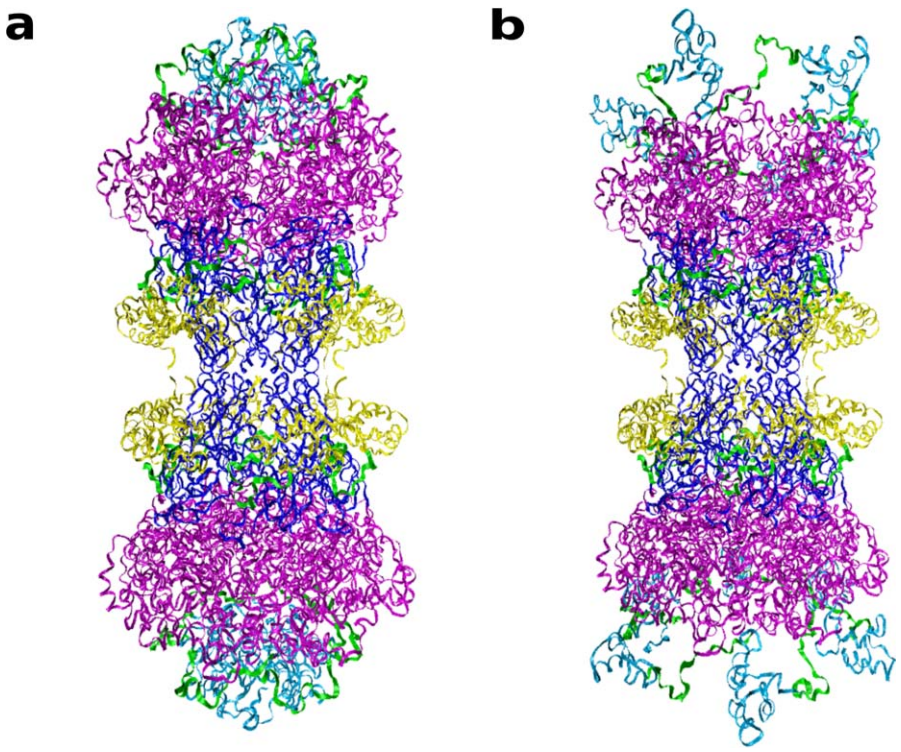

Figure 2

Starting structures for the full-length mtMCM dodecamer that were used for Monte Carlo modeling. The domains and unstructured regions are colored as in Figure 1. (**a**) The closed starting structure in which the HTH region was positioned inward such that the ends of the hole in the center of the dodecamer were obstructed. (**b**) The open starting structure in which the HTH region was positioned outward so that the central hole was unobstructed. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

the N-terminal mtMCM monomer (PDB accession number 1LTL),[12] which contains residues 5–243 of the mtMCM sequence, and from the coordinates of the near-full-length monomer of ssoMCM (PDB accession number 3F9V),[27] which contains residues 7–601 of the ssoMCM sequence. The full-length sequence of mtMCM contains 666 residues and ssoMCM contains 686 residues. The N-terminal part of mtMCM (1LTL) was aligned with that of ssoMCM (3F9V) using PHYRE2.[36] Making use of the coordinates of the catalytic region of ssoMCM, the catalytic region of mtMCM was then added to 1LTL using the program PSFGEN, which is part of the NAMD[37] software suite.

HHPRED[38] was used to find the best-matched homologous structure for the HTH region of mtMCM (residues 600–666). A model structure for the HTH region of mtMCM was built from the resulting best-aligned structure, which was from a H1 histone (PDB accession number 1UHM).[39] This region was positioned close enough to the catalytic domain to accommodate the linker residues 580–599, which were added using PSFGEN. The full mtMCM monomer was built by adding the final unstructured region consisting of residues 1–4 to the N-terminus using PSFGEN. The resultant full monomer structure was energy minimized for 5000 steps using NAMD, first constraining all but the linker region between domains AAA+ and HTH (residues 580–599) and then a second time with no constraints. The resultant structure was subjected to a 10 ps dynamics run using NAMD to confirm the integrity of the starting structure.

Finally, the initial mtMCM double hexamer structures were built by positioning the N-terminal part of the full mtMCM monomer in the same locations as those of the N-terminal mtMCM dodecamer model structure (1LTL). The resultant dodecamer structures were energy minimized for 5000 steps and then subjected to a 10 ps dynamics run using NAMD. Two different starting dodecamer structures were built in this manner: one in which the HTH region was positioned inward such that the ends of the hole in the center of the dodecamer were obstructed (closed starting structure) and a second in which the HTH region was positioned outward so that the hole was unobstructed (open starting structure). These structures are shown in Figure 2(a,b), respectively, with the domain colors corresponding to those shown in Figure 1.

