

Proteins. Author manuscript, available in PMC 2010 January 1

Published in final edited form as:

Proteins. 2009 January; 74(1): 1-5. doi:10.1002/prot.22232.

Identifying polymer-forming SAM domains

Alejandro D. Meruelo¹ and James U. Bowie^{2,*}

¹Medical Scientist Training Program UCLA-DOE Institute for Genomics and Proteomics Molecular Biology Institute Boyer Hall UCLA 611 Charles E. Young Dr. E. Los Angeles, CA 90095-1570

²Department of Chemistry and Biochemistry UCLA-DOE Institute for Genomics and Proteomics Molecular Biology Institute Boyer Hall UCLA 611 Charles E. Young Dr. E. Los Angeles, CA 90095-1570

Abstract

Sterile Alpha Motif (SAM) domains are common protein modules in eukaryotic cells. It has not been possible to assign functions to uncharacterized SAM domains because they have been found to participate in diverse functions ranging from protein-protein interactions to RNA binding. Here we computationally identify likely members of the subclass of SAM domains that form polymers. Sequences were virtually threaded onto known polymer structures and then evaluated for compatibility with the polymer. We find that known SAM polymers score better than the vast majority of known non-polymers: 100% (7 of 7) of known polymers and only 8% of known non-polymers (1 of 12) score above a defined threshold value. Of 2901 SAM family members, we find 694 that score above the threshold and are likely polymers, including SAM domains from the proteins Lethal Malignant Brain Tumor, Bicaudal-C, Liprin- β , Adenylate Cyclase and Atherin.

Keywords

protein-complex threading; protein-protein interaction; polymer prediction; Sterile Alpha Motif; Pointed domain; scaffolding proteins

INTRODUCTION

Many common protein modules have conserved functions. For example, SH2 domains almost invariably bind to phosphotyrosines and SH3 domains almost invariably bind to proline-containing sequence motifs [1]. Thus, the presence of particular domains in a protein can often immediately suggest a functional hypothesis. On the other hand, some protein modules, such as SAM domains, perform a variety of different functions [2]. Different SAM domains can self-associate [3], bind to other SAM domains [4], bind to other non-SAM proteins [5], bind to RNA or DNA [6,7], or even bind lipids [8]. Because of their diverse functions, the presence of a SAM domain does not imply a particular function, but rather a collection of possible functions. The challenge then becomes identifying which of the possible functions a particular SAM domain performs.

Among the SAM domains that self-associate, many of them form polymeric structures. Polymeric SAM domains have been characterized in transcriptional repressors [9-12], scaffolding proteins [13,14] and regulatory enzymes [15]. So far all the SAM polymers known to be biologically relevant have common features: they are all left-handed helices with six subunits per turn. One SAM polymer is shown in Fig. 1. They differ in the precise

^{*}Corresponding author Phone: (310) 825-3754 Fax: (310) 206-3914 bowie@mbi.ucla.edu

residues used in the interface and the helical pitch, but the inter-subunit contacts are made from two similar patches on the SAM domains, which have been called the Mid-Loop (ML) and End-Helix (EH) surfaces.

While it is clear that many SAM domains form polymers, the polymerization function is difficult to characterize because polymeric SAM domains are heterogeneous assemblies that are often insoluble. It would therefore be useful to identify likely SAM polymers computationally. Here we sought to build on our current knowledge (six polymeric structures) to identify those SAM domains that are compatible with the already known polymer structures.

MATERIALS AND METHODS

SAM Polymer Template database

We constructed a database of template polymer structures initially from a collection of five known SAM polymer structures from the proteins Shank3 [13], Diacylglycerol Kinase $\delta 1$ [15], Sex-Comb on Midleg [11], TEL [10], and Polyhomeotic [9]. We then built eleven pseudo-polymers from non-polymeric SAM domain structures by first performing a structural alignment and then substituting the sequence of the polymeric SAM with the non-polymeric SAM. In this manner, sequences that only align to a non-polymeric SAM can be placed into a polymeric context. The SAM domains of known structure used for the pseudo-polymers were from the proteins STE11 [16], EPHA4 [17], EPHB2 (possibly forms an alternative type of polymer) [18], ETS1 [19], Mae [12], p73 [20], GABP α [21], Erg [21], Fli-1 (unpublished, PDB code 1X66), Smaug [22], and VTS1 [7]. Structural alignment was performed using the combinatorial extension method (http://cl.sdsc.edu/ce/ce_align.html) [23]. Each non-polymer structure was aligned with every polymer structure and the best alignment was chosen as the template.

