

UNIVERSITY OF CALIFORNIA

Los Angeles

The Structures, Functions, and Evolution of Sm-like Archaeal Proteins (SmAPs)

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy in
Biochemistry and Molecular Biology

by

Cameron Mura

2002

© Copyright by

Cameron Mura

2002

The dissertation of Cameron Mura is approved.

James U. Bowie

Allan Tobin

David Eisenberg, Committee Chair

University of California, Los Angeles

2002

*To anyone
with whom
I've discussed
science*

TABLE OF CONTENTS

List of Figures and Tables	vii
Acknowledgements	x
Vita	xiii
Abstract of the Dissertation	xv
A Synopsis of the Dissertation	1
Chapter 1: Introduction to RNA metabolism and Sm proteins	6
Overview of RNA metabolism.....	7
Spliceosome-mediated excision of introns and mRNA processing.....	9
Archetypal RNPs of the spliceosome: The assembly and structure of U snRNPs.....	11
The biological functions of Sm proteins: formation of U snRNP cores.....	13
Phylogenetic conservation of Sm proteins and their broader significance.....	14
What is known about RNA processing in the archaea?.....	15
Seeking an atomic-resolution understanding of snRNP cores and Sm proteins.....	16
References.....	19
Chapter 2: The crystal structure of a heptameric archaeal Sm protein: Implications for the eukaryotic snRNP core	34
Abstract and Introduction.....	35
Materials and Methods.....	35
Results.....	36

Discussion.....	38
References.....	40
Chapter 3: The oligomerization and ligand-binding properties of Sm-like archaeal proteins (SmAPs).....	41
Abstract.....	42
Introduction.....	43
Materials and Methods.....	47
Results.....	56
Discussion.....	68
References.....	75
Chapter 4: The structure and potential function of an archaeal homolog of survival protein E (SurEα).....	94
Abstract.....	95
Introduction.....	96
Materials and Methods.....	99
Results.....	107
Discussion.....	111
Conclusions.....	121
References.....	122

Chapter 5: The crystal structure of a Nudix protein from <i>Pyrobaculum</i>	
<i>aerophilum</i> reveals a dimer with two intersubunit β -sheets.....	144
Abstract.....	145
Introduction.....	146
Materials and Methods.....	148
Results.....	153
Discussion.....	160
References.....	163
Appendix to the Dissertation: Various UNIX and Perl scripts.....	180

LIST OF FIGURES AND TABLES

Chapter 1: Introduction to RNA metabolism and Sm proteins

Fig. 1.1: A modular approach to RNA metabolism.....	28
Fig. 1.2: The spliceosomal cycle and the complexity of pre-mRNA processing <i>via</i> intron excision.....	29
Fig. 1.3: An example of the central role of Sm proteins in RNP assembly.....	30
Fig. 1.4: Secondary structures of some snRNAs, and their Sm binding sites.....	31
Fig. 1.5: Progress towards an atomic-resolution understanding of eukaryotic Sm proteins and snRNPs.....	32
Table 1.1: A summary of all known Sm and SmAP structures.....	33

Chapter 2: The crystal structure of a heptameric archaeal Sm protein: Implications for the eukaryotic snRNP core

Table 1: Crystallographic statistics.....	36
Fig. 1: Sequence analysis of Sm-like archaeal proteins (SmAPs).....	36
Fig. 2: <i>Pae</i> SmAP monomer and dimer structures.....	37
Fig. 3: Structure of the <i>Pae</i> SmAP heptamer and the cationic pore.....	38
Fig. 4: Electrostatic properties of the SmAP heptamer surface.....	39
Fig. 5: Model of a SmAP•ssRNA complex.....	39

Chapter 3: The oligomerization and ligand-binding properties of Sm-like archaeal proteins (SmAPs)