The Complex Monte Carlo module of SASSIE[31] was used to generate ensembles of structures from these two starting structures for comparison to SANS data. Accepted (nonoverlapping) configurations were generated by sampling backbone dihedral angles using Charmm-22

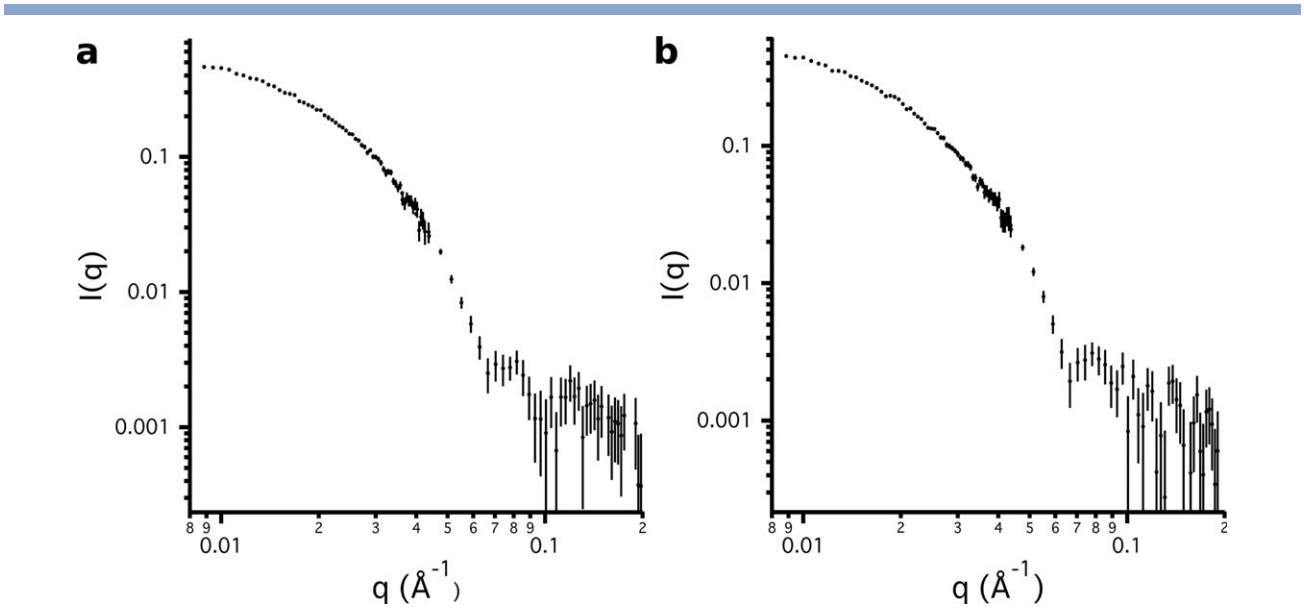

Figure 3

Log(I) versus log(q) plots of the SANS data from (**a**) the (–DNA) sample and (**b**) the (+DNA) sample. Error bars represent plus and minus the combined standard uncertainty of the data collection.

all-atom protein force-field parameters.[40] The new configuration was checked for overlap of basis atoms, which were chosen as α-carbon atoms in this case. If the overlap distance between basis atoms was ≥3Å, the new structure was accepted. Accepted configurations for each starting structure were energy minimized using NAMD. SANS profiles were calculated from the ensembles using the program Xtal2Sas[41,42] within SASSIE and the calculated data (model SANS curves) were compared to the experimental data and scored for quality using the χ^2 equation, $\chi^2 = \frac{1}{(N-1)} \frac{\sum_q (I_{exp}(q) - I_{calc}(q))^2}{\sigma_{exp}(q)^2}$, where $I_{exp}(q)$ is the experimentally determined SANS intensity curve, $I_{calc}(q)$ is the calculated intensity curve from the model structure and $\sigma_{exp}(q)$ is the q-dependent error of the $I_{exp}(q)$ values. The sum was taken over $N = 60$ data points.

Examination of a plot of χ^2 versus R_g provides an idea of how well the individual structures generated from each starting structure fit the data as well as which starting structure produces the overall best fits to the data. The best-fit (lowest χ^2) and worst fit (highest χ^2) model SANS curves are noted for each case, along with the average model SANS curve from the entire ensemble of accepted structures. These curves were plotted along with the experimental SANS data to aid in the visualization of the quality of the fits to the data. Surface plots representing the extreme positions within the total configuration space examined by all of the accepted structures were generated and compared to that representing the best-fit family of structures in each case. The best-fit family of structures was chosen based on those giving the lowest ≈10% of the χ^2 values for each series. (This cutoff was chosen arbitrarily based on the shape of the χ^2 vs. R_g curves and used to analyze the results further.) These structures were energy minimized using the program NAMD[37] and compared to the data a second time to insure that the results were the same.

RESULTS AND DISCUSSION

SANS data

The SANS data obtained for the two different mtMCM samples are shown on a log(I) versus log(q) scale in Figure 3. SANS data for the (–DNA) sample are shown in Figure 3(a) and the corresponding data the (+DNA) sample are represented in Figure 3(b). Guinier fits to the (–DNA) and the (+DNA) resulted in R_g values of (89.0 ± 2.6) Å and (93.0 ± 2.5) Å, respectively. $I(0)$ values were (0.57 ± 0.01) cm^{-1} for both, and Kratky plots confirmed that both have compact structures. The Guinier and Kratky plots can be found in Figures S1 and S2, respectively, in the Supporting Information.