SAM domain sequences

2801 SAM domain sequences were obtained from the SMART database (http://smart.embl-heidelberg.de) [24], using the keywords SAM and SAM-PNT.

Sequence alignment

All query SAM domain sequences were aligned to the sequences of the polymer template database in two steps. First, similar sequences to the polymer template sequences were identified using PSI-BLAST [25] using 12 iterations. Sequences were selected if they aligned with a p-value less than 10^{-3} and if at least 70% of the residues could be aligned relative to the shorter of the two sequences. Similar sequences were then combined in a multiple sequence alignment using the program CLUSTALW (http://www.ebi.ac.uk/clustalw/) [26]. Using this alignment, residues from similar sequences could be mapped onto the polymer structure.

Sequence-Structure Scoring

The interface energy of two neighboring chains, ΔG_{bind} , was computed using a dimer interface potential [27]. We selected interacting pairs of residues between subunits from the known polymer structures if any of the heavy atoms were within 4.5 Å and both residues had a solvent accessibility of less than 0.4. The identity of the residues used for calculation of ΔG_{bind} was from the sequence alignment (discussed above).

To adjust for sequence composition effects, the energy scores were compared to the energies of decoy alignments and a Z-score determined using the expression (4):

$$Z - score = \left(< \Delta G^{d}_{bind} > - \Delta G_{bind} \right) / \left(< \left(\Delta G^{d}_{bind} \right)^{2} \right) > - < \Delta G^{d}_{bind} >^{2} \right)^{1/2}$$

where ΔG^d_{bind} refers to the interface energy of the decoy alignments. Decoy alignments were created by circular permutation of the query sequence. If a sequence could be mapped to more than one template, the most positive Z-score was selected. To reduce false positives, we averaged the Z-scores of each sequence with the Z-scores of very close homologs, defined by a PSI-BLAST p-value $< 10^{-18}$, to obtain the final score, Z-final.

RESULTS AND DISCUSSION

The algorithm

Our algorithm is based on the protein-complex threading idea originally developed by Skolnick and co-workers [27]. All known SAM domain sequences were threaded onto all known polymer structures (templates) and the compatibility of the aligned sequence with the polymer interface evaluated as outlined in Fig. 2. This was accomplished by first identifying all sequences similar to each template sequence using PSI-BLAST [25]. The similar query sequences were then combined in a multiple sequence alignment using ClustalW [26], thereby mapping positions of the query sequences onto the template structure. The amino acid type in the polymer template structure was then replaced with the amino acid type in each related query sequence. The energy of the replaced sequence in the polymer structure was then evaluated using a statistical potential function that measures the likelihood of finding different contacting amino acid pairs in a protein-protein interface [27]. To account for sequence composition effects, the energy was compared to the energies of 100 circularly permuted alignments to obtain a Z-Score, which is the number of standard deviations away from the mean of circularly permuted alignments. The best Z-Score for a given query sequence on all template polymers was selected. Finally, to reduce noise, a Z-final score was calculated, which is an average of the best Z-scores of the query sequence and all very closely related sequences.

Score distribution and polymer identification

Starting with 2801 SAM domain sequences, we were able to find 1885 significant alignments to the known polymers. The distribution of Z-final scores for the 1885 aligned query SAM domain sequences is shown in Fig. 3. The Z-final scores for known polymers and presumed non-polymeric SAM domains are noted in Fig. 3 and are listed in Table I. For testing the scoring of the known polymers and non-polymers, we did not allow alignments of known polymers to their own structures, providing a more realistic scenario. As can be seen the vast majority of known polymers score better than known non-polymers. All of the seven known polymers score higher than a Z-final value of 0.65 and all but one of the twelve non-polymers scored below 0.56. We therefore picked the mean of these Z-final values, 0.61, as an arbitrary threshold defining likely polymers. As we are using a previously defined energy function, the only variable parameter here is, in fact, the threshold.

We find that 694 SAM domain sequences of the 1885 that could be aligned to the known polymers or pseudo-polymers score higher than the cutoff of 0.61 and are potentially polymers. A list of all the SAM domains that could be aligned and their Z-final scores are given in Supplemental Materials. In addition to family members of known SAM domain polymers, a number of other protein families were found consistently among the high scoring proteins and may therefore contain polymeric SAM domains.

