



LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

**FY05 LDRD Final Report Investigation
of AAA+ protein machines that
participate in DNA replication,
recombination, and in response to
DNA damage LDRD project tracking
code: 04-LW-049**

D. Sawicka, M. Simoes de Carvalho-Kavanagh, D.
Barsky, C. Venclovas

December 4, 2006

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

Auspices Statement

This work was performed under the auspices of the U. S. Department of Energy (DOE) by the University of California, Lawrence Livermore National Laboratory (LLNL) under Contract No. W-7405-Eng-48. The project (04-LW-049) was funded by the Laboratory Directed Research and Development Program at LLNL.

FY05 LDRD Final Report
**Investigation of AAA+ protein machines that participate in
DNA replication, recombination, and in response to DNA
damage**
LDRD Project Tracking Code: 04-LW-049
PI: Dorota Sawicka

Research team

Dorota Sawicka, post doctoral fellow (Ph.D. in chemistry) - expansion of the current system for continuous update of sequence-based searches and application of comparative modeling

Marianne Simoes de Carvalho-Kavanagh, staff scientist (Ph. D. in biology) - analysis of protein-protein interactions using both yeast-two hybrid and co-immunoprecipitation techniques.

Daniel Barsky, staff scientist (Ph. D. in biophysics) - application of molecular dynamics simulations and analysis of protein surface properties.

Collaborators

Česlovas Venclovas, staff scientist (Ph. D. in chemistry) - application of comparative modeling
Institute of Biotechnology, LT-02241 Vilnius, Lithuania

Vladimir Bermudez and Jerard Hurwitz
Memorial Sloan Kettering Cancer Center, New York, NY 10021, USA

Abstract

The AAA+ proteins are remarkable macromolecules that are able to self-assemble into nanoscale machines. These protein machines play critical roles in many cellular processes, including the processes that manage a cell's genetic material, but the mechanism at the molecular level has remained elusive. We applied computational molecular modeling, combined with advanced sequence analysis and available biochemical and genetic data, to structurally characterize eukaryotic AAA+ proteins and the protein machines they form. With these models we have examined intermolecular interactions in three-dimensions (3D), including both interactions between the components of the AAA+ complexes and the interactions of these protein machines with their partners. These computational studies have provided new insights into the molecular structure and the mechanism of action for AAA+ protein machines, thereby facilitating a deeper understanding of processes involved in DNA metabolism.

Background and significance

Biochemical evolution is driven by selection within “trial and error” events. It is now becoming apparent that evolution has selected highly-ordered protein systems to power many cellular processes. The AAA+ class of proteins (ATPases Associated with various cellular Activities) represent perhaps one of the best examples of such ordered systems. The proteins belonging to this ancient class play critical roles in essentially every step of DNA replication (copying of DNA prior to each cell division), recombination (exchanges between different DNA molecules in a cell) and response to DNA damage (delay of the cell cycle, repair of damaged DNA sites). For example, protein complexes containing AAA+ proteins are critical for initiation (e.g., ORC, Cdc6, Mcm2-7) and progress of DNA replication (e.g., RFC). Altogether, there are at least twenty protein families within the AAA+ class involved in DNA metabolism in eukaryotes identified to date. The AAA+ protein class is an excellent illustration of how a once-found solution is reused by nature again and again. AAA+ proteins share a conserved two-domain structure of approximately 220 amino acids. This structure contains several conserved sequence patterns including “Walker A” and “Walker B” motifs that form an ATP binding pocket. ATP binding, hydrolysis, and subsequent dissociation of the product (ADP) are coupled with the conformational changes between the two conserved structural domains. It is these ATP-fueled motions that make an AAA+ protein a molecular motor. Just as in everyday life where motors run different machinery, AAA+ proteins make a variety of protein machines. An assembly of individual AAA+ protein subunits into protein complexes and the addition of structural domains are the two major ways contributing to their diversity. Variation along these two themes in the course of evolution has produced a large number of different AAA+ protein machines interacting with different partners and performing different molecular functions. The prerequisite for elucidation of the molecular mechanisms of AAA+ proteins is the knowledge of the three-dimensional (3D) structure of individual components and higher-order complexes as well as the intermolecular interactions. Many of the AAA+ proteins involved in DNA replication, recombination and in response to DNA damage have been studied by genetic and biochemical methods. However, only few proteins and even fewer functional complexes within these functional categories have experimentally determined (X-ray or NMR) 3D structures. Moreover, to our knowledge, no experimentally determined structures to date are available for the eukaryotic AAA+ complexes functioning in DNA metabolism.

