Stephanie M. Patterson was born and raised in Corpus Christi, Texas. She received her Associate's degree in mathematics from Del Mar College in May 2007. During the summers of 2006 and 2007, Stephanie participated in the FaST internship at the Lawrence Berkeley National Laboratory where she has continued her research through the fall of 2007 and spring of 2008. She plans to receive a Bachelor's Degree in Physics and a PhD in Biophysics. Her interests include philosophy and active participation in the University Students Cooperative Association.

Steve Yannone is currently a staff scientist at Lawrence Berkeley National Laboratory in the department of Molecular Biology. He completed his undergraduate studies in biochemistry at the University of California at Riverside, his Ph.D. in biochemistry at UC Irvine and spent two years of postdoctoral training at Los Alamos National Laboratory. Steve began his research career with an undergraduate research scholarship to study insect kairomones in the department of entomology at UC Riverside. His graduate studies focused on the roles of metalloproteins in mediating cellular stress responses and redox status in the nitrogen fixing bacteria Azotobacter vinelandii. His graduate studies lead him into the field of DNA repair and he now leads an active research program at LBNL studying the biochemistry of DNA repair enzymes.

Expression, Purification, and Small Angle X-Ray Scattering of DNA Replication and Repair Proteins from the Hyperthermophile *Sulfolobus solfataricus*

Stephanie M. Patterson, J. Robert Hatherill, Michal Hammel, Greg L. Hura, John A. Tainer, and Steven M. Yannone

ABSTRACT

Vital molecular processes such as DNA replication, transcription, translation, and maintenance occur through transient protein interactions. Elucidating the mechanisms by which these protein complexes and interactions function could lead to treatments for diseases related to DNA damage and cell division control. In the recent decades since its introduction as a third domain, Archaea have shown to be simpler models for complicated eukaryotic processes such as DNA replication, repair, transcription, and translation. *Sulfolobus solfataricus* is one such model organism. A hyperthermophile with an optimal growth temperature of 80°C, *Sulfolobus* protein-protein complexes and transient protein interactions should be more stable at moderate temperatures, providing a means to isolate and study their structure and function. Here we provide the initial steps towards characterizing three DNA-related *Sulfolobus* proteins with small angle X-ray scattering (SAXS): Sso0257, a cell division control and origin recognition complex homolog, Sso0768, the small subunit of the replication factor C, and Sso3167, a Mut-T like protein. SAXS analysis was performed at multiple concentrations for both short and long exposure times. The Sso0257 sample was determined to be either a mixture of monomeric and dimeric states or a population of dynamic monomers in various conformational states in solution, consistent with a flexible winged helix domain. Sso0768 was found to be a complex mixture of multimeric states in solution. Finally, molecular envelope reconstruction from SAXS data for Sso3167 revealed a novel structural component which may function as a disordered to ordered region in the presence of its substrates and/or protein partners.

Introduction

Transient protein complexes and assemblies perform many essential biological processes in organisms. Elucidating the precise mechanism by which these processes are carried out will give valuable insight into how cells function. Of particular interest is how cells efficiently replicate, repair, translate and transcribe their DNA. In humans and other mesophiles, these processes involve dynamic protein complexes, which can be difficult to study due to their transient nature. Instead, researchers have looked to hyperthermophilic Archaea to provide more tractable model systems for the more complicated eukaryotic processes.

A growing number of archaeal genomes have recently been sequenced, revealing that Archaea process information similarly to Eukaryotes [1,2,3,4]. Archaeal DNA translation, transcription, and replication resemble eukaryotic processes, but with Archaea-specific characteristics. In particular, the archaeal transcription initiation complex shares many important characteristics with the more complicated eukaryotic complexes [5,6,7,8]. In addition to providing a simpler model, many Archaea are hyperthermophiles implying that their transient protein complexes will be more physicochemically stable at moderate temperatures, potentially allowing researchers to thermally-trap protein complexes that are unstable in their eukaryotic homologs.

Sulfolobus solfataricus (*S. so*) is a broadly researched crenarchaeon, whose proteins have been used to model molecular mechanisms in the cell cycle, chromosomal maintenance, transcription, translation, RNA processing, and DNA replication [9, 10]. *S. so* has an optimal growth temperature of 80°C, is aerobic and heterotrophic, can be

grown on liquid or solid media, and is relatively easy to work with in the lab.

Here we present the expression, purification, and low-resolution structural characterization using small angle X-ray scattering (SAXS) for three *S. so* proteins involved in nucleic acid metabolism.

Sso0257, Cell Division Control 6 (CDC6), Origin Recognition Complex 1 (ORC1) homolog

S. so contains open reading frames (ORFs) encoding eukaryotic-like initiation factors including three cell division control 6 (CDC6)/ origin recognition complex 1 (ORC1) homologs: CDC6-1, CDC6-2, and CDC6-3 [4,10]. In Eukaryotes, ORC1 recognizes and binds to origins of replications in an ATP-dependent mechanism, then recruits the prereplication complexes, including CDC6, to the replication origin. CDC6 helps to assemble the minichromosome maintenance (MCM) proteins. Based on sequence similarity, She *et al*. have suggested that the three *S. so* CDC6 homologs combine the functions of the eukaryotic CDC6 and ORC1 proteins [4]. Sso0257 is the 45 kilodalton (kDa) CDC6-1 protein. Sso0257 has been biochemically characterized reporting the formation of monomers in solution and the ability to auto-phosphorylate *in vitro* [11]. Additionally, Sso0257 binds DNA, with a preference for bubble and fork substrates and stimulates *S. so* MCM protein binding to oligonucleotides with bubble or fork structure [11]. Recently, Sso0257 was crystallized in complex with CDC6-3 and replication origin DNA displaying an independent yet consorted deformation of the DNA where both the ATPase and the winged helix domain bind DNA [12].

Sso0768 Replication Factor C (RFC), Activator 1, small subunit

Replication factor C is a clamp-loader protein known to load the polymerase processivity factor (Proliferating Cell Nuclear Antigen, PCNA in Eukaryotes) onto DNA. In *S. so*, two RFC proteins were identified: a large (47 kDa) and a small subunit (38 kDa) [4]. Sso0768, the small subunit, cannot alone bind DNA effectively or initiate DNA polymerase B1 in the presence of either *S. so* PCNA-like proteins, 039p and 048p [13].

