Title
Computational prediction and analysis of protein structure

Author
Meruelo, Alejandro Daniel

Publication Date
2012

UNIVERSITY OF CALIFORNIA

Los Angeles

Computational prediction and analysis of protein structure

A dissertation submitted in partial satisfaction of the requirements for the degree

Doctor of Philosophy in Molecular Biology

by

Alejandro Daniel Meruelo

2012

ABSTRACT OF THE DISSERTATION

Computational prediction and analysis of protein structure

by

Alejandro Daniel Meruelo
Doctor of Philosophy in Molecular Biology
University of California, Los Angeles, 2012
Professor James U. Bowie, Chair

Identifying polymer-forming SAM domains

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. In polymerization experiments, all of these predictions (except Adenylate Cyclase) were confirmed. As a result, the original SAM database was updated and additional predictions were obtained.

TMKink: A method to predict transmembrane helix kinks

A hallmark of membrane protein structure is the large number of distorted transmembrane helices. Because of the prevalence of bends, it is important to not only understand how they are generated but also to learn how to predict their occurrence. Here, we find that there are local sequence preferences in kinked helices, most notably a higher abundance of proline, which can be exploited to identify bends from sequence information. A neural network predictor identifies over two-thirds of all bends (sensitivity 0.70) with high reliability (specificity 0.89). It is likely that more structural data will allow for better helix distortion predictors with increased coverage in the future. The kink predictor, TMKink, is available at http://tmkinkpredictor.mbi.ucla.edu/.

Structural differences between mesophilic and thermophilic membrane proteins

Protein thermostability remains a focal point of interest for protein scientists. The differences in thermostability between mesophilic and thermophilic soluble proteins have been extensively studied. No differences in packing values have been found in soluble proteins. Membrane protein packing is different from soluble protein packing; thermophilic adaptation may be different as a result. Surprisingly, burial and packing values appear to be shared between mesophiles and thermophiles in both soluble and

membrane proteins. We created a non-redundant database of unpaired and paired structures for the study of thermophile-mesophile structural differences in membrane proteins. We found little or no differences in burial or packing values in both the soluble and transmembrane regions of membrane proteins.

The dissertation of Alejandro Daniel Meruelo is approved.

Feng Guo

Todd O. Yeates

Matteo Pellegrini

James U. Bowie, Committee Chair

University of California, Los Angeles

2012

To: Samantha, Dad, Mom, and Natasha

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

Page

Chapter 1

Chapter 2

Chapter 3

ACKNOWLEDGMENTS

Chapter 1 is a version of “Identifying polymer-forming SAM domains”, *Proteins* 74,1-5, Copyright (2009) Wiley. Permission to reproduce this work has been granted by Wiley/Proteins. I would like to acknowledge James U. Bowie, the other author of this paper, and would like to 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. This work was supported by NIH grant R01 GM063919 and NIH MSTP grant T32-GM08042. I would also like to thank Mary Jane Knight for her experimental work confirming the predicted polymers from this paper that is described in Appendix 1 of Chapter 1.

Chapter 2 is a version of “TMKink: a method to predict transmembrane helix kinks,” *Protein Science* 20,1256-64, Copyright (2011) Wiley. Permission to reproduce this work has been granted by Wiley/Protein Science. I would like to acknowledge Ilan Samish and James U. Bowie, the other authors of this paper, and would like to thank Zheng Cao for helpful feedback on web program development, Duilio Cascio for helpful program testing suggestions, Luki Goldschmidt for assistance with the cluster, Tom Holton and Alex Lisker for help setting up the MBI webspace, Amit Oberai for helpful discussion about program development, and members of the Bowie lab for careful reading of the manuscript. A.D.M. is a Ruth L. Kirschstein NRSA Predoctoral fellow and a Molecular Biology Whitecome stipend recipient. This work was supported by NIH grant RO1 GM063919.

Chapter 3 is a manuscript in preparation. I would like to acknowledge Seong-Kyu Han, Sanguk Kim, and James U. Bowie, the other authors of this paper, and would like to thank members of the Bowie lab for careful reading of the manuscript. A.D.M. is a Ruth L. Kirschstein NRSA Predoctoral fellow. This work was supported by NIH grant RO1 GM063919.

VITA

2002-2006	B.S., Physics with Honors California Institute of Technology Pasadena, California
2006-2008	Medical Student Medical Scientist Training Program David Geffen School of Medicine at UCLA Los Angeles, California
2008-2012	Graduate Student Molecular Biology Interdepartmental Ph.D. Program University of California, Los Angeles Los Angeles, California

PUBLICATIONS AND PRESENTATIONS

Meruelo AD, Samish I, and Bowie JU. TMKink: A method to predict transmembrane helix kinks. *Protein Sci*. 20:1256-64, 2011.

Meruelo AD, Samish I, and Bowie JU. Kink prediction in membrane proteins. Invited talk given at *MBI Lake Arrowhead Retreat*, Lake Arrowhead, CA, October, 2010.

Meruelo AD and Bowie JU. Kinks in membrane proteins. Talk given as part of *UCLA MSTP tutorial series* held in MacDonald Research Labs, Los Angeles, CA, April, 2010.

Meruelo AD and Bowie JU. A computational approach to classifying transmembrane helical kinks. Invited talk given at *MBI Lake Arrowhead Retreat*, Lake Arrowhead, CA, October, 2009.

Meruelo AD and Bowie JU. Identifying polymer-forming SAM domains. *Proteins.* 74:1-5, 2009.

Hwang BJ, **Meruelo AD**, and Sternberg PW. *C. elegans* EVI1 proto-oncogene, EGL-43, is necessary for Notch-mediated cell fate specification and regulates cell invasion. *Development*. 134:669-679, 2007.

Meruelo AD, Hwang BJ, and Sternberg PW. A bioinformatical technique for the identification and exploration of gene interactions using *C. elegans*, *C. briggsae,* and *C. remanei.* Presented at *SURF Summer Seminar Day*, Pasadena, CA, August, 2004.

Meruelo AD, Hwang BJ, and Sternberg PW. A bioinformatical approach to genome-wide identification of lin-3 regulatory sites across *C. elegans* and *C. briggsae*. Presented at *SURF Summer Seminar Day*, Pasadena, CA, August, 2003.

Chapter 1

Identifying polymer-forming SAM domains

PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

SHORT COMMUNICATION

Identifying polymer-forming SAM domains

Alejandro D. Meruelo[1] **and James U. Bowie**[2*]

[1] Medical Scientist Training Program, UCLA-DOE Institute for Genomics and Proteomics, Molecular Biology Institute, UCLA, Los Angeles, California 90095-1570

[2] Department of Chemistry and Biochemistry, UCLA-DOE Institute for Genomics and Proteomics, Molecular Biology Institute, UCLA, Los Angeles, California 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 nonpolymers: 100% (7 of 7) of known polymers and only 8% of known nonpolymers (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.

Proteins 2009; 74:1–5.

Key words: 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 sterile alpha motif (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 Figure 1. They differ in the precise residues used in the interface and the helical pitch, but the intersubunit contacts are made from two similar patches on the SAM domains, which have been called the Mid-Loop (ML) and End-Helix (EH) surfaces.

Although 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.

Additional Supporting Information may be found in the online version of this article.
Grant sponsor: NIH grant; Grant number: R01 GM063919; Grant sponsor: NIH MSTP grant; Grant number: T32-GM08042.
*Correspondence to: James U. Bowie, Department of Chemistry and Biochemistry, UCLA-DOE Institute for Genomics and Proteomics, Molecular Biology Institute, UCLA, Los Angeles, California 90095-1570.
E-mail: bowie@mbi.ucla.edu
Received 5 May 2008; Revised 10 July 2008; Accepted 27 July 2008

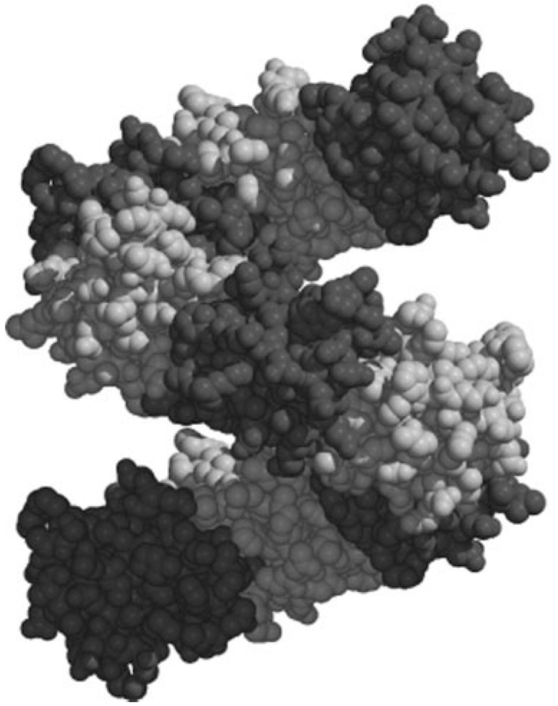


Figure 1

SAM polymer architecture. The structure shown is the SAM domain from TEL. Every other SAM domain is shaded dark or light.

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 δ1,[15] Sex-Comb on Midleg,[11] Translocation Ets Leukemia (TEL),[10] and Polyhomeotic.[9] We then built 11 pseudopolymers from nonpolymeric SAM domain structures by first performing a structural alignment and then substituting the sequence of the polymeric SAM with the nonpolymeric SAM. In this manner, sequences that only align to a nonpolymeric SAM can be placed into a polymeric context. The SAM domains of known structure used for the pseudopolymers were from the proteins STE11,[16] EPHA4,[17] EPHB2 (possibly forms an alternative type of polymer),[18] ETS1,[19] Mae,[12] p73,[20] GABPα,[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 nonpolymer structure was aligned with every polymer structure and the best alignment was chosen as the template.

SAM domain sequences

About 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 with 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 earlier).

To adjust for sequence composition effects, the energy scores were compared with the energies of decoy alignments and a *Z*-score determined using the expression[4]:

$$Z\text{-score} = (\langle \Delta G^{\mathrm{d}}_{\mathrm{bind}} \rangle - \Delta G_{\mathrm{bind}})/(\langle (\Delta G^{\mathrm{d}}_{\mathrm{bind}})^{2} \rangle - \langle \Delta G^{\mathrm{d}}_{\mathrm{bind}} \rangle^{2})^{1/2}$$

where $\Delta G^{\mathrm{d}}_{\mathrm{bind}}$ 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 coworkers.[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 Figure 2. This

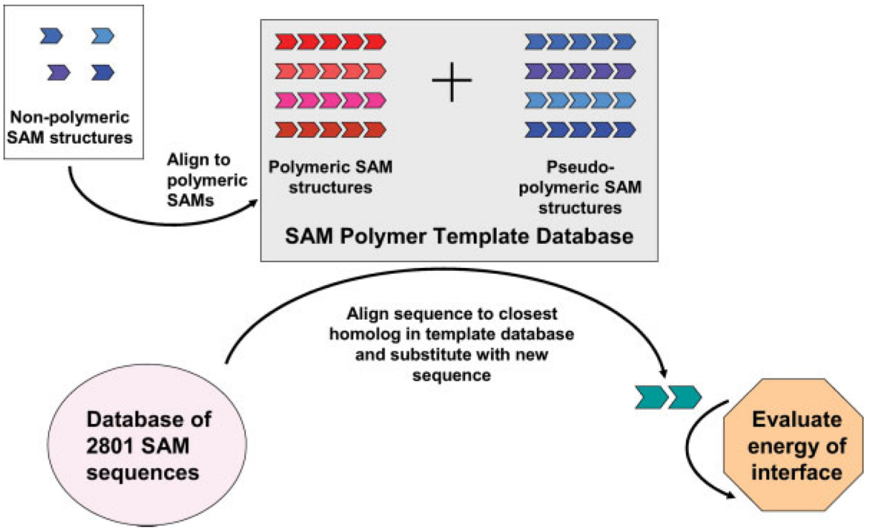


Figure 2

Outline of the polymer compatibility test algorithm. See text.

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 with 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 Figure 3. The *Z*-final scores for known polymers and presumed nonpolymeric SAM domains are noted in Figure 3 and are listed in Table I. For testing the scoring of the known polymers and nonpolymers, 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 nonpolymers. All of the seven known polymers score higher than a *Z*-final value of 0.65 and all but one of the 12 nonpolymers 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

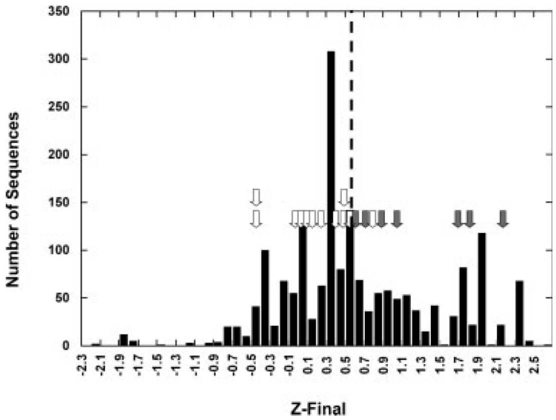


Figure 3

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 nonpolymers are indicated by the positions of the white arrows. The dotted line indicates the *Z*-final cutoff used to separate polymers from nonpolymers.

Table I
Z-Scores of known Polymers and Nonpolymers

Polymers	*Z*-score	Nonpolymers	*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

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 pseudopolymers 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 Supporting Information. 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 copolymer 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 copolymers, 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 antiatherosclerotic therapy.

CONCLUSIONS

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.

ACKNOWLEDGMENTS

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.
2. Qiao F, Bowie JU. The many faces of SAM. Sci STKE 2005; 2005:re7.
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.

4. 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.
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.
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.
7. 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.
8. 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 α-motif domain of the deleted in liver cancer 2. Proteins 2007;67:1154–1166.
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.
10. 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.
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.
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.
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.
14. De Rycker M, Venkatesan RN, Wei C, Price CM. Vertebrate tankyrase domain structure and sterile alpha motif (SAM)-mediated multimerization. Biochem J 2003;372(Part 1):87–96.
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.
16. Kwan JJ, Warner N, Pawson T, Donaldson LW. The solution structure of the *S. cerevisiae* Ste11 MAPKKK SAM domain and its partnership with Ste50. J Mol Biol 2004;342:681–693.
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.
18. Thanos C, Goodwill K, Bowie J. Oligomeric structure of the human EphB2 receptor SAM domain. Science 1999;283:833–836.
19. 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 USA 1998;95:12129–12134.
20. 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.
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.
22. Green JB, Gardner CD, Wharton RP, Aggarwal AK. RNA recognition via the SAM domain of Smaug. Mol Cell 2003;11:1537–1548.
23. Shindyalov IN, Bourne PE. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. Protein Eng 1998;11:739–747.
24. 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–D260.
25. 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.
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.
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.
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.
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.
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.
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.
32. Olsen O, Moore KA, Nicoll RA, Bredt DS. Synaptic transmission regulated by a presynaptic MALS/Liprin-α protein complex. Curr Opin Cell Biol 2006;18:223–227.
33. Stryker E, Johnson KG. LAR, liprin α and the regulation of active zone morphogenesis. J Cell Sci 2007;120(Part 21):3723–3728.
34. 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.

Appendix 1

While many previously uncharacterized SAMs were predicted to polymerize, no experimental characterization had been completed by the time of publication of the above paper. Experimental validation was later done by Mary Jane Knight *et al*[1]. In her work, she developed a method to rapidly screen SAM polymers. SAM domains were fused to super-negatively charged green fluorescent protein. Using these fusions, a large library of human SAM domains were screened for polymerization using a native gel screen. All experimentally tested predictions were found to be correct as has been summarized below in Table 1-2[1]; Adenylate Cyclase could not be cloned out for expression and so was not able to be experimentally studied[2].

Table 1-2: Experimental validation of SAM polymer predictions

Predicted polymer	**Experimentally confirmed**
Atherin	Yes
Bicaudal-C	Yes
Liprin-ß	Yes
Lethal Malignant Brain Tumor	Yes
Adenylate Cyclase	Not tested

Because of these experimental successes, we updated our SAM database. We added 21 new structures to our mapping database to make for a total of 37 (provided below). We used the non-redundancy criteria of requiring that all minimum P-value structures with > 90% identity be thrown out. In addition, the SAM database was expanded to include 3500 candidates from the original 3000. A number of new predicted polymers (e.g., Synaptotagmin-3) were found as a result and are tabulated in Table 1-3 below.

New structures (21): ARAP2, BIFUNCTIONAL_APOPTOSIS_REG, DCIP, EGFPS8, ELF3, EPHA2, EPHA8, EPHB1, EPHB4, FLJ21935, HCNK1, KIAA0902, LCP2, LOC64762, Neurabin, PDEF, PUTATIVE_47_KDA, SAMSN1, SASH1, Ship2, and SPHINGOMYELIN_SYNTHASE

Table 1-3: Summary of new predicted polymers

Protein	**Function**
UPI000155B956/44-126	PREDICTED: similar to B-cell lineage specific activator/TEL oncogene fusion protein, partial
UPI0000E4781C/1217-1284	PREDICTED: similar to mitogen-activated kinase kinase kinase 5
UPI0000F2AE6A/869-935	PREDICTED: similar to serologically defined colon cancer antigen 3
Q6BKM8_DEBHA/264-332	Similar to CA5428|CaBOI2 Candida albicans CaBOI2 budding protein
UPI000155472E/1-65	PREDICTED: similar to N-acylaminoacyl-peptide hydrolase
Q96X32_USTMA/14-80	MAP kinase pathway-interacting Ubc2
UPI00004D61B3/1152-1218	Synaptotagmin-3 (Ca sensor involved in both (i) early synaptic vesicle docking to the presynaptic membrane (ii) late steps of Ca-evoked synaptic vesicle fusion with the presynaptic membrane)
UPI0000E80BD6/971-1038	PREDICTED: similar to early development regulator 3
Q5KJP5_CRYNE/17-83	Protein kinase regulator, putative

References

1. Knight MJ, Leettola C, Gingery M, Li H, Bowie JU (2011 [cited 2012]) A human sterile alpha motif domain polymerizome. Protein Science 20:1697–1706.