Families containing possible polymeric SAM domains

Of the 21 Lethal Malignant Brain Tumor (LMBT) homologs, all but two had Z-final scores greater than 1.197, much higher than the threshold of score of 0.61. Like the polymeric SAM-containing proteins Scm and Ph, LMBT proteins are polycomb group proteins, involved in the maintenance of transcriptional repression. The SAM domain of L(3)MBT is known to self-associate [28]. L(3)MBT was found to work with in conjunction with another polymeric SAM-containing protein, TEL, and the SAM domains of TEL and L(3)MBT were found to bind to each other [28]. It is therefore possible that the SAM domain of L(3)MBT forms a polymer that could extend the polymer formed by TEL-SAM or could even form a mixed co-polymer with TEL-SAM. A similar mechanism has been proposed for the SAM polymer interactions of Ph and Scm [11]. It may be possible to adapt our algorithm to look specifically for co-polymers, but we may need to find effective ways to reduce the combinatorial complexity.

Another protein family important for development, Bicaudal-C, was also found among the top scoring SAM domains. Out of 11 Bicaudal-C homologs, the lowest Z-final was 1.321. Bicaudal-C is an RNA binding proteins that represses expression by regulating polyadenylation [29]. It is possible that SAM polymerization could enable spreading along the RNA transcripts or the SAM domain could provide a scaffolding role similar to the Shank [13] and Tankyrase [14] SAM domains.

Liprins were originally identified as proteins that bind to protein tyrosine phosphatases at focal adhesions [30]. There are two classes of liprins α and β [31]. Liprin- α is involved in organizing the active zone of neural synapses and is known to bind to a number of other proteins [32]. Less is know about the liprin- β class, but they interact directly with liprin- α proteins [31]. The liprins are somewhat unusual in that they contain three tandem SAM domains. Some of the SAM domains in the liprin- α class bind to other proteins [33]. We find here that the second SAM domain of the liprin- β class scores highly for polymer formation, suggesting that liprin- β could create a scaffold for organizing liprin- α and its binding partners.

We also found that a family of adenylate cyclases contains potential polymeric SAM domains. This family is predicted to be expressed as multiple splice variants, one that contains the SAM domain and one that does not (see http://dylan.embl-heidelberg.de/). The SAM-containing splice variant does not include the adenylate cyclase catalytic domain. It is therefore possible that the SAM-containing splice variant could be a scaffolding protein.

The SAM domain from the protein Atherin scores highly for polymer potential. Atherin is associated with atherosclerotic plaques and may recruit low-density lipoprotein complexes by direct binding [34]. Atherin polymerization could create a matrix for plaque formation. If so, inhibition of polymer formation could be a potential anti-atherosclerotic therapy.

CONCLUSION

In this work we have assigned a possible polymerization function to hundreds of uncharacterized SAM domains, providing a testable functional hypothesis for these proteins. The methodology could be applied to other domain types to predict polymerization.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by NIH grant R01 GM063919 and NIH MSTP grant T32-GM08042. We thank Yungok Ihm and Frank Pettit for help in developing the program and fruitful discussion, Alex Lisker for assistance using the computer cluster and members of the lab for helpful comments on the manuscript.