Table 3.1: Statistics for several crystal forms.....	85
Table 3.2: Pairwise RMSDs between <i>Pae</i> , <i>Mth</i> , and <i>Afu</i> SmAP1.....	86
Fig. 3.1: 3D structural alignment of <i>Pae</i> , <i>Mth</i> , and <i>Afu</i> SmAP1 dimers.....	86
Fig. 3.2: The oligomeric states of <i>Pae</i> and <i>Mth</i> SmAP1 in solution.....	87
Fig. 3.3: Polymerization of SmAP1s into polar fibers.....	88
Fig. 3.4: Various crystalline oligomers of <i>Pae</i> and <i>Mth</i> SmAP1.....	89
Fig. 3.5: Ligand-binding sites in the structure of the <i>Mth</i> 14-mer bound to UMP.....	90
Fig. 3.6: Ligand-binding sites in the structure of the <i>Pae</i> 14-mer bound to UMP.....	91
Fig. 3.7: Conserved mode of uridine recognition by <i>Mth</i> and <i>Afu</i> SmAP1.....	92
Fig. 3.8: Gel-shift of supercoiled DNA by <i>Mth</i> and <i>Pae</i> SmAP1.....	93

Chapter 4: The structure and potential function of an archaeal homolog of survival protein E (SurE α)

Table 4.1: Crystallographic statistics for <i>Pae</i> SurE α	130
Table 4.2: Buried surface area statistics for <i>Pae</i> and <i>Tma</i> SurEs.....	131
Table 4.3: All 43 known SurE homologs.....	133
Fig. 4.1: Structure validation and sample electron density for the refined <i>Pae</i> SurE α model.....	134
Fig. 4.2: Sequence analysis of all archaeal SurEs.....	135
Fig. 4.3: <i>Pae</i> SurE α is a Rossmann-like fold with an extended C-terminal domain.....	136

Fig. 4.4: Structural comparison of <i>Pae</i> and <i>Tma</i> SurE reveals conformationally invariant regions.....	137
Fig. 4.5: Crystalline <i>Pae</i> SurE α is predominantly non-domain swapped.....	138
Fig. 4.6: <i>Pae</i> SurE α has a much less extensive dimer interface than <i>Tma</i> SurE.....	139
Fig. 4.7: The putative SurE active site is highly acidic and strongly conserved.....	140
Fig. 4.8: Hypothetical model of GMP bound to the conserved SurE active site.....	141
Fig. 4.9: Phylogenetic distribution and genomic organization of <i>surE</i> genes.....	142
Fig. 4.10: Representative reactions catalyzed by homologs of <i>surE</i> gene neighbors.....	143
Chapter 5: The crystal structure of a Nudix protein from <i>Pyrobaculum aerophilum</i> reveals a dimer with two intersubunit β-sheets	
Table 5.1: Statistics of data collection for <i>Pae</i> Nudix.....	170
Table 5.2: Statistics for atomic refinement of <i>Pae</i> Nudix.....	171
Fig. 5.1: A section of the initial electron density map superimposed on the final coordinates.....	172
Fig. 5.2: Ribbon diagrams of the <i>Pae</i> Nudix structure.....	173
Fig. 5.3: The topology of the <i>Pae</i> Nudix structure.....	174
Fig. 5.4: A comparison of the Nudix crystal structures in the $P2_12_12_1$ (Native-1) and $P2_1$ (Native-2) crystal forms.....	175
Fig. 5.5: Sequence alignment between the <i>E. coli</i> MutT and the <i>Pae</i> Nudix.....	176
Fig. 5.6: Conserved Nudix box residues.....	177
Fig. 5.7: Structure superposition between the subunit A of <i>Pae</i> Nudix and the <i>E. coli</i> MutT.....	178
Fig. 5.8: RMSDs of the main chain atomic positions for the superimposed subunits of the Native-1 dimer.....	179

ACKNOWLEDGEMENTS

All of the research in this dissertation was conducted in the laboratory of my thesis advisor – David Eisenberg. I am greatly indebted to him for continuous support of my scientific endeavors, and allowing me the room to freely investigate proteins that I find interesting. His encouragement was obvious from my first contact with him (as a rotation student), and has been constant over the past several years. I would also like to thank the other two members of my dissertation committee – Jim Bowie and Allan Tobin. Their enthusiasm and interest in biochemistry has always been apparent and inspirational.