The distance distribution functions, $P(r)$ versus r, were calculated from the SANS data keeping in mind that the positions of the HTH regions are not fixed because of the flexibility of the unstructured loop between the AAA+ and HTH regions. Thus, the maximum possible distance between two points in the molecule, D_{max}, is not the same for each molecule in the solution. To encompass the possible maxima, D_{max} values of 250 Å, 275 Å, and 300 Å were used in separate calculations, and the resultant R_g and $I(0)$ values are shown in Table I. The R_g values vary from 83 Å to 94 Å depending on the choice of D_{max}. The $P(r)$ versus r

Table I
Distance Distribution, $P(r)$ Versus r, Parameters

Sample	R_g (Å)	$I(0)$ cm^{-1}	D_{max} (Å)
(–DNA)	83 ± 1	0.54 ± 0.01	250
	88 ± 1	0.56 ± 0.01	275
	92 ± 1	0.58 ± 0.01	300
(+DNA)	85 ± 1	0.52 ± 0.01	250
	88 ± 1	0.54 ± 0.01	275
	94 ± 1	0.56 ± 0.01	300

curves from both the (–DNA) and (+DNA) data, using a D_{max} value of 275 Å, are shown in Figure 4. The peaks in Figure 4 occur at slightly different distances, and the curves show the greatest differences between them for distances between 125 Å and 200 Å. These differences are independent of the choice of D_{max}, as illustrated in Figure S3 in the Supporting Information, where the $P(r)$ versus r curves for all three values of D_{max} are shown. Taken together, the distance distribution and Guinier results suggest that the (+DNA) sample has a slightly higher R_g value than the (–DNA) sample and that the molecule exhibits some variation in the structures between the length scales of 125 Å and 200 Å.

Because the SANS data are on an absolute scale, the $I(0)$ values are related to the molecular mass of the molecule. The predicted $I(0)$ value of the (–DNA) sample at 1.74 mg mL^{-1} in 95% D_2O buffer is 1.15 cm^{-1}, assuming a molecular mass of 920 kDA as calculated directly from the full-length sequence. It is more difficult to predict the $I(0)$ value of the (+DNA) sample since the MCM:DNA stoichiometry is not known. The average $I(0)$ value for the (–DNA) sample of 0.56 ± 0.02 cm^{-1} from Table I suggests a concentration of 0.85 ± 0.03 mg mL^{-1}. It is likely that the concentration of the measured mtMCM samples were closer to this value due to the loss of some sample as a result of the formation of filaments,[19,23] which were removed during the centrifugation step prior to the SANS measurements. Thus, $I(0)$ is not a reliable means of corroborating the integrity of the mtMCM sample. A similar result was obtained during the structural studies of the N-terminal part of the enzyme.[14] In this case, the integrity of the (−DNA) and (+DNA) samples was corroborated by direct comparison of experimental data with model SANS curves of both single hexamer and dodecamer mtMCM complexes, since the predicted $I(0)$ values for the single hexamer are approximately that of the measured values for both samples. The single hexamer was immediately ruled out since R_g values for those structures were in the 40–45 Å range, which is much smaller than the R_g values listed in Table I, and the model SANS curves were very poor fits to the SANS data.

Finding the mtMCM structures best fitting the SANS data

Allowing flexibility between the A and B/C and between the AAA+ and HTH domains

The two different starting structures shown in Figure 2(a,b) were used to generate subsequent structures for

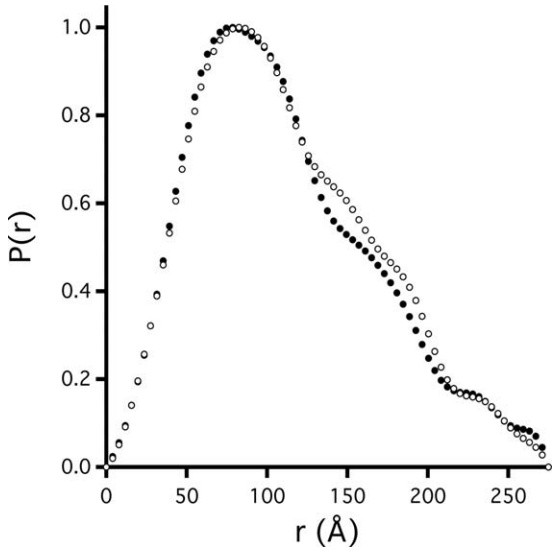

Figure 4
$P(r)$ versus r, for the (–DNA) sample (solid circles) and the (+DNA) sample (open circles). The curves have been scaled such that $P(r)_{max}$ = 1. A D_{max} value of 275 Å was used to obtain these curves. Additional $P(r)$ versus r curves obtained using different values of D_{max} can be found in the Supporting Information.

comparison to the SANS data. The closed starting structure in Figure 2(a) has the HTH regions positioned inward such that the hole in the center of the dodecamer is obstructed, whereas the open starting structure in Figure 2(b) has the HTH regions positioned outward so that the central hole is unobstructed. The Monte Carlo simulation only allowed variation in the backbone dihedral angles of residues 89–108 between the A and B/C domains and residues 580–600 between the AAA+ and HTH domains. These are the unstructured loop regions that are shown in green in Figures 1 and 2. The variations were made independently for each monomer in the full dodecamer structure.