Methods Used

At the cornerstone of this proposal is the application of comparative modeling to generate structures of AAA+ proteins and their complexes. The basic concept of comparative modeling relies on the observation that structural features of proteins are conserved during evolution to a much higher degree than sequences, and proteins related even by distant sequence similarity can be expected to have similar 3D structures [1]. Therefore, once a 3D structure is determined for at least one representative of a protein family, models for other family members can be derived using the known structure as a template. To a large degree, structural conservation also applies to protein complexes. In other words, very often protein complexes formed by evolutionary related proteins share a similar higher-order arrangement.

Taking all this into account, comparative modeling is well suited to study AAA+ proteins. Despite ancient roots, and therefore very divergent protein sequences, at least the evolutionary conserved core can be modeled with sufficient reliability for essentially every AAA+ protein. In addition, the PDB now holds structures representing ~1000 of estimated few thousand folds that exist in nature [2]. Therefore, it is reasonable to expect that most of the additional divergent domains extending beyond the conserved core of AAA+ proteins can be related to the database structures and modeled by comparison to them. Along with the experimental structures for individual proteins, the PDB contains a large number of biologically relevant protein complexes, including those formed by AAA+ proteins. The structures of these complexes can be used in comparative modeling as higher-order templates enabling the exploration of various possible arrangements of individual subunits.

There are also prerequisites to address the dynamics of the AAA+ protein “machines” using molecular dynamics simulations. The PDB has examples of AAA+ structures in various functional states, such as nucleotide (ATP or ADP) bound or cofactor free. These different functional states coupled with molecular modeling can serve as starting points for the elucidation of the nature and dynamics of conformational changes that are inherent attributes of AAA+ protein “machines”.

The difficulties of studying the AAA+ protein “machines” using experimental methods is another motivation to use computational methods. Usually it is difficult or impossible to study AAA+ proteins as monomers, since most of them have a strong tendency to associate into complexes. The large size of the AAA+ complexes and their inherent instability complicates both purification and crystallization studies. Thus it is not surprising that there has been limited success in determining experimental 3D structures for AAA+ proteins despite high interest in these proteins worldwide.

The comparative modeling approach applied here has been developed by one of us (CV), and is one of the most effective in the field. Importantly, this comparative modeling approach proved to be efficient in the case of distant evolutionary relationship (sequence identity reaching as low as 12%), greatly expanding the general applicability of protein modeling. During the last two world-wide competitions of protein structure prediction methods conducted in a “blind” mode, the results obtained using this approach [3, 4] were given top ranks by independent experts [5, 6]. It is noteworthy, that superior results achieved in these competitions are considered to represent the current state-of-the-art in protein structure prediction. Furthermore, our comparative modeling approach has been used to predict a PCNA-like fold for the Rad1 family of proteins involved in both DNA repair and cell cycle control [7]. We have also shown that comparative modeling can be successfully extended to predict structures of protein complexes. Our recent prediction of the structure for the “9-1-1” checkpoint complex, which is important in repairing DNA damage and cell cycle control [8], is an example of such an extension. These molecular models have provided both new insights and a structural framework for the design of new experiments. Not surprisingly, the articles detailing these molecular models are among the most frequently cited papers that originated from the Biosciences Directorate in recent years. Most recently we have used comparative modeling to uncover novel putative interactions and to propose a mechanism of action for Replication Factor C, a complex formed by five AAA+ proteins [9]. The steps taken in that study (Fig. 1) formed the basis for the research in this project (see below). We also have extensive expertise in applying molecular dynamics simulations to biomolecules, and have considered the effects DNA damage to its structure and stability [10] and to dynamics of proteins bound to DNA [11].

Results

We have used molecular modeling and other computational techniques methods combined with the available experimental data to understand the structure, interactions and the molecular mechanisms for AAA+ proteins involved in various aspects of DNA metabolism, including DNA replication, recombination and response to DNA damage.