Sso3167 Mut-T like protein

A member of the base excision repair family, Mut-T proteins repair endogenous oxidative damage to dGTP nucleotides prior to incorporation in the genome during replication. Mut-T was originally discovered in bacterial cells; however a similar protein has been partially purified from human cells [14]. Sso3167 is an 18 kDa polypeptide.

Materials and Methods

One liter LB media containing 0.1 mg/ml ampicillin was inoculated from frozen stocks of *E. coli* Rosetta cells (Novagen) transformed with the pet21a vector (Novagen) containing the *S. so* gene of interest fused to a C-terminal six-histidine tag. Cultures were grown at 37°C while shaking at 250 revolutions per minute (rpm) to an optical density of 0.5 at 600 nm (logarithmic phase). Protein expression was induced by the addition of isopropyl ß-D-1-thiogalactopyranoside (IPTG) to a final concentration of 1 mM at 20°C while shaking at 225 rpm for 8 hours.

Cells were harvested by centrifugation at 4500 x g at 4°C. Packed cell volume (PCV) was approximately 7 ml and cell pellet mass was about 7 g. Cell pellets were resuspended and lysed by sonication in four to five times the PCV of 25 mM TRIS, pH 8.0, 10% glycerol, 250 mM NaCl (buffer A). Samples were heated at 65°C for 25 minutes to denature *E. coli* proteins, and then centrifuged at 14,500 x g and soluble extract was retained for purification. Protein expression was confirmed by visualization of appropriately sized bands on sodium dodecyl sulfate — polyacrylamide gel electrophoresis (SDS-PAGE) stained with Coomassie Brilliant Blue (CBB) dye.

Ni-NTA Superflow beads (QIAGEN) were equilibrated with buffer A in a one to one ratio. Three milliliters of the buffer A and bead resin was added to each soluble protein extract and tumbled gently at 4°C for one hour. The beads were washed three times with buffer A, then loaded onto a column (15 mm x 150 mm, KONTES) previously rinsed with buffer A. Protein was eluted using a linear imidazole gradient from 0 to 250 mM over 30 ml in a Fast Protein Liquid Chromatography (FPLC) system (Pharmacia). Protein elution was monitored by UV absorbance at 280 nm (Pharmacia LKB, Optical Unit UV-1) and collected in 1 ml fractions. Purity and yield were determined by protein visualization on CBB stained SDS-PAGE. Uncontaminated fractions were pooled and concentrated using Amicon Ultra-4 centrifugal filter devices (Millipore).

Experimental SAXS data of protein samples in buffer A were collected at the SIBYLS Beamline 12.3.1 of the Advanced Light Source (Lawrence Berkeley National Laboratory). Three proteins, lysozyme (14.3 kDa), bovine serum albumin (66.2 kDa), and glucose isomerase (172 kDa), were used to estimate molecular weights with respect to zero angle scattering intensity and concentration. All samples were exposed at 12 KeV for 5 seconds, 50 seconds, and then 5 seconds again to check for radiation damage. Sso0257 data were collected at concentrations of 6.9, 3.5, and 1.8 mg/ml. Sso0768 data were collected at concentrations of 2.3, 1.1, and 0.6 mg/ml. Sso3167 data were collected at concentrations of 6.0, 3.0, and 1.5 mg/ml. CRYSOL [15] software was used to generate theoretical scattering curves from crystal structures of homologs. OLIGOMER [16] was used to predict multimeric composition from solved structures to best fit the experimental data. DAMMIN [17] was used to generate *ab initio* envelope predictions.

Results

Recombinant Protein Expression

Induced protein expression and solubility was evaluated for all three proteins by visualization of soluble extract fractions on a CBB stained SDS-PAGE (see Figure 1). Other heat-stable proteins were observed, most of which were common to all samples; however, the over-expression of an approximately 30 kDa polypeptide was clearly evident and unique to Sso0257. Another contaminating polypeptide of about 20 kDa was observed exclusively in the

over-expression of Sso0768 (see Figure 1). The over-expression of Sso3167 was visualized by SDS-PAGE as a small band; however, because of the low molecular weight (18 kDa), the protein ran with the dye front, making it difficult to resolve (see Figure 1). All three recombinant proteins (Sso0257, 0768, and 3167) were abundantly over-expressed in *E. coli* and resulted in heat stable and soluble recombinant proteins.

Figure 1. SDS-PAGE shows heat stable protein extracts for uninduced and induced cultures. Target proteins are circled in red.

Sso0257 CDC6/ORC1

Large scale purification of Sso0257 was carried out and purity and abundance were determined by CBB SDS-PAGE. Peak elution fractions were pooled and concentrated to 6.9 mg/ml. Dilutions were then made using buffer A to concentrations of 3.5 mg/ml and 1.8 mg/ml.

Scattering curves for the different concentrations suggest non-specific aggregation at higher concentrations (see Figure 2A). This is also evident in the Guinier plots. At a low concentration (1.8 mg/ml) the Guinier plot is linear (see Figure 2B) but at the highest concentration (6.9 mg/ml) the plot lacks linearity (see Figure 2C). Scattering curves for multiple exposure times show no X-ray induced damage at the lower angles (see Figure 2D) nor is there damage evident in the higher angles (see Figures 4A–B). The radius of gyration determined by the linear range of the Guinier plot at a concentration of 1.8 mg/ml is 32.0 Å. The molecular weight of Sso0257 was experimentally determined from SAXS data by extrapolation on a standard curve generated from three known proteins with respect to zero angle intensity and sample concentration (see Figure 3). This experimental data is most consistent with the monomeric state of the recombinant Sso0257, which has a molecular weight of 46 kDa, including six-histidine tag (see Figure 3). The crystal structure of the homologous *Pyrobaculum aerophilum* CDC6p protein (1FNN.pdb) was used for comparison with experimental data [18]. This homolog shares 27% sequence identities with the Sso0257 protein and forms a dimer in the crystal structure. CRYSOL was used to generate scattering curves from the entire structure as well as the single chain A to provide dimer and monomer theoretical scattering curves to compare with the Sso0257 scattering data. However, neither theoretical curve was in good agreement with the experimental data (see Figures 4A–B). Therefore, the OLIGOMER program was used to estimate mulitimeric composition. The best fit to the experimental data was a 50/50 mixture of monomer and dimer states (see Figure 5). DAMMIN was used to construct an *ab initio* shape prediction for the SAXS data, which was superimposed on the CDC6p dimer structure as well as the single chain A (monomer) structure; however, the *ab initio* shape did not agree with either form of the crystal structure (see Figures 4C–D).