Also, I thank Jim Bowie, William Gelbart, and Todd Yeates for fielding various structural biology, physical chemistry, or math questions I have had over the years. Doug Black and Guillaume Chanfreau have been immensely helpful with my forays into the RNA world of Sm proteins. I appreciate the receptiveness of these professors. The dedicated crystallization efforts of Anna Kozhukhovsky cannot go unmentioned. The crystallographic expertise of Dan Anderson, Duilio Cascio, and Mike Sawaya has been indispensable – without their help I would have been cutoff at the knees. Other members of the UCLA structural biology community who have provided encouragement or scientific advice include: Mari Gingery, Celia C.W. Goulding, Rob Grothe, Magdalena Ivanova, Gary Kleiger, Sangho Lee, Yanshun Liu, Jeanne Perry, and Ioannis Xenarios. I especially thank Celia, Duilio, Mari, and Marcín Apostol for providing an indescribable work environment over the past couple years.

For their pivotal roles in my education prior to graduate school, I would like to thank my high school chemistry teacher (Dorothy Flanagan) and the chemistry advisor at

Georgia Tech who introduced me to crystallography (Loren D. Williams). Also, all of my family has been supportive of my academic interests. I especially thank my parents (Aubin and Pari Mura) for promotion of my interests in nature since roughly age seven, and my brother (Cambiz Mura) for sharing my interests in and enthusiasm for math and science. My wife (Linda Columbus) is a talented biochemist who also provided much needed encouragement during my polyglutamine era.

I had the pleasure of collecting most of the crystallographic data for this dissertation at beamlines 5.0.2 of the Advanced Light Source (Berkeley National Lab) and X8C of the National Synchrotron Light Source (Brookhaven National Lab). The funding agencies that supported various portions of this research include a National Science Foundation Graduate Research Fellowship (NSF 01-146) and a UCLA Dissertation Year Fellowship. Their financial support made this work possible.

Chapter 2 of this dissertation is used with permission of the National Academy of Sciences, and is a reprint of the following article: *The crystal structure of a heptameric archaeal Sm protein: Implications for the eukaryotic snRNP core* by Mura, C., Cascio, D., Sawaya, M.R., and Eisenberg, D. *Proc. Natl. Acad. Sci. USA* (2001), 98, 5532-5537.

Chapter 3 of this dissertation is an adapted pre-print of the following manuscript: *The oligomerization and ligand-binding properties of Sm-like archaeal proteins (SmAPs)* by Mura, C., Kozhukhovsky, A., Gingery, M., Phillips, M., and Eisenberg, D. (prepared for submission to the journal *Protein Science*). I would like to thank Drs. Duilio Cascio and Michael Sawaya (UCLA) for diffraction data collection at the synchrotron.

Chapter 4 of this dissertation is an adapted pre-print of the following manuscript: *The structure and potential function of an archaeal homolog of survival protein E (SurE α)* by Mura, C., Katz, J., Clarke, S.G. and Eisenberg, D. (prepared for submission to the *Journal of Molecular Biology*). The following people were helpful in this work: Drs. Duilio Cascio and Michael Sawaya (UCLA) provided crystallographic advice; Dr. Thomas Schneider (University of Göttingen) assisted with the ESCET program; and Dr. Sorel Fitz-Gibbon (UCLA) provided the phosmid vector containing the *Pae surE* gene. Prof. Eisenberg directed and supervised the research that forms the basis of this chapter.

Chapter 5 is an adaptation of the following published article: *Structure of a Nudix protein from Pyrobaculum aerophilum reveals a dimer with two intersubunit β -sheets* by Wang, S., Mura, C., Sawaya, M.R., and Eisenberg, D. *Acta Crystallographica* (2002), *D58*, 571-578. I thank Dr. Hanjing Yang (UCLA) for helpful Nudix protein discussions.