The research has the following major components: 1) Advanced protein sequence database searches and sequence analysis using profile-based methods; 2) Comparative modeling of individual AAA+ proteins; 3) Combining models with data from structural databases as well as available biochemical and genetic data to assemble AAA+ protein machines 4) Structure-function analysis using the obtained models.

- 1) Initially, we performed advanced protein amino acid sequence searches against available sequence databases and analyze all families of eukaryotic AAA+ proteins. This analysis was used to achieve several goals: a) to determine approximate domain boundaries for each AAA+ protein family, b) to identify related proteins with known experimental 3D structures that could be used as structural templates for modeling of either domains or complete structures of query proteins, c) to possibly identify new protein families that have not yet been assigned to the AAA+ class.
- 2) The results of the initial sequence-based analysis determined how molecular modeling was applied for each individual AAA+ protein family. If the whole query protein sequence could be matched to one or more structural templates, the model for the complete structure was built. The domains beyond the conserved core that did not produce statistically significant matches to known structures were further explored. A diverse set of fold recognition methods (methods that aim at detecting similarity of protein folding patterns in the absence of clear sequence signal) was used to identify related structures followed by full-atom model building.
- 3) Models for individual proteins were assembled into complexes based on comparison with multi-subunit structures available in the structural database (PDB). Available experimental data (such as point mutation, protein-protein interaction data) was used to both validate and refine the assembly of the individual protein molecules into multi-subunit complexes.
- 4) The properties of the obtained models both of the individual proteins and their complexes were comprehensively interrogated. This includes characterization of functionally important structural motifs, surface properties, and conformational transitions between different functional states.

Year One Accomplishments

System for Weekly Updates

A real-world problem in homology modeling is the need to monitor the emergence of new sequences and structures which may improve models in preparation. Typically all searches are done painstakingly by hand and may not be repeated for months. We have established a system that provides a continuous update of sequence-based searches by sending a weekly email for each protein corresponding to new structural templates that have been identified in the sequence search or to sequences that have increased in statistical significance. These structural templates can be used as a starting point for modeling the AAA+ proteins. Three dimensional models, derived from the identified structural templates, and conserved sequence patterns facilitate the prediction of key

residues that may be important structurally (i.e. in protein-protein interactions) or functionally (active site residues). We have created a webpage summarizing the top hits for each protein or protein domain of interest with links to current BLAST and PSIBLAST output files containing the detailed alignment information.

Comparative Modeling

Our initial modeling efforts have focused on the RFC1, Rad17, and Ctf18 members of the AAA+ family. RFC1, Rad17, and Ctf18 are eukaryotic proteins that can each form a distinct protein complex with the small RFC subunits (RFC2-5). These proteins are referred to as clamp loading complexes since they interact with proteins known as DNA clamps. One function of the RFC1-containing clamp loading complex is to load PCNA, a DNA sliding clamp onto primed DNA sites. PCNA is critical for processive DNA replication and also functions in many other processes of DNA metabolism. Recently, it has been shown that the Ctf18 clamp loading complex also interacts with PCNA. It is thought that the Ctf18 role in sister chromatid cohesion involves a specialized replication step. The third, Rad17-RFC complex, upon DNA damage, specifies the loading of a different sliding clamp, the Rad9-Rad1-Hus1 (9-1-1) heterotrimer.

We have conducted a phylogenetic analysis of RFC1, Ctf18, Rad17, PCNA, and the proteins of the 9-1-1 complex. Rad17 and the 9-1-1 complex are not universally conserved. *P. falciparum* apparently lacks both Rad17 and the 9-1-1 complex. The 9-1-1 proteins are missing in another eukaryotic parasite, *E. cuniculi*, but we have detected a putative Rad17 protein suggesting that Rad17 might have additional important molecular functions beyond clamp loading. RFC1, Ctf18, and their interacting clamp unlike Rad17 and 9-1-1 proteins were present in all analyzed species.