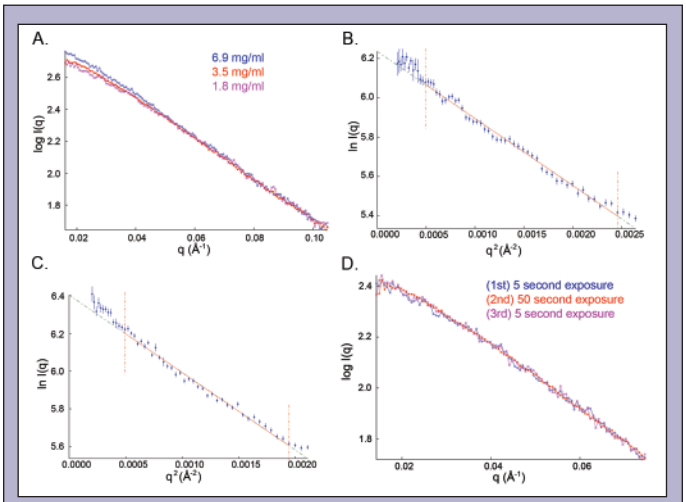


Figure 2. Sso0257 SAXS Data A) Scattering curve shows non-specific aggregation at higher concentrations. B) Linear range of Guinier plot at a concentration of 1.8 mg/ml suggests no aggregation. C) Guinier plot at highest concentration (6.9 mg/ml) shows aggregation. D) Scattering curve (3.5 mg/ml) at various exposure times indicates no X-ray induced damage.

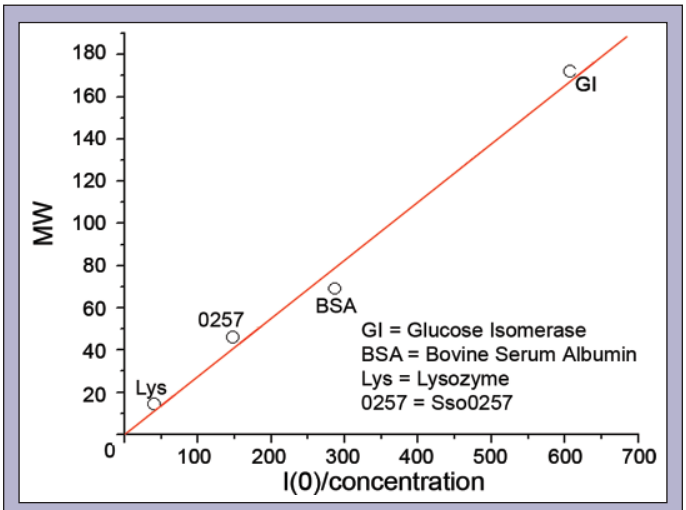


Figure 3. Sso0257 molecular weight estimation. Red line indicates standard curve based on three proteins with respect to known molecular weight, zero angle intensity and sample concentration.

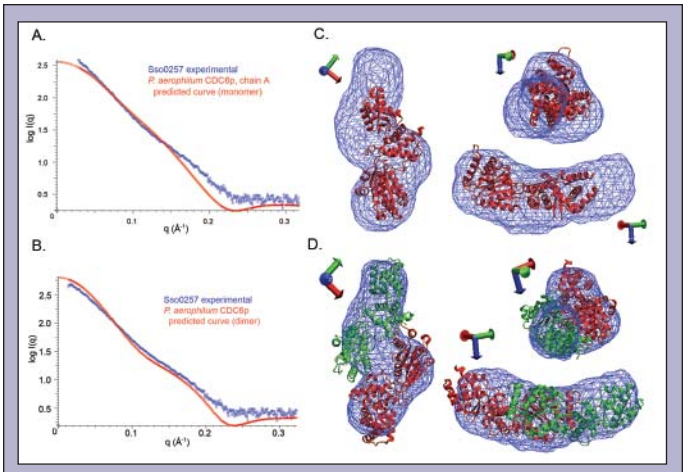


Figure 4. Sso0257 SAXS data compared with homologous crystal structure. A) Experimental scattering curve superimposed on theoretical scattering curve for *P. aerophilum* CDC6p, chain A (monomer). B) Experimental scattering curve superimposed on theoretical scattering curve for *P. aerophilum* CDC6p dimer structure. C) Sso0257 DAMMIN *ab initio* shape prediction superimposed on *P. aerophilum*, chain A crystal structure in three orientations. D) Sso0257 DAMMIN *ab initio* shape prediction superimposed on *P. aerophilum* CDC6p dimer crystal structure in three orientations.

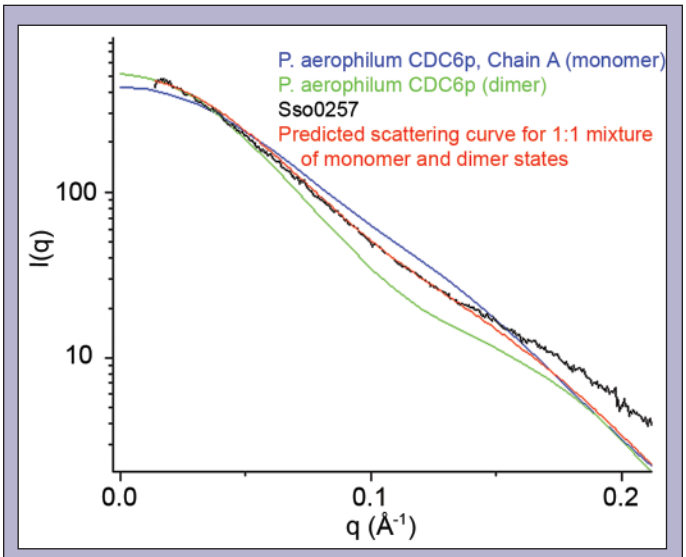


Figure 5. Scattering curves for Sso0257 experimental data, *P. aerophilum* CDC6p theoretical curves for monomer and dimer, and OLIGOMER best fit curve.

Sso0768 Replication factor C (RFC), Activator 1, small subunit

Protein purity and abundance from the over-expression of Sso0768 were determined by CBB SDS-PAGE, and then peak elution fractions were pooled and concentrated to 2.3 mg/ml. Dilutions were then prepared in buffer A to concentrations of 1.1 mg/ml and 0.6 mg/ml.