REFERENCES

- 1. Pawson T, Nash P. Assembly of cell regulatory systems through protein interaction domains. Science. 2003; 300:445–452. [PubMed: 12702867]
- 2. Qiao F, Bowie JU. The many faces of SAM. Sci STKE. 2005; 2005:re7. [PubMed: 15928333]
- 3. Golub T, Goga A, Barker G, Afar D, McLaughlin J, Bohlander S, Rowley J, Witte O, Gilliland D. Oligomerization of the ABL tyrosine kinase by the Ets protein TEL in human leukemia. Mol Cell Biol. 1996; 16:4107–4116. [PubMed: 8754809]
- Peterson A, Kyba M, Bornemann D, Morgan K, Brock H, Simon J. A domain shared by the Polycomb group proteins Scm and ph mediates heterotypic and homotypic interactions. Mol Cell Biol. 1997; 17:6683–6692. [PubMed: 9343432]
- 5. Seidel JJ, Graves BJ. An ERK2 docking site in the Pointed domain distinguishes a subset of ETS transcription factors. Genes Dev. 2002; 16:127–137. [PubMed: 11782450]
- 6. Aviv T, Lin Z, Lau S, Rendl LM, Sicheri F, Smibert CA. The RNA-binding SAM domain of Smaug defines a new family of post-transcriptional regulators. Nat Struct Biol. 2003; 10:614–621. [PubMed: 12858164]
- Oberstrass FC, Lee A, Stefl R, Janis M, Chanfreau G, Allain FH. Shape-specific recognition in the structure of the Vts1p SAM domain with RNA. Nat Struct Mol Biol. 2006; 13:160–167. [PubMed: 16429156]
- Li H, Fung KL, Jin DY, Chung SS, Ching YP, Ng IO, Sze KH, Ko BC, Sun H. Solution structures, dynamics, and lipid-binding of the sterile alpha-motif domain of the deleted in liver cancer 2. Proteins. 2007; 67:1154–1166. [PubMed: 17380510]
- 9. Kim CA, Gingery M, Pilpa RM, Bowie JU. The SAM domain of polyhomeotic forms a helical polymer. Nat Struct Biol. 2002; 9:453–457. [PubMed: 11992127]
- Kim CA, Phillips ML, Kim W, Gingery M, Tran HH, Robinson MA, Faham S, Bowie JU. Polymerization of the SAM domain of TEL in leukemogenesis and transcriptional repression. Embo J. 2001; 20:4173–4182. [PubMed: 11483520]
- 11. Kim CA, Sawaya MR, Cascio D, Kim W, Bowie JU. Structural organization of a Sex-comb-on-midleg/polyhomeotic copolymer. J Biol Chem. 2005; 280:27769–27775. [PubMed: 15905166]
- 12. Qiao F, Song H, Kim CA, Sawaya MR, Hunter JB, Gingery M, Rebay I, Courey AJ, Bowie JU. Derepression by depolymerization: structural insights into the regulation of Yan by Mae. Cell. 2004; 118:163–173. [PubMed: 15260987]
- 13. Baron MK, Boeckers TM, Vaida B, Faham S, Gingery M, Sawaya MR, Salyer D, Gundelfinger ED, Bowie JU. An architectural framework that may lie at the core of the postsynaptic density. Science. 2006; 311:531–535. [PubMed: 16439662]
- De Rycker M, Venkatesan RN, Wei C, Price CM. Vertebrate tankyrase domain structure and sterile alpha motif (SAM)-mediated multimerization. Biochem J. 2003; 372(Pt 1):87–96.
 [PubMed: 12589701]
- 15. Harada BT, Knight MJ, Imai S, Qiao F, Ramachander R, Sawaya MR, Gingery M, Sakane F, Bowie JU. Regulation of enzyme localization by polymerization: polymer formation by the SAM domain of diacylglycerol kinase delta1. Structure. 2008; 16:380–387. [PubMed: 18334213]
- Kwan JJ, Warner N, Pawson T, Donaldson LW. The solution structure of the S.cerevisiae Stel1 MAPKKK SAM domain and its partnership with Ste50. J Mol Biol. 2004; 342:681–693.
 [PubMed: 15327964]
- 17. Stapleton D, Balan I, Pawson T, Sicheri F. The crystal structure of an Eph receptor SAM domain reveals a mechanism for modular dimerization. Nat Struct Biol. 1999; 6:44–49. [PubMed: 9886291]

18. Thanos C, Goodwill K, Bowie J. Oligomeric structure of the human EphB2 receptor SAM domain. Science. 1999; 283:833–836. [PubMed: 9933164]