2. Knight MJ (2010) Re: Adenylate cyclase SAM polymer.

Appendix 2

The list of new polymer and non-polymer predictions has been included below. The first line consists of the protein identifier and residue range. The second line provides the species description. The third line provides the amino acid sequence for the SAM. The last (and summary) line provides the residue range, simple protein identifier, and Z-score for polymer prediction. Z-scores with a value of 0.61 or higher were predicted as polymers and those below this value were predicted as nonpolymers.

```
>smart|SAM-uniprot|Q29KC7|Q29KC7_DROPS/849-915 GA18459-PA
[Drosophila pseudoobscura]
MQLSQHKDIQTLLTSLGLEHYIKIFVLNEIDLEMFTTLTEENLMELGITAFGARKKLLAAIHTL
LAN
849-915_Q29KC7 -2.297
>smart|SAM-uniprot|Q7QYZ7|Q7QYZ7_GIALA/5-63 GLP_164_5127_2983
[Giardia intestinalis]
VYQWLESANLQQYYPAFEQQGITPQRFITITIQDYGALGIQALPDKQKLFRLITTLKSR
5-63_Q7QYZ7    -2.160
>smart|SAM-uniprot|Q5TRB3|Q5TRB3_ANOGA/1-73 ENSANGP00000029645
[Anopheles gambiae]
VLGWSCENVSDWARKEGLSRCIIDWIAREDIDGRCLLAITEQDVHDLPQHCASPLRFGDIKRFW
FATRLLQRQ
1-73_Q5TRB3    -2.032
>smart|SAM-uniprot|UPI0000F20118|UPI0000F20118/578-645
PREDICTED: SAM domain, SH3 domain and nuclear localisation
signals, 1 [Danio rerio]
CKKPRPNTLLELLERLHLEEYASSLLLNGYQTVDDLRHLKERHLIELNVTDPEHRHRLLAASDC
LYVT
578-645_UPI0000F20118    -1.842
>smart|SAM-uniprot|Q4RQC8|Q4RQC8_TETNG/8-77 Chromosome 17
SCAF15006, whole genome shotgun sequence [Tetraodon
nigroviridis]
VSAWSCVQVAHWLQKEGFGEYVDLLCARHRLDGLGLLALTEADLRGPPLGLAVLGDIKRLVIAL
RRLQRR
8-77_Q4RQC8    -1.817
>smart|SAM-uniprot|UPI0000584492|UPI0000584492/3-72 PREDICTED:
hypothetical protein isoform 4 [Strongylocentrotus purpuratus]
```

```
VESWTCKDVAKWLSTKGFGQYARQLCTNHTIDGQALLLLTEADLRSPPVSMKVLGDIKRLIRQI
DLLKHQ
3-72_UPI0000584492  -1.812
>smart|SAM-uniprot|UPI00005A0829|UPI00005A0829/72-141 PREDICTED:
similar to sterile alpha motif domain containing 8 (predicted)
[Canis lupus familiaris]
IRRWTTKHVAVWLKDEGFFEYVDVLCNKHRLDGITLLTLTEYDLRSPPLEIKILGDIKRLMLSV
RKLQKI
72-141_UPI00005A0829     -1.800
>smart|SAM-uniprot|UPI000069F923|UPI000069F923/608-675 SAM and
SH3 domain-containing protein 1 (Proline-glutamate repeat-
containing protein). [Xenopus tropicalis]
KGLLKPKSVEDLLERINLKEHMPTFLFNGYEDLDAFKLLEEEDLDELNIQNPEHKAVLLTAIEL
LQEY
608-675_UPI000069F923    -1.763
>smart|SAM-uniprot|UPI0000E80117|UPI0000E80117/819-886
PREDICTED: similar to putative adapter and scaffold protein
[Gallus gallus]
SRPPQPKSVEDLLDRINLKEHMPTFLFNGYEDLDTFKLLEEEDLDELNIRDPEHRAVLLTAVEH
LQEY
819-886_UPI0000E80117    -1.751
>smart|SAM-uniprot|Q4REI4|Q4REI4_TETNG/584-651 Chromosome 10
SCAF15123, whole genome shotgun sequence [Tetraodon
nigroviridis]
ARQPKPISVEELLERINLKEHLPTFLFNGYEDLDTFKLLEEEDLDELNIRDSQHRAVLLTAVEL
LQEY
584-651_Q4REI4 -1.749
>smart|SAM-uniprot|UPI000065E67A|UPI000065E67A/973-1034 Homolog
of Homo sapiens &quot;SAM and SH3 domain containing protein 1
[Takifugu rubripes]
FGSWLKRTVSFAQEHMPTFLFNGYEDLDTFKLLEEEDLDELNIRDPQHRAVLLTAVELLQEY
973-1034_UPI000065E67A   -1.698
>smart|SAM-uniprot|Q4S703|Q4S703_TETNG/1001-1074 Chromosome 14
SCAF14723, whole genome shotgun sequence [Tetraodon
nigroviridis]
RPEVLLSGLSEHMRSLEIRDVSFTQEHMPTFLFNGYEDLDTFKLLEEEDLDELNIRDPQHRAVL
LTAVELLQEY
1001-1074_Q4S703    -1.690
>smart|SAM-uniprot|UPI00006605E2|UPI00006605E2/506-573 Homolog
of Homo sapiens &quot;SAM and SH3 domain containing protein 1
[Takifugu rubripes]
```

```
ARQPKPTSVEELLVRINLKEHLPTFLFNGYEDLDTFKLLEEEDLDELNIKDSQHRAVLLTAVEL
LQED
506-573_UPI00006605E2    -1.670
>smart|SAM-uniprot|Q299I9|Q299I9_DROPS/1-65 GA14584-PA
[Drosophila pseudoobscura]
MCPHAVREWLVLLTMEMYMDRFMKNGYDTIAQCKLIVPRDLKILGIYHPGHRKLLVDGVQLMNN
A
1-65_Q299I9    -1.669
>smart|SAM-uniprot|UPI0000F20302|UPI0000F20302/609-676
PREDICTED: hypothetical protein [Danio rerio]
SRPPKPTSVEELLERISLKEHMPTFLFNGYEDLDTFKLLEEEDLDELNISDPQHRAVLLTAVEL
LQEC
609-676_UPI0000F20302    -1.663
>smart|SAM-uniprot|UPI0000E21203|UPI0000E21203/805-872
PREDICTED: SAM and SH3 domain containing 1 [Pan troglodytes]
GRPPQPKSVEDLLDRINLKEHMPTFLFNGYEDLDTFKLLEEEDLDELNIRDPEHRAVLLTAVEL
LQED
805-872_UPI0000E21203    -1.657
>smart|SAM-uniprot|Q5TZ33|Q5TZ33_DANRE/625-692 Novel protein
similar to vertebrate SAM and SH3 domain containing 1 [Danio
rerio]
GRPPKPSSVEELLERINLKEHMPTFLFNGYEDLDTFKLLEEEDLDELNIKDPQHRAVLMTAVEL
LQDY
625-692_Q5TZ33 -1.584
>smart|SAM-uniprot|Q7Q002|Q7Q002_ANOGA/22-92 ENSANGP00000014193
[Anopheles gambiae]
VYMWTVADTLKWLWRHCNERCVKYQELFQRHEITGRALLRITDEQTLYRMGIHDEKDRGDILRE
IIKQRLK
22-92_Q7Q002   -1.575
>smart|SAM-uniprot|P59808|SASH1_MOUSE/623-690 SAM and SH3
domain-containing protein 1 [Mus musculus]
GRPSQPKSVEDLLDRINLKEHMPTFLFNGYEDLDTFKLLEEEDLDELNIRDPEHRAVLLTAVEL
LQEY
623-690_P59808 -1.574
>smart|SAM-uniprot|UPI0000EBCFE4|UPI0000EBCFE4/801-868
PREDICTED: similar to putative adapter and scaffold protein [Bos
taurus]
GRPPQPKSVEDLLDRINLKEHMPTFLFNGYEDLDTFKLLEEEDLDELNIRDPEHRAVLLTAVEL
LQEY
801-868_UPI0000EBCFE4    -1.562
>smart|SAM-uniprot|UPI00006A0C3A|UPI00006A0C3A/243-310 SAM
domain-containing protein SAMSN-1 (SAM domain, SH3 domain and
```

```
nuclear localization signals protein 1) (SH3-SAM adaptor
protein) (Hematopoietic adaptor containing SH3 and SAM domains
1). [Xenopus tropicalis]
SKRPKPKTLQELMERLNLQDYISSLLLNGYEKLEDLKDLKESHLHELNITSSEERARLLAAIEN
LQDC
243-310_UPI00006A0C3A    -1.555
>smart|SAM-uniprot|UPI0000D5574F|UPI0000D5574F/663-729
PREDICTED: similar to CG31163-PB, isoform B [Tribolium
castaneum]
IKGKPISVEDLLQKINLQEYISVFVLNGYEDLELFKEIEPSDLDYLGIVNSDHRSKILTAVQLL
HDL
663-729_UPI0000D5574F    -1.476
>smart|SAM-uniprot|Q7QBB0|Q7QBB0_ANOGA/610-677
ENSANGP00000020309 [Anopheles gambiae]
LPTSCPTSVEELLLRIGLKEYTSVFVLNGYEDLELFKELEPSDLDYLGIGNTEHRAKILAAVQL
LHDL
610-677_Q7QBB0 -1.473
>smart|SAM-uniprot|Q4S3R6|Q4S3R6_TETNG/622-687 Chromosome 17
SCAF14747, whole genome shotgun sequence [Tetraodon
nigroviridis]
EDGEDFEDLLSVLQHLGLSEYKTTFDDEKIDIESFLLCTIEDLKEMGIPLGPRKKIAKFVKERV
NK
622-687_Q4S3R6 -1.459
>smart|SAM-uniprot|UPI000051A64D|UPI000051A64D/427-494
PREDICTED: similar to CG31163-PB, isoform B, partial [Apis
mellifera]
RYRQKPGSVQELLQRMNLQEHIPVFVFNGYEDLELFREIEAADLDYLRIHQPEHRAKILTAVQL
LNDL
427-494_UPI000051A64D    -1.448
>smart|SAM-uniprot|Q8SZS0|Q8SZS0_DROME/737-804 LD31046p
[Drosophila melanogaster]
SHNGGPCSVEDLLIRIGLKEYTSVFVLNGYEDLELFKELEPADLDYLGILNQEHRAKLLTAVQL
LHDI
737-804_Q8SZS0 -1.446
>smart|SAM-uniprot|UPI000065F73B|UPI000065F73B/75-142 Homolog of
Homo sapiens &quot;Splice Isoform 1 of SAM-domain protein SAMSN-
1 [Takifugu rubripes]
CKGAPPQSLLELLQSLNLEDYASALLPEGCQTVQDLLNPQKKQLIELNIKDPEHRYRLLAAAEY
FNTE
75-142_UPI000065F73B     -1.425
>smart|SAM-uniprot|Q3U754|Q3U754_MOUSE/1-60 Bone marrow
macrophage cDNA, RIKEN full-length enriched library,
```

```
clone:I830086L09 product:SAM domain and HD domain, 1, full
insert sequence [Mus musculus]
CSFLENRGFREKKVLDIFRDNKIAGSFLPFLDEDRLEDLGVSSLEERKKMIECIQQLSQS
1-60_Q3U754    -1.418
>smart|SAM-uniprot|A0CUC0|A0CUC0_PARTE/87-154 Chromosome
undetermined scaffold_28, whole genome shotgun sequence
[Paramecium tetraurelia]
PKKWSLEDVSQWLNFVGLQQLQTTFTKNSIDGSCLELIEENDLIEDLGITNKIVRKKLMHWLKT
GLKE
87-154_A0CUC0  -1.410
>smart|SAM-uniprot|Q8ML92|AVE_DROME/23-91 Protein aveugle
[Drosophila melanogaster]
VYLWTVSDVLKWYRRHCGEYTQYEQLFAQHDITGRALLRITDSSLQRMGVTDNRDREAIWREIV
KQRLK
23-91_Q8ML92   -1.408
>smart|SAM-uniprot|Q5TZ33|Q5TZ33_DANRE/1120-1187 Novel protein
similar to vertebrate SAM and SH3 domain containing 1 [Danio
rerio]
VSLVHVSSVSDWLVSIGMPMYSAPLLAAGFDTLSRVSSLNEAQARDAGLKEEHHIRILLSEAQL
VSAA
1120-1187_Q5TZ33    -1.397
>smart|SAM-uniprot|UPI0000DA2491|UPI0000DA2491/41-109 PREDICTED:
similar to SAM domain and HD domain-containing protein 1
(Interferon-gamma-inducible protein Mg11) [Rattus norvegicus]
LRTWGPEDVCSFLEKRGFREKKVLDIFRENKISGSFLPFLDENLLESLGVSSLEQRKKMIECIQ
KLNQS
41-109_UPI0000DA2491     -1.382
>smart|SAM-uniprot|A6NCH1|A6NCH1_HUMAN/249-316 Uncharacterized
protein CXorf9 [Homo sapiens]
GKRPKPKTLHELLERIGLEEHTSTLLLNGYQTLEDFKELRETHLNELNIMDPQHRAKLLTAAEL
LLDY
249-316_A6NCH1 -1.380
>smart|SAM-uniprot|A5D8Q2|A5D8Q2_XENLA/240-307 LOC100049772
protein [Xenopus laevis]
SKRQKPKTVQELLERINMQEHIPTLLLNGYETIEDFKDLSENHLIELNITDPQHRVKLLTAAEF
LLDY
240-307_A5D8Q2 -1.376
>smart|SAM-uniprot|UPI0000D9C6DD|UPI0000D9C6DD/144-212
PREDICTED: SAM domain- and HD domain-containing protein 1
[Macaca mulatta]
YKTWGPEQVCFFLRRGGFGEPALLKNIRENKITGALLPCLDESHFENLGVSSLGERKKLLSYIQ
RSGQI
```

```
144-212_UPI0000D9C6DD    -1.374
>smart|SAM-uniprot|UPI00005A459C|UPI00005A459C/80-148 PREDICTED:
similar to SAM domain and HD domain-containing protein 1
(Dendritic cell-derived IFNG-induced protein) (DCIP) (Monocyte
protein 5) (MOP-5) [Canis lupus familiaris]
HQSWGPEQVCCFLKRSGFGDPRLLERFRENKITGLTLLHLNEADLERLGISTLGDRKKLQACIQ
QLSEI
80-148_UPI00005A459C     -1.365
>smart|SAM-uniprot|Q291B9|Q291B9_DROPS/21-89 GA15868-PA
[Drosophila pseudoobscura]
VYQWTVSDVLKWYRRHCGEYSQYDQLFVQHDITGRALLRITDTSLQRMGVTDNRDREAIWREIV
KQRLK
21-89_Q291B9   -1.352
>smart|SAM-uniprot|UPI0000F2B3C2|UPI0000F2B3C2/40-108 PREDICTED:
similar to MOP-5 [Monodelphis domestica]
PQSWGPEQLCAFLNHQGFSDPALLSRFRENKIRGSQLLHLDESLLENLGVRSSEDRQKLLDCVY
KLNQS
40-108_UPI0000F2B3C2     -1.349
>smart|SAM-uniprot|A1L2P0|A1L2P0_XENLA/637-702 LOC100036906
protein [Xenopus laevis]
EQQKPLPSLREALESLSLAEYSSVFEKEQIDMESLLMCTVDDLKEMGIPLGPRKKIAKFVQTKR
KA
637-702_A1L2P0 -1.324
>smart|SAM-uniprot|UPI0000E47174|UPI0000E47174/3-70 PREDICTED:
similar to centaurin delta 2 isoform a variant
[Strongylocentrotus purpuratus]
QSGNGPSSVQEWLQSLHLNQFLNDFLFSGYTDLSKCCHITEADLVRLGITQVGYQKRILSRLPS
KADL
3-70_UPI0000E47174  -1.322
>smart|SAM-uniprot|Q68FK2|Q68FK2_XENLA/644-709 MGC84165 protein
[Xenopus laevis]
EQQKPLPSLQEALELLSLAEYSSVFEKEQIDMESLLMCTVDDLKEMGIPLGPRKKIAKFVQERA
VT
644-709_Q68FK2 -1.321
>smart|SAM-uniprot|UPI0000D77D07|UPI0000D77D07/623-688
hypothetical protein LOC560750 [Danio rerio]
EEEEETGSLSSVLEAIGLSEYLSIFENEKIDVESLLLCTVDDLKEMSIPLGPRKKLANFIKEKS
AK
623-688_UPI0000D77D07    -1.301
>smart|SAM-uniprot|UPI000155C77E|UPI000155C77E/646-711
PREDICTED: hypothetical protein [Ornithorhynchus anatinus]
```

```
SIDNGISTLQETLEMLSLSEYISTFEKEKIDMESLLMCTVDDLKEMGIPLGPRKKIANFVKDKA
TE
646-711_UPI000155C77E    -1.299
>smart|SAM-uniprot|Q9Y6Y8|S23IP_HUMAN/641-706 SEC23-interacting
protein [Homo sapiens]
VENKEVLTLQETLEALSLSEYFSTFEKEKIDMESLLMCTVDDLKEMGIPLGPRKKIANFVEHKA
AK
641-706_Q9Y6Y8 -1.296
>smart|SAM-uniprot|UPI0000F2B03E|UPI0000F2B03E/638-702
PREDICTED: hypothetical protein [Monodelphis domestica]
VENEVLTLQETLEELSLSEYISTFEKEKIDMESLLMCTIDDLKEMGIPLGPRKKIANFVKDKAT
K
638-702_UPI0000F2B03E    -1.293
>smart|SAM-uniprot|UPI0000D8CDA0|UPI0000D8CDA0/227-294
hypothetical protein LOC450018 [Danio rerio]
SRRPRPKTLQELLERLNLEEHISSLLLNGYQTVEDLRDLKEQHLVELNMTDPEHRHRLLAAAEN
LLDS
227-294_UPI0000D8CDA0    -1.263
>smart|SAM-uniprot|Q4S4R5|Q4S4R5_TETNG/351-417 Chromosome 2
SCAF14738, whole genome shotgun sequence [Tetraodon
nigroviridis]
STPVFLQDLAEFLEQIGFSKYLPLLQEQDVDLRIFLTLTENDLKEIGITLFGPKRKMTSAIARW
HSS
351-417_Q4S4R5 -1.255
>smart|SAM-uniprot|UPI0000F213F2|UPI0000F213F2/406-472
PREDICTED: hypothetical protein [Danio rerio]
PSYTGPKDLSEFLEQIGFSKYLPLLEEQDIDLRIFLTLTENDLKEIGITLFGPKRRMTSAIARW
HSN
406-472_UPI0000F213F2    -1.254
>smart|SAM-uniprot|UPI00005E9A1F|UPI00005E9A1F/423-489
PREDICTED: hypothetical protein [Monodelphis domestica]
PTYSGPQDLATLLEQIGCLKYLQVFEEQDVDLRIFLTLTESDLKEIGITLFGPKRKMTSAIARW
HSS
423-489_UPI00005E9A1F    -1.229
>smart|SAM-uniprot|UPI0000ECD45F|UPI0000ECD45F/373-440 SAM
domain, SH3 domain and nuclear localisation signals, 1 [Gallus
gallus]
SKRNKPKTLQELLERVHLQEYTSTLLLNGYETLEDLKDLHESHLIELNISNPEDRARLLSAIEN
LQDN
373-440_UPI0000ECD45F    -1.228
>smart|SAM-uniprot|Q9D4W6|Q9D4W6_MOUSE/406-473 Adult male testis
cDNA, RIKEN full-length enriched library, clone:4930548L11
```

```
product:SAM domain, SH3 domain and nuclear localisation signals,
1, full insert sequence [Mus musculus]
RRRENHQTIQEFLERIHLQEYTSTLLLNGYETLDDLKDIKESHLIELNIADPEDRARLLSAAES
LLDE
406-473_Q9D4W6 -1.226
>smart|SAM-uniprot|UPI0000F2CDF8|UPI0000F2CDF8/310-377
PREDICTED: similar to protein with homology to KIAA0790
[Monodelphis domestica]
SKMAKPQTVQEFLERILLQEYASTLLLNGYETLEDLKDLKESHLIELNIKNPEDRKRLLSAAEN
LLDD
310-377_UPI0000F2CDF8    -1.221
>smart|SAM-uniprot|UPI000155CCD0|UPI000155CCD0/237-304
PREDICTED: similar to protein with homology to KIAA0790
[Ornithorhynchus anatinus]
SKRAKPKTLQEFLERIHLQEYASTLLLNGYESLEDLKYITESHLVELNIVNPEDRMRLLSAAEN
LLDD
237-304_UPI000155CCD0    -1.218
>smart|SAM-uniprot|UPI000155D10F|UPI000155D10F/371-437
PREDICTED: hypothetical protein [Ornithorhynchus anatinus]
SHSAGPQDLASFLEQIGCLKYLQVFEEQDVDLRIFLTLTESDLKEIGITLFGPKRKMTSAIARW
HSS
371-437_UPI000155D10F    -1.216
>smart|SAM-uniprot|UPI0000F1F822|UPI0000F1F822/113-179
PREDICTED: hypothetical protein [Danio rerio]
LAELNMDEVCQWFSNIGLEKCLPFIREAEFSGSHIASIDLKTLEILQVSNLEDRERLLSAIYHE
LHP
113-179_UPI0000F1F822    -1.215