SAXS curves for the lower concentrations (1.1 mg/ml and 0.6 mg/ml) show different oligomerization states for different concentrations (see Figure 6A). A Guinier plot is not linear for the 1.1 mg/ml sample concentration suggesting aggregation (see Figure 6B). The radius of gyration approximated from the Guinier plot is 34.3 Å. An archaeon RFC small subunit crystal structure from the hyperthermophile *Pyrococcus furiosus* [19] with 59% sequence identities to Sso0768 was available for comparison with the experimental scattering curves. The *P. furiosus* RFC small subunit crystal structure (1IQP.pdb) is a hexamer, therefore theoretical scattering curves were predicted using CRYSOL from the six chains of the structure (monomer based on chain A, dimer based on chains A and B, trimer based on chains A, B, and C, tetramer based on chains A, B, C, and D, pentamer based on chains A, B, C, D, and E, hexamer based on complete crystal structure) [19]. Neither the monomer nor the hexamer theoretical curves coincide with the experimental curve (see Figures 7A–B). OLIGOMER was used to estimate the composition of oligomeric states. The best fit to the experimental data was to a mixture of 69% monomer, 13% trimer, 3% tetramer, and 15% hexamer (see Figure 7C).

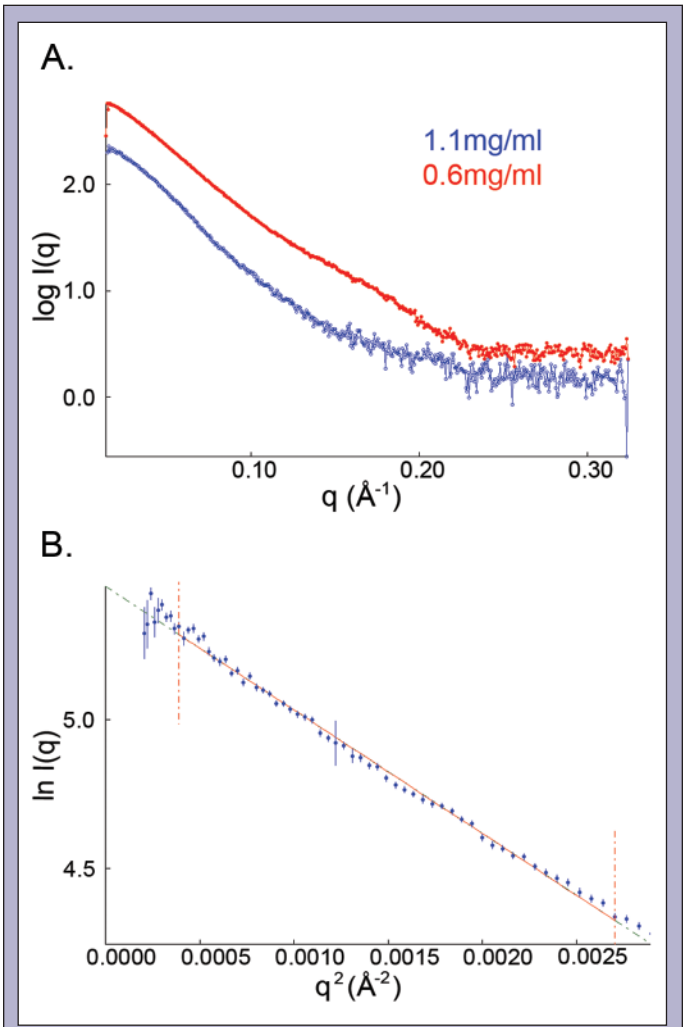


Figure 6. Sso0768 SAXS data A) Scattering curves at different concentrations display increased oligomerization at higher concentrations. B) Guinier plot at a concentration of 1.1 mg/ml shows slight aggregation.