- Slupsky CM, Gentile LN, Donaldson LW, Mackereth CD, Seidel JJ, Graves BJ, McIntosh LP. Structure of the Ets-1 pointed domain and mitogen-activated protein kinase phosphorylation site. Proc Natl Acad Sci U S A. 1998; 95:12129–12134. [PubMed: 9770451]
- Chi SW, Ayed A, Arrowsmith CH. Solution structure of a conserved C-terminal domain of p73 with structural homology to the SAM domain. Embo J. 1999; 18:4438–4445. [PubMed: 10449409]
- 21. Mackereth CD, Scharpf M, Gentile LN, MacIntosh SE, Slupsky CM, McIntosh LP. Diversity in structure and function of the Ets family PNT domains. J Mol Biol. 2004; 342:1249–1264. [PubMed: 15351649]
- Green JB, Gardner CD, Wharton RP, Aggarwal AK. RNA recognition via the SAM domain of Smaug. Mol Cell. 2003; 11:1537–1548. [PubMed: 12820967]
- 23. Shindyalov IN, Bourne PE. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. Protein Eng. 1998; 11:739–747. [PubMed: 9796821]
- Letunic I, Copley RR, Pils B, Pinkert S, Schultz J, Bork P. SMART 5: domains in the context of genomes and networks. Nucleic Acids Res. 2006; 34(Database issue):D257–260. [PubMed: 16381859]
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997; 25:3389–3402. [PubMed: 9254694]
- 26. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W. improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994; 22:4673–4680. [PubMed: 7984417]
- 27. Lu L, Lu H, Skolnick J. MULTIPROSPECTOR: an algorithm for the prediction of protein-protein interactions by multimeric threading. Proteins. 2002; 49:350–364. [PubMed: 12360525]
- 28. Boccuni P, MacGrogan D, Scandura JM, Nimer SD. The human L(3)MBT polycomb group protein is a transcriptional repressor and interacts physically and functionally with TEL (ETV6). J Biol Chem. 2003; 278:15412–15420. [PubMed: 12588862]
- 29. Chicoine J, Benoit P, Gamberi C, Paliouras M, Simonelig M, Lasko P. Bicaudal-C recruits CCR4-NOT deadenylase to target mRNAs and regulates oogenesis, cytoskeletal organization, and its own expression. Dev Cell. 2007; 13:691–704. [PubMed: 17981137]
- 30. Serra-Pages C, Kedersha N, Fazikas L, Medley Q, Debant A, Streuli M. The LAR transmembrane protein tyrosine phosphatase and a coiled-coil LAR-interacting protein co-localize at focal adhesions. Embo J. 1995; 14:2827–2838. [PubMed: 7796809]
- 31. Serra-Pages C, Medley QG, Tang M, Hart A, Streuli M. Liprins, a family of LAR transmembrane protein-tyrosine phosphatase- interacting proteins. J Biol Chem. 1998; 273:15611–15620. [PubMed: 9624153]
- Olsen O, Moore KA, Nicoll RA, Bredt DS. Synaptic transmission regulated by a presynaptic MALS/Liprin-alpha protein complex. Curr Opin Cell Biol. 2006; 18:223–227. [PubMed: 16504495]
- 33. Stryker E, Johnson KG. LAR, liprin alpha and the regulation of active zone morphogenesis. J Cell Sci. 2007; 120(Pt 21):3723–3728. [PubMed: 17959628]
- Lees AM, Deconinck AE, Campbell BD, Lees RS. Atherin: a newly identified, lesion-specific, LDL-binding protein in human atherosclerosis. Atherosclerosis. 2005; 182:219–230. [PubMed: 16159594]

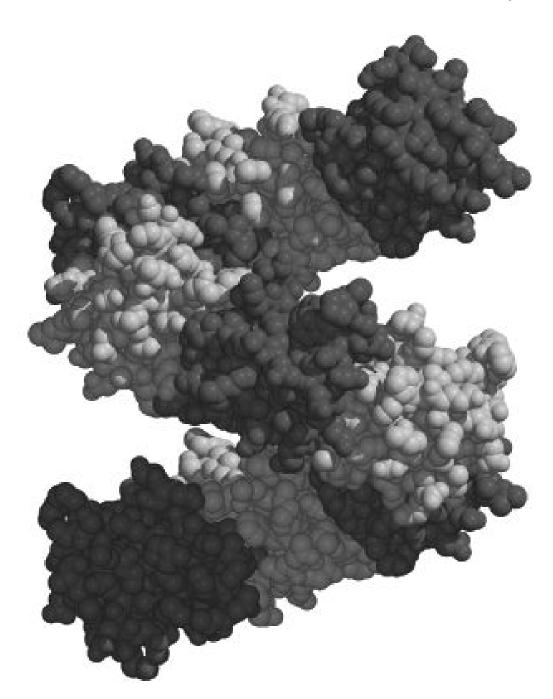


Figure 1. SAM polymer architectureThe structure shown is the SAM domain from TEL. Every other SAM domain is shaded dark or light.

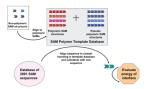


Figure 2. Outline of the polymer compatibility test algorithm See text.

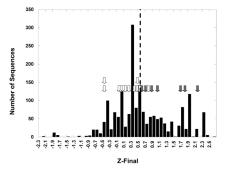


Figure 3. Histogram of the Z-final score distribution

The scores for the known polymers are indicated by the positions of the shaded arrows and the scores for the known non-polymers are indicated by the positions of the white arrows. The dotted line indicates the Z-final cutoff used to separate polymers from non-polymers.

Table I

Z-scores of known polymers and non-polymers

polymers	Z-score	non-polymers	Z-score
Scm	2.253	Erg	-0.410
Ph	1.848	Fli-1	-0.406
Shank	1.687	BYR2	-0.061
Tankyrase-1	0.976	p73	0.004
DGK	0.835	Mae	0.064
TEL	0.716	Smaug	0.129
Yan	0.653	EPHA4	0.290
		GABPA	0.460
		Pnt-P2	0.483
		ETS1	0.493
		STE11	0.563
		VTS1	0.792