A total of 1732 accepted structures (as defined in Materials and Methods) were generated from the closed starting structure, and 2980 accepted structures were generated from the open starting structure. R_g was calculated for each accepted structure, along with a model SANS curve. Each model SANS curve was compared to the SANS data from the (–DNA) and (+DNA) samples and the overall quality of the fit to the experimental data, χ^2, was determined. The resultant χ^2 versus R_g plot for the (–DNA) sample is shown in Figure 5. The χ^2 versus R_g plot for the (+DNA) sample is similar and is not shown. In the graph, each black square represents a member of the ensemble of structures generated from the closed starting structure. The yellow square shows the closed starting structure position. The blue circles are from the ensemble of structures generated from the open starting structure, which is represented by the yellow

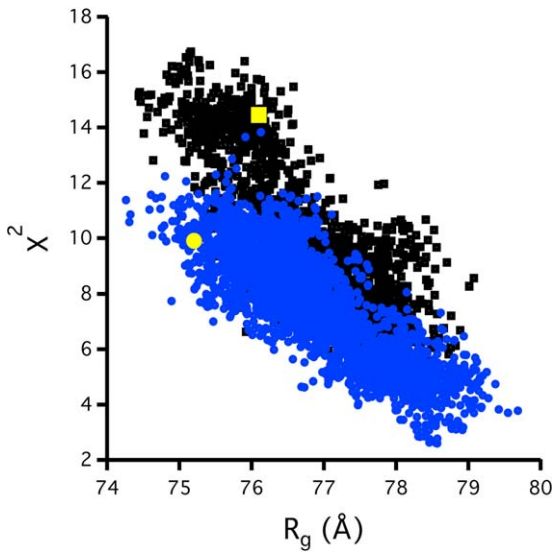

Figure 5

χ^2 versus R_g plots for the (–DNA) sample for the 1732 structures generated from the closed starting structure (black squares) and for the 2980 structures generated from the open starting structure (blue circles). The value for the closed starting structure is shown by the yellow square and for the open starting structure by the yellow circle. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

circle. The generated structures all have R_g values less than 80 Å, and there is no clear minimum in the χ^2 versus R_g plots.

Figure 6 shows the (–DNA) SANS data along with the SANS curves for the best-fit models from both the open and closed starting structures. The χ^2 value for the best-fit curve derived from the open starting structure is 2.6 while that for the best-fit curve derived from the closed starting structure is 5.1. The corresponding plot for the (+DNA) sample is similar and is not shown. However, a comparison of the best-fit model SANS curves to those calculated from the original open and closed starting structures can be found for both the (–DNA) and (+DNA) samples in Figure S4.

The best-fit model SANS curves in Figure 6 fit the data better below $q = 0.065$ Å^{-1} than those from the original starting structures. However, they do not have the same shape as the data below $q = 0.065$ Å^{-1}, and, in addition, the location of the subsidiary maxima at $q > 0.065$ Å^{-1} do not match that of the data. The worst match occurs for the closed structure (green curve in Fig. 6). Furthermore, the R_g values from the model SANS curves are several Å smaller than those obtained from Guinier and $P(r)$ analysis of the SANS data.

The lack of agreement between the R_g values obtained from the experimental SANS curves and those calculated from the best-fit model structures, the mismatch in positions of the subsidiary maxima between the experimental and the model SANS curves, and the absence of a minimum in the χ^2 versus R_g graphs all indicate that allowing only the N-terminal and C-terminal unstructured loops to vary during the model generating process is not sufficient to obtain structures that fit the SANS data. Rather, further flexibility must be allowed in the Monte Carlo simulations.

Allowing flexibility between the A and B/C, between the B/C and AAA+, and between the AAA+ and HTH domains

To approach structures that better fit the SANS data, the backbone dihedral angles of residues 244–246 between the B/C and AAA+ domains in the original open starting structure were allowed to energetically sample different conformations, in addition to those of residues 89–108 between the A and B/C domains and residues 580–600 between the AAA+ and HTH domains. The same sampling strategy was used simultaneously for each monomer in the dodecamer in order to generate 1077 symmetric structures with the AAA+ catalytic region further away from the center of the dodecamer. These structures are represented in Figure S5(a) by a surface plot representing the total configuration space examined by all 1077 accepted structures in the ensemble. Figure S5(b) shows the (–DNA) SANS data along with the best-fit model SANS curve. The corresponding best-fit ($\chi^2 = 1.7$, $R_g = 80.9$ Å), energy-minimized symmetric structure [inset, Fig. S5(b)] then was used as the improved starting structure for further modeling.

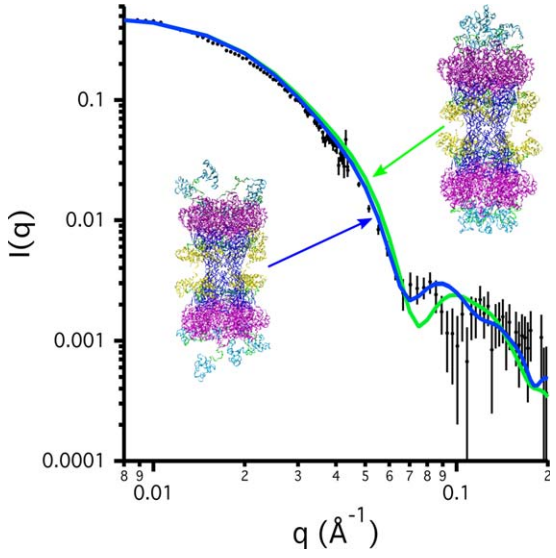

Figure 6

SANS data for the (–DNA) sample along with best-fit model SANS curves generated from the closed starting structure (green solid line) and from the open starting structure (blue solid line). The corresponding best-fit structures are shown as insets. Error bars on the data represent plus and minus the combined standard uncertainty of the data collection. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