VITA

- 1975 Born in Shiraz, Iran
- 1996 B.S. in Chemistry with Highest Honors
Georgia Institute of Technology
- 1996-1997 Teaching Assistant
Department of Chemistry & Biochemistry
University of California, Los Angeles
- 1996-1999 Predoctoral Fellow
National Science Foundation Graduate Fellowship
Department of Chemistry & Biochemistry
University of California, Los Angeles
- 2001-2002 Predoctoral Fellow
UCLA Dissertation Year Fellowship
Molecular Biology Institute
University of California, Los Angeles

PUBLICATIONS

Oxanion-mediated inhibition of serine proteases

S.R. Presnell, G.S. Patil, C. Mura, K.M. Jude, J.M. Conley, J.A. Bertrand, C.-M. Kam, J.C. Powers, & L.D. Williams. *Biochemistry* (1998), *37*, 17068-17081

The crystal structure of a heptameric archaeal Sm protein: Implications for the eukaryotic snRNP core

C. Mura, D. Cascio, M.R. Sawaya, & D. Eisenberg. *Proc. Natl. Acad. Sci. USA* (2001), *98*, 5532-5537

Structure of a Nudix protein from *Pyrobaculum aerophilum* reveals a dimer with two intersubunit β -sheets

S. Wang, C. Mura, M. Sawaya, & D. Eisenberg. *Acta Crystallographica Sect. D* (2002), *D58*, 571-578

The structure and potential function of an archaeal homolog of survival protein E (SurE α)

C. Mura, J. Katz, S. Clarke, & D. Eisenberg. (prepared for submission to *J. Mol. Biol.*)

The oligomerization and ligand-binding properties of Sm-like archaeal proteins

C. Mura, A. Kozhoukhovsky, M. Gingery, M. Phillips, & D. Eisenberg.
(prepared for submission to *Prot. Sci.*)

Synthetic *versus* biological catalysis of an unnatural transformation: Comparisons of chemzyme, ribozyme, and abzyme catalysis of a biphenyl stereoisomerization

X. Wang, C. Mura, A. Shvets, S.C. Zimmerman, & K.N. Houk.
(in preparation for *J. Org. Chem.*)

ABSTRACT OF THE DISSERTATION

The Structures, Functions, and Evolution of Sm-like Archaeal Proteins

by

Cameron Mura

Doctor of Philosophy in Biochemistry and Molecular Biology

University of California, Los Angeles, 2002

Professor David Eisenberg, Chair

Sm proteins were discovered nearly 20 years ago as a group of small antigenic proteins (\approx 90-120 residues). Since then, an extensive amount of biochemical and genetic data have illuminated the crucial roles of these proteins in forming ribonucleoprotein (RNP) complexes that are used in RNA processing, *e.g.*, spliceosomal removal of introns from pre-mRNAs. Spliceosomes are large macromolecular machines that are comparable to ribosomes in size and complexity, and are composed of uridine-rich small nuclear RNPs (U snRNPs). Various sets of seven different Sm proteins form the cores of most snRNPs. Despite their importance, very little is known about the atomic-resolution structure of snRNPs or their Sm cores. As a first step towards a high-resolution image of snRNPs and their hierarchic assembly, we have determined the crystal structures of archaeal homologs of Sm proteins, which we term Sm-like archaeal proteins (SmAPs).

Beginning with a *Pyrobaculum aerophilum* (*Pae*) structural genomics pilot project, we determined the structure of *Pae* SmAP1. This structure provided the first

direct evidence for a toroid-shaped Sm homoheptamer at the snRNP core, and provided many insights and implications for SmAP evolution and RNA binding in Sm cores. Then, in order to extend these results, we solved the structure of *Pae* SmAP1 and a heptameric methanobacterial SmAP (*Mth* SmAP1) bound to uridine-5'-monophosphate (UMP); the uracil bases line the heptamer pore in the *Mth* ligand-bound structure, and suggest a more specific model for RNA binding than we were able to propose earlier.