>smart|SAM-uniprot|UPI0000500FE5|UPI0000500FE5/238-305 SAM
domain, SH3 domain and nuclear localization signals, 1 [Rattus
norvegicus]
SRRENPQTLQEFLERIHLQEYTSAFLLNGYEALDDLKDIKESHLIELNIANPEDRARLLSAAES
LLDE
238-305_UPI0000500FE5    -1.212
>smart|SAM-uniprot|UPI0001560616|UPI0001560616/388-455
PREDICTED: similar to SAM domain, SH3 domain and nuclear
localisation signals, 1 [Equus caballus]
SEREKSKTLQEFLERIQLQEYASTLLLNGYETVEDLKDITESHLIELNIKNPEDRMRLLSAAEN
LLDE
388-455_UPI0001560616    -1.199
>smart|SAM-uniprot|UPI00005A53A0|UPI00005A53A0/389-456
PREDICTED: similar to SAM-domain protein SAMSN-1 (SAM domain,
SH3 domain and nuclear localisation signals protein 1) (SH3-SAM
```

```
adaptor protein) (Hematopoietic adaptor containing SH3 and SAM
domains 1) [Canis lupus familiaris]
SGKEKPKTLQEFLERIHLQEYTSTLLLNGYETLEDLKDIKESHLIELNIKNPEDRMRLLSAAEN
LLDE
389-456_UPI00005A53A0    -1.198
>smart|SAM-uniprot|A6NJW8|A6NJW8_HUMAN/406-473 Uncharacterized
protein SAMSN1 [Homo sapiens]
SNSKKSKTLQEFLERIHLQEYTSTLLLNGYETLEDLKDIKESHLIELNIENPDDRRRLLSAAEN
FLEE
406-473_A6NJW8 -1.185
>smart|SAM-uniprot|UPI0000434EA0|UPI0000434EA0/1-68 centaurin,
delta 3 [Mus musculus]
MAAPQDLDIAVWLALVHLEQYADTFRRHGLATAGAAQHLGHEELRHLGISATGHRKRILRLLRA
GSAE
1-68_UPI0000434EA0  -1.181
>smart|SAM-uniprot|UPI0000F32B5F|UPI0000F32B5F/1-68 Centaurin-
delta 3 (Cnt-d3) (Arf-GAP, Rho-GAP, ankyrin repeat and
pleckstrin homology domain-containing protein 3). [Bos taurus]
MAAPQDLDIAVWLALVHLEQYADAFRQQGLATAGAARGLGHGELRQLGVNATGHRKRILRLLQA
GTAE
1-68_UPI0000F32B5F  -1.179
>smart|SAM-uniprot|Q3ZBW8|Q3ZBW8_BOVIN/237-304 Similar to SAM-
domain protein SAMSN-1 [Bos taurus]
SGVEKPETLQEFLERIHLQEYTSTLLLNGYETLEDLKDIKESHLIELNIKNPEDRMRLLSAAEN
LLDE
237-304_Q3ZBW8 -1.177
>smart|SAM-uniprot|UPI0000F2B52E|UPI0000F2B52E/1-68 PREDICTED:
similar to phosphoinositide-binding proteins [Monodelphis
domestica]
MESPRDLDIEVWLAGVHLERYADTFRRHGVSTAAAAQSLSQEDLRRMGIWATGHRKRILRLLQG
ASLE
1-68_UPI0000F2B52E  -1.175
>smart|SAM-uniprot|A0C082|A0C082_PARTE/14-78 Chromosome
undetermined scaffold_14, whole genome shotgun sequence
[Paramecium tetraurelia]
PKKWKLDDVTIWLKFIGLQDLDVNFRQNSVDGALLSTLDDNDLKEMGIAESSLKIKKLVQWIKI
G
14-78_A0C082   -1.162
>smart|SAM-uniprot|UPI0000519A8B|UPI0000519A8B/407-473
PREDICTED: similar to Phosphatidylinositol 3-kinase regulatory
subunit alpha (PI3-kinase p85-subunit alpha) (PtdIns-3-kinase
p85-alpha) (PI3K) [Apis mellifera]
```

```
VSEWTSLNVVEWMAALNLYRYADVFKSKDIKGSDLPFLDRDKLMNMGIKDDFHQKTILACIEEL
CQT
407-473_UPI0000519A8B    -1.141
>smart|SAM-uniprot|UPI00005C14DE|UPI00005C14DE/4-70 PREDICTED:
similar to phosphatidylcholine:ceramide
cholinephosphotransferase 1 isoform 2 [Bos taurus]
VVYWSPEKVADWLQENAMPEYCEPLEHFTGRDLINLTQEDFTKPPLCLVTSDNGQRLLHMIETL
KME
4-70_UPI00005C14DE  -1.097
>smart|SAM-uniprot|UPI0000D56B70|UPI0000D56B70/25-94 PREDICTED:
similar to CG30476-PA [Tribolium castaneum]
VYSWTVVDVQKWLRRHCTDYYALYVDKFIQNDVTGRTLLRINDNSLLRLGITDVDHREAIWREI
LKLRLK
25-94_UPI0000D56B70 -1.089
>smart|SAM-uniprot|Q7PIK6|Q7PIK6_ANOGA/822-887
ENSANGP00000024928 [Anopheles gambiae]
VEEWSNEQVCQWLLGIGLDHHIPAFMQHSVEGGALLQLDKPDFKILNVGGDDKKLLKRNIKELR
RL
822-887_Q7PIK6 -1.086
>smart|SAM-uniprot|Q4RG97|Q4RG97_TETNG/909-974 Chromosome 12
SCAF15104, whole genome shotgun sequence [Tetraodon
nigroviridis]
MTMLDMNISQFLKSLGLDHLRDIFEREQITLDVLADMGHEELKEIGINAYGHRHKLIKGVERLL
GG
909-974_Q4RG97 -1.018
>smart|SAM-uniprot|Q9V4E5|Q9V4E5_DROME/1028-1095 CG1511-PC,
isoform C [Drosophila melanogaster]
RGQNIFISTDLWLEHIKMSRYCHHFKEANLINAQQISRLTAQQLSDMGITLVGHQKKILHQARQ
LDTI
1028-1095_Q9V4E5    -0.998
>smart|SAM-uniprot|A0D4Q1|A0D4Q1_PARTE/14-80 Chromosome
undetermined scaffold_38, whole genome shotgun sequence
[Paramecium tetraurelia]
PRRWTQCDVESWLEFIGLQDLSDVFQRNAIDGACLEVLNDDDLNELGISSNVKKKKILQWIQNG
FIE
14-80_A0D4Q1   -0.981
>smart|SAM-uniprot|Q7PRV9|Q7PRV9_ANOGA/965-1032
ENSANGP00000007421 [Anopheles gambiae]
GNPGIFISTDLWLEGIKMSRYSQHFKEAGLVTAQQLSRLTAQQLSDMGITLVGHQKKILHQARQ
IDTI
965-1032_Q7PRV9     -0.980
```

```
>smart|SAM-uniprot|Q4H3M5|Q4H3M5_CIOIN/900-967 Ephrin receptor
[Ciona intestinalis]
NSIQDVSTLDEWLDMVKLGRYRRSFHNNGINDLESLAHISESELDRLGIASPSHRTRLQGGINT
LRQH
900-967_Q4H3M5 -0.970
>smart|SAM-uniprot|Q0JDT6|Q0JDT6_ORYSJ/152-215 Os04g0378200
protein [Oryza sativa]
SQPKADSPLDIFLKSLGLEKYSITFQAEEVDMAALRHMTDSDLKALGIPMGPRKKIMLALESRA
152-215_Q0JDT6 -0.956
>smart|SAM-uniprot|Q5BT66|Q5BT66_SCHJA/12-81 SJCHGC02625 protein
[Schistosoma japonicum]
VLNWTPEDVVKWLHKYASPVIAKYSELFEANSINGMCLRLMTDEWLLRLGIADQSDRSALMSHI
YRMRLK
12-81_Q5BT66   -0.948
>smart|SAM-uniprot|A0JMN0|A0JMN0_DANRE/4-70 Zgc:154039 [Danio
rerio]
VAAWSEDEVSLWLTDQGLQEYSEALRKYDGLALLNLTTEDFKRTPLSRVTSDGGRLLLEKIETL
KIE
4-70_A0JMN0    -0.946
>smart|SAM-uniprot|Q16S46|Q16S46_AEDAE/91-148 Eph receptor
tyrosine kinase [Aedes aegypti]
HSRIDGRTDQWLESIKMSRYSQHFKEAGLVTAQQISRLTAQQLSDMGITLVGHQKKIL
91-148_Q16S46  -0.942
>smart|SAM-uniprot|UPI0000F20BB6|UPI0000F20BB6/949-1014
PREDICTED: similar to TRF1-interacting ankyrin-related ADP-
ribose polymerase [Danio rerio]
ITILDMNISQFLKSLGLEHLRDIFEREQITLDVLADMGHEELKEIGINAYGHRHKLIKGIERLL
GG
949-1014_UPI0000F20BB6   -0.934
>smart|SAM-uniprot|UPI0000ECB6F9|UPI0000ECB6F9/663-719 KS5
protein [Gallus gallus]
ISEWTTQQVCHWLMGMNMEQYITEFTAMNIDGQQLMQLDSDKLKALGVSSQNDRSTI
663-719_UPI0000ECB6F9    -0.931
>smart|SAM-uniprot|UPI0000D93040|UPI0000D93040/874-939
PREDICTED: similar to tankyrase-like protein [Monodelphis
domestica]
VPGIDFSINQFVRNLGLEHLIDIFEREQITLDVLVEMGHKELKEIGINAYGHRHKIIKGVERLI
SG
874-939_UPI0000D93040    -0.921
>smart|SAM-uniprot|UPI00006A0F4F|UPI00006A0F4F/113-178 Tankyrase
2 (EC 2.4.2.30) (TANK2) (Tankyrase II) (TNKS-2) (TRF1-
```

```
interacting ankyrin-related ADP-ribose polymerase 2) (Tankyrase-
like protein) (Tankyrase-related protein). [Xenopus tropicalis]
VPGVELSINQFLQNLGLEHLIDIFEREQITLDVLIEMGHKEFKEIGINAYGHRHKIIKGVERLI
SG
113-178_UPI00006A0F4F    -0.916
>smart|SAM-uniprot|Q4QPU2|Q4QPU2_DROME/16-83 IP04461p
[Drosophila melanogaster]
RTRMCGRAVRDWLVLLTMEKYIGKFLERGYDSIERCKLIIVSDLIMLGVDNPAHRKLLLEGVRF
LVNA
16-83_Q4QPU2   -0.905
>smart|SAM-uniprot|Q4SML6|Q4SML6_TETNG/180-250 Chromosome 18
SCAF14547, whole genome shotgun sequence [Tetraodon
nigroviridis]
VSRWTPEEVLSWLDQLGTWAQLYREPFQQENVNGRLLLMLGEEELLKPPFNVENQAHRRAVLAE
LDRIKLL
180-250_Q4SML6 -0.887
>smart|SAM-uniprot|UPI000065DC62|UPI000065DC62/180-250 Homolog
of Homo sapiens &quot;Apoptosis regulator [Takifugu rubripes]
VSRWTPEEVLSWLDQLGPWAHLYREPFQQENVNGRLLLMLGEEELLKPPYNIGNQAHQRAVLAE
LNRIKLL
180-250_UPI000065DC62    -0.883
>smart|SAM-uniprot|UPI0000DB6E0B|UPI0000DB6E0B/642-708
PREDICTED: similar to Bicaudal C CG4824-PA, isoform A [Apis
mellifera]
ITSNRYDDLTSMLTSVGLEKYIRLFTSHEVDMATFPSLTEKDLCEIGITAWGARRKIMLLIAEM
NKR
642-708_UPI0000DB6E0B    -0.871
>smart|SAM-uniprot|Q5XGM0|Q5XGM0_XENLA/175-245 LOC495189 protein
[Xenopus laevis]
VAKWTAEEVVLWLEQLGPWASHYKEHFLSGNVNGRLLLTLAEEEFSKNPYSIENSSHRKAILLE
LDRVKTL
175-245_Q5XGM0 -0.870
>smart|SAM-uniprot|UPI000044908D|UPI000044908D/167-237
PREDICTED: similar to apoptosis regulator [Gallus gallus]
VAKWTAEEVALWLEQLGPWASLYKERFLMEKVNGRLLLTLTEEDFTKEPYSIENNNHKKAITTE
LECVKTL
167-237_UPI000044908D    -0.869
>smart|SAM-uniprot|Q20696|SMSR1_CAEEL/11-80 Sphingomyelin
synthase-related 1 [Caenorhabditis elegans]
PNEWRCEDVGNWLKKIGMAKYADLIAMKHKVDGKCLLALTDTDLKDPPVSINCLGDIKKILFAI
EFLSQK
11-80_Q20696   -0.868
```

```
>smart|SAM-uniprot|Q5CVX5|Q5CVX5_CRYPV/289-355 Ser/Thr protein
kinase with MORN repeats at the N-terminus and a sterile alpha
motif [Cryptosporidium parvum]
CLSWNMYQVAYFVKCIGLSKYVKNFILNEIDGCTLPYIGFEELKEMGITEITHIKYIRMSLSLL
LKL
289-355_Q5CVX5 -0.862
>smart|SAM_PNT-uniprot|Q5TW03|Q5TW03_ANOGA/48-132
ENSANGP00000027568 [Anopheles gambiae]
KNGMQATCTELGISSDPKQWTTAKVHQWLELAMGKYFLPRLDNLAALFPENGAQLAALPLEEFV
RRIPQGGDKLHGYLDL
48-132_Q5TW03  -0.856
>smart|SAM-uniprot|Q9VBP3|Q9VBP3_DROME/886-952 CG4719-PA
[Drosophila melanogaster]
PDADTITNVSGFLSSQQLHHLIELFEREQITLDILAEMGHDDLKQVGVSAYGFRHKILKGIAQL
RST
886-952_Q9VBP3 -0.842
>smart|SAM-uniprot|Q295L9|Q295L9_DROPS/894-960 GA18382-PA
[Drosophila pseudoobscura]
PDAESITNVSGFLSSQQLNHLIELFEREQITLDILAEMGHDDLKQVGVSAYGFRHKILKGIAQL
RAT
894-960_Q295L9 -0.828
>smart|SAM-uniprot|Q7PYH8|Q7PYH8_ANOGA/860-926
ENSANGP00000018360 [Anopheles gambiae]
SVPSIESSVSVFLTSLQLEHLIDLFEREQITMDILAEMGHEDLKQVGVSAYGFRHKILKGIATL
RAT
860-926_Q7PYH8 -0.819
>smart|SAM-uniprot|UPI000044AACC|UPI000044AACC/371-436
PREDICTED: similar to Ddhd2 protein [Gallus gallus]
SSSRGAEEVKEILKKLDLSEYCPVFEKEKIDGEALFLCAERNLEEMGIPLGPRMKLLHYISSRR
EV
371-436_UPI000044AACC    -0.810
>smart|SAM-uniprot|UPI00005472AC|UPI00005472AC/247-314
PREDICTED: hypothetical protein [Danio rerio]
HQKAKPKTLEEVLERIGLTELGSLLSMHGFQSLEDFTGLKESHLNDLNITDPEQRVKILKATEL
IHDS
247-314_UPI00005472AC    -0.806
>smart|SAM-uniprot|Q4SBI8|Q4SBI8_TETNG/1-59 Chromosome 15
SCAF14667, whole genome shotgun sequence [Tetraodon
nigroviridis]
VGLWLVSMDMDQYASEFTARGVDGPQLLSLDGEKLKALGVCSHSDRAVLKKKLKEIKKM
1-59_Q4SBI8    -0.803
```

```
>smart|SAM-uniprot|UPI000023EA8E|UPI000023EA8E/160-227
hypothetical protein FG07068.1 [Gibberella zeae]
LASWTTEMVAQAMLNAGIELSVADRFMENDISGAILITLKFEDLKELSIQSFGIRTKVWHQIQA
LRDS
160-227_UPI000023EA8E    -0.802
>smart|SAM-uniprot|Q4H3M4|Q4H3M4_CIOIN/961-1028 Ephrin receptor
[Ciona intestinalis]
LRHQVPSSLGVWLDGLGLGNYKENFVSAGYSSLDQVLQMSPSDLASLGIVETNQQNAVMESISS
IRYK
961-1028_Q4H3M4     -0.754
>smart|SAM-uniprot|UPI0000D554FF|UPI0000D554FF/868-934
PREDICTED: similar to CG4719-PA [Tribolium castaneum]
DLIGNITNIPSFLASLNLEQLIDIFEREQITLDILAEMNHDDLKQIGISAYGFRHKLIKGMERL
MSS
868-934_UPI0000D554FF    -0.753
>smart|SAM-uniprot|UPI0000F1EB95|UPI0000F1EB95/88-157 PREDICTED:
hypothetical protein [Danio rerio]
VVLWSQQDVCRWLKKHCPHNYLTYVEAFSHHAITGRALLRLNGEKLERMGLVQETLRQELLQQV
LQLQVQ
88-157_UPI0000F1EB95     -0.751
>smart|SAM-uniprot|UPI000155CAB3|UPI000155CAB3/86-155 PREDICTED:
similar to sterile alpha motif domain containing 10
[Ornithorhynchus anatinus]
VVLWSQQDVCKWLKKHCPHNYLVYVEAFAHHAITGRALLRLNAEKLQRMGISQESQRQEVLQQV
LQLQVR
86-155_UPI000155CAB3     -0.744
>smart|SAM_PNT-uniprot|P29775|ETS4_DROME/260-344 DNA-binding
protein D-ETS-4 [Drosophila melanogaster]
KREADAICAELQISQDPNGWSPAQVHAWLRSTLAQFRLPPVADLELHFCENGAALALLSEEEFV
RRLPESGSTLHAQLEI
260-344_P29775 -0.737
>smart|SAM-uniprot|Q9NGC7|Q9NGC7_MYAAR/485-551 P73-like protein
[Mya arenaria]
QNDMQDNSVSTWLNALGLGAYIDGFHEQNLYSLLQLDDFSLDDLAKMKIGNSHRNKIWKSLLEL
RNQ
485-551_Q9NGC7 -0.733
>smart|SAM-uniprot|UPI0000E48DF4|UPI0000E48DF4/406-472
PREDICTED: similar to Viral A-type inclusion protein repeat
containing protein [Strongylocentrotus purpuratus]
VSKWTERDIIHWLDSIHLGHLAQNFIDHDITGKELMEIDMSLLDEMDIDSSLEREKILSTLYDL
TNP
406-472_UPI0000E48DF4    -0.702
```

```
>smart|SAM-uniprot|UPI0000EBD589|UPI0000EBD589/74-143 PREDICTED:
similar to sterile alpha motif domain containing 12 isoform 2
[Bos taurus]
VALWTQQDVCKWLKKHCPNQYQIYSESFKQHDITGRALLRLTDKKLERMGIAQENLRQLILQQV
LQLKVR
74-143_UPI0000EBD589     -0.692
>smart|SAM-uniprot|UPI00001CFD37|UPI00001CFD37/74-143 PREDICTED:
similar to sterile alpha motif domain containing 12 [Rattus
norvegicus]
VALWTQQDVCKWLKKHCPNQYHIYSESFKQHDITGRALLRLTDKKLERMGVAQENQRQHILQQV
LQLKVR
74-143_UPI00001CFD37     -0.684
>smart|SAM-uniprot|UPI0000F20A91|UPI0000F20A91/78-147 PREDICTED:
similar to Sterile alpha motif domain containing 12, partial
[Danio rerio]
VALWSQQDVCKWLKKHCPNQHQVYSDSFKQHDITGRALMRLTDRKLERMGIMQESQRQFILQQV
LQLRVR
78-147_UPI0000F20A91     -0.676
>smart|SAM-uniprot|UPI00005844F3|UPI00005844F3/226-298
PREDICTED: similar to apoptosis regulator [Strongylocentrotus
purpuratus]
LSRWRPADVSQWLSELGAWAADGYSRQFNASEINGVLLEGLQSEGELEKEPFNINNSLHRQAIL
IALEKVKQL
226-298_UPI00005844F3    -0.651
>smart|SAM-uniprot|Q9SFU7|Q9SFU7_ARATH/140-203 T1B9.17 protein
[Arabidopsis thaliana]
KSQQADSSVDSFLESLGLEKYSTAFQVEEVDMDALMHMTDDDLKAMLIPMGPRKKILLALGSKP
140-203_Q9SFU7 -0.642
>smart|SAM-uniprot|A7A1P0|A7A1P0_YEAST/17-84 MEK kinase
[Saccharomyces cerevisiae YJM789]
EKTNDLPFVQLFLEEIGCTQYLDSFIQCNLVTEEEIKYLDKDILIALGVNKIGDRLKILRKSKS
FQRD
17-84_A7A1P0   -0.627
>smart|SAM-uniprot|UPI0000E4A182|UPI0000E4A182/794-860
PREDICTED: similar to neurabin [Strongylocentrotus purpuratus]
VWDWNTTHVSQWLMANNLAQYINDFSANSINGQMLLQLDGARLKTIGVTNTNDKNTFKKKVKEL
KAC
794-860_UPI0000E4A182    -0.626
>smart|SAM-uniprot|A6RE25|A6RE25_AJECA/157-224 Predicted protein
[Ajellomyces capsulatus]
VRSWTPTRVAEWMYESGFEDAIVERFIRNDISGSVLLDLQMDDLKELDISSFGIRHRLMNMIQG
FRAE
```

```
157-224_A6RE25 -0.607
>smart|SAM-uniprot|Q2UBM0|Q2UBM0_ASPOR/181-248 Predicted protein
[Aspergillus oryzae]
IRGWMPTQVAHWMHIAGYDDYVIEKFMVNDITGSVLLSLQIDDLKELGIKSFGKRHQLMSSINH
LRTT
181-248_Q2UBM0 -0.601
>smart|SAM-uniprot|A1CE91|A1CE91_ASPCL/188-255 SAM and PH domain
protein [Aspergillus clavatus]
VAKWSPQDVVKWMQTLGFEDGIVEKFFINDISGSILLELQVEDLKELDIQSFGKRHQLMGAIRQ
LRNG
188-255_A1CE91 -0.581
>smart|SAM-uniprot|A0FVA6|A0FVA6_9BURK/17-82 Adenylate/guanylate
cyclase [Burkholderia phymatum]