Since less is known about the specific interactions of Ctf18 with PCNA than in RFC1 with PCNA, we decided to focus our modeling efforts on the Ctf18-PCNA interaction. We have identified that amino acids, Val422 and Leu423, will potentially interact with the human DNA clamp protein, PCNA. Using curated family alignments with highly sensitive sequence profile comparison methods, we have discovered that Ctf18 shares a sequence similarity to two other AAA+ proteins, RFC1 and Rad17 in the C-terminal region beyond the AAA+ module. Ctf18, in contrast to RFC1 and Rad17, has an extended C terminus (approximately 60 amino acids). Since only Ctf18, and not RFC1 nor Rad17, has been found to form a complex with two additional proteins, DCC1 and Ctf8, this extended C-terminus could mediate an additional protein interaction between Ctf18 and the DCC1 or Ctf8 proteins.

Year Two Accomplishments

Due to our success in generating a discrete and testable hypothesis for the Ctf18 protein, we broadened the scope of our proposal to include an experimental effort to verify our computational predictions using Ctf18. Protein-protein interactions between computationally designed point mutants, Val422 and Leu423 of Ctf18 and PCNA were tested using yeast-two hybrid and co-immunoprecipitation techniques. We created mutants of Ctf18, designed using computational modeling, that we predicted would diminish the strength of binding of Ctf18 to PCNA, a replication protein, and to Dcc1/Ctf8, two proteins that may be involved in sister chromatid cohesion. We expressed mutants of Ctf18 in a baculovirus system and tested for protein-protein interactions with PCNA by co-immunoprecipitation. We tested the ability of these Ctf18 mutants to catalyze ATP-dependent loading of PCNA onto DNA. We found that replacement of valine with alanine at position 422 of Ctf18 decreases PCNA-dependent DNA replication by about 70%; 1664 U/mg specific

activity for mutant Ctf18(V422A) compared to 5007 U/mg specific activity for wild type Ctf18 (Vladimir Bermudez and Jerard Hurwitz, Sloan Kettering Cancer Center, unpublished data). Another experiment guided by our model for Ctf18, involved truncation of the extended C-terminal domain of Ctf18 and a check for the absence of protein-protein interactions between Ctf18-Dcc1 and Ctf18-Ctf8 proteins (Vladimir Bermudez and Jerard Hurwitz, Sloan Kettering Cancer Center, unpublished data). We found that a C-terminal truncation mutant of Ctf18 reduces binding to Dcc1/Ctf8 as expressed in an in-vitro transcription-translation system[12]. We further investigated these protein-protein interactions by-Biacore analysis of peptide fragments of Ctf18 binding to PCNA. This combined experimental and computational approach is general and can be applied to any of the AAA+ proteins.

We also tested how these AAA+ clamp loader proteins open their respective clamps. We explored the effect that RFC1 binding has on its clamp (PCNA). It has been known that many proteins interact with PCNA, but the RFC1 clamp loader is only one of two known proteins that can open PCNA. For this we used molecular dynamics to compare the differences in conformation between free PCNA, PCNA bound to our modeled peptide fragment of RFC1 and PCNA bound to the p21 peptide. The p21 peptide is similar to the RFC1 fragment, but its interaction with PCNA involves more residues, suggesting that these additional interactions may contribute to the difference in the PCNA ring stability.

Publication

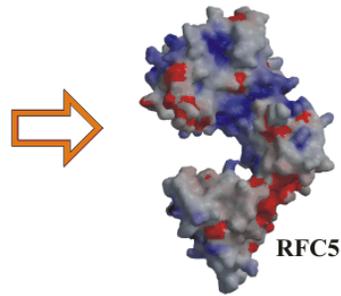
Processivity clamp interactions differentially alter the dual activities of UmuC (2006) *Molecular Microbiology*, 59(2): 460-474. UCRL-JRNL-216586

Modeling individual proteins

Sequence search and analysis;
Identification of the matching structural templates;

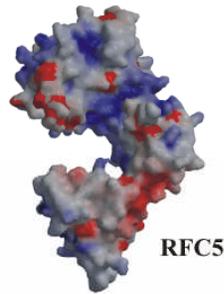


Functional structural motifs;
Surface properties;
Dynamic properties;



Assembling the RFC protein complex

Analysis of similar structural complexes;
Protein-protein interaction data;
Point mutation data;
Other biochemical/genetic data;



RFC complex



Proposed putative molecular mechanism for the RFC-dependent loading of a DNA clamp