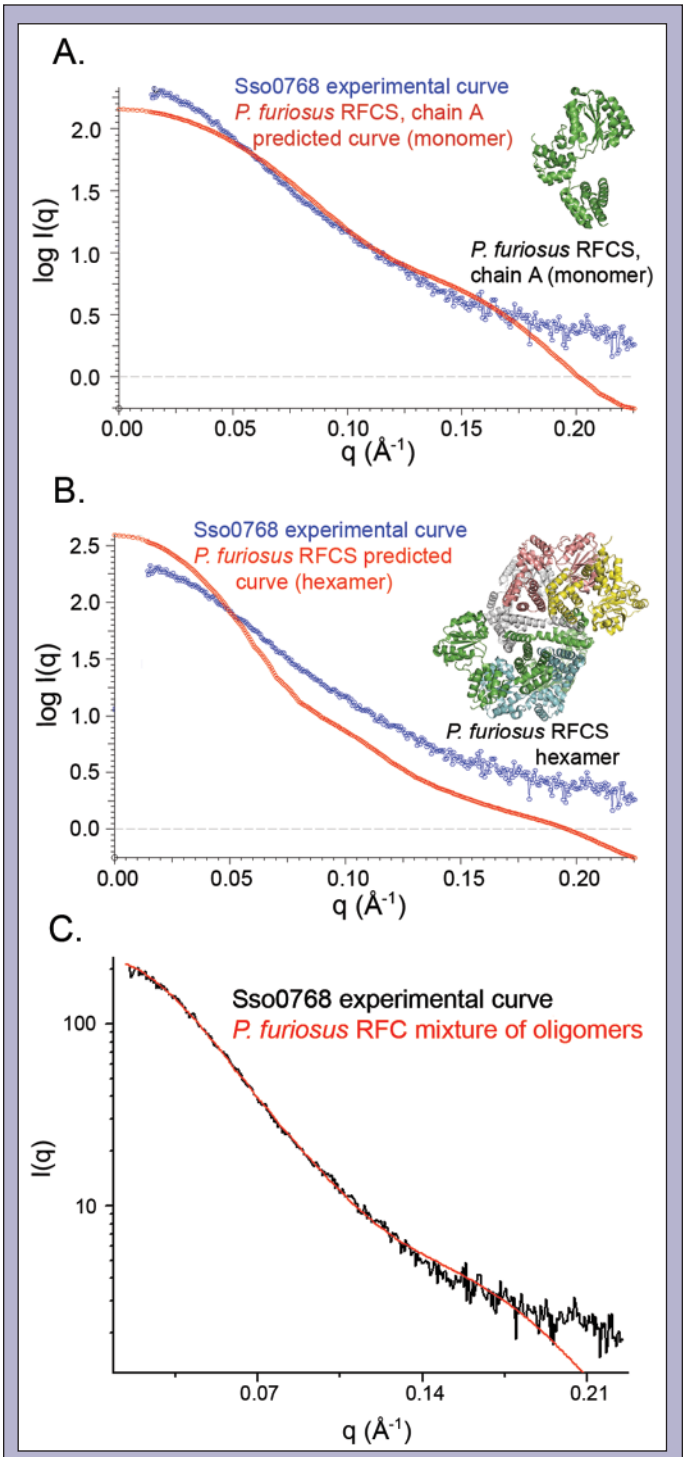


Figure 7. Sso0768 SAXS data compared with *P. furiosus* homologous crystal structure. A) Experimental data superimposed on the theoretical scattering curve for the *P. furiosus* RFC small subunit, chain A (monomer). Also shown is the crystal structure from which the theoretical curve was derived. B) Experimental data superimposed on the theoretical scattering curve for the *P. furiosus* RFC small subunit hexamer. Also shown is the crystal structure used to generate the theoretical curve. C) Sso0768 experimental data overlaid with OLIGOMER best fit for *P. furiosus* RFC small subunit complex mixture of oligomeric states.

Sso3167 Mut-T like protein

Protein purity and abundance were determined by CBB SDS-PAGE after large scale protein preparations. Fractions corresponding to the elution peak were pooled and concentrated to 6.0 mg/ml and dilutions were then prepared in buffer A to concentrations of 3.0 mg/ml and 1.5 mg/ml.

Experimental scattering curves for various concentrations and exposure times are in agreement suggesting no concentration dependence or radiation damage (see Figures 8A–B). At a low concentration (1.5 mg/ml), the experimental data fits linearly on a Guinier plot of the lower angles indicating sample homogeneity (see Figure 8C), but at high concentration (6.0 mg/ml), the data begins to lack linearity suggesting slight aggregation or an elongation within the protein (see Figure 8D). The radius of gyration given by the linear range of the Guinier plot is 18.7 Å at a concentration of 1.5 mg/ml and 19.7 Å at a concentration of 6 mg/ml.

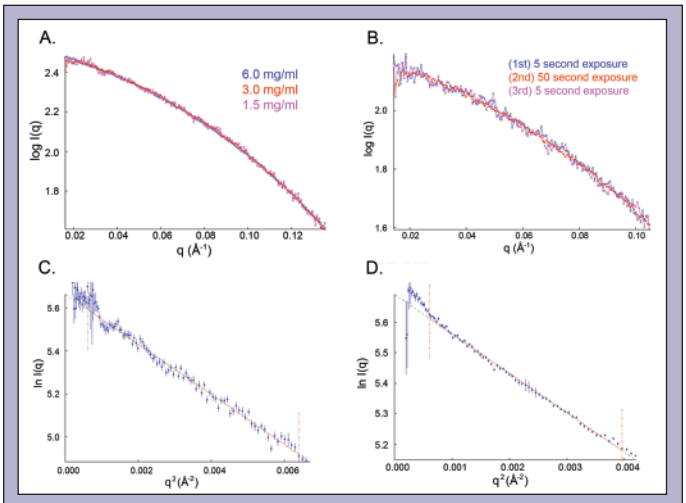


Figure 8. Sso3167 SAXS data. A) Scattering curves for various concentrations are consistent with each other suggesting no concentration dependence. B) Scattering curve of low angles for consecutive exposure times shows no X-ray induced damage. C) At a low concentration (1.5 mg/ml) the Guinier plot is linear suggesting sample homogeneity. D) At a higher concentration (6 mg/ml), the Guinier plot begins to lack linearity, which may correspond to slight aggregation or an elongated portion of the protein.

The *Deinococcus radiodurans* Mut-T/NADIX, chain A crystal structure [20] was used for comparison with the Sso3167 experimental data. The *D. radiodurans* homolog (1NQY.pdb) shares 27% sequence identities to Sso3167 including a similar large region of disorder. A theoretical scattering curve was generated from the crystal structure using CRYSOL. When overlaid with the experimental data, the theoretical curve is similar, but offset at higher angles indicating a deviation in overall protein shape (see Figure 9A).

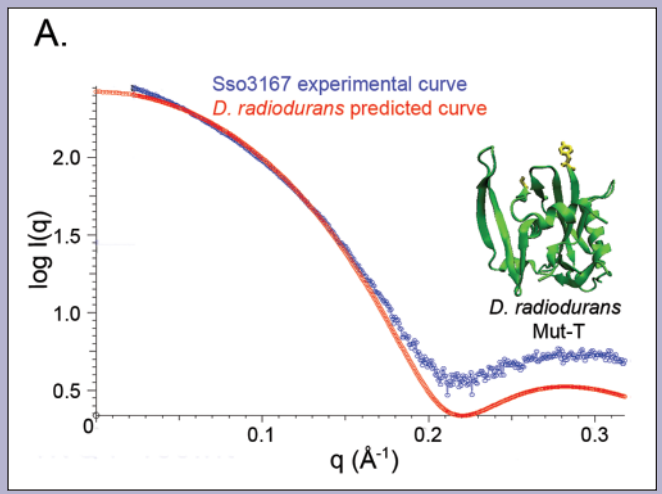
DAMMIN was used to generate *ab initio* shape predictions from the experimental scattering curves. Ten individual runs using the q range up to 0.21 Å consistently predicted an extension in various orientations. These ten individual runs were averaged to generate the final shape prediction (See Figure 9B). The *D. radioduran* crystal structure was superimposed on the shape prediction for

the Sso3167 data. The crystal structure is missing eleven residues which may fit in the elongated potion of the envelope (see Figure 9B). Although normalized spatial discrepancy has previously been suggested to determine goodness of fit for three-dimensional objects [21], statistical methods have proven unreliable in accessing the fit of a model. Instead, best fit of the shape prediction is based on experimental repetition and visually comparing the final models [22].