QRSTMTDDLTTWLAQIGLGAYAAKFISQGIDWDVLSDLSEGDLKELGLTLGDRKRLAKGLAALN
GV
17-82_A0FVA6   -0.578
>smart|SAM-uniprot|Q0CRF1|Q0CRF1_ASPTN/192-259 Predicted protein
[Aspergillus terreus]
IRKWSPSQVAHWMHIAGYDGTVIDQFLINDITGSVLLSLQIDDLKELGIQSFGKRHQIMTSIDH
LRNT
192-259_Q0CRF1 -0.568
>smart|SAM-uniprot|Q9Y1Y3|Q9Y1Y3_9METZ/817-884 Protein tyrosine
kinase [Ephydatia fluviatilis]
KTPLDYQSIEAWLEAIKMDRYVQNFHSQGLKSIQDVLLLAEEDLRGMGVAIPGHVNKIMSSVRG
AHQQ
817-884_Q9Y1Y3 -0.565
>smart|SAM-uniprot|Q4WNA7|Q4WNA7_ASPFU/189-256 SAM and PH domain
protein (Boi1), putative [Aspergillus fumigatus]
VASWSPQDVVMWMLQLGFEESIVEKFFINDISGAILLELEANDLKELDIQSFGKRHHLMNCIRQ
LKSS
189-256_Q4WNA7 -0.557
>smart|SAM-uniprot|A1DMA0|A1DMA0_NEOFI/194-261 SAM and PH domain
protein [Neosartorya fischeri]
IRAWSPSQVAHWMFIAGYDDSVIEKFIMNDITGSVLLSLQADDLKELGIKSFGKRHQLMGSIDY
LRNT
194-261_A1DMA0 -0.553
>smart|SAM-uniprot|UPI0000E4719B|UPI0000E4719B/694-761
PREDICTED: similar to cask-interacting protein 2
[Strongylocentrotus purpuratus]
VPGYIPGDVGTWLRLLDLSQYHNTLNTNRYGVMEEVFHITWEDLQDIGIDMLGHQKKIMLAVRR
LKDL
694-761_UPI0000E4719B    -0.519
```

```
>smart|SAM-uniprot|Q8BRB1|Q8BRB1_MOUSE/909-976 Ephrin receptor
[Mus musculus]
VDIATFHTTGDWLNGMRTAHCKEIFTGVEYSSCDTIAKISTDDMKKVGVTVVGPQKKIISTIKA
LETQ
909-976_Q8BRB1 -0.512
>smart|SAM-uniprot|P29320|EPHA3_HUMAN/908-975 Ephrin type-A
receptor 3 precursor [Homo sapiens]
VDITTFRTTGDWLNGVWTAHCKEIFTGVEYSSCDTIAKISTDDMKKVGVTVVGPQKKIISSIKA
LETQ
908-975_P29320 -0.511
>smart|SAM-uniprot|UPI000066023D|UPI000066023D/903-970 Homolog
of Brachydanio rerio &quot;EphA4 protein. [Takifugu rubripes]
SEVTSVGTVGDWMDGMRTLPCKEAFSGVSYSSCDTLAKTSTEDLKKVGVTIVGPQKKIVSSLKA
LESH
903-970_UPI000066023D    -0.504
>smart|SAM-uniprot|UPI00001D04ED|UPI00001D04ED/409-476
PREDICTED: similar to sterile alpha and TIR motif containing 1
[Rattus norvegicus]
VASWKEAEVQTWLQQIGFSQYCENFRDQQVDGDLLLRLTDEELQTDLGMKSSITRKRFFRELTE
LKTF
409-476_UPI00001D04ED    -0.503
>smart|SAM-uniprot|Q4SB53|Q4SB53_TETNG/748-851 Chromosome
undetermined SCAF14677, whole genome shotgun sequence [Tetraodon
nigroviridis]
IHQWGTEEVGAWLDFLCLTEYKDIFTGHDVRGAELLHLERRDLKVPTPRLLPQWNLHFLSLEPG
PSRRLGGAGLTVSLLL
748-851_Q4SB53 -0.498
>smart|SAM-uniprot|Q1L8W1|Q1L8W1_DANRE/636-703 Novel protein
similar to vertebrate EPH receptor A3 [Danio rerio]
TDITSFHTMGDWLESARTLPCKDAFSGVSYSSCDTLAKTSAEDFKKVGVTIVGPQKKIVSSLTT
LETH
636-703_Q1L8W1 -0.496
>smart|SAM-uniprot|UPI000155F20A|UPI000155F20A/402-469
PREDICTED: similar to sterile alpha and TIR motif containing
protein 1 [Equus caballus]
VASWKEAEVQTWLQQIGFSQYCQSFREQQVDGDLLLRLTEEELQTDLGMKSGITRKRFFRELTE
LKTF
402-469_UPI000155F20A    -0.493
>smart|SAM-uniprot|UPI00005A1DDA|UPI00005A1DDA/409-476
PREDICTED: similar to sterile alpha and TIR motif containing 1
[Canis lupus familiaris]
```

```
VASWKEAEVQTWLQQIGFSHYCESFREQQVDGDLLLRLTEEELQTDLGMKSGITRKRFFRELTE
LKTF
409-476_UPI00005A1DDA    -0.492
>smart|SAM-uniprot|Q6GP30|Q6GP30_XENLA/12-78 MGC80721 protein
[Xenopus laevis]
VSCWSPEALIEYFRDHNLRDCEKLVKKQGITGRRFLEMTENEIQKFPKLAVPMLIKIQQHMNKK
DEK
12-78_Q6GP30   -0.491
>smart|SAM-uniprot|UPI0000D9E230|UPI0000D9E230/598-665
PREDICTED: similar to sterile alpha and TIR motif containing 1
[Macaca mulatta]
VPSWKEAEVQTWLQQIGFSKYCESFREQQVDGDLLLRLTEEELQTDLGMKSGITRKRFFRELTE
LRTF
598-665_UPI0000D9E230    -0.488
>smart|SAM-uniprot|UPI0000DB748B|UPI0000DB748B/1786-1851
PREDICTED: similar to Spinophilin CG16757-PA [Apis mellifera]
VTDWSKEQVCQWLSGIGLERYASRFLETGINGGNLLRLESRDLKAFGICGEEKSHLKRKLKELR
AQ
1786-1851_UPI0000DB748B  -0.487
>smart|SAM-uniprot|UPI0000D5741F|UPI0000D5741F/587-654
PREDICTED: similar to l(3)mbt-like 3 isoform a [Tribolium
castaneum]
PLRWSVNVVSEFVSTLPGLEHSSDHFRNSEIDGDAFLSLNQRDLIDILHVKVGPAIKLYNVIVL
LRRK
587-654_UPI0000D5741F    -0.478
>smart|SAM-uniprot|Q9VR43|Q9VR43_DROME/77-142 CG15625-PA
[Drosophila melanogaster]
FEWDDMDIRRWINGYGYPQYMNTFRVNMITGRKLLLLDASALCAMNIKNFDHIRHISYGIRMLF
HF
77-142_Q9VR43  -0.470
>smart|SAM-uniprot|UPI000065D140|UPI000065D140/453-520 Homolog
of Homo sapiens &quot;Sterile alpha and TIR motif containing
protein 1 isoform a [Takifugu rubripes]
VPNWKTCEVQTWLQQVGFSAFCNRFQELQVDGDLLLNITDQDLSADLGMTAGLTRKRFLRDLRV
LKTY
453-520_UPI000065D140    -0.457
>smart|SAM-uniprot|Q4SYZ5|Q4SYZ5_TETNG/464-531 Chromosome 10
SCAF11883, whole genome shotgun sequence [Tetraodon
nigroviridis]
VPNWKTCEVQTWLQQVGFSAFCNRFQELQVDGDLLLSITDQDLSADLGMTAALTRKRFLRDLRV
LKTY
464-531_Q4SYZ5 -0.445
```

```
>smart|SAM-uniprot|Q4RA58|Q4RA58_TETNG/3-55 Chromosome
undetermined SCAF24437, whole genome shotgun sequence [Tetraodon
nigroviridis]
DDSLQQYTASFQRQQVDGDKLLKMSHQELLSLGVMRVGHQELILEACDLLCAL
3-55_Q4RA58    -0.441
>smart|SAM-uniprot|Q01JD5|Q01JD5_ORYSA/279-344 OSIGBa0130P02.4
protein [Oryza sativa]
VMPEVPPTVTGLLNSLGLEKYVFLFHAEEVDMAALSQMGDSDLKEIGVPMGPRKKILQAVAPYS
KR
279-344_Q01JD5 -0.438
>smart|SAM-uniprot|Q00Z16|Q00Z16_OSTTA/199-266 Protein kinase
family protein [Ostreococcus tauri]
VDAWTVQDVSEWLRSIELEELVERFAKAKINGYELLRLTEKDLRESLHLERNLERVRAIRAINV
LRAS
199-266_Q00Z16 -0.433
>smart|SAM_PNT-uniprot|Q4S5P4|Q4S5P4_TETNG/492-574 Chromosome 9
SCAF14729, whole genome shotgun sequence [Tetraodon
nigroviridis]
LKDVDTACKLLSIAADPLDWSCVQVQKWLLWTEHLYRLPQVSTMFQELTGRDLCSMTEADFRQR
SSQFGDVLYAHLDIWR
492-574_Q4S5P4 -0.419
>smart|SAM_PNT-uniprot|UPI0000D55E79|UPI0000D55E79/305-387
PREDICTED: similar to CG5583-PA [Tribolium castaneum]
RKDVDTTCAALGISPDPTQWSAADVLSWLQWTSRQFGLTEPVPDHWDMNGPSLAALSEEDFTRR
APQGGMILHAQLEIWK
305-387_UPI0000D55E79    -0.414
>smart|SAM-uniprot|UPI0000D55DD1|UPI0000D55DD1/895-964
PREDICTED: similar to CG4393-PA [Tribolium castaneum]
QNSLNADSIEKWLQTLGMQDYVGLFMSNGFDDVNFLVSKVIEDCDLSEMGITSELERQIILDAV
KQLPLK
895-964_UPI0000D55DD1    -0.412
>smart|SAM-uniprot|Q5TVD8|Q5TVD8_ANOGA/37-101 ENSANGP00000029329
[Anopheles gambiae]
TSLPHMELGEWLQRLDLAEYGDKFRSFNGVEELLFLSESDIKKLGIRNNAHRARIVSSLVALRE
K
37-101_Q5TVD8  -0.408
>smart|SAM_PNT-uniprot|UPI0000D91D5E|UPI0000D91D5E/126-208
PREDICTED: similar to Ets transcription factor PDEF [Monodelphis
domestica]
LKDIETACKLLNITADPNDWNPGNVQKWLLWTEHQYRLPPVGKAFQELGGKELCAMSEEQFRQR
SPLCGDILHAHLDIWK
126-208_UPI0000D91D5E    -0.405
```

```
>smart|SAM-uniprot|UPI0000EBC729|UPI0000EBC729/33-99 PREDICTED:
similar to KIAA2004 protein [Bos taurus]
TDDWTKEDVNQWLESHKIEQKHRAILTAQDVSGAILKYLTKDHLVAMGITFGPAIQIEQLFKEL
LET
33-99_UPI0000EBC729 -0.375
>smart|SAM-uniprot|P18160|KYK1_DICDI/905-972 Non-receptor
tyrosine kinase spore lysis A [Dictyostelium discoideum]
FKNWAPNDVAIWLESFNYGQYRKNFRDNNISGRHLEGITHAMLKNDLGIEPYGHREDIINRLNR
MIQI
905-972_P18160 -0.374
>smart|SAM-uniprot|Q4JFH3|Q4JFH3_CAEEL/1169-1235 Temporarily
assigned gene name protein 137, isoform b [Caenorhabditis
elegans]
IPYWTSEEVCAWLSSIGMSEYGSTFRKNDIQGSELMHLERSDIMDIGITKIGHVKRLQSAIVDL
RAQ
1169-1235_Q4JFH3    -0.333
>smart|SAM-uniprot|UPI000065EE58|UPI000065EE58/3-55 Homolog of
Homo sapiens &quot;membrane associated guanylate kinase
interacting protein-like 1 [Takifugu rubripes]
DDSLQQYISSFQRQQVDGDKLLKMSHQELLSLGVMRVGHQELILEAVDLLCAL
3-55_UPI000065EE58  -0.332
>smart|SAM-uniprot|UPI0000D66E0F|UPI0000D66E0F/397-463
PREDICTED: hypothetical protein [Mus musculus]
HLTWSPERVAEWISDLGFPQYKECFTENFINGQKLIHVNCSNLPQMGITDFEDMKAISYHTRVL
LGI
397-463_UPI0000D66E0F    -0.321
>smart|SAM-uniprot|UPI000065CC80|UPI000065CC80/264-330 FRANK1
protein (GASZ). [Takifugu rubripes]
ECVKKLDEIQLLLHGLDLGYLTDIMTENDITWSHLLSMEKEDLEKIGVTDPGDQQKVMNALQQM
TLD
264-330_UPI000065CC80    -0.319
>smart|SAM-uniprot|UPI0000D9BD12|UPI0000D9BD12/535-598
PREDICTED: similar to CG3108-PA isoform 2 [Macaca mulatta]
HLNWDPEKVAEWISQLGFPQYKECFITNFISGRKLIHVNCSNLPQMGITNFEDMKTEVMDLGET
535-598_UPI0000D9BD12    -0.315
>smart|SAM-uniprot|UPI00006112B0|UPI00006112B0/13-79 PREDICTED:
hypothetical protein [Gallus gallus]
VLAWGTAEVAEWVSGLGFPQYQECFRANGITGRRLILVNCSSLPAMGITDFKHMQEISRHVREL
LGI
13-79_UPI00006112B0 -0.311
```

```
>smart|SAM-uniprot|UPI00004A4D8E|UPI00004A4D8E/518-584
PREDICTED: similar to Muscle M-line assembly protein unc-89
(Uncoordinated protein 89) [Canis lupus familiaris]
YLKWSPEKVAEWISELGFPQYKECFTTNFVCGRKLIHVNCSNLPQMGITDFEDMKVISRHTRAL
LGI
518-584_UPI00004A4D8E    -0.303
>smart|SAM-uniprot|UPI000069E2B4|UPI000069E2B4/55-121 SAM domain
and HD domain-containing protein 1 (Dendritic cell-derived IFNG-
induced protein) (DCIP) (Monocyte protein 5) (MOP-5). [Xenopus
tropicalis]
YREWDVEEVCLFLGSQGLGELEGVFRENKIKGRILEYLTESHLIDLQISSVALRLDLLHCLRML
CQN
55-121_UPI000069E2B4     -0.284
>smart|SAM-uniprot|Q7QIZ0|Q7QIZ0_ANOGA/495-559
ENSANGP00000007551 [Anopheles gambiae]
GHTWVASEWLPDIGLPQYGESFMHSLVDARMLDTLSKKELEKYLGVTRKFHQASIVHGIHVLRI
M
495-559_Q7QIZ0 -0.280
>smart|SAM-uniprot|Q75DM8|Q75DM8_ASHGO/1-67 ABL011Cp
[Eremothecium gossypii]
MATVGASAEQLLQDINCEGYLGAFRKYEISSVDQLQHLDHEILEQVGIRSIGDRIRILNKSRTL
QAG
1-67_Q75DM8    -0.274
>smart|SAM-uniprot|UPI0000D55A44|UPI0000D55A44/1441-1506
PREDICTED: similar to CG16757-PA [Tribolium castaneum]
ISEWSKDQVSQWLLGIGMENHIPKFVELDVNGQALLLLTSADFKILGIASDDKSYLKRKIKELK
IQ
1441-1506_UPI0000D55A44  -0.265
>smart|SAM_PNT-uniprot|Q4H3K3|Q4H3K3_CIOIN/202-287 Transcription
factor protein [Ciona intestinalis]
SDHHDTKELEVFVPSELSSWTNSHVVQWLTWVVHEYKLQKVDCSRFKGITGIDLSKMTVKDFKN
LTSTEKDADVFFSHLQ
202-287_Q4H3K3 -0.260
>smart|SAM-uniprot|Q6DD51|CSKI2_XENLA/535-601 Caskin-2 [Xenopus
laevis]
PQQIPVDLWDWLSQLGLPEYHKQLSENGYESLSTVTELTWEGLQEIGIHRLGHQKKLLLGVKRL
LDL
535-601_Q6DD51 -0.242
>smart|SAM-uniprot|UPI0000EBC36D|UPI0000EBC36D/328-395
PREDICTED: hypothetical protein [Bos taurus]
TENWSEDDVSNWLCAEGLEDLVDIFKMNNIDGRELLNLTKESLADDLKIGFSQWPEIIFEKVVE
KKQI
```

```
328-395_UPI0000EBC36D    -0.240
>smart|SAM-uniprot|Q4SUQ5|Q4SUQ5_TETNG/632-699 Chromosome
undetermined SCAF13844, whole genome shotgun sequence [Tetraodon
nigroviridis]
LPEFIPSDLGEWLSAIGLSQYHKKLSENGYDSISIVKDLTWEDLQEIGITKLGHQKKLMLAVKK
LCDI
632-699_Q4SUQ5 -0.238
>smart|SAM-uniprot|UPI00006A0F24|UPI00006A0F24/536-603 Caskin-2.
[Xenopus tropicalis]
LPQQTPVDLWDWLSQLGLPEYHKQLSENGYESLSTVTELTWEGLQEIGIHRLGHQKKLLLGVKR
LTDL
536-603_UPI00006A0F24    -0.233
>smart|SAM-uniprot|A2AAI8|A2AAI8_MOUSE/554-621 Cask-interacting
protein 2 [Mus musculus]
LPNYIPVDLLEWLCALGLPQYHKQLVSSGYDSMGLVADLTWEELQEIGVNKLGHQKKLMLGVKR
LAEL
554-621_A2AAI8 -0.231
>smart|SAM-uniprot|Q4TAH3|Q4TAH3_TETNG/388-455 Chromosome
undetermined SCAF7326, whole genome shotgun sequence [Tetraodon
nigroviridis]
LPNYIPPDLGEWLSAIGLPQYQQRLCDNGYDSITIVKDITWEDLQEIGITKLGHQKKLMLAVKR
LCDL
388-455_Q4TAH3 -0.230
>smart|SAM-uniprot|UPI0000140BCB|UPI0000140BCB/555-622 Caskin-2.
[Homo sapiens]
LPSYIPTDLLEWLCALGLPQYHKQLVSSGYDSMGLVADLTWEELQEIGVNKLGHQKKLMLGVKR
LAEL
555-622_UPI0000140BCB    -0.229
>smart|SAM-uniprot|UPI0000F2091C|UPI0000F2091C/702-769
PREDICTED: hypothetical protein [Danio rerio]
LPDYIPSDIGKWLSAIGLPQYQKKLAENGYDSISIVQDITWEDLQEIGITKLGHQKKLMLAVKR
LSDV
702-769_UPI0000F2091C    -0.228
>smart|SAM-uniprot|Q4T9Q1|Q4T9Q1_TETNG/159-219 Chromosome
undetermined SCAF7511, whole genome shotgun sequence [Tetraodon
nigroviridis]
VSSWTTQQVCHWLRGLSMDRYVPEFSARDVDGQELVQMDGKKLKVNKRQLHLKEEEHKQGQ
159-219_Q4T9Q1 -0.226
>smart|SAM-uniprot|UPI0000EBDED2|UPI0000EBDED2/555-622
PREDICTED: similar to cask-interacting protein 2 [Bos taurus]
LPNYIPADLLEWLCALGLPQYHKQLVSSGYDSMGLVADLTWEELQEIGVNKLGHQKKLMLGVKR
LAEL
```

```
555-622_UPI0000EBDED2    -0.225
>smart|SAM-uniprot|UPI00004A5EB5|UPI00004A5EB5/555-622
PREDICTED: similar to cask-interacting protein 2 [Canis lupus
familiaris]
LPNYIPADLREWLCALGLPQYHKQLVSSGYDSMGLVADLTWEELQEIGVNKLGHQKKLMLGVKR
LAEL
555-622_UPI00004A5EB5    -0.223
>smart|SAM-uniprot|UPI0000E49437|UPI0000E49437/1934-2000
PREDICTED: hypothetical protein [Strongylocentrotus purpuratus]
PWTWGIAEVVQFLAETGEGSCAECFCRQNINGQKLMSLSKEQLVKLTGMKVAPSLKIYEQIVKL
RSQ
1934-2000_UPI0000E49437  -0.222
>smart|SAM-uniprot|UPI000069DF96|UPI000069DF96/514-581 RNA-
binding protein with serine-rich domain 1 (SR-related protein
LDC2). [Xenopus tropicalis]
LPEYKPANLSLWLTMIGLIQYYKVLVENGYENIDFITDITWEDLQEIGITKLGHQKKLMLAVKK
LAEI
514-581_UPI000069DF96    -0.221
>smart|SAM-uniprot|UPI000065DCB1|UPI000065DCB1/483-550 Homolog
of Homo sapiens &quot;CASK interacting protein 1 [Takifugu
rubripes]
LPEQKPANLGEWLSAIGLSQYHQVLVQNGYENIEFITDITWEDLQEIGITKLGHQKKLMLAVKR
LAEM
483-550_UPI000065DCB1    -0.220
>smart|SAM-uniprot|Q4S059|Q4S059_TETNG/551-618 Chromosome
undetermined SCAF14784, whole genome shotgun sequence [Tetraodon
nigroviridis]
LPEHKPANLGEWLSAIGLSQYHQVLVQNGYENIEFITDITWEDLQEIGITKLGHQKKLMLAVKR
LAEM
551-618_Q4S059 -0.219
>smart|SAM-uniprot|UPI0000F1D271|UPI0000F1D271/549-616
PREDICTED: similar to cask-interacting protein 1 [Danio rerio]
IPQEKPSSLAEWLSAIGLNQYYQTLVQNGYDNMDFISDITLEDLKEIGFTKLGHQKKLMLAIKR
LSES
549-616_UPI0000F1D271    -0.218
>smart|SAM-uniprot|Q4S703|Q4S703_TETNG/1550-1617 Chromosome 14
SCAF14723, whole genome shotgun sequence [Tetraodon
nigroviridis]
QPSGNISSVSDWLTSIGLPMYASPLAAAGVHTLSGVSSLTESGAREAGVRDERHLRRLVGEARL
VGTE
1550-1617_Q4S703    -0.216
```

```