Further characterization of the oligomerization and ligand-binding properties of *Mth* and *Pae* SmAP1s has allowed us to conclude that: (i) SmAPs form several oligomers besides the archetypal heptamer, including 14-mers and fibrillar polymers; (ii) *Mth* SmAP1 and *Archaeoglobus fulgidus* (*Afu*) SmAP1 recognize uracil bases in a nearly identical manner, suggesting a conserved RNA-binding mode for SmAPs; and (iii) *Pae* and *Mth* SmAP1s gel-shift supercoiled DNA, perhaps by nonspecific binding to single-stranded DNA. Our sequence analyses shed light on the evolution of Sm proteins: the SmAP module is a phylogenetically well-conserved domain that probably gave rise to modern (eukaryotic) Sm *heteroheptamers* via gene duplication and neutral drift. Crystal structure determinations for *Pae* SmAP2 and SmAP3 proteins are currently in progress, and will deepen our understanding of Sm protein function and evolution.

As part of the same *Pae* structural genomics project, we solved two structures that are unrelated to the SmAP work: an archaeal homolog of survival protein E (*Pae* SurE α) and a putative *Pae* Nudix protein. The final two chapters of this dissertation describe these structures and their significance.

A synopsis of the dissertation

The research reported in this dissertation is primarily concerned with Sm-like archaeal proteins (SmAP), which are archaeal homologs of eukaryotic Sm proteins. The Sm family is a broad, phylogenetically well-conserved set of proteins that function in several aspects of RNA processing, as described in an introductory chapter on the background and significance of Sm proteins in RNA metabolism (Chapter 1). Eukaryotic Sm and Sm-like (Lsm) proteins bind to small nuclear RNAs to form the core of several ribonucleoprotein (RNP) particles, most notably the uridine-rich small nuclear RNPs (U snRNP) that constitute the spliceosome (which in turn splices introns out of pre-mRNA). Since the biological function of Sm proteins has been most characterized in the context of U snRNPs and their role in spliceosome assembly and intron excision, the emphasis in Chapter 1 is on what is known about the canonical human and yeast Sm proteins and their involvement in intron splicing. However, the vast diversity of RNA metabolism – together with the central role of Sm proteins in forming the cores of U snRNPs – makes it likely that Sm, Lsm, and SmAP proteins are involved in a wide range of RNP complexes in addition to U snRNPs, and may justify the generalization of our SmAP results to include non-splicing functions (*e.g.*, Sm-like proteins as RNA chaperones).

Over a decade of biochemical and genetic data suggests that a heteroheptamer of Sm proteins is the biologically relevant species at the core of eukaryotic snRNPs. In order to obtain a high-resolution image of the structure and function of Sm proteins, we determined the crystal structure of SmAP1 from *Pyrobaculum aerophilum* (*Pae*) at a resolution of 1.75 Å (Chapter 2 is a reprint of the article which describes this structure). *Pae* SmAP1 was found to form a homoheptameric ring perforated by a cationic pore, thus

providing the first direct evidence for such an assembly in eukaryotic snRNPs. Additionally, the structure: (i) showed that *Pae* SmAP homodimers are structurally similar to two human Sm heterodimers; (ii) supported a gene duplication model of Sm protein evolution; and (iii) suggested features that may be important in RNA binding (such as the cationic pore). In order to extend or generalize these results, we then studied another archaeal Sm protein – *Methanobacterium thermautotrophicum* (*Mth*) SmAP1.