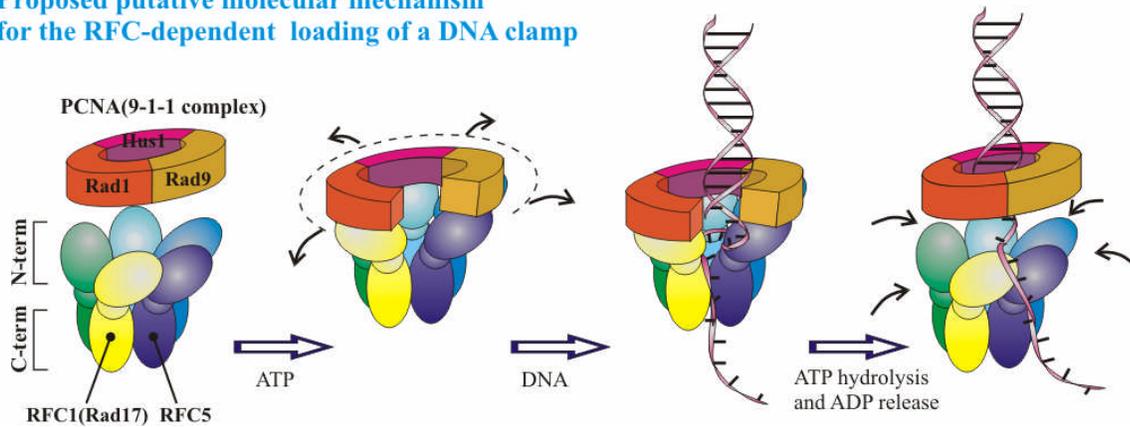


Figure 1. A flow chart for the computational modeling and analysis illustrating the main steps in the research plan. The scheme presented here has been successfully applied for the RFC complex, one of the AAA+ protein “machines” [9]. The three panels represent increasing levels of complexity in the computational study of AAA+ protein families.

References

1. Chothia C, Lesk AM. The relation between the divergence of sequence and structure in proteins. *Embo J* 1986;5:823-826.
2. Govindarajan S, Recabarren R, Goldstein RA. Estimating the total number of protein folds. *Proteins* 1999;35:408-414.
3. Venclovas C. Comparative modeling of CASP4 target proteins: Combining results of sequence search with three-dimensional structure assessment. *Proteins* 2001;Suppl. 5:47-54.
4. Venclovas C. Comparative modeling in CASP5: progress is evident, but alignment errors remain a significant hindrance. *Proteins* 2003.
5. Tramontano A, Morea V, Leplae R. Analysis and assessment of comparative modeling predictions in CASP4. *Proteins* 2001;Suppl. 5:22-38.
6. Tramontano A. Of men and machines. *Nat Struct Biol* 2003;10:87-90.
7. Thelen MP, Venclovas C, Fidelis K. A sliding clamp model for the Rad1 family of cell cycle checkpoint proteins. *Cell* 1999;96:769-770.
8. Venclovas C, Thelen MP. Structure-based predictions of Rad1, Rad9, Hus1 and Rad17 participation in sliding clamp and clamp-loading complexes. *Nucleic Acids Res* 2000;28:2481-2493.
9. Venclovas C, Colvin ME, Thelen MP. Molecular modeling-based analysis of interactions in the RFC-dependent clamp-loading process. *Protein Sci* 2002;11:2403-2416.
10. Barsky D, Foloppe N, Ahmadi S, Wilson DM, 3rd, MacKerell AD, Jr. New insights into the structure of abasic DNA from molecular dynamics simulations. *Nucleic Acids Res* 2000;28:2613-2626.
11. Nguyen LH, Barsky D, Erzberger JP, Wilson DM, 3rd. Mapping the protein-DNA interface and the metal-binding site of the major human apurinic/apyrimidinic endonuclease. *J Mol Biol* 2000;298:447-459.
12. Bermudez, V. P., Maniwa, Y., Tappin, I., Ozato, K., Yokomori, K., Hurwitz. The alternative Ctf18-Dcc1-Ctf8-replication factor C complex required for sister chromatid cohesion loads proliferating cell nuclear antigen onto DNA