Discussion and Conclusions

Sso0257 CDC6-1

Scattering data at higher concentrations (6.9 mg/ml) shows non-specific aggregation (see Figure 2A). Although the homologous *P. aerophilum* CDC6p crystal structure is dimeric, its theoretical scattering profile is not in good agreement with the experimental data (see Figure 4A). Interestingly, the theoretical curve for the single chain A (monomer) also did not agree with the SAXS data (see Figure 4B). Using a standard curve based on three proteins, the estimated molecular weight for Sso0257 is 46kDa, which is consistent with a monomeric state (see Figure 3); however the *ab initio* shapes appear



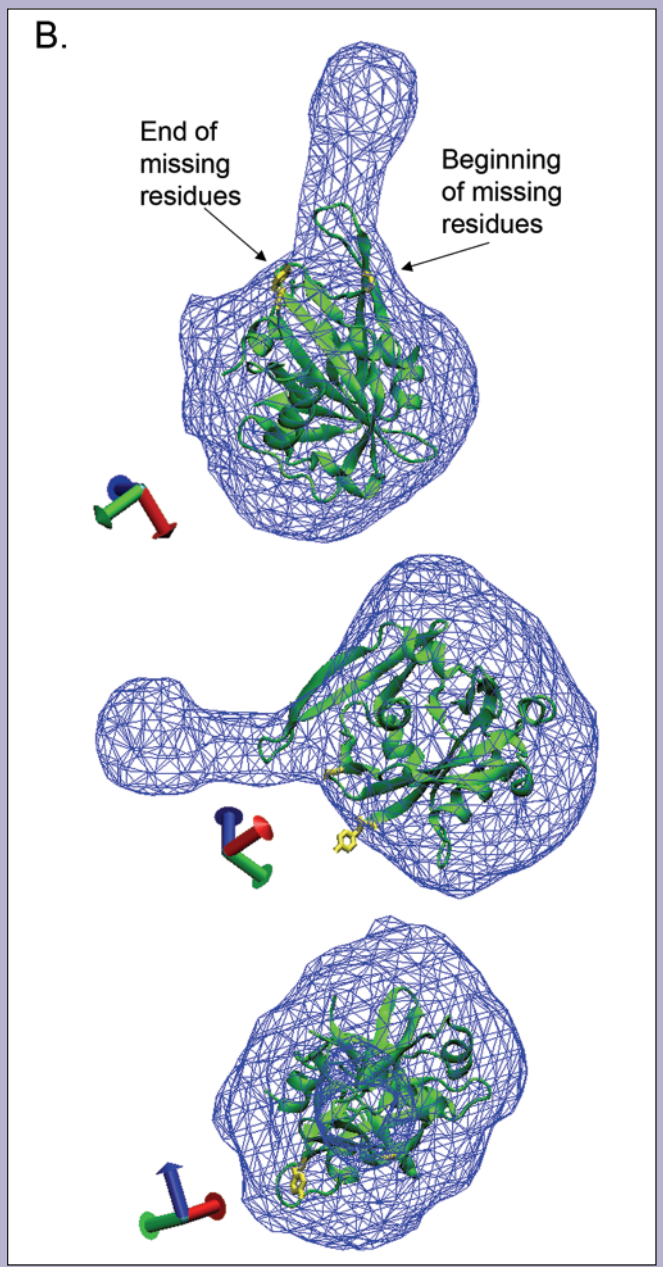


Figure 9. Sso3167 experimental data compared with homologous crystal structure. A) Sso3167 scattering curve superimposed on the theoretical curve for the *D. radiodurans* Mut-T crystal strucutre shows slight deviation at the higher angles. B) Sso3167 DAMMIN *ab initio* shape prediction superimposed on the *D. radiodurans* Mut-T crystal structure in three orientations.

too large to be monomeric when compared with the *P. aerophilum* monomer structure (see Figure 4C). The shape also conflicts with the *P. aerophilum* dimer structure (see Figure 4D). It is probable that the discrepancy in the predicted low resolution shape and the homologous crystal structures is attributed to a flexible winged helix domain within the protein. The flexibility of this domain in solution may cause the shape reconstructions to be larger than the monomeric crystal structure. The functional relevance of this domain is evident in the recently published crystal structure of CDC6-1 complexed with CDC6-3 and origin DNA (2QBY.pdb) where the winged helix domain is deeply inserted into the DNA grooves and effectively deforms the duplex [12]. Another possibility for the inconsistency in the shape predictions and the crystal structures is that the sample is a mixture of monomer and dimer states. This is supported by the best fit curve from the OLIGOMER analysis based on predicted scattering profiles for the *P. aerophilum* monomer and dimer (see figure 5). In order to discriminate between these two possibilities, the protein sample will be further purified by size exclusion. This will ensure monodispersity in the sample and will eliminate the possibility of a mixture of oligomeric states. The protein will then be analyzed with SAXS and again compared to the homologous crystal structure to determine whether the size difference between *ab initio* shape and the crystal structure is attributed to the flexibility of the winged helix domain.

In a separate article published at the same time as the CDC6-1, CDC6-3 and origin DNA complex, the *A. pernix* ORC1 protein was crystallized with a 22bp canonical origin of replication sequence [23] showing DNA-binding activity for both the expected winged helix domain, but also, surprisingly, for the ATPase domain as well. Therefore, further studies will be conducted using SAXS on the conformational changes assumed in presence of bubble or forked DNA and ATP.

Sso0768 RFC6, Activator 1, small subunit

Comparison of scattering data at 1.1 mg/ml and 0.6 mg/ml shows different scattering curves, which indicates aggregation and higher oligomerization of this protein at higher concentrations (see Figure 6A). The *P. furiosus* RFC small subunit crystal structure is a hexamer, but neither the theoretical scattering curve derived from the hexamer nor the monomer state agree with the experimental scattering curve for Sso0768 (see Figures 7A–B). The hexameric crystal structure has two-fold symmetry and can be viewed as two trimer states (see Figure 10). This view supports the OLIGOMER fit of mostly monomers, trimers and hexamers to the SAXS data (see Figure 7C). It is likely that at higher concentrations the Sso0768 protein forms trimers, which then assemble into hexamers because of interactions between the proteins. As a clamp loader, this oligomeric state may also serve a functional purpose loading the polymerase onto DNA. Sso0768 will be further purified by size exclusion and will be analyzed by SAXS at increased temperature (40°C and 60°C) to study what affect temperatures closer to biological conditions will have on the oligomerization state. These methods will elucidate the functional and physiological relevance of the oligomeric state and should resolve whether trimers and hexamers are only formed at high concentrations or if they are functionally significant. Also, Sso0768 will be co-expressed with the RFC large subunit to investigate any conformational or oligomeric changes brought about by interactions between the subunits in solution.