>smart|SAM-uniprot|UPI0000F1DA4B|UPI0000F1DA4B/534-601
PREDICTED: similar to cask-interacting protein 1 [Danio rerio]
LPDQKPASLGEWLAMIGLSQYHQVLVQNGYENIDFITDITWEDLQEIGITKLGHQKKLMLAVKK
LAEI
534-601_UPI0000F1DA4B    -0.215
>smart|SAM-uniprot|UPI00006CD8A5|UPI00006CD8A5/320-387 protein
kinase domain [Tetrahymena thermophila]
LREWNTGHICKWLDSIGLGEFSPSFESNQICGQSICMLTEQELEQSIGMKALGKRKIFLKYQEM
LKKY
320-387_UPI00006CD8A5    -0.211
>smart|SAM-uniprot|UPI0000DC1A24|UPI0000DC1A24/472-539 Caskin-1
(CASK-interacting protein 1). [Rattus norvegicus]
SASEGKANLAVWLSMIGLAQYYKVLVDNGYENIDFITDITWEDLQEIGITKLGHQKKLMLAVRK
LAEL
472-539_UPI0000DC1A24    -0.210
>smart|SAM-uniprot|Q4SW70|Q4SW70_TETNG/33-99 Chromosome 9
SCAF13686, whole genome shotgun sequence [Tetraodon
nigroviridis]
FPQWGEEETCQYLRKEGFGEWQEAFREREIDGARLRSLRDADLEQMGIRCPDERLRLLRSLRKL
WQI
33-99_Q4SW70   -0.196
>smart|SAM-uniprot|UPI0001554AA4|UPI0001554AA4/12-78 PREDICTED:
similar to SLP-76 [Ornithorhynchus anatinus]
VLGWDPEDLAEYFKKLNFKDCEKAVRKFHIDGERFLNLSDNDIQKFPKLRVPILSKLSQDINKN
EER
12-78_UPI0001554AA4 -0.195
>smart|SAM-uniprot|UPI0000141AD9|UPI0000141AD9/12-78 Lymphocyte
cytosolic protein 2 (SH2 domain-containing leukocyte protein of
76 kDa) (SLP-76 tyrosine phosphoprotein) (SLP76). [Homo sapiens]
VLGWDPDSLADYFKKLNYKDCEKAVKKYHIDGARFLNLTENDIQKFPKLRVPILSKLSQEINKN
EER
12-78_UPI0000141AD9 -0.191
>smart|SAM-uniprot|Q08DS6|Q08DS6_BOVIN/12-78 Similar to
Lymphocyte cytosolic protein 2 [Bos taurus]
VLSWDPNTLADYFKKLNYKDCEKVVRKYDIDGPRFLNLAENDIQKFPKLRVPILSKLSQEINKN
EER
12-78_Q08DS6   -0.187
>smart|SAM_PNT-uniprot|A2BIB5|A2BIB5_MOUSE/31-115 Ets homologous
factor [Mus musculus]
CNVSSGFFGSQWHEIHPQYWTKYQVWEWLQHLLDTNQLDASCIPFQEFDISGEHLCSMSLQEFT
RAAGSAGQLLYSNLQH
31-115_A2BIB5  -0.182
```

```
>smart|SAM_PNT-uniprot|UPI00015613AC|UPI00015613AC/31-115
PREDICTED: similar to Ets homologous factor [Equus caballus]
CNVSSGFFGGQWHEIHPQYWTKYQVWEWLQHLLDTNQLDASCIPFQEFDINGEHLCGMSLQEFT
RAAGTAGQLLYSNLQH
31-115_UPI00015613AC     -0.181
>smart|SAM-uniprot|Q4T305|Q4T305_TETNG/64-130 Chromosome 18
SCAF10147, whole genome shotgun sequence [Tetraodon
nigroviridis]
LPDHKPANLADWLAHLGLSQYYQVLVQNGYENIDFISDISLEDLQEIGITKLAVEAVEHKIKEE
DAE
64-130_Q4T305  -0.180
>smart|SAM-uniprot|Q5ZMC3|WSDU1_CHICK/330-396 WD repeat, SAM and
U-box domain-containing protein 1 [Gallus gallus]
AENWSEDDVSAWLCAQGFAELVGLFKANNIDGKELVNLTRESLIHELKMSLWLRSKILQKIEEL
RMK
330-396_Q5ZMC3 -0.173
>smart|SAM-uniprot|Q8C4H2|SAMD3_MOUSE/1-66 Sterile alpha motif
domain-containing protein 3 [Mus musculus]
METWSVDQVCKWLVEKNLGELVPRFQEEEVSGATLLALNDRMVQQLVKKIGHQAVLMDFIKKYK
QG
1-66_Q8C4H2    -0.170
>smart|SAM-uniprot|Q1AMZ6|Q1AMZ6_MYTTR/520-586 P63/p73-like
protein [Mytilus trossulus]
LMANGDHSISNWLTTLGLSAYIDNFHQQNLFTMEQLDDFTVEDLQKMRIGTSHRNKIWKALVEF
HSE
520-586_Q1AMZ6 -0.161
>smart|SAM-uniprot|Q27937|Q27937_LOLFO/451-517 P53 tumor
suppressor homolog [Loligo forbesi]
KCEPTENTIAQWLTKLGLQAYIDNFQQKGLHNMFQLDEFTLEDLQSMRIGTGHRNKIWKSLLDY
RRL
451-517_Q27937 -0.156
>smart|SAM_PNT-uniprot|UPI0000F21820|UPI0000F21820/35-120
PREDICTED: hypothetical protein [Danio rerio]
SMMTSNFTSRLWTQDPHPQNWTKFQVWEWLQQTLDMHQIDATSIPFQNFDLDGRQLCNMSFQDF
TRAAGSVGSILFQSLT
35-120_UPI0000F21820     -0.154
>smart|SAM_PNT-uniprot|UPI00006A0047|UPI00006A0047/1-74 ETS
homologous factor (hEHF) (ETS domain-containing transcription
factor) (Epithelium-specific Ets transcription factor 3) (ESE-
3). [Xenopus tropicalis]
WHDIHPQYWSKFQVWEWLQQLVDTNQLDANCIPFHDFDIAGEHLCNMTLPDFTQAAGSAGHILY
NNLQNLKWND
```

```
1-74_UPI00006A0047  -0.151
>smart|SAM_PNT-uniprot|UPI00006A0045|UPI00006A0045/33-117 ETS
homologous factor (hEHF) (ETS domain-containing transcription
factor) (Epithelium-specific Ets transcription factor 3) (ESE-
3). [Xenopus tropicalis]
ALPTTGYFGGQWHDIHPQYWSKFQVWEWLQQLVDTNQLDANCIPFHDFDIAGEHLCNMTLPDFT
QAAGSAGHILYNNLQN
33-117_UPI00006A0045     -0.150
>smart|SAM_PNT-uniprot|UPI00005A35DB|UPI00005A35DB/78-162
PREDICTED: similar to E74-like factor 5 ESE-2a [Canis lupus
familiaris]
AFEHQTACDSYWTSVHPEYWTKRHVWEWLQFCCDQYKLDANCISFCHFNVSGLQLCSMTQEEFV
EAAGICGEYLYFILQN
78-162_UPI00005A35DB     -0.149
>smart|SAM_PNT-uniprot|UPI00006A004B|UPI00006A004B/1-74 ETS-
related transcription factor Elf-5 (E74-like factor 5)
(Epithelium-specific Ets transcription factor 2) (ESE-2)
(Epithelium- restricted ESE-1-related Ets factor). [Xenopus
tropicalis]
WTSLLPEHWTKDHVCEWLQYCCDSYKLDANCIPFSHFNVSGLQLCNMTRDDFTEAAGLCGYYLY
SLLQEIRTQA
1-74_UPI00006A004B  -0.146
>smart|SAM-uniprot|Q0JRM9|Q0JRM9_9VEST/481-547 P73-like protein
[Haliotis tuberculata]
GQTLVDNSVASWLTSIGLSAYIDNFHEKNFLNMFQLDEFSLEDLSAMKIGTSHRNKIWKSLVEY
KQA
481-547_Q0JRM9 -0.141
>smart|SAM_PNT-uniprot|Q8VDK3|ELF5_MOUSE/35-119 ETS-related
transcription factor Elf-5 [Mus musculus]
AFEHQTACDSYWTSVHPEYWTKRHVWEWLQFCCDQYKLDANCISFCHFNISGLQLCSMTQEEFI
EAAGICGEYLYFILQN
35-119_Q8VDK3  -0.138
>smart|SAM_PNT-uniprot|Q5TRG1|Q5TRG1_ANOGA/64-148
ENSANGP00000027450 [Anopheles gambiae]
TSAKKVQYGADGLPVDPRDWTRANVWTWLINLAQSEGLDISPELAQKFPMNGKALCLMSLDMYL
SRVPIGGKMLYRDFRV
64-148_Q5TRG1  -0.128
>smart|SAM_PNT-uniprot|Q703F5|Q703F5_TRICA/16-102 ETS activity
modulator [Tribolium castaneum]
GPTERWQRLSFDDEDNLPKDPRQWTREHVAQWINLVTQQHGLPEVPSSRFLMNGKALCLMSLGM
FLSRVPLGGKLLYKDF
16-102_Q703F5  -0.124
```

```
>smart|SAM-uniprot|Q4P965|VTS1_USTMA/608-672 Protein VTS1
[Ustilago maydis]
ADLSTLEDVPAWLRQLRLHKYTPNFETSHWKEMVMMGDKQLEDKGVAALGARRKMLKTFELVRK
K
608-672_Q4P965 -0.123
>smart|SAM_PNT-uniprot|Q9VB52|Q9VB52_DROME/1278-1367 CG5954-PA,
isoform A [Drosophila melanogaster]
LPNYRLWQLNSAFKLDDVRTNPLHWTSWDVCEYIERALDSTDIAKVIFEQDIDGRALLMLGRKE
LDTYLKLKVGPAVKLY
1278-1367_Q9VB52    -0.121
>smart|SAM-uniprot|UPI0000EFCDA7|UPI0000EFCDA7/60-128
hypothetical protein An17g01280 [Aspergillus niger]
VRSWDEKQVIAWLHSINCGQYEALFKAPANNFNGNNLIECDQKILQEMGIKKIGDRVRIFVAIK
QLRNK
60-128_UPI0000EFCDA7     -0.119
>smart|SAM-uniprot|UPI00001AF4B3|UPI00001AF4B3/237-304 WD
repeat, SAM and U-box domain-containing protein 1. [Homo
sapiens]
TEDWSEEDVSTWLCAQDLKDLVGIFKMNNIDGKELLNLTKESLADDLKIESLGLRSKVLRKIEE
LRTK
237-304_UPI00001AF4B3    -0.117
>smart|SAM-uniprot|Q55FH8|Q55FH8_DICDI/446-512 Kelch repeat-
containing protein [Dictyostelium discoideum]
LTESQKKDVKYFLNSIGMSKYVNRFIEEEIDSTVFHLLTENHLKEIGIKSLGERLSIIKGVREY
SLS
446-512_Q55FH8 -0.116
>smart|SAM-uniprot|UPI000056694E|UPI000056694E/327-394 WD
repeat, SAM and U-box domain containing 1 [Mus musculus]
TEEWSEEDVSVWLRAQGLEDLVGIFRANNIDGKELLHLTKESLAGDLKIESLGLRSKVLRSIEE
LRAK
327-394_UPI000056694E    -0.113
>smart|SAM-uniprot|UPI0000EE0148|UPI0000EE0148/329-396
PREDICTED: similar to WD repeat, sterile alpha motif and U-box
domain containing 1 [Ornithorhynchus anatinus]
IENWSEEEVSVWLCAQGLKDLVGVFKMNNIDGKELLSLTKESLTNDLKIESLGLRSKILRKIEE
LNIK
329-396_UPI0000EE0148    -0.112
>smart|SAM_PNT-uniprot|Q4H3K2|Q4H3K2_CIOIN/201-285 Transcription
factor protein [Ciona intestinalis]
TSEENKKNKRVLVPADPLIWSPCHVTAWLEWVVNEYGLHHIDLTKFSSVTGTELCRMNVEDLTR
YTTRYNSEVLVQHLKF
201-285_Q4H3K2 -0.111
```

```
>smart|SAM_PNT-uniprot|Q4H3K4|Q4H3K4_CIOIN/333-417 Transcription
factor protein [Ciona intestinalis]
QGNGEAETKRVIVPADPLVWTEEHVQEWVEWTITEYNLSDVDRSSFSNINGRALCEMTKEHFHR
ITSAKNADVFMSHLNY
333-417_Q4H3K4 -0.110
>smart|SAM_PNT-uniprot|UPI00005A0990|UPI00005A0990/114-198
PREDICTED: similar to Friend leukemia integration 1
transcription factor (Fli-1 proto-oncogene) (ERGB transcription
factor) isoform 7 [Canis lupus familiaris]
PPNMTTNERRVIVPADPTLWTQEHVRQWLEWAIKEYGLMEIDTSFFQNMDGKELCKMNKEDFLR
ATSLYNTEVLLSHLSY
114-198_UPI00005A0990    -0.109
>smart|SAM_PNT-uniprot|Q4SAF7|Q4SAF7_TETNG/27-113 Chromosome 13
SCAF14688, whole genome shotgun sequence [Tetraodon
nigroviridis]
PMVVSGYFSPLWSGCESQPQFWTKFQVWEWLQQVLDMHQMDTSNFPFQNFNVDGHQLCNLSHQD
YVCAAGSVGSILFQRV
27-113_Q4SAF7  -0.108
>smart|SAM_PNT-uniprot|Q4T1Z3|Q4T1Z3_TETNG/112-195 Chromosome
undetermined SCAF10411, whole genome shotgun sequence [Tetraodon
nigroviridis]
PSSSATEEKRVIVPADPEVWTQDHVRQWLDWAIKEYVLEEVDVVLFQALDGKALCKMTKDDMMR
LTSAYNADILLSHLNY
112-195_Q4T1Z3 -0.107
>smart|SAM_PNT-uniprot|UPI00005A5398|UPI00005A5398/115-199
PREDICTED: similar to v-ets erythroblastosis virus E26 oncogene
like isoform 1 isoform 1 [Canis lupus familiaris]
PPNMTTNERRVIVPADPTLWSTDHVRQWLEWAVKEYGLPDVDILLFQNIDGKELCKMTKDDFQR
LTPSYNADILLSHLHY
115-199_UPI00005A5398    -0.106
>smart|SAM-uniprot|Q5FVN8|WSDU1_RAT/329-396 WD repeat, SAM and
U-box domain-containing protein 1 [Rattus norvegicus]
IEDWSEEDVSKWLRAQGLEDLVDIFRTNNIDGKELLHLTKESLAGDMKIESLGLRSKVLRSIDE
LRTR
329-396_Q5FVN8 -0.105
>smart|SAM_PNT-uniprot|Q4Z8P1|Q4Z8P1_RAT/114-198 Friend leukemia
integration 1 [Rattus norvegicus]
PPNMTTNERRVIVPADPTLWTQEHVRQWLEWAIKEYGLMEIDTSFFQNMDGKELCKMSKEDFLR
ATSLYNTEVLLSHLSY
114-198_Q4Z8P1 -0.104
>smart|SAM_PNT-uniprot|Q1L8J3|Q1L8J3_DANRE/137-200 Friend
leukemia integration 1 [Danio rerio]
```

```
PPNMTTNERRVIVPADPSLWSPDHVRQWLDWAIKEYGLQEIDTAMFHSTDGKELCKMSKDDFLR
137-200_Q1L8J3 -0.103
>smart|SAM_PNT-uniprot|A2T6E0|A2T6E0_PAPHA/1-77 ELF5 [Papio
hamadryas]
CDSYWTSVHPEYWTKRHVWEWLQFCCDQYKLDINCISFCHFNISGLQLCSMTQEEFVEAAGLCG
EYLYFILQNIRTQ
1-77_A2T6E0    -0.102
>smart|SAM_PNT-uniprot|Q6GNZ9|Q6GNZ9_XENLA/124-208 MGC80765
protein [Xenopus laevis]
PPNMTTNERRVIVPADPTLWSTDHVRQWLEWAIKEYGLPDVDVLLFQNIDGKELCKMTKEDFHR
LTPSYNADILLSHLHY
124-208_Q6GNZ9 -0.100
>smart|SAM_PNT-uniprot|P41157|FLI1_XENLA/113-197 Retroviral
integration site protein Fli-1 homolog [Xenopus laevis]
PPNMTTNERRVIVPADPALWSQDHVRQWLEWAIKEYGLVEIDCSLFQNIDGKELCKMSKEDFLR
STSIYNTEVLLSHLNY
113-197_P41157 -0.099
>smart|SAM_PNT-uniprot|UPI0000DC187D|UPI0000DC187D/111-195 v-ets
erythroblastosis virus E26 oncogene like [Rattus norvegicus]
PPNMTTNERRVIVPADPTLWSTDHVRQWLEWAVKEYGLLDVDVLLFQNIDGKELCKMTKDDFQR
LTPSYNADILLSHLHY
111-195_UPI0000DC187D     -0.098
>smart|SAM_PNT-uniprot|UPI000155479C|UPI000155479C/35-119
PREDICTED: similar to E74-like factor 5 ESE-2a [Ornithorhynchus
anatinus]
TYDHQTACDSYWTSVHPEYWTKRHVWEWLQFCCDQYKLDANCISFCHFNINGLQLCSMTQEEFI
DAAGICGEYLYFILQN
35-119_UPI000155479C      -0.097
>smart|SAM-uniprot|UPI0000D55DEA|UPI0000D55DEA/136-206
PREDICTED: similar to Stromal interaction molecule homolog
precursor [Tribolium castaneum]
VHNWTVEQTTDWLVSSVDLPQYVPSFISHKVTGANLPRLAANNVNYLNHLGIKDPIHKQKIVLK
AMDVVLF
136-206_UPI0000D55DEA     -0.094
>smart|SAM_PNT-uniprot|UPI000065F421|UPI000065F421/53-135
Homolog of Homo sapiens &quot;ESE-1a [Takifugu rubripes]
ISNLSGQSYRQIEPQYWTVDNVLEWISDHVESNKFDAHSLSLALCSMDGYSLCQMAQDQMIEAF
GSQLGSHLYHSLQEHK
53-135_UPI000065F421      -0.093
>smart|SAM-uniprot|UPI0000DB6D0E|UPI0000DB6D0E/6-74 PREDICTED:
similar to SMSr CG32380-PA, isoform A [Apis mellifera]
```

```
IIEWTQKDVAKWLSESGHEKFVDLFFDQEIDGKVLLTLKEDDLKLISMNIKKIGDIKRLYISIK
QLQRA
6-74_UPI0000DB6D0E  -0.090
>smart|SAM-uniprot|UPI000065CB47|UPI000065CB47/326-393 Homolog
of Brachydanio rerio &quot;Similar to U5 snRNP-specific 40 kDa
protein (HPrp8-binding). [Takifugu rubripes]
VSDWSEQDVSSWLLEEGLEGLVDKFRANNIDGTELLNLTKETLASELHIESVGLRNKLLRKVEE
LKNE
326-393_UPI000065CB47    -0.081
>smart|SAM-uniprot|Q7KUZ1|Q7KUZ1_DROME/117-188 CG9126-PC,
isoform C [Drosophila melanogaster]
VHNWTIEQTTDWLAQSVQLPQYVDLFKLHKVTGAALPRLAVNNLQYVGNVLGIKDPIHKQKISL
KAMDVVLF
117-188_Q7KUZ1 -0.080
>smart|SAM_PNT-uniprot|Q4TA74|Q4TA74_TETNG/63-151 Chromosome
undetermined SCAF7420, whole genome shotgun sequence [Tetraodon
nigroviridis]
VSNTAYMSCLSGQSYRQTEPQYWTVENVLEWISDHVESNKIDAHSLSLALCSMNGYSLCQMAQD
QMMEVFGPQLGSHLYQ
63-151_Q4TA74  -0.074
>smart|SAM-uniprot|UPI000069E562|UPI000069E562/333-400 WD
repeat, SAM and U-box domain-containing protein 1. [Xenopus
tropicalis]
VEAWSEEEVATWLSQEGLNEVEHIFKANNIDGKELLHLTKDSLLNDLKIESLGLRNKLVRRIED
LKSS
333-400_UPI000069E562    -0.073
>smart|SAM-uniprot|A0AUS0|WSDU1_DANRE/344-411 WD repeat, SAM and
U-box domain-containing protein 1 [Danio rerio]
VSEWSEEEVLAWLREEGLEAVTDAFKSNNIDGEELLSLSKETLSSDLHIESLGLRSKVMKKIEE
LKMV
344-411_A0AUS0 -0.065
>smart|SAM-uniprot|UPI0000F2BF15|UPI0000F2BF15/1120-1186
PREDICTED: similar to protein phosphatase 1, regulatory subunit
9B, spinophilin, [Monodelphis domestica]
ANHWTSQQVGQWLESLNLDQYTAEFAARQVDGQQLLQLDGGKLKSLGLSNSHDRALVKRKLKEL
VAA
1120-1186_UPI0000F2BF15  -0.054
>smart|SAM-uniprot|Q8J218|Q8J218_EMENI/60-126 MAPKK kinase
[Emericella nidulans]
VRSWDENQVISWLHSINCQQYEPLFRANNFNGNNLIECDQKILQEMGIKKIGDRVRIFVAIKQL
RNK
60-126_Q8J218  -0.053
```

```
>smart|SAM-uniprot|Q4WTX3|Q4WTX3_ASPFU/61-127 MAP kinase kinase
kinase SteC [Aspergillus fumigatus]
VRSWDEKQVINWLHSINCGQYESLFRANNFNGNNLIECDQKILQEMGIKKVGDRVRIFVAIKKL
RNK
61-127_Q4WTX3  -0.048
>smart|SAM-uniprot|UPI000023D1D8|UPI000023D1D8/60-126
hypothetical protein FG05484.1 [Gibberella zeae]
VKNWDEDQVCEYLRTVKCGEYEKIFRKNHINGENLLEMDKDVLKEMGVEKVGDRVRLFLSIKKL
RTR
60-126_UPI000023D1D8     -0.047
>smart|SAM-uniprot|A1C9Y4|A1C9Y4_ASPCL/60-126 MAP kinase kinase
kinase Ste11/SteC [Aspergillus clavatus]
VRSWDEKQVIAWLHSINCGQYESLFRANNFNGDNLIECDQKILQEMGVKKIGDRVRIFVAIKKL
RNK
60-126_A1C9Y4  -0.046