An additional 10,892 accepted structures were generated from this improved symmetric starting structure, allowing residues 89–108, 244–246, and 580–600 to vary independently for each monomer in the dodecamer. Figure 7(a) shows the χ^2 versus R_g plot for the (–DNA) sample for the ensembles of structures generated from the original closed and original open starting structures (black squares and blue circles, respectively, from Fig. 5) as well as from the ensemble of 10,892 structures generated from the improved starting structure (gray triangles). A clear minimum in the χ^2 versus R_g plot can now be seen. Figure 7(b) shows the above mentioned gray triangles, along with the 1391 best-fit structures with $\chi^2 \leq 1.4$ (red symbols), which represent ≈10% of the structures with the lowest χ^2 values. The yellow triangle indicates the improved symmetric starting structure [Fig. S5(b), inset].

The same information contained in the χ^2 versus R_g plots in Figure 7(b) is illustrated in another way in Figure 7(c) using surface plots to represent configuration space. The surface plot representing the configuration space sampled by the entire ensemble of 10,892 structures generated from the improved starting structure, gray triangles in Figure 7(b), is shown in transparent gray. The configuration surface plot for the ensemble of 1391 best-fit structures with $\chi^2 \leq 1.4$, which corresponds to the structures represented by the red symbols in Figure 7(b), is shown in red.

The best-fit (χ^2 = 1.1, R_g = 81.0 Å) and worst-fit (χ^2 = 4.3, R_g = 84.7 Å) individual model structures from the entire ensemble of 10,892 structures are shown in Figure S6 in the Supporting Information for comparison. The AAA+ catalytic domains in both structures show large movements with respect to their positions in the original starting structures shown in Figure 4. The unstructured loop regions linking the AAA+ and HTH regions (residues 580–600) are in positions resembling those in the original open starting structure. A similar

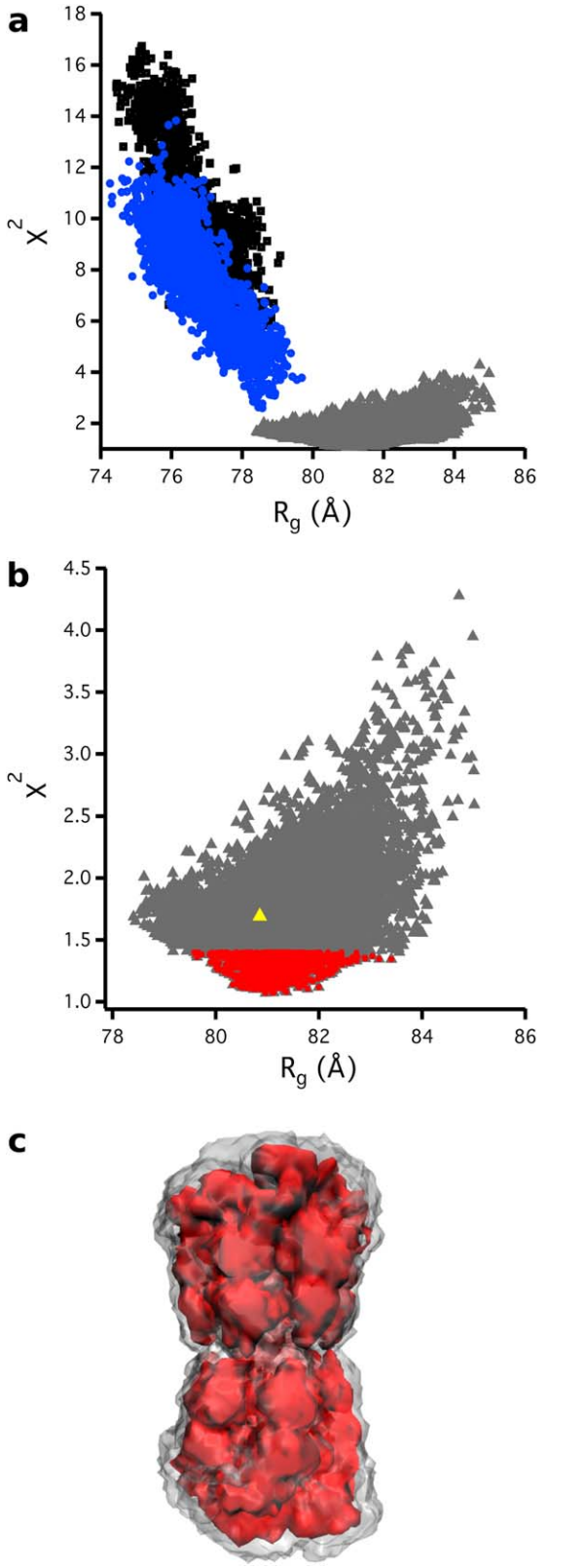

Figure 7.