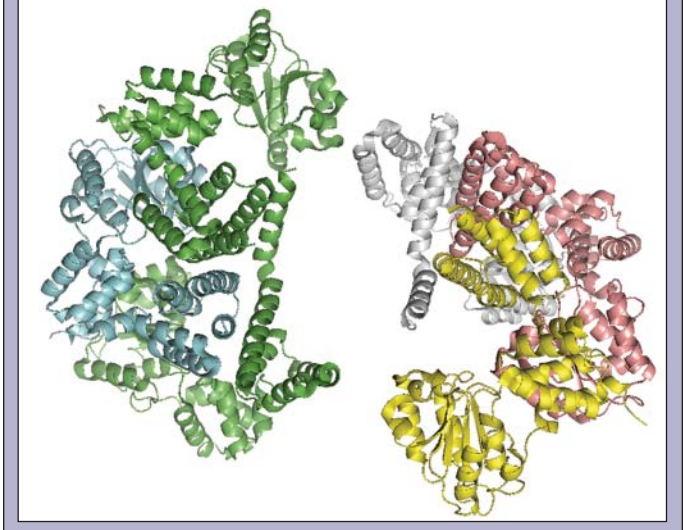


Figure 10. *P. furiosus* RFC small subunit hexamer arranged into two trimers.

Sso3167 Mut-T like protein

The predicted scattering curve from the *D. radiodurans* homolog and the experimental scattering curve are analogous but do not overlay completely at larger angles (see Figure 9A). This discrepancy may be caused by an extended region of the protein that is not resolved in the crystal structure. This possibility is further supported by the *ab initio* shape prediction for Sso3167, which contains a protrusion that may correspond to the eleven missing residues in the crystal structure of the *D.radiodurans* homolog (see Figures 9B). Both Sso3167 and the *D. radiodurans* Mut-T protein contain large regions of disorder based on sequence as predicted by PONDR [24,25,26] (data not shown). It is likely that these disordered regions occupy the extension in the shape prediction, transitioning from a floppy domain in solution to a functional region in the presence of a substrate. This feature may also help to recognize oxidized dGTP and catalyze the reaction to remove it from the free nucleotide pool. Possible changes in conformation and oligomeric states in the presence of oxidized dGTP and other nucleotides will be studied using SAXS to evaluate the functional role of the elongation in enzymatic activity.

These three archaeal proteins were chosen because of their involvement in processes that are similar to and representative of more complicated processes in Eukaryotes and their preliminary characterization emphasizes the utility of hyperthermophilic proteins as models for higher eukaryotic processes. Sso0257 is an essential component in origin recognition and has also been implicated in cell division control [18]. Understanding how this protein recognizes origins and initiates replication will help to shed light on how and why mutations occur during this process. Sso0768 is a RFC subunit thought to help load the processivity factor onto DNA. The mechanism for this is still unclear as is the role of the large

subunit and its possible interaction with Sso0768, the processivity factors and DNA. Sso3167 is a Mut-T like protein. This family of proteins is important in preventing mutated nucleotides from being incorporated into DNA. The expression, purification and initial SAXS analysis of these proteins demonstrates the applicability of thermo-stable archaeal proteins to provide simplified model systems for transitory eukaryotic processes.

Acknowlegements

I would like to thank my mentor, Dr. Steven M. Yannone, for his patient guidance. Thanks also to Greg Langland, Imran Khan, Jill Fuss, Susan Tsutakawa, Denise Muñoz, and everyone in the Cooper lab, SIBYLS Beamline 12.3.1, CSEE, and Bio-Link. Lastly, thanks to the Lawrence Berkeley National Laboratory for hosting me over the summer and the Department of Energy and the Office of Science for this fantastic opportunity to participate in the Faculty and Student Team (FaST) program.

References

[1] C.J. Bult, O. White, G.J. Olsen, L. Zhou, R.D. Fleischmann, G.G. Sutton, J.A. Blake, L.M. Fitzgerald, R.A. Clayton, J.D. Gocayne, A.R. Kerlavage, B.A. Dougherty, J.F. Tomb, M.D. Adams, C.I. Reich, R. Overbeek, E.F. Kirkness, K.G. Weinstock, J.M. Merrick, A. Glodek, J.L. Scott, N.S. Geoghagen, and J.C. Venter, "Complete Genome Sequence of the Methanogenic Archaeon, *Methanococcus jannaschii*," Science, vol. 273, no. 5278, Aug., 1058–1073, 1996.

[2] H.P. Klenk, R.A. Clayton, J.F. Tomb, O. White, K.E. Nelson, K.A. Ketchum, R.J. Dodson, M. Gwinn, E.K. Hickey, J.D. Peterson, D.L. Richardson, A.R. Kerlavage, D.E. Graham, N.C. Kyrpides, R.D. Fleischmann, J. Quackenbush, N.H. Lee, G.G. Sutton, S. Gill, E.F. Kirkness, B.A. Dougherty, K. McKenney, M.D. Adams, B. Loftus, S. Peterson, C.I. Reich, L.K. McNeil, J.H. Badger, A. Glodek, L. Zhou, R. Overbeek, J.D. Gocayne, J.F. Weidman, L. McDonald, T. Utterback, M.D. Cotton, T. Spriggs, P. Artiach, B.P. Kaine, S.M. Sykes, P.W. Sadow, K.P. D'Andrea, C. Bowman, C. Fujii, S.A. Garland, T.M. Mason, G.J. Olsen, C.M. Fraser, H.O. Smith, C.R. Woese, and J.C. Venter, "The Complete Genome Sequence of the Hyperthermophilic, Sulphate-reducing Archaeon *Archaeoglobus fulgidus*," Nature, vol. 390, Nov., 364–370, 1997.

[3] Y. Kawarabayasi, Y. Hino, H. Horikawa, K. Jin-no, M. Takahashi, M. Sekine, S. Baba, A. Ankai, H. Kosugi, A. Hosoyama, S. Fukui, Y. Nagai, K. Nishijima, R. Otsuka, H. Nakazawa, M. Takamiya, Y. Kato, T. Yoshizawa, T. Tanaka, Y. Kudoh, J. Yamazaki, N. Kushida, A. Oguchi, K. Aoki, S. Masuda, M. Yanagii, M. Nishimura, A. Yamagishi, T. Oshima, and H. Kikuchi, "Complete Genome Sequence of an Aerobic Thermoacidophilic Crenarchaeon, *Sulfolobus tokodaii* Strain 7," DNA Research, vol. 8, 123–140, 2001.