Chapter 3 describes the oligomerization and ligand-binding properties of *Mth* and *Pae* SmAP1s (in preparation for publication). The *Mth* SmAP1 structure was determined in three crystal forms, each with different heptamer packings. In one of the forms an *Mth* SmAP1 14-mer co-crystallized with uridine-5'-monophosphate (UMP), and showed that our earlier, speculative model for RNA binding is probably incorrect (intercalation of uracil bases between conserved pore side chains suggests that RNA may wrap *around* the pore, not thread *through* it). The five *Pae* and *Mth* crystal structures contain various small molecules bound in what appear to be conserved ligand-binding sites. We fortuitously discovered that *Pae* and *Mth* SmAP1 gel-shift negatively supercoiled DNA. In addition to presenting these ligand binding properties, Chapter 3 describes the following features of the oligomerization of SmAPs: (i) *Pae* SmAP1 forms disulfide-bonded 14-mers, whereas *Mth* SmAP1 is almost exclusively heptameric *in vitro*; (ii) *Pae* SmAP1 forms sub-heptameric states when its inter-subunit disulfide bonds are reduced; (iii) both *Pae* and *Mth* SmAP1 polymerize into polar fibers by the head-to-tail stacking of heptamers. Our crystal structures of *Pae* SmAP1 in two crystal forms and *Mth* SmAP1 in three crystal forms corroborate these novel oligomerization and polymerization properties of SmAPs.

The final two chapters describe two other crystal structures that we determined – an archaeal homolog of survival protein E (*Pae* SurE α , Chapter 4) and an archaeal homolog of Nudix proteins (Chapter 5). These proteins are unrelated to the SmAP work, and were solved in the course of a *P. aerophilum* structural genomics pilot project (although a recent report of the Sm-like properties of *E. coli* Hfq protein suggests a weak link between Sm proteins and SurE – see the introduction in Chapter 3). One of the interesting findings of this work was that crystalline *Pae* SurE α is an inhomogeneous mixture of domain swapped and non-domain swapped dimers. The account of the SurE α structure in Chapter 4 is an adaptation of a manuscript submitted for publication, and the description of the *Pae* Nudix structure in Chapter 5 is adapted from a published article (see chapter title pages for citations). The work reported in Chapter 5 was done in collaboration with Dr. Shuishu Wang of UCLA. After creating a *Pae* Nudix M16L point mutant by site-directed mutagenesis, I over-expressed, purified, and crystallized this protein in a $P2_1$ form (“Native-2” in Chapter 5). I then used the structure of Dr. Wang’s $P2_12_12_1$ “Native-1” dimer as a molecular replacement search model for the Nudix tetramer found in the asymmetric unit of the $P2_1$ form.

The Appendix provides some of the more useful scripts that were used in the research of this dissertation. All of these utilities were written in either the UNIX C shell or the Perl scripting language, and a brief description of each script is provided at the beginning of the Appendix. Two notes regarding these programs are: (i) the **scripted_glr.sh**, **alter.pl**, and **process_bigrun.pl** trio uses a published program (GLRF) to calculate cross-rotation functions that are systematically varied over the integration

radius and resolution limits of diffraction data, and processes the output in a user-friendly format; and (ii) the rare codon calculator accepts query DNA sequences at the following URL: <http://www.doe-mbi.ucla.edu/cgi/cam/racc.html>.

As a final note on the SmAP work, sequence analysis suggests that *P. aerophilum* has two other Sm-like proteins, *Pae* SmAP2 and SmAP3. We have cloned and purified these proteins, and their structure determinations are in progress for the following reasons: (i) SmAP2 and SmAP3 co-crystallized with UMP, so their likely RNA-binding sites may be revealed and compared to known uracil binding sites; (ii) due to its Loop-4 insertion, the SmAP3 sequence more closely resembles certain eukaryotic Sm proteins than does any other SmAP; (iii) the SmAP2 and SmAP3 paralogs may provide insight into gene duplication as a mechanism for the evolution of eukaryotic Sm heteroheptamers; (iv) the collection of *Pae* SmAP1, SmAP2, and SmAP3 crystal structures will provide a high-resolution picture of the entire Sm protein complement of *Pae*, and serve as a starting point for further biochemical and biophysical experiments that will address the function of SmAPs in archaea (which presumably do not contain snRNPs). Progress towards the crystal structures of these two new SmAPs is not reported in this dissertation, although we are far along: diffraction data to 2.0 Å have been processed from native SmAP2 crystals, and phases have been calculated to 2.0 Å for derivatized SmAP3 crystals (see pg. 32). Preliminary results suggest that SmAP2 and SmAP3 are also heptamers, and these structures are anticipated within the next few months.