>smart|SAM-uniprot|Q0MQZ5|Q0MQZ5_PENMA/20-86 STE11-like protein
[Penicillium marneffei]
VRSWDEKKVADWLHSIRCAQYEPLFKANNFNGDNLLECDQKILQEIGIKKIGDRVRIFVAIKQL
RNK
20-86_Q0MQZ5   -0.019
>smart|SAM_PNT-uniprot|UPI000065EC68|UPI000065EC68/57-145
Homolog of Gallus gallus &quot;Transforming protein p54/c-ets-1.
[Takifugu rubripes]
GETFTGAVLLYTLYICIYADPTLWSEWEVNYWLDWCQSEFGLHSLDSELRCLQGRDLCALDKDA
FLALISDCTAGEILWE
57-145_UPI000065EC68     -0.017
>smart|SAM-uniprot|UPI0000EBD8A0|UPI0000EBD8A0/68-135 PREDICTED:
similar to centaurin delta 2, partial [Bos taurus]
EAGDGALSVAEWLRALHLEQYTRLFEQHGLVWATDCQGLSDTRLVDMGMVLPGHRRRILAGLLR
AHTP
68-135_UPI0000EBD8A0     -0.013
>smart|SAM-uniprot|UPI0000F2D237|UPI0000F2D237/3-70 PREDICTED:
similar to centaurin, delta 2, [Monodelphis domestica]
EEVGAMLPVAEWLRALHLEQYAGHFEQYGLVWAADCRGLSDGQLLNMGVTMPGHRRRILSGLQR
AYAT
3-70_UPI0000F2D237  -0.012
>smart|SAM-uniprot|UPI0000F1D36C|UPI0000F1D36C/396-462
PREDICTED: hypothetical protein [Danio rerio]
LSEWTNQQVCHWLMGMNMEQYIAEFTAKGVDGQHLLSMDSSKLKELGVSSQRDRATIKRRLKDM
KKA
396-462_UPI0000F1D36C    -0.004
```

```
>smart|SAM-uniprot|Q7JQG5|Q7JQG5_DROME/321-388 LD09801p
[Drosophila melanogaster]
LPHFVPGSIEEWLQLLRLEEYIQPLLEQNYKTVRDVTQVTWEDLEDIGIVKLGHQKKILLAIKR
VKDI
321-388_Q7JQG5 -0.003
>smart|SAM-uniprot|UPI000155CEFD|UPI000155CEFD/657-723
PREDICTED: similar to KS5 protein [Ornithorhynchus anatinus]
ISEWTTQQVCHWLMGMNMEQYITEFTAKNIDGQQLMLLDSDKLKTLGVTSQNDRSSIKKKIKDM
KKT
657-723_UPI000155CEFD    -0.001
>smart|SAM-uniprot|Q4RNS2|Q4RNS2_TETNG/682-748 Chromosome 2
SCAF15010, whole genome shotgun sequence [Tetraodon
nigroviridis]
VSEWTSVQVCHWLMGMNMDQYTQEFTAKGVDGQQLLHLDSDRLKALGVSSQSDRSTLKKKMKEL
RKA
682-748_Q4RNS2 0.000
>smart|SAM-uniprot|Q4S122|Q4S122_TETNG/426-492 Chromosome 1
SCAF14770, whole genome shotgun sequence [Tetraodon
nigroviridis]
PPYPMDSSISSFLLRLGCAGCLDYFTAQGLTNIYQIENYNMEDLSRLKIPSEFQHIIWKGIMEH
RQA
426-492_Q4S122 0.007
>smart|SAM-uniprot|UPI0000E489DC|UPI0000E489DC/1262-1325
PREDICTED: similar to inositol polyphosphate phosphatase-like 1
[Strongylocentrotus purpuratus]
RPITITEWLSNLGLMQYKQQLYDNGWDNVAFLDTITDQDLISARVESIKHRRHMLESITCIMAM
1262-1325_UPI0000E489DC  0.012
>smart|SAM-uniprot|Q7QIQ9|Q7QIQ9_ANOGA/433-498
ENSANGP00000014993 [Anopheles gambiae]
NEENDSLAILRFLAAFKLEDYYPIFQKNEIDMETLMMLTETDVKSLGLPLGPYRRLCNAIQERR
EA
433-498_Q7QIQ9 0.018
>smart|SAM-uniprot|UPI000069F923|UPI000069F923/1076-1143 SAM and
SH3 domain-containing protein 1 (Proline-glutamate repeat-
containing protein). [Xenopus tropicalis]
VSPRHISSVTDWLVSVGLPMYARPLAEAGITTLSTLPSLSLTHLQKAGIKEERHINKLLSATRI
LKSM
1076-1143_UPI000069F923  0.023
>smart|SAM-uniprot|Q7Q0R1|Q7Q0R1_ANOGA/849-915
ENSANGP00000006183 [Anopheles gambiae]
PESWEVKDVATFLTINDCAVHAEQFVQKDIDGKRLLELSKDDIITLLNLKVGPALKIFDLIQQL
KCK
```

```
849-915_Q7Q0R1 0.034
>smart|SAM-uniprot|Q9HFR3|Q9HFR3_PNECA/7-73 Mekk [Pneumocystis
carinii]
VRLWSEEEVGEWLESNNFGDYMDIFKENNINGDILLECNAAVLKELGVKKLGDRIRLSVCIKGL
REK
7-73_Q9HFR3    0.035
>smart|SAM-uniprot|UPI000065EAA8|UPI000065EAA8/695-761 Homolog
of Gallus gallus &quot;KS5 protein. [Takifugu rubripes]
VSEWTNQQVCHWLMGMNMDQYTQEFTAKAVDGQQLLHLDSDKLKALGVSSQSDRSTIKKKLKDL
RKA
695-761_UPI000065EAA8    0.040
>smart|SAM-uniprot|Q29L50|Q29L50_DROPS/1201-1267 GA14249-PA
[Drosophila pseudoobscura]
PDSWNVYDVSQFLRVNDCTAYCDTFSRSKIDGKRLLQLTKDDIMPLLGMKVGPALIISDLITQL
KCK
1201-1267_Q29L50    0.044
>smart|SAM-uniprot|UPI0000D56EC4|UPI0000D56EC4/647-713
PREDICTED: similar to CG4824-PA, isoform A [Tribolium castaneum]
IANTTWSDLPSLLSALGLDRYISLFIRHEVDLPTFTTLSDKDLMTIGVTAFGSRRKMLLAISEL
NKR
647-713_UPI0000D56EC4    0.047
>smart|SAM-uniprot|UPI0000ECB15E|UPI0000ECB15E/521-587 tumor
protein p73-like [Gallus gallus]
PPYPTDCSIVSFLARLGCSSCVDYFTTQGLTTIYQIEHYSMDDLVSLKIPEQFRHAIWKGILDH
RQL
521-587_UPI0000ECB15E    0.053
>smart|SAM-uniprot|P59808|SASH1_MOUSE/1157-1224 SAM and SH3
domain-containing protein 1 [Mus musculus]
LSPGCVASMSDWLISIGLPMYTSTLSDAGFSTLSQVPSLSHSCLQEAGITEERHIRKLITAARL
FKLP
1157-1224_P59808    0.054
>smart|SAM-uniprot|Q7Z8J5|Q7Z8J5_YARLI/189-255 MAP kinase kinase
kinase [Yarrowia lipolytica]
VTSWDKTKVAEWLRSLKCAEYIPMFEQNEICGEVLLEIDQKLLKDMGVVKVGDRVRINVAIKDL
RSQ
189-255_Q7Z8J5 0.057
>smart|SAM-uniprot|UPI0000DA1BC8|UPI0000DA1BC8/1162-1229
PREDICTED: similar to putative adapter and scaffold protein 1
[Rattus norvegicus]
LSPGCVTSVSDWLISIGLPMYTSTLSDAGFSALSQVPSLSHTCLQEAGITEERHIRKLVTAARL
FKLP
1162-1229_UPI0000DA1BC8  0.058
```

```
>smart|SAM-uniprot|UPI000155FD3D|UPI000155FD3D/541-607
PREDICTED: similar to TA p63 alpha isoform 1 [Equus caballus]
PPYPTDCSLVSFLARLGCSSCLDYFTTQGLTTIYQIEHYSMDDLASLKIPEQFRHAIWKGILDH
RQL
541-607_UPI000155FD3D    0.071
>smart|SAM-uniprot|UPI0000D55CC9|UPI0000D55CC9/803-869
PREDICTED: similar to CG16975-PB, isoform B [Tribolium
castaneum]
PDEWNIYDVAQFLRVNDCANYCDSFSKQKVDGKTLMNLSKEDILEYTGGKVGPSLKIFDLIQQL
KIK
803-869_UPI0000D55CC9    0.076
>smart|SAM-uniprot|UPI0000DB6BAA|UPI0000DB6BAA/807-873
PREDICTED: similar to CG16975-PB, isoform B [Apis mellifera]
PSEWNVFDVAQFLRVNDCATYCDNFSKRKIDGKTLLTLTKDQIIDLTGFKVGPSLKIYDLIQQL
KIK
807-873_UPI0000DB6BAA    0.077
>smart|SAM-uniprot|Q5XJ13|Q5XJ13_DANRE/640-706 Zgc:85964 [Danio
rerio]
GTITDEDELSGILRKLSLEKYQPIFEEQEVDMEAFLTLTDGDLQELGIRTDGPRQQILAAISEL
NAG
640-706_Q5XJ13 0.078
>smart|SAM-uniprot|UPI0000E7FE83|UPI0000E7FE83/773-839
PREDICTED: similar to sterile alpha motif domain containing 6
[Gallus gallus]
GTITEDDELTGILKKLSLEKYQPIFEEQEVDMEAFLTLTDGDLKELGIKTDGSRQQILAAISEL
NAG
773-839_UPI0000E7FE83    0.080
>smart|SAM-uniprot|UPI00006A1675|UPI00006A1675/613-682 Ankyrin
repeat and SAM domain-containing protein 1 (Odin). [Xenopus
tropicalis]
GSYVPEQSVAQWLKSIGLLQYESKLVLNGFDDVRFLADTMMDDQDLLDIGITDLQHRHQILRLA
QSLPKP
613-682_UPI00006A1675    0.081
>smart|SAM-uniprot|UPI0000F1EC7A|UPI0000F1EC7A/336-407
PREDICTED: hypothetical protein [Danio rerio]
IHSWTEEHVHFWMEQIFEGDGELQQYANVFRQNHITGQRLLLLSESDMRDLGVTSKGHIIHLKT
EIEKLTHD
336-407_UPI0000F1EC7A    0.097
>smart|SAM-uniprot|UPI0000E814A6|UPI0000E814A6/506-572
PREDICTED: similar to P53-like transcription factor [Gallus
gallus]
```

```
PPYNPDPSLVSFLTGLGCPNCIDYFTSQGLQNIYHLQNLSIEDLGALKIPEQYRMIIWRGLQEL
KQS
506-572_UPI0000E814A6    0.101
>smart|SAM-uniprot|Q4RNR9|Q4RNR9_TETNG/336-410 Chromosome 2
SCAF15010, whole genome shotgun sequence [Tetraodon
nigroviridis]
IYTWTEEHVYFWMQQIFGEGEGPSGMQVYADLFKENHITGRRLLLLTEGDMRDMGVKSKGHVMH
LQAAIHGEIEK
336-410_Q4RNR9 0.103
>smart|SAM_PNT-uniprot|UPI0000548281|UPI0000548281/33-122
PREDICTED: hypothetical protein [Danio rerio]
HLSEITFNSNSTLGSWDQVNPQMWSRQNVLEWIGFHVEESRFDAGLLNLNYCTMDGLTLCATSK
EALMSMFGPELGNLLH
33-122_UPI0000548281     0.104
>smart|SAM-uniprot|UPI0000EBED4F|UPI0000EBED4F/453-519
PREDICTED: similar to P73 alpha protein [Bos taurus]
PPYHADSSLVSFLTGLGCPNCIEYFTSQGLQNIYHLQNLTIEDLGALKIPDQYRMTIWRGLQDL
KQS
453-519_UPI0000EBED4F    0.115
>smart|SAM-uniprot|UPI0000364E65|UPI0000364E65/328-394 Homolog
of Brachydanio rerio &quot;Similar to U5 snRNP-specific 40 kDa
protein (HPrp8-binding). [Takifugu rubripes]
FSDWSEEDVQAWLRQEGLQDLVGTFKSNNIDGAELSRLNKDTAAELGIESVGLRCRLMRKIEAL
KAE
328-394_UPI0000364E65    0.116
>smart|SAM-uniprot|UPI0000DA2DE3|UPI0000DA2DE3/807-876
PREDICTED: similar to cajalin 2 isoform a [Rattus norvegicus]
RPRCPVQTVGQWLESIGLPQYENHLTANGFDNVQFMGSNVMEDQDLLEIGILNSGHRQRILQAI
QLLPKM
807-876_UPI0000DA2DE3    0.122
>smart|SAM-uniprot|UPI000060934C|UPI000060934C/806-875
PREDICTED: ankyrin repeat and sterile alpha motif domain
containing 1B isoform 2 [Mus musculus]
RPRCPVQTVGQWLESIGLPQYENHLMANGFDSVQFMGSNVMEDQDLLEIGILNSGHRQRILQAI
QLLPKM
806-875_UPI000060934C    0.126
>smart|SAM-uniprot|Q6UNX2|Q6UNX2_DANRE/491-557 Transcription
factor TAp73 alpha [Danio rerio]
PPYNPDPSLVSFLTSVGCQNCIDYFTSQGLQSVYHLQTLTMEDLGALKIPEQFRLAIWRGLQEM
KQG
491-557_Q6UNX2 0.127
```

```
>smart|SAM-uniprot|Q9NRX7|Q9NRX7_HUMAN/33-102 Putative 47 kDa
protein [Homo sapiens]
GPRCPVQTVGQWLESIGLPQYENHLMANGFDNVQFMGSNVMEDQDLLEIGILNSGHRQRILQAI
QLLPKM
33-102_Q9NRX7  0.130
>smart|SAM-uniprot|UPI0000E46C34|UPI0000E46C34/1048-1117
PREDICTED: similar to Ankyrin repeat and sterile alpha motif
domain containing 1B [Strongylocentrotus purpuratus]
DGPGGPQSIGAWLDSLNLGQYENMLIANGFDHLDFIGGGIIEEQDLMDIGIVEEIHRKMILEAT
RCLPIM
1048-1117_UPI0000E46C34  0.131
>smart|SAM-uniprot|UPI0000660959|UPI0000660959/33-102 Homolog of
Homo sapiens &quot;Ankyrin repeat and SAM domain-containing
protein 1 [Takifugu rubripes]
CGKALDQPVGEWLEHVGLLQYESKFLLNGFDDLRFMGSNVMEDQDLRDVGITDPGHRKKILHAA
RSLPKV
33-102_UPI0000660959     0.134
>smart|SAM-uniprot|Q6FXI1|Q6FXI1_CANGA/4-62 Candida glabrata
strain CBS138 chromosome B complete sequence [Candida glabrata]
IEFLERAGCSQYADLFEEYSLTDINDLMHIDRDILLEIGVKKLGDRIRILKESKKLQEP
4-62_Q6FXI1    0.137
>smart|SAM-uniprot|UPI0000E7F896|UPI0000E7F896/803-872
PREDICTED: similar to cajalin 2 [Gallus gallus]
RPRCPVQTVGQWLENIGLPQYENHLLANGFDNVQFMGSNVMEDQDLLEIGILNSGHRQRILQAI
QLLPKM
803-872_UPI0000E7F896    0.139
>smart|SAM-uniprot|UPI0000F2C069|UPI0000F2C069/716-785
PREDICTED: similar to KIAA0229 [Monodelphis domestica]
GSRMLEQSVGEWLESIGLQQYESKLLLNGFDDVHFLGSNVMEDQDLREIGISDPQHRRKLLQAA
RSLPKV
716-785_UPI0000F2C069    0.140
>smart|SAM-uniprot|UPI0000D9AC74|UPI0000D9AC74/773-842
PREDICTED: similar to ankyrin repeat and sterile alpha motif
domain containing 1 [Macaca mulatta]
GSRTLEQSVGEWLESIGLQQYESKLLLNGFDDVRFLGSNVMEEQDLRDIGISDPQHRRKLLQAA
RSLPKV
773-842_UPI0000D9AC74    0.144
>smart|SAM-uniprot|UPI0001555BA2|UPI0001555BA2/315-384
PREDICTED: similar to AIDA1C transcript, partial
[Ornithorhynchus anatinus]
GPRPSERSVGEWLGSVGLQQYESKLLLNGFDDVRFLGSNVMEEQDLREIGITDPQHRRKLLQAA
RSLPKV
```

```
315-384_UPI0001555BA2    0.145
>smart|SAM-uniprot|Q9VMJ2|Q9VMJ2_DROME/84-150 CG13996-PA
[Drosophila melanogaster]
VFKWSINDVTDWLRNFGYPEYEQTFRENYIDGHKLLNLDAVALVALNVRNFEHIRHLGRGIRAL
YRK
84-150_Q9VMJ2  0.148
>smart|SAM-uniprot|Q22CV7|Q22CV7_TETTH/73-140 FHA domain protein
[Tetrahymena thermophila]
PLIWSQEDVLEWLKFIGLDKYAPQFKENTIDGSCLNTLTSDDLQNHLKIDNAIQRKKILNWISV
GIQT
73-140_Q22CV7  0.150
>smart|SAM-uniprot|Q1LX04|Q1LX04_DANRE/18-84 Novel protein
similar to vertebrate protein phosphatase 1, regulatory
(Inhibitor) subunit 9A [Danio rerio]
VTEWTCQQVSCWLMGLNLEQYVSLFTAKNLDGEQLLKLDSTALKALGIESSSDRALIKKKLKDL
KVL
18-84_Q1LX04   0.152
>smart|SAM-uniprot|UPI000065D045|UPI000065D045/611-680 Homolog
of Homo sapiens &quot;AIDA-1a [Takifugu rubripes]
GPRCPMQSVGQWLDSIGLVQYENHLLANGFDNVQFMGSNVVEDQDLLEIGILNSAHRQRLLQAI
RLLPRV
611-680_UPI000065D045    0.160
>smart|SAM-uniprot|P28829|BYR2_SCHPO/1-67 Protein kinase byr2
[Schizosaccharomyces pombe]
MEYYTSKEVAEWLKSIGLEKYIEQFSQNNIEGRHLNHLTLPLLKDLGIENTAKGKQFLKQRDYL
REF
1-67_P28829    0.163
>smart|SAM-uniprot|UPI0000F210A8|UPI0000F210A8/685-754
PREDICTED: similar to KIAA0229 [Danio rerio]
SGRLLEQPVGDWLEHIGLPQYESKLLLNGFDDLRYMGNDVMEEQDLREIGITDPGHRRKILSAA
RSLPKV
685-754_UPI0000F210A8    0.167
>smart|SAM-uniprot|A2AM22|A2AM22_MOUSE/486-552 Transformation
related protein 73 [Mus musculus]
PPYHADPSLVSFLTGLGCPNCIECFTSQGLQSIYHLQNLTIEDLGALKVPDQYRMTIWRGLQDL
KQS
486-552_A2AM22 0.168
>smart|SAM-uniprot|UPI00005A0D1F|UPI00005A0D1F/933-999
PREDICTED: similar to tumor protein p73 [Canis lupus familiaris]
PPYHADPSLVSFLTGLGCPNCIEYFTSQGLQNIYHLQNLTIEDLGALKIPDQYRMTIWRGLQDL
KQS
933-999_UPI00005A0D1F    0.173
```

```
>smart|SAM-uniprot|UPI000065D5D3|UPI000065D5D3/496-562 Homolog
of Barbus barbus &quot;P73. [Takifugu rubripes]
PPYNPDPSLVSFLTSLGCQNFIEYFASQGLQSVYHLQTLSMEDLGAMKIPEQSRLAIWRGLQDM
KQG
496-562_UPI000065D5D3    0.181
>smart|SAM-uniprot|Q8J0Z5|Q8J0Z5_CRYNV/94-160 STE11
[Cryptococcus neoformans var. grubii]
LGDWGSQELFTFLDIHKCGQYLAIFQENDINGKILLDLDVTALKSMGIDKISERVRLIGGIKDL
RKR
94-160_Q8J0Z5  0.186
>smart|SAM-uniprot|Q4RQJ9|Q4RQJ9_TETNG/342-408 Chromosome 2
SCAF15004, whole genome shotgun sequence [Tetraodon
nigroviridis]
FSDWSEEDVQAWLREEGLQDLIGIFTTHNIDGAELTRLDKDTAAELGIESLGLRSRLMRKIEAL
KAE
342-408_Q4RQJ9 0.189
>smart|SAM-uniprot|Q4S837|Q4S837_TETNG/488-554 Chromosome 9
SCAF14710, whole genome shotgun sequence [Tetraodon
nigroviridis]
PPYNPDPSLVSFLTSLGCQNFIEYFASQGLQSVYHLQTLSMEDLGAMKIPEQSRLAIWRALQDL
KQD
488-554_Q4S837 0.193
>smart|SAM-uniprot|UPI0000DB6E48|UPI0000DB6E48/469-536
PREDICTED: similar to Ect4 CG7915-PB, isoform B, partial [Apis
mellifera]
VPLWSTEDVREWVKQIGFAECAQNFVESRVDGDLLLQLTEENLKEDIGLANGIRRKRFTRELQN
LKKM
469-536_UPI0000DB6E48    0.196
>smart|SAM-uniprot|UPI0000F51D0A|UPI0000F51D0A/4-70 Neurabin-1
[Rattus norvegicus]
VHEWSVQQVSHWLVGLSLDQYVSEFSAQNISGEQLLQLDGNKLKALGMTSSQDRALVKKKLKEM
KMS
4-70_UPI0000F51D0A  0.198
>smart|SAM-uniprot|Q9P6R7|VTS1_SCHPO/592-656 Protein vts1
[Schizosaccharomyces pombe]
VGNELPQDIPSWLRSLRLHKYTNNLKDTDWDALVSLSDLDLQNRGIMALGARRKLLKSFQEVAP
L
592-656_Q9P6R7 0.202
>smart|SAM-uniprot|UPI0000F20BC0|UPI0000F20BC0/8-76 PREDICTED:
similar to connector enhancer of kinase suppressor of Ras 2,
[Danio rerio]
```

```
VSRWSTSQVVDWMKGLDDCLQQYIKNFEQEKVGGEQLLRITHQELEDLGVSRIGHQELILEAVD
LLCAL