[4] Q. She, R.K. Singh, F. Confalonieri, Y. Zivanovic, G. Allard, M.J. Awayez, C.C.Y. Chan-Weiher, I.G. Clausen, B.A. Curtis, A. De Moors, G. Erauso, C. Fletcher, P.M.K. Gordon, I. Heikamp-de Jong, A.C. Jeffries, C.J. Kozera, N. Medina, X. Peng, W.F. Doolittle, M. Duguet, T. Gaasterland, R.A. Garrett, M.A. Ragan, C.W. Sensen and J. Van der Oost, "The Complete Genome of the Crenarchaeon *Sulfolobus solfataricus* P2," Proceedings of the National Academy of Sciences, vol. 98, no. 14, July, 7835–7840, 2001.

[5] W. Zillig, K.O. Stetter, and D. Janekovic, "DNA-dependent RNA Polymerase from the Archaebacterium *Sulfolobus acidocaldarius*," European Journal of Biochemistry, vol. 96, June, 597–604, 1979.

[6] J.N. Reeve, K. Sandman, and C.J. Daniels, "Archaeal Histones, Nucleosomes, and Transcription Initiation," Cell, vol. 89, June, 999–1002, 1997.

[7] M. Thomm, "Archaeal Transcription Factors and Their Role in Transcription Initiation," FEMS Microbiology Review, vol. 18, May, 159–171, 1996.

[8] S.A. Qureshi and S.P. Jackson, "Sequence-specific DNA Binding by the *S. shibatae* TFIIB Homolog, TFB, and its Effect on Promoter Strength," Molecular Cell, vol. 1, Feb., 389–400, 1998.

[9] F. Pfeifer, K.H. Schleifer, P. Palm, Eds., Molecular Biology of Archaea, Munich: John Wiley & Sons, 1994.

[10] M. Ciaramella, F.M. Pisani, and M. Rossi, "Molecular Biology of Extremophiles: Recent Progress on the Hyperthermophilic Archaeon *Sulfolobus*," Antonie van Leeuwenhoek, vol. 81, 85–97, 2002.

[11] M. De Felice, L. Esposito, M. Rossi, and F. Pisani, "Biochemical Characterization of Two Cdc6/ORC1-like Proteins from the Crenarchaeon *Sulfolobus solfataricus*," Extremophiles, vol. 10, 61–70, 2006.

[12] E.L. Cunningham Dueber, J.E. Corn, S.D. Bell, J.M. Berger, "Replication Origin Recognition and Deformation by a Heterodimeric Archaeal Orc1 Complex," Science, vol. 317, 1210–1213, 2007.

[13] F. Pisani, M. De Felice, Fl. Carpentieri, and M. Rossi, "Biochemical Characterization of a Clamp-loader Complex Homologous to Eukaryotic Replication Factor C from the Hyperthermophilic Archaeon *Sulfolobus solfataricus*," Journal of Molecular Biology, vol. 301, 61–73, 2000.

[14] J.Y. Mo, H. Maki, M. Sekiguchi, "Hydrolytic Elimination of a Mutagenic Nucleotide, 8-oxoGTP, by Human 18-kilodalton protein: Sanitization of Nucleotide Pool," Proceedings of

the National Academy of Sciences, vol. 89, 11021–11025, 1992.

[15] D.I. Svergun, C. Barberato, and M.H.J. Koch, “CRYSOL — a Program to Evaluate X-Ray Solution Scattering of Biological Macromolecules from Atomic Coordinates,” Journal of Applied Crystallography, vol. 28, 768–773, 1995.

[16] P.V. Konarev, V.V. Volkov, A.V. Sokolova, M.H.J. Koch, and D.I. Svergun, “PRIMUS: a Windows PC-based system for Small-Angle Scattering Data Analysis,” Journal of Applied Crystallography, vol. 36, 1277–1282, 2003.

[17] D.I. Svergun, “Restoring Low Resolution Structure of Biological Macromolecules from Solution Scattering Using Simulated Annealing,” Biophysical Journal, vol. 76, June, 2879–2886, 1999.

[18] J. Liu, C.L. Smith, D. DeRyckere, K. DeAngelis, G.S. Martin, J.M. Berger, “ Structure and Function of Cdc6/ Cdc18: Implications for Origin Recognition and Checkpoint Control”, Molecular Cell, vol. 6, 637–648, 2000.

[19] T. Oyama, Y. Ishino, I.K.O. Cann, S. Ishino, and K. Morikawa, “Atomic Structure of the Clamp Loader Small Subunit for *Pyrococcus furiosus*,” Molecular Cell, vol. 8, Aug., 455–463, 2001.

[20] L. Kang, S.B. Gabelli, M.A. Bianchet, W.L. Xu, M.J. Bessman, and L.M. Amzel, “Structure of a Coenyme A Pyrophosphatase from *Deinococcus radiodurans*: a Member of the Nudix Family,” Journal of Bacteriology, vol. 185, no. 14, July, 4110–4118, 2003.

[21] M.B. Kozin and D.I. Svergun, “Automated matching of low- and high-resolution structural models,” Journal of Applied Crystallography, vol. 34, no. 1, Feb., 33–41, 2001.

[22] C.D. Putnam, M. Hammel, G.L. Hura, and J.A. Tainer, “Solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution,” Quarterly Reviews in Biophysics, vol. 40, no. 3, Aug., 191–285, 2007.

[23] M. Gaudier, B.S. Schuwirth, S.L. Westcott, D.B. Wigley, “Structural Basis of DNA Replication by an ORC protein,“ Science, vol. 317, Aug., 1213–1216, 2007.

[24] X. Li, P. Romero, M. Rani, A.K. Dunker, and Z. Obradovic, “Predicting protein disorder for N-, C-, and internal regions,” Genome Informatics, vol. 10, 30–40, 1999.

[25] P. Romero, Z. Obradovic, X. Li, E. Garner, C. Brown, and A.K. Dunker, “Sequence complexity of disordered protein,” Proteins: Structure, Function, and Genetics, vol. 42, no. 1, Jan., 38–48, 2001.

[26] P. Romero, Z. Obradovic, and A.K. Dunker, “Sequence data analysis for long disordered regions prediction in the calcineurin family,” Genome Informatics, vol. 8, 110–124, 1997.