Figure 7

(**a**) χ^2 versus R_g plot for the (–DNA) sample for the ensembles of structures generated from the original closed and open starting structures (black squares and blue circles, respectively, from Fig. 5) and the minimized improved starting structure with the more open AAA+ catalytic region (gray triangles). (**b**) χ^2 versus R_g plot for the ensemble of structures generated from the minimized improved starting structure (gray triangles), along with the 1391 best-fit structures with $\chi^2 \leq 1.4$ (red symbols). The yellow triangle represents the minimized improved starting structure. (**c**) Surface plots representing the configuration space sampled by the entire ensemble of 10,892 structures generated from the improved starting structure (transparent gray), along with that for the ensemble of 1391 best-fit structures to the (–DNA) data with $\chi^2 \leq 1.4$ (opaque red). The former surface plot corresponds to the structures represented by the gray triangles in (b) and the latter corresponds those represented by the red symbols in the same panel. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

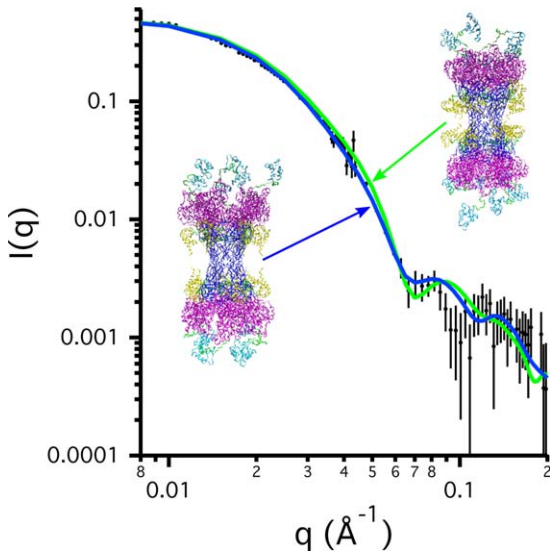

Figure 8

Log(I) versus log(q) plots of the SANS data from the (–DNA) sample, along with the best-fit model SANS curves generated from the original open starting structure (solid green line) and from the improved symmetric starting structure with the more open AAA+ domains (solid blue line). The corresponding best-fit structures are shown as insets. Error bars represent plus and minus the combined standard uncertainty of the data collection. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

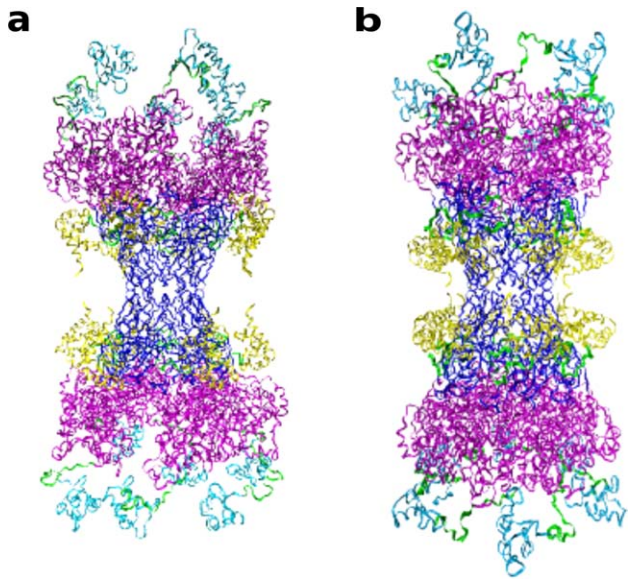

Figure 9

(**a**) Best-fit model structure to the SANS data from the (–DNA) sample generated from the improved symmetric starting structure [inset, Fig. S5(b)]. (**b**) Original symmetric open starting structure [Fig. 2(b)]. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

result was obtained for the (+DNA) sample (data not shown).

To further analyze the two structure ensembles represented in Figure 7(b,c), the SANS curves for the (–DNA) and (+DNA) samples are compared in Figure S7 to the model SANS curves from the best-fit and worst-fit structures to the experimental data as well as the average SANS curve for each ensemble. For the smaller best-fit ensemble [red symbols in Fig. 7(b) or red surface in Fig. 7(c)], all three curves fit the data equally well. This indicates that not only is each individual structure in this ensemble a good representation of the SANS data, but also the entire ensemble of best-fit structures would reproduce the measured SANS curve.

The best-fit structure ensembles found for both the (–DNA) and (+DNA) samples can be compared visually using the configuration surface plots in Figure S8. It is apparent that the full extent of configuration space sampled is similar for both ensembles. So, while the Guinier analysis suggests that the (+DNA) structure has a slightly higher R_g value than the (–DNA) structure, the corresponding structural differences suggested by the $P(r)$ analysis cannot be localized to a particular region or regions of the dodecamer.

To illustrate the significantly better agreement with the SANS data that was obtained with the additional flexibility of residues 244–246, Figure 8 shows the best-fit model SANS curves for the (–DNA) sample generated from the original open starting structure (solid green line) together with that from the improved symmetric starting structure that has the more open AAA+ domains (solid blue line). Their corresponding model structures are seen as insets. The best-fit structure from the improved starting structure has a larger R_g value (81.0 Å) than that for the best-fit structure generated from the original open starting structure (78.6 Å). By allowing the AAA+ domains to move away from the center of the molecule, structures with larger R_g values were generated that were better fits to the SANS data. A similar result was obtained for the (+DNA) sample (data not shown).