8-76_UPI0000F20BC0  0.211
>smart|SAM-uniprot|UPI000156080B|UPI000156080B/85-155 PREDICTED:
similar to CNKSR family member 3 [Equus caballus]
VLDEKHLQYVEFVRGCGLDDCLQQYVHKFEREKINGEQLLQISHQDLEELGVTRIGHQELVLEA
VDLLCAL
85-155_UPI000156080B     0.212
>smart|SAM-uniprot|UPI0000EBE5C0|UPI0000EBE5C0/1-73 PREDICTED:
similar to connector enhancer of KSR2A [Bos taurus]
MPINTFCETGHKENAKKRGLDDCLQQYIKNFEREKISGDQLLRITHQELEDLGVSRIGHQELIL
EAVDLLCAL
1-73_UPI0000EBE5C0  0.214
>smart|SAM-uniprot|UPI0000E80130|UPI0000E80130/102-170
PREDICTED: similar to CNKSR family member 3 [Gallus gallus]
VTKWSPKQVVDWTKGLDDCLQQYVHKFEREKINGEQLLQISHQDLEDMGITRIGHQELVLEAVD
LLCAL
102-170_UPI0000E80130    0.215
>smart|SAM-uniprot|UPI000065CEF8|UPI000065CEF8/8-76 Homolog of
Homo sapiens &quot;Splice Isoform 1 of Connector enhancer of
kinase suppressor of ras 2 [Takifugu rubripes]
VSKWSSGQVVDWMKGLDDCLQQYIKTFEREKVGGDQLLRITHQELEDLGVSRIGHQELILEAVD
LLCAL
8-76_UPI000065CEF8  0.216
>smart|SAM-uniprot|UPI00005A5BFB|UPI00005A5BFB/8-76 PREDICTED:
similar to connector enhancer of kinase suppressor of Ras 2
isoform 2 [Canis lupus familiaris]
VSKWSPSQVVDWMKGLDDCLQQYIKNFEREKISGDQLLRITHQELEDLGVSRIGHQELILEAVD
LLCAL
8-76_UPI00005A5BFB  0.217
>smart|SAM-uniprot|Q6GQI9|Q6GQI9_XENLA/8-76 MGC79120 protein
[Xenopus laevis]
VSKWTPSQVVDWMKGLDDCLQQYIKNFEREKIAGDQLLRITHQELEELGVTRIGHQELILEAVD
LLCAL
8-76_Q6GQI9    0.218
>smart|SAM-uniprot|UPI0000E7FAB9|UPI0000E7FAB9/8-76 PREDICTED:
similar to connector enhancer of KSR2A [Gallus gallus]
VSKWSPSQVVDWMAGLDDCLQQYIKNFEREKISGDQLLRITHQELEDLGVTRIGHQELILEAVD
LLCAL
8-76_UPI0000E7FAB9  0.219
>smart|SAM-uniprot|UPI0000F2C1E1|UPI0000F2C1E1/8-69 PREDICTED:
similar to CNKSR family member 3 [Monodelphis domestica]
```

```
LAASSKKGLDDCLQQYVHKFEREKINGEQLLQISHQDLEELGVSRIGHQELVLEAVDLLCAL
8-69_UPI0000F2C1E1  0.220
>smart|SAM-uniprot|UPI000155E0FD|UPI000155E0FD/984-1050
PREDICTED: similar to PPP1R9A protein isoform 1 [Equus caballus]
VHEWSVQQVSHWLMSLNLEQYVSEFSAQNITGEQLLQLDGNKLKALGMTSSQDRAVVKKKLKEM
KVS
984-1050_UPI000155E0FD   0.221
>smart|SAM-uniprot|UPI0000F208C4|UPI0000F208C4/4-72 PREDICTED:
similar to maguin-like protein, partial [Danio rerio]
ITKWTAAQVVDWIRGLDDSLQQYISNFEREKISGERLLKISHQELEELSVTRVGHQELILEAVD
LLCAL
4-72_UPI0000F208C4  0.222
>smart|SAM-uniprot|UPI0000DB7389|UPI0000DB7389/6-74 PREDICTED:
similar to connector enhancer of ksr CG6556-PA [Apis mellifera]
VAEWKTEQVCEWLKGLDNSVLPYVHSFTNHAVNGQQLLSLRPEDLEHLGVLKLGHQEIILEAVE
YLRNF
6-74_UPI0000DB7389  0.223
>smart|SAM-uniprot|UPI000065F984|UPI000065F984/8-76 Homolog of
Homo sapiens &quot;Splice Isoform 1 of Connector enhancer of
kinase suppressor of ras 2 [Takifugu rubripes]
VSKWTTGQVVDWMKGLDDCLQQYVCVFERGGVCGERLLRISHAELEELGVSRIGHQELILEAVD
LLCAL
8-76_UPI000065F984  0.224
>smart|SAM-uniprot|Q4S8W9|Q4S8W9_TETNG/8-77 Chromosome 7
SCAF14703, whole genome shotgun sequence [Tetraodon
nigroviridis]
VSKWSSGQVVDWMKGLDDCLQQYIKTFEKEKVGGDQLLRITHQELEDLGVSRIGHQELILEAVG
PTLCSG
8-77_Q4S8W9    0.225
>smart|SAM-uniprot|UPI0000D57459|UPI0000D57459/6-74 PREDICTED:
similar to CG6556-PA [Tribolium castaneum]
VRDWSVDQVTDWLKGLDSIILQYTTSFLNNGVTGHQLLNLRADDLDNLGVKSLGHQELILESVE
HLRHF
6-74_UPI0000D57459  0.226
>smart|SAM-uniprot|UPI000060F8C6|UPI000060F8C6/1007-1073
Neurabin-1 (Neurabin-I) (Neural tissue-specific F-actin-binding
protein I) (Protein phosphatase 1 regulatory subunit 9A).
[Gallus gallus]
VQEWSVQQVSHWLMSLNLEQYVSEFSAQNINGEHLLQLDGSKLKALGMTSSQDRAIVKKKLKEM
KAS
1007-1073_UPI000060F8C6  0.233
```

```
>smart|SAM_PNT-uniprot|P51023|PNT2_DROME/166-250 ETS-like
protein pointed, isoform P2 [Drosophila melanogaster]
FASWEKEVQKCNITKDPREWTEEHVIYWLNWAKNEFSLVSMNLDPFYKMKGRAMVDLGKEKFLA
ITPPFTGDILWEHLDI
166-250_P51023 0.234
>smart|SAM-uniprot|UPI00006A1675|UPI00006A1675/683-751 Ankyrin
repeat and SAM domain-containing protein 1 (Odin). [Xenopus
tropicalis]
PVLSADLSVPSWLRSLALHQYAPNFISSGYSSMETLRNLWELEIVNVLKVNLLGHRKRILASLA
ERPYE
683-751_UPI00006A1675    0.238
>smart|SAM-uniprot|Q4RP02|Q4RP02_TETNG/3-55 Chromosome 10
SCAF15009, whole genome shotgun sequence [Tetraodon
nigroviridis]
DDSLQQYVSNFEREKISGEQLLKITHQDLEELGLARIGHQELVLEAVDLLCAL
3-55_Q4RP02    0.239
>smart|SAM-uniprot|Q7PT20|Q7PT20_ANOGA/536-599
ENSANGP00000021615 [Anopheles gambiae]
QPNVGMSSIAHWLKSLRLHKYVWLFSNLTYDKMLGITEEYLQSLGVTKGARHKLAICIQKLKER
536-599_Q7PT20 0.242
>smart|SAM_PNT-uniprot|Q16VF8|Q16VF8_AEDAE/149-232 DNA-binding
protein elg, putative [Aedes aegypti]
DTNFKRDQQRLNIPDDPLEWTVAQVKHWIQWAVKIFQLTSIKLQDWSISGKELCDMDHAEFKQK
VPSDPGDLFWTHLELL
149-232_Q16VF8 0.246
>smart|SAM_PNT-uniprot|UPI0000DA36BA|UPI0000DA36BA/297-380
PREDICTED: similar to C-ets-2 protein [Rattus norvegicus]
FSGFHKEQRRLGIPKNPWLWNEQQVCQWLHWATNEFSLVNVNLQRFGMNGQMLCNLGKERFLEL
APDFVGDILWEHLEQM
297-380_UPI0000DA36BA    0.248
>smart|SAM-uniprot|Q16F20|Q16F20_AEDAE/769-836 Sarm1 [Aedes
aegypti]
VPLWSTEDVREWVKQIGFNEYEHNFFDSKVDGDLLLQLTEDNLREDIGMTNGIRRKRFERELQN
LKRM
769-836_Q16F20 0.249
>smart|SAM_PNT-uniprot|UPI0000D55EDE|UPI0000D55EDE/215-300
PREDICTED: similar to DNA-binding protein D-ETS-6 [Tribolium
castaneum]
DSDSDIDNEMVLVPSDPLEWTNTHIKSWLSWCSRKFSLNPKPDFEKFPTTGKELCELTRTDFET
KAGCERTGTILAKHIA
215-300_UPI0000D55EDE    0.251
```

```
>smart|SAM_PNT-uniprot|UPI0000DB74A2|UPI0000DB74A2/209-293
PREDICTED: similar to pointed CG17077-PD, isoform D [Apis
mellifera]
FASWEKERVRLNITKDPRQWSEAAVAHWLHWAIGEFSLAGVAIQPWQNMTGKQICAMGKESFLA
RAPAFMGDILWEHLEI
209-293_UPI0000DB74A2    0.252
>smart|SAM-uniprot|Q174C4|Q174C4_AEDAE/712-775 Smaug protein
[Aedes aegypti]
SPNVGMSSIAHWLKSLRLHKYVWLFSNLTYEQMMEMTEEYLANLGVTKGARHKLVLCIHKLKER
712-775_Q174C4 0.253
>smart|SAM_PNT-uniprot|Q5TRW4|Q5TRW4_ANOGA/27-111
ENSANGP00000025810 [Anopheles gambiae]
FASWEKEVQTFKITKDPRQWTAEHVLIWLNWSIKEFSLEGVNKEPFQKMSGRDIVGLGREGFLA
IAPPFTGDILWEHLEI
27-111_Q5TRW4  0.254
>smart|SAM_PNT-uniprot|Q26645|Q26645_STRPU/122-205 ETS homologue
[Strongylocentrotus purpuratus]
FSSFKKEQERLGIPKDPSRWSECQVVAWVHWSIKEFSLEGVSINNFRISGRDLCTLPKTDFLSR
APPFMGDILWEHIDML
122-205_Q26645 0.255
>smart|SAM_PNT-uniprot|Q330K7|Q330K7_PARLI/122-205 ETS1
[Paracentrotus lividus]
FSSFKKEQERLGIPKDPSGWAECQVVAWVLWAIKEFSLEGVSVNNFRINGRDLCTLPKTDFLSR
APPFMGDILWEHIDML
122-205_Q330K7 0.256
>smart|SAM-uniprot|Q6BRC3|Q6BRC3_DEBHA/10-79 Similar to
CA4000|IPF4262 Candida albicans IPF4262 [Debaryomyces hansenii]
FLQWDPAQVCNYINSLNLGDQTSLGDCFLDNNIEGSLLPFITTEHLKEIGVTKLGTRLQIKKAT
SELIAN
10-79_Q6BRC3   0.259
>smart|SAM_PNT-uniprot|UPI00005E967B|UPI00005E967B/53-136
PREDICTED: hypothetical protein [Monodelphis domestica]
FSGFTKEQQRLGIPKDPRQWTETHVRDWVMWAVNEFSLKGVNFQKFCMSGAALCALGKDCFLEL
APDFVGDILWEHLEIL
53-136_UPI00005E967B     0.260
>smart|SAM-uniprot|UPI0000F2D6E7|UPI0000F2D6E7/4-70 PREDICTED:
similar to ARAP2 [Monodelphis domestica]
VSEVNPELRELLTNINLEQYLGCFQQSGYHTLKDCAQINDGILQKIGVLPTGHRRRILKQLEMI
LSR
4-70_UPI0000F2D6E7  0.265
>smart|SAM-uniprot|UPI0000D9B349|UPI0000D9B349/4-70 PREDICTED:
similar to centaurin delta 1 isoform a, partial [Macaca mulatta]
```

```
VSEVNVDIKDFLMSINLEQYLLHFHESGFTTVKDCAVINDSLLQKIGISPTGHRRRILKQLQII
LSK
4-70_UPI0000D9B349  0.267
>smart|SAM_PNT-uniprot|P01105|MYBE_AVILE/278-361 p135Gag-Myb-
Ets-transforming protein [Contains: Transforming protein v-Myb;
Transforming protein v-Ets] [Avian leukemia virus]
FSGFAKEQQRLGIPKDPQQWTETHVRDWVMWAVNEFSLKGVDFQKFCMNGAALCALGKECFLEL
APDFVGDILWEHLEIL
278-361_P01105 0.268
>smart|SAM_PNT-uniprot|Q8K3Q9|Q8K3Q9_MOUSE/46-129 Ets-1 [Mus
musculus]
FSGFTKEQQRLGIPKDPRQWTETHVRDWVMWAVNEFSLKGVDFQKFCMSGAALCALGKECFLEL
APDFVGDILWEHLEIL
46-129_Q8K3Q9  0.270
>smart|SAM_PNT-uniprot|P13474|ETS1A_CHICK/53-136 Transforming
protein p54/c-ets-1 [Gallus gallus]
FSGFAKEQQRLGIPKDPQQWTETHVRDWVMWAVNEFSLKGVDFQKFCMNGAALCALGKECFLEL
RPDFVGDILWEHLEIL
53-136_P13474  0.271
>smart|SAM-uniprot|UPI000155E0D0|UPI000155E0D0/4-70 PREDICTED:
similar to ARAP2 [Equus caballus]
VSEVNVDIKDFLMSINLEQYLLHFREFGFNTVKDCAAINDSVLHKIGISPTGHRRRILKQLQII
LSK
4-70_UPI000155E0D0  0.272
>smart|SAM_PNT-uniprot|UPI0000D90565|UPI0000D90565/90-173
PREDICTED: similar to v-ets erythroblastosis virus E26 oncogene
homolog 2 (avian), [Monodelphis domestica]
FSGFTKEQRRLGIPNNPWLWTEEQVCQWLFWATNEFSLMDVNFQKFIMNGQVLCNLGKERFLEL
APDFVGDILWEHLEQM
90-173_UPI0000D90565     0.273
>smart|SAM_PNT-uniprot|O62803|O62803_SHEEP/43-126 Transcription
factor Ets-2 [Ovis aries]
FSGFKKEQRRLGIPKSPWLWTEQQVCQWLLWATNEFSLVDVNLQRFGMTGQVLCNLGKERFLEL
APDFVGDILWEHLEQM
43-126_O62803  0.274
>smart|SAM_PNT-uniprot|UPI0000EDFF36|UPI0000EDFF36/90-173
PREDICTED: hypothetical protein [Ornithorhynchus anatinus]
FNGFTKEQCRLGIPNNPWLWTEEQVCQWLFWATNEFSLMDVNFQRFVMNGQVLCNLGKERFLEL
APDFVGDILWEHLEQM
90-173_UPI0000EDFF36     0.275
>smart|SAM-uniprot|UPI000155D1CF|UPI000155D1CF/3-70 PREDICTED:
similar to ARAP2 [Ornithorhynchus anatinus]
```

```
SSPEAHFGIQELLTSINLGQYLPCFLESGYVTIQDCAGINNSVLLKIGVFPTGHRRRILKQLEI
ISSK
3-70_UPI000155D1CF  0.276
>smart|SAM_PNT-uniprot|UPI00006A08C2|UPI00006A08C2/87-170 C-ets-
2 protein. [Xenopus tropicalis]
FNGFTKERCRLGIPGNPWLWDENHVFQWLLWAAKEFSLENVNFQKFLMNGHELCSLGKERFLAL
APDFVGDILWEHLEEM
87-170_UPI00006A08C2     0.277
>smart|SAM-uniprot|UPI0000E8047B|UPI0000E8047B/15-82 PREDICTED:
similar to ARAP2 [Gallus gallus]
SSTGSDIDIEEFLVNINLGQYLPNFKEHGYNVVTDCVGINNSALQQMGVLPTGHRRRILKQLDT
ALSK
15-82_UPI0000E8047B 0.281
>smart|SAM-uniprot|UPI0000F2C069|UPI0000F2C069/790-858
PREDICTED: similar to KIAA0229 [Monodelphis domestica]
CDGNAQPSVPSWLDSLGLQDYIQSFLSSGYSSIDTVKNLWELELVNVLKVHLLGHRKRIIASLA
DRPYE
790-858_UPI0000F2C069    0.282
>smart|SAM-uniprot|O08644|EPHB6_MOUSE/938-1005 Ephrin type-B
receptor 6 precursor [Mus musculus]
LDFPCLDSPQAWLSAIGLECYQDNFSKFGLSTFSDVAQLSLEDLPGLGITLAGHQKKLLHNIQL
LQQH
938-1005_O08644     0.284
>smart|SAM-uniprot|UPI0000E21804|UPI0000E21804/944-1011
PREDICTED: ephrin receptor EphB6 isoform 2 [Pan troglodytes]
LDFPCLDSPQAWLSAIGLECYQDNFSKFGLCTFSDVAQLSLEDLPALGITLAGHQKKLLHHIQL
LQQH
944-1011_UPI0000E21804   0.285
>smart|SAM-uniprot|UPI000061005E|UPI000061005E/873-940 Ephrin
type-A receptor 8 precursor (EC 2.7.10.1) (Tyrosine-protein
kinase receptor EEK) (EPH- and ELK-related kinase) (HEK3).
[Gallus gallus]
LDFSSCLTVEDWLDSIRLGHYRDNFAMAGYSSLGMVMSMNIEDVRSLGITMMGHQKKILSSIQA
MRSQ
873-940_UPI000061005E    0.286
>smart|SAM-uniprot|UPI000155E747|UPI000155E747/908-975
PREDICTED: similar to receptor protein-tyrosine kinase [Equus
caballus]
PEFSAVVSVGDWLQAIKMDRYKDNFTAAGYTTLEAVVHVNQDDLARIGITAVTHQNKILSSVQA
MRTQ
908-975_UPI000155E747    0.287
```

```
>smart|SAM-uniprot|UPI0000F200FF|UPI0000F200FF/910-977
PREDICTED: eph-like receptor tyrosine kinase 4 [Danio rerio]
CSAPSLASVDDWLKLIGLEQYRENFNTAGYSSIESVIPMNHEDLAKMGISCSAHQRKILSSVED
LLSG
910-977_UPI0000F200FF    0.289
>smart|SAM-uniprot|Q4SMF1|Q4SMF1_TETNG/641-708 Chromosome 8
SCAF14552, whole genome shotgun sequence [Tetraodon
nigroviridis]
TDLSAVSTVHEWLRALRMERYQEEFDRAHLDTLDRVSLLTMEDIQNLGVTLLGHQRKILNAAQQ
LRAY
641-708_Q4SMF1 0.290
>smart|SAM-uniprot|UPI0000E1F9B8|UPI0000E1F9B8/907-974
PREDICTED: ephrin receptor EphA4 isoform 2 [Pan troglodytes]
PEFSAVVSVGDWLQAIKMDRYKDNFTAAGYTTLEAVVHVNQEDLARIGITAITHQNKILSSVQA
MRTQ
907-974_UPI0000E1F9B8    0.291
>smart|SAM-uniprot|UPI0000EDE459|UPI0000EDE459/908-975
PREDICTED: similar to chromosome 10 open reading frame 53
[Ornithorhynchus anatinus]
PEFSAVVSVTDWLQAIKMERYKDNFTAAGYTTLEAVVHMNQDDLTRIGITAITHQNKILGSVQS
MRTQ
908-975_UPI0000EDE459    0.292
>smart|SAM-uniprot|Q91845|EPA4A_XENLA/908-975 Ephrin type-A
receptor 4-A precursor [Xenopus laevis]
PEWSQVASVLDWLQAIKMERYKDNFTAAGYTSLEAVVHVNQDDLTRIGISSPSHQNKILSSVQG
MRTQ
908-975_Q91845 0.293
>smart|SAM-uniprot|UPI0000F21D41|UPI0000F21D41/278-345
PREDICTED: hypothetical protein [Danio rerio]
PSFPSFSSVGEWLEAVDMGRYKDNFTAAGYCYLESVARMTVQDVLSLGITCLEHQKQILSAIQT
LRAQ
278-345_UPI0000F21D41    0.294
>smart|SAM-uniprot|Q6VU50|Q6VU50_MANSE/933-1000 Eph receptor
[Manduca sexta]
PDLIQFASVEEWLECIKMSRYVEKFRAGGVTDMDAVVDLTVLQLASLGVTLVGHQKKIMNSVQS
MRAH
933-1000_Q6VU50     0.295
>smart|SAM-uniprot|Q9U8V9|Q9U8V9_EPTBU/275-342 EphA [Eptatretus
burgeri]
TDNPAVRTVPEWLEAIKMGRYADNFAAAGYTTLDSVAQLDADGLVRIGITLIGHQKKILSNIQT
MKAE
275-342_Q9U8V9 0.296
```

```
>smart|SAM-uniprot|UPI000155CF4D|UPI000155CF4D/500-567
PREDICTED: similar to EPH receptor A10 [Ornithorhynchus
anatinus]
SAFPCFSSVGEWLEAIELERYKDNFAAAGYRSLEFVARLTIQDVANLGITTMEHQRQLLAGIQA
LRTQ
500-567_UPI000155CF4D    0.297
>smart|SAM-uniprot|Q4SFK0|Q4SFK0_TETNG/917-984 Chromosome 7
SCAF14601, whole genome shotgun sequence [Tetraodon
nigroviridis]
LALSGCASVGEWLRTIKMERYEDSFLQAGLSTVDQLSQISTQDLLHMGVTLAGHQRKILSSIQS
MTLR
917-984_Q4SFK0 0.298
>smart|SAM-uniprot|UPI000065CF31|UPI000065CF31/904-971 Homolog
of Brachydanio rerio &quot;Eph-like receptor tyrosine kinase
rtk5. [Takifugu rubripes]
LALSGCASVGEWLRAIKMERYEDSFLQAGLSTVDQLAQISTQDLLHMGVTLAGHQRKILSSIQT
MTFR
904-971_UPI000065CF31    0.299
>smart|SAM-uniprot|UPI000155DD70|UPI000155DD70/947-1014
PREDICTED: similar to EPH receptor A8 [Equus caballus]
RGGSGGLTVGDWLDSIRMGRYRDHFAAGGYSSLGMVLRMNAQDVRALGITLMGHQKKILGSIQT