Furthermore, the flexible linker near the N-terminal of each monomer (residues 89–108) is less extended in the best-fit structure obtained from the improved starting structure. Although this linker region is free to vary independently for each monomer in the dodecamer, the best-fit structures to the (–DNA) SANS data all show this linker folded such that domain A (residues 1–88) is tucked closely to the AAA+ domain. Figure 9 compares the best-fit structure with the overall lowest χ^2 value directly to the original open starting structure. While the former structure [Fig. 9(a)] shows the AAA+ domains (magenta) positioned outward from the center of the molecule and in close proximity to the A domains (yellow), the latter structure [Fig. 9(b)] shows the AAA+ domains closer to the center of the molecule and the A

domains folded down and inward toward the midline between the two hexamers. As a result, the locations of the A and AAA+ domains appear to be important for the model structures to produce scattering consistent with the SANS data. A similar result was found for the (+DNA) sample (data not shown). This result is consistent with the domain-push model,[12,20] which suggests that movement of domain A is necessary for mtMCM function. The flexibility of the loose connection of domains A to B/C appears to be necessary to facilitate the domain A motion.

CONCLUSIONS

SANS was used to measure the solution structure of the full-length mtMCM in the presence and absence of 50-meric ssDNA. The SANS data showed that the MCM protein forms dodecamers and that the DNA-bound sample has a higher R_g than the sample without DNA. Comparison of the calculated model SANS curves to the data revealed that the best-fit structures to the SANS data were those in which the positions of the AAA+ catalytic domains were tilted away from the center of the molecule and those of domain A were tucked closely to the AAA+ domain. Surface plots representing the total configuration space examined by the best-fit ensembles of structures of both samples were equivalent. Therefore, the differences in R_g and $P(r)$ between the two samples cannot be localized to any particular region of the MCM dodecamer. Nevertheless, the data fitting and modeling show that ssDNA does not induce a fixed structure to the AAA+ catalytic domains. This motion leaves them with the flexibility to move both for function and for binding along the ssDNA.

REFERENCES

1. Sakakibara N, Kelman LM, Kelman Z. Unwinding the structure and function of the archaeal MCM helicase. Mol Microbiol 2009;72: 286–296.
2. Costa A, Onesti S. Structural biology of MCM helicases. Crit Rev Biochem Mol Biol 2009;44:326–342.
3. Brewster AS, Chen XS. Insights into the MCM functional mechanism: lessons learned from the archaeal MCM complex. Crit Rev Biochem Mol Biol 2010;45:243–256.
4. Krupovic M, Gribaldo S, Bamford DH, Forterre P. The evolutionary history of archaeal MCM helicases: a case study of vertical evolution combined with hitchhiking of mobile genetic elements. Mol Biol Evol 2010;27:2716–2732.
5. Slaymaker IM, Chen XS. MCM structure and mechanics: what we have learned from archaeal MCM. In: MacNeill S, editor. Subcellular biochemistry, Vol. 62. Dordrecht: Springer Netherlands; 2012. pp 89–111.
6. Medagli B, Onesti S. Structure and mechanism of hexameric helicases. In: Spies M, editor. Advances in experimental medicine and biology, Vol. 973. New York, NY: Springer; 2013. pp 75–95.
7. Pan M, Santangelo TJ, Li Z, Reeve JN, Kelman Z. *Thermococcus kodakarensis* encodes three MCM homologs but only one is essential. Nucleic Acids Res 2011;39:9671–9680.
8. Ishino S, Fujino S, Tomita H, Ogino H, Takao K, Daiyasu H, Kanai T, Atomi H, Ishino Y. Biochemical and genetical analyses of the three MCM genes from the hyperthermophilic archaeon, *Thermococcus kodakarensis*. Genes Cells 2011;16:1176–1189.
9. Kelman Z, Lee JK, Hurwitz J. The single minichromosome maintenance protein of *Methanobacterium thermoautotrophicum* ΔH contains DNA helicase activity. Proc Natl Acad Sci USA 1999;96:14783–14788.
10. Chong JP, Hayashi MK, Simon MN, Xu RM, Stillman B. A double-hexamer archaeal minichromosome maintenance protein is an ATP-dependent DNA helicase. Proc Natl Acad Sci USA 2000;97:1530–1535.
11. Shechter DF, Ying CY, Gautier J. The intrinsic DNA helicase activity of *Methanobacterium thermoautotrophicum* ΔH minichromosome maintenance protein. J Biol Chem 2000;275:15049–15059.
12. Fletcher RJ, Bishop BE, Leon RP, Sclafani RA, Ogata CM, Chen XS. The structure and function of MCM from archaeal *M. thermoautotrophicum*. Nat Struct Biol 2003;10:160–167.
13. Liu W, Pucci B, Rossi M, Pisani FM, Ladenstein R. Structural analysis of the *Sulfolobus solfataricus* MCM protein N-terminal domain. Nucleic Acids Res 2008;36:3235–3243.
14. Krueger S, Shin J-H, Raghunandan S, Curtis JE, Kelman Z. Atomistic ensemble modeling and small-angle neutron scattering of intrinsically disordered protein complexes: applied to minichromosome maintenance protein. Biophys J 2011;101:2999–3007.
15. Remus D, Beuron F, Tolun G, Griffith JD, Morris EP, Diffley JFX. Concerted loading of Mcm2–7 double hexamers around DNA during DNA replication origin licensing. Cell 2009;139:719–730.
16. Stelter M, Gutsche I, Kapp U, Bazin A, Bajic G, Goret G, Jamin M, Timmins J, Terradot L. Architecture of a dodecameric bacterial replicative helicase. Structure 2012;20:554–564.
17. Takahashi TS, Wigley DB, Walter JC. Pumps, paradoxes and ploughshares: mechanism of the MCM2-7 DNA helicase. Trends Biochem Sci 2005;30:437–444.
18. Poplawski A, Grabowski B, Long SE, Kelman Z. The zinc finger domain of the archaeal minichromosome maintenance protein is required for helicase activity. J Biol Chem 2001;276:49371–49377.
19. Chen Y-J, Yu X, Kasiviswanathan R, Shin J-H, Kelman Z, Egelman EH. Structural Polymorphism of *Methanothermobacter thermautotrophicus* MCM. J Mol Biol 2005;346:389–394.
20. Fletcher RJ, Chen XS. Biochemical activities of the BOB1 mutant in *Methanobacterium thermoautotrophicum* MCM. Biochemistry (Mosc) 2006;45:462–467.
21. Kasiviswanathan R, Shin J-H, Melamud E, Kelman Z. Biochemical characterization of the *Methanothermobacter thermautotrophicus* minichromosome maintenance (MCM) helicase N-terminal domains. J Biol Chem 2004;279:28358–28366.
22. Fletcher RJ, Shen J, Gómez-Llorente Y, Martín CS, Carazo JM, Chen XS. Double hexamer disruption and biochemical activities of *Methanobacterium thermoautotrophicum* MCM. J Biol Chem 2005;280: 42405–42410.
23. Shin J-H, Heo G-Y, Kelman Z. The *Methanothermobacter thermautotrophicus* MCM helicase is active as a hexameric ring. J Biol Chem 2009;284:540–546.
24. Jenkinson ER, Costa A, Leech AP, Patwardhan A, Onesti S, Chong JPJ. Mutations in subdomain B of the minichromosome maintenance (MCM) helicase affect DNA binding and modulate conformational transitions. J Biol Chem 2009;284:5654–5661.
25. Sakakibara N, Kasiviswanathan R, Melamud E, Han M, Schwarz FP, Kelman Z. Coupling of DNA binding and helicase activity is mediated by a conserved loop in the MCM protein. Nucleic Acids Res 2008;36:1309–1320.
26. Barry ER, Lovett JE, Costa A, Lea SM, Bell SD. Intersubunit allosteric communication mediated by a conserved loop in the MCM helicase. Proc Natl Acad Sci 2009;106:1051–1056.
27. Brewster AS, Wang G, Yu X, Greenleaf WB, Carazo JM, Tjajadi M, Klein MG, Chen XS. Crystal structure of a near-full-length archaeal

MCM: functional insights for an AAA+ hexameric helicase. Proc Natl Acad Sci 2008;105:20191–20196.
28. Duderstadt KE, Berger JM. AAA+ ATPases in the initiation of DNA replication. Crit Rev Biochem Mol Biol 2008;43:163–187.
29. Aravind L, Koonin EV. DNA-binding proteins and evolution of transcription regulation in the archaea. Nucleic Acids Res 1999;27: 4658–4670.
30. Barry ER, McGeoch AT, Kelman Z, Bell SD. Archaeal MCM has separable processivity, substrate choice and helicase domains. Nucleic Acids Res 2007;35:988–998.
31. Curtis JE, Raghunandan S, Nanda H, Krueger S. SASSIE: a program to study intrinsically disordered biological molecules and macromolecular ensembles using experimental scattering restraints. Comput Phys Commun 2012;183:382–389.
32. Glinka CJ, Barker JG, Hammouda B, Krueger S, Moyer JJ, Orts WJ. The 30 m small-angle neutron scattering instruments at the National Institute of Standards and Technology. J Appl Crystallogr 1998;31:430–445.
33. Kline SR. Reduction and analysis of SANS and USANS data using IGOR Pro. J Appl Crystallogr 2006;39:895–900.
34. Receveur-Brechot V, Durand D. How random are intrinsically disordered proteins. A small angle scattering perspective. Curr Protein Pept Sci 2012;13:55–75.
35. Semenyuk AV, Svergun DI. GNOM—a program package for small-angle scattering data processing. J Appl Crystallogr 1991;24:537–540.
36. Kelley LA, Sternberg MJE. Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc 2009;4:363–371.
37. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L, Schulten K. Scalable molecular dynamics with NAMD. J Comput Chem 2005;26:1781–1802.
38. Soding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res 2005;33:W244–W248.
39. Ono K. The linker histone homolog Hho1p from *Saccharomyces cerevisiae* represents a winged helix–turn–helix fold as determined by NMR spectroscopy. Nucleic Acids Res 2003;31:7199–7207.
40. MacKerell, AD, Bashford D, Bellott, Dunbrack, RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 1998;102:3586–3616.
41. Heidorn DB, Trewhella J. Comparison of the crystal and solution structures of calmodulin and troponin C. Biochemistry (Mosc) 1988;27:909–915.
42. Krueger S, Gorshkova I, Brown J, Hoskins J, McKenney KH, Schwarz FP. Determination of the conformations of cAMP receptor protein and its T127L, S128A mutant with and without cAMP from small angle neutron scattering measurements. J Biol Chem 1998; 273:20001–20006.