MRAQ
947-1014_UPI000155DD70   0.300
>smart|SAM-uniprot|Q4SRF4|Q4SRF4_TETNG/450-517 Chromosome
undetermined SCAF14527, whole genome shotgun sequence [Tetraodon
nigroviridis]
PDFSSFCSVGDWLEAIKMERYRDNFTAAGYTALESVARMSFEDVMNLGITLVGHQKKIMNSIQT
MRAQ
450-517_Q4SRF4 0.301
>smart|SAM-uniprot|UPI0000F2D7AF|UPI0000F2D7AF/975-1042
PREDICTED: similar to EPH receptor A5, [Monodelphis domestica]
LGTGAYRSVGEWLEAIKMGRYTEIFMENGYSSMDAVAQVTLEDLRRLGVTLVGHQKKIMNSLQE
MKVQ
975-1042_UPI0000F2D7AF   0.302
>smart|SAM-uniprot|UPI0000E2040F|UPI0000E2040F/914-981
PREDICTED: ephrin receptor EphA5 isoform 1 [Pan troglodytes]
LGSGAYRSVGEWLEAIKMGRYTEIFMENGYSSMDAVAQVTLEDLRRLGVTLVGHQKKIMNSLQE
MKVQ
914-981_UPI0000E2040F    0.303
>smart|SAM-uniprot|Q1MT83|Q1MT83_DANRE/587-654 Novel protein
similar to vertebrate Eph receptor A7 [Danio rerio]
```

```
PDFTSFRLVSEWLEAIKMERYMDNFTAAGYSSLESVARMTIEDVMSLGISLVGHQKKIMSSIQT
MRAQ
587-654_Q1MT83 0.304
>smart|SAM-uniprot|Q4T1U5|Q4T1U5_TETNG/19-86 Chromosome
undetermined SCAF10475, whole genome shotgun sequence [Tetraodon
nigroviridis]
PPPSACGSVSEWLRAIKMERYQQDFLQAGFTSLDVVAHLNTEDLLRVGVTLAGHQKKILSSIQT
LRIH
19-86_Q4T1U5   0.305
>smart|SAM-uniprot|A3KG07|A3KG07_MOUSE/926-993 Ephrin receptor
[Mus musculus]
GSGNGDLTVGDWLDSIRMGRYRDHFAAGGYSSLGMVLRMNAQDVRALGITLMGHQKKILGSIQT
MRAQ
926-993_A3KG07 0.306
>smart|SAM-uniprot|Q9U8V7|Q9U8V7_BRABE/267-334 Eph2
[Branchiostoma belcheri]
ALQPVLPTVPQWLDSMKMGRYRDNFTNGGYSSMDLVMKMNIRDLQGIGVTMLAHQKKILNSIQN
LHTQ
267-334_Q9U8V7 0.307
>smart|SAM-uniprot|UPI0000F21110|UPI0000F21110/898-965
PREDICTED: similar to Cek10 protein [Danio rerio]
PDYTTFTTVGDWLDAIKMSRYRDNFRKGCFGSSGPVDPRVAEDLLRIGVTLAGHQKKILGSIQD
MRLQ
898-965_UPI0000F21110    0.308
>smart|SAM-uniprot|UPI000065E0B8|UPI000065E0B8/980-1042 Homolog
of Brachydanio rerio &quot;EphA4 protein. [Takifugu rubripes]
PDLSSVSSVEDWLAALKMTQYRDSFLGSGFTSLPLVTQITAEDLQRIGVSLAGHQKKILTSVQ
980-1042_UPI000065E0B8   0.309
>smart|SAM-uniprot|Q4SFJ9|Q4SFJ9_TETNG/699-761 Chromosome 7
SCAF14601, whole genome shotgun sequence [Tetraodon
nigroviridis]
LALSGCASVGEWLRTIKMERYEDSFLQAGLSTVDQLAQISTQDLLHMGVTLAGHQRKILSSIQ
699-761_Q4SFJ9 0.310
>smart|SAM-uniprot|UPI0000E21660|UPI0000E21660/808-875
PREDICTED: ephrin receptor EphB4 [Pan troglodytes]
PHYSAFGSVGEWLRAIKMGRYEESFAAAGFGSFELVSQISAEDLLRIGVTLAGHQKKILASVQH
MKSQ
808-875_UPI0000E21660    0.311
>smart|SAM-uniprot|UPI0000EBE37D|UPI0000EBE37D/923-990
PREDICTED: similar to ephrin receptor EphB4 precursor isoform 2
[Bos taurus]
```

```
PHYSAFGSVGEWLRAIKMGRYEESFAAAGFGSFELVSQISTEDLLRIGVTLAGHQKKILASVQH
MKSQ
923-990_UPI0000EBE37D    0.312
>smart|SAM-uniprot|UPI00004A6713|UPI00004A6713/208-275
PREDICTED: similar to Ephrin type-A receptor 6 precursor
(Tyrosine-protein kinase receptor EHK-2) (EPH homology kinase-2)
[Canis lupus familiaris]
PEYPLFVTVGDWLDSIKMGQYKNNFMAAGFTTFDLISRMSIDDIRRTGVLLIGHQRRIVSSIQT
LRLH
208-275_UPI00004A6713    0.314
>smart|SAM-uniprot|Q9U8W1|Q9U8W1_EPTBU/275-342 EphC1 [Eptatretus
burgeri]
PEYTMFCSVSEWLEAIKMGRYKENFTNAGYVAWTSVAQMTLEDLQRVGVTLAGHQKRILTSIQA
LRAQ
275-342_Q9U8W1 0.315
>smart|SAM-uniprot|P54758|EPHA6_RAT/957-1024 Ephrin type-A
receptor 6 precursor [Rattus norvegicus]
PEYPLFVTVGDWLDSIKMGQYKSNFMAAGFTTFDLISRMSIEDIRRIGVILIGHQRRIVSSIQT
LRLH
957-1024_P54758     0.316
>smart|SAM-uniprot|UPI00004D424D|UPI00004D424D/898-965 Ephrin
type-B receptor 3 precursor (EC 2.7.10.1) (Tyrosine-protein
kinase receptor HEK-2). [Xenopus tropicalis]
PDYTTFPTVSDWLEAIKMGQYQENFLSAGFTSFHLVAQMTAEDLLRIGVTLAGHQKKLLNSVQD
MRLQ
898-965_UPI00004D424D    0.317
>smart|SAM-uniprot|Q60629|EPHA5_MOUSE/801-869 Ephrin type-A
receptor 5 precursor [Mus musculus]
LGSGAYRSVGEWLEAIKMGRYTEIFMENGYSSMDAVAQVTLEDLRRLGVTLVGHQKKIMSSLQ
EMKVQ
801-869_Q60629 0.318
>smart|SAM-uniprot|UPI000069DB6A|UPI000069DB6A/874-941 Ephrin
type-A receptor 6 precursor (EC 2.7.10.1) (Tyrosine-protein
kinase receptor EHK-2) (EPH homology kinase 2). [Xenopus
tropicalis]
PEYHFVVTVSEWLESIKMGQYAHNFMASGFTTLDMVSRMTIDDIRRIGITLIGHQRRIVTSIQT
LRLQ
874-941_UPI000069DB6A    0.319
>smart|SAM-uniprot|UPI0000D9CE5D|UPI0000D9CE5D/443-511
PREDICTED: similar to cajalin 2 isoform a isoform 2 [Macaca
mulatta]
```

```
HDGYHPTSVAEWLDSIELGDYTKAFLINGYTSMDLLKKIWEVELINVLKINLIGHRKRILASLG
DRLHD
443-511_UPI0000D9CE5D    0.322
>smart|SAM-uniprot|Q89T12|Q89T12_BRAJA/6-71 Blr2238 protein
[Bradyrhizobium japonicum]
PGTGAEGGLKRWLEGIGLAHYSDLFAQHRLDLDVMADLTESDLVELGLPLGDRKRLQRAMAALF
QA
6-71_Q89T12    0.323
>smart|SAM-uniprot|UPI000155CCE6|UPI000155CCE6/882-950
PREDICTED: similar to cajalin 2 [Ornithorhynchus anatinus]
QDGYHPTSVAEWLDSIELGDYTKPFLINGYTSMDLLKKIWEVELINVLKISLIGHRKRILASLG
DRLHE
882-950_UPI000155CCE6    0.325
>smart|SAM-uniprot|Q9PWR5|Q9PWR5_XENLA/902-969 Eph receptor
tyrosine kinase precursor [Xenopus laevis]
SEGMPFRTIAEWLDSIKMQQYTEFFMASPYNSMDKIILMNQEDIKHLGIRQTGHQKRIAFSILG
LKEH
902-969_Q9PWR5 0.327
>smart|SAM-uniprot|UPI0000E1E682|UPI0000E1E682/901-968
PREDICTED: ephrin receptor EphA2 [Pan troglodytes]
SEGVPFRTVSEWLESIKMQQYTEHFMAAGYTAIEKVVQMTNDDIKRIGVRLPGHQKRIAYSLLG
LKDQ
901-968_UPI0000E1E682    0.328
>smart|SAM-uniprot|Q4RXA0|Q4RXA0_TETNG/969-1036 Chromosome 11
SCAF14979, whole genome shotgun sequence [Tetraodon
nigroviridis]
CDGTLFRSVPEWLESIRMSQYSESFARAGITSMEQVLALRHEDIKNIGVRLPGHMKRIAYSILG
LKDE
969-1036_Q4RXA0     0.332
>smart|SAM-uniprot|Q3UNI2|Q3UNI2_MOUSE/902-969 Ephrin receptor
[Mus musculus]
SEGVPFRTVSEWLESIKMQQYTEHFMVAGYTAIEKVVQMSNEDIKRIGVRLPGHQKRIAYSLLG
LKDQ
902-969_Q3UNI2 0.333
>smart|SAM-uniprot|UPI0000E48B74|UPI0000E48B74/6-75 PREDICTED:
similar to connector enhancer of KSR2A [Strongylocentrotus
purpuratus]
IPRWSAKDVADWMKGLDFVLQPYINTILTKKVTGDRLLLMTTSQDLEDLGVLKVGHQEIMLDAI
SLLCNL
6-75_UPI0000E48B74  0.335
```

```
>smart|SAM-uniprot|UPI0000F2B347|UPI0000F2B347/233-296
PREDICTED: similar to sterile alpha motif domain containing 4
isoform 2 [Monodelphis domestica]
GTSSSSTNVPAWLKSLRLHKYAALFSQMTYEEMMALTECQLEAQNVTKGARHKIVISIQKLKER
233-296_UPI0000F2B347    0.338
>smart|SAM-uniprot|Q0VA96|Q0VA96_HUMAN/231-294 SAMD4A protein
[Homo sapiens]
IGTSTSTNVPAWLKSLRLHKYAALFSQMTYEEMMALTECQLEAQNVTKGARHKIVISIQKLKER
231-294_Q0VA96 0.340
>smart|SAM-uniprot|UPI000065E100|UPI000065E100/971-1038 Homolog
of Brachydanio rerio &quot;Eph-like receptor tyrosine kinase 6.
[Takifugu rubripes]
CDGTVFRTVPEWLESIRMSQYSESFARAGITSMEQVLGLRHEDIRNIGVRLPGHLKRIAYSILG
LKDE
971-1038_UPI000065E100   0.341
>smart|SAM-uniprot|UPI000155DE6C|UPI000155DE6C/900-967
PREDICTED: similar to EPH receptor A2 [Equus caballus]
SEGVPFHTVSEWLESIKMQQYTEHFLAAGYTAIEKVVQMTNDDIKRIGVRLPGHQKRIAYSLLG
LKDQ
900-967_UPI000155DE6C    0.342
>smart|SAM-uniprot|UPI000065F26A|UPI000065F26A/946-1013 Homolog
of Brachydanio rerio &quot;Eph-like receptor tyrosine kinase 6.
[Takifugu rubripes]
SDGSPFRSVSEWLESIKMSQYGENFSRAGVVTMDQVLQMKNEDIKNIGVRLPGHLKRIAYSILG
LKDQ
946-1013_UPI000065F26A   0.343
>smart|SAM-uniprot|Q5Y251|Q5Y251_CRYGA/74-140 STE11p
[Cryptococcus gattii]
LDKWGPEQVMGFLAIHGCEHHSGIFIKHSIDGKLLLDFDMSVLTSLGITKVGERIRLMGGIKCL
RQR
74-140_Q5Y251  0.344
>smart|SAM-uniprot|UPI00005E8655|UPI00005E8655/914-980
PREDICTED: similar to Ephrin type-A receptor 1 precursor
(Tyrosine-protein kinase receptor EPH) [Monodelphis domestica]
SDGIPYRSVSEWLESIRMKRYILHFRSAGLDTMECVLELTTEDLKQMGITQAGHQKRILCSIQG
FKD
914-980_UPI00005E8655    0.345
>smart|SAM-uniprot|UPI0000D9A9F5|UPI0000D9A9F5/917-983
PREDICTED: ephrin receptor EphA1 isoform 2 [Macaca mulatta]
SDGIPYRSVSEWLESIRMKRYILHFHSAGLDTMECVLELTAEDLTQMGITLPGHQKRILCSIQG
FKD
917-983_UPI0000D9A9F5    0.346
```

```
>smart|SAM_PNT-uniprot|UPI0000E49D06|UPI0000E49D06/256-340
PREDICTED: hypothetical protein [Strongylocentrotus purpuratus]
CSNYRDEQDRLNIPYDPELWTEDHTLHWLIWASGEFANEALDHVKDIVMTGSEMCALTKEEFCE
RFQKESEDLFWTHLEL
256-340_UPI0000E49D06    0.362
>smart|SAM_PNT-uniprot|Q4S0J3|Q4S0J3_TETNG/163-244 Chromosome 2
SCAF14781, whole genome shotgun sequence [Tetraodon
nigroviridis]
LEGYRKEQVRLGIPYDPVLWSADQVIHWAVWVMKEFNIDEMEIGSIHIPGRDLCAFSQEEFLQK
VPNGEILWSHLELLRK
163-244_Q4S0J3 0.367
>smart|SAM_PNT-uniprot|UPI000001B0B6|UPI000001B0B6/163-244
Transcription factor E4TF1. [Takifugu rubripes]
LEGYRKEQVRLGIPYDPVLWSADQVIHWAVWVMKEFNIDEMEIGSIHIPGRDLCSFSQEEFLQK
VPNGEILWSHLELLRK
163-244_UPI000001B0B6    0.371
>smart|SAM_PNT-uniprot|UPI00005EA6C9|UPI00005EA6C9/170-251
PREDICTED: similar to transcription factor E4TF1-60 [Monodelphis
domestica]
LEGYRKEQERLGIPYDPIQWSTDQVLHWVVWVMKEFSMTDIDLPTLSISGRELCSLNQEDFFQR
VPRGEILWSHLELLRK
170-251_UPI00005EA6C9    0.372
>smart|SAM_PNT-uniprot|UPI000056C69D|UPI000056C69D/187-268 GA-
binding protein transcription factor, alpha subunit [Danio
rerio]
LEGYRKEQVRLNIPYDPVQWTADQVIHWAVWVMKEFGIDEMEVGGIHIPGRQLCGFSQEEFLQR
VPSGEILWSHLELLRK
187-268_UPI000056C69D    0.373
>smart|SAM-uniprot|Q0DLR9|Q0DLR9_ORYSJ/234-300 Os03g0850300
protein [Oryza sativa]
PGSMGYGGVKAWLDGLGLSRYAPVFEIHEVDDEVLPLLTLEDLKDMGIGAVGSRRKLYAAIQKL
QRS
234-300_Q0DLR9 0.374
>smart|SAM-uniprot|Q94071|LIPB_CAEEL/683-749 Liprin-beta
[Caenorhabditis elegans]
ANKWDVHQTLRWLDDIGLPQYKDVFAENVVDGPLLLSLTANDAVEMKVVNAHHYATLARSIQFL
KKA
683-749_Q94071 0.382
>smart|SAM_PNT-uniprot|UPI000069F541|UPI000069F541/170-251 GA-
binding protein alpha chain (GABP-subunit alpha) (Transcription
factor E4TF1-60) (Nuclear respiratory factor 2 subunit alpha).
[Xenopus tropicalis]
```

```
LEGYRKEQERLGIPYDPLQWSVDQVLHWVLWVMKEFCLTEINVNSLGITGRELCNLNQEDFFQR
VPRGEILWSHLELLRK
170-251_UPI000069F541    0.384
>smart|SAM-uniprot|UPI0000EBD618|UPI0000EBD618/1199-1264
PREDICTED: similar to inositol polyphosphate 5-phosphatase [Bos
taurus]
AGGLGEAGMGAWLRAIGLERYEEGLVHNGWDDLEFLSDITEEDLEEAGVQDPAHKRLLLDTLQL
SK
1199-1264_UPI0000EBD618  0.388
>smart|SAM-uniprot|Q69P98|Q69P98_ORYSJ/545-611 Putative SNM1
[Oryza sativa]
ESNVEKRRVMEWLRNLGLSKYEEIFIKEEVDWETLQWLTEEDLLGMGITSLGPRKKIAHALCEL
RKK
545-611_Q69P98 0.396
>smart|SAM-uniprot|UPI0000F2D250|UPI0000F2D250/1171-1236
PREDICTED: similar to inositol polyphosphate phosphatase-like 1,
[Monodelphis domestica]
AGSLGEGSMGAWLRALGMECYEEGLVHNGWDDLEFLSDITEEDLEEAGVQDPAHKRLLLDSLQH
SK
1171-1236_UPI0000F2D250  0.412
>smart|SAM-uniprot|UPI00004DA19F|UPI00004DA19F/226-291 inositol
polyphosphate phosphatase-like 1 [Xenopus tropicalis]
SSSLRESSVGEWLRAIGLERYEEGLIQNGWDDLEFLSDIVEEDLEEAGVLDPTHKRVILESLQA
RK
226-291_UPI00004DA19F    0.429
>smart|SAM-uniprot|Q2I6J1|SHP2A_DANRE/1200-1266
Phosphatidylinositol-3,4,5-trisphosphate 5-phosphatase 2A [Danio
rerio]
SSLSVDCSVGEWLQKLGLQHYEEGLLHNGWDDLEFLSDITEEDLEEAGVRDPAHKKILLASLKQ
QQK
1200-1266_Q2I6J1    0.435
>smart|SAM-uniprot|UPI000065F5C0|UPI000065F5C0/1140-1202 Homolog
of Homo sapiens &quot;Inositol polyphosphate 5-phosphatase
[Takifugu rubripes]
SLSVGESSVGEWLQRLGLERYEHGLLHNGWDDLEFLSDITEEDLEEAGVLDPVHKQILLQSLK
1140-1202_UPI000065F5C0  0.438
>smart|SAM-uniprot|Q86DA5|SARM1_CAEEL/611-678 Sterile alpha and
TIR motif-containing protein tir-1 [Caenorhabditis elegans]
VPGWTCADVQYWVKKIGFEEYVEKFAKQMVDGDLLLQLTENDLKHDVGMISGLHRKRFLRELQT
LKVA
611-678_Q86DA5 0.443
```

```
>smart|SAM-uniprot|UPI0000D55C4D|UPI0000D55C4D/266-329
PREDICTED: similar to Sterile alpha motif domain-containing
protein 4 [Tribolium castaneum]
SNLPVMKSILSWLKSLRLHKYSWVFHNLTYQQMLDLSDESLQTIGITKGARHKLLLSIAKLKER
266-329_UPI0000D55C4D    0.445
>smart|SAM-uniprot|Q2QQY1|Q2QQY1_ORYSJ/139-205 SAM domain family
protein [Oryza sativa]
AAEGEVVDVVEWLWGIGMGRYAAAFEAHEVDGEVLPCLTMDDLRDMGIGAVGARRKLYCAIQRL
PPP
139-205_Q2QQY1 0.459
>smart|SAM-uniprot|UPI0000D9CBA4|UPI0000D9CBA4/638-705
PREDICTED: similar to PTPRF interacting protein binding protein
1 isoform 1 [Macaca mulatta]
FAKWTKEQVCNWLMEQGLGSYLNSGKHWIASGQTLLQASQQDLEKELGIKHSLHRKKLQLALQA
LGSE
638-705_UPI0000D9CBA4    0.463
>smart|SAM-uniprot|UPI0000DA2632|UPI0000DA2632/637-704
PREDICTED: similar to PTPRF interacting protein, binding protein
1 (liprin beta 1) [Rattus norvegicus]
FAKWTKEQVCSWLAEQGLGSYLSSGKHWIMSGQTLLQASQQDLEKELGIKHSLHRKKLQLALQA
LGSE
637-704_UPI0000DA2632    0.464
>smart|SAM-uniprot|UPI0000F2E1B7|UPI0000F2E1B7/601-668
PREDICTED: similar to PTPRF interacting protein binding protein
1 isoform 3 [Monodelphis domestica]
FAKWTKEQVCNWLLDQGLGSYMNSGKHWIASGQTLLQASQQDLEKELGIKHSLHRKKLQLALQA
LGSE
601-668_UPI0000F2E1B7    0.465
>smart|SAM_PNT-uniprot|UPI0000E48771|UPI0000E48771/101-182
PREDICTED: similar to modulator of activity of ets genes
[Strongylocentrotus purpuratus]
KSPSSPLPIMSHVPADPRKWAALHVGRWLEAVSAKYALQVNKTDFVMNGRALCLMKREGFLDRV
PENGAILFEDFRRRLR
101-182_UPI0000E48771    0.466
>smart|SAM-uniprot|UPI000065EBB0|UPI000065EBB0/659-726 Homolog
of Homo sapiens &quot;Splice Isoform 4 of Liprin-beta 1
[Takifugu rubripes]
FARWSKDQVCDWLQEQGLGLYVNMARVWISSGQTLLQASQTDLERELGIKHPLHRKKLQLALQA
LGSE
659-726_UPI000065EBB0    0.469
>smart|SAM-uniprot|Q6DCK5|Q6DCK5_XENLA/598-665 Ppfibp2-prov
protein [Xenopus laevis]
```

```
FAKWTKDQVCSWLRDQGLGGYVNSCKQWIVSGQTLLHASQQDLEKELGIKQPLHRKKLQLALQS
LGSE
598-665_Q6DCK5 0.471
>smart|SAM-uniprot|UPI00006A1F7E|UPI00006A1F7E/605-672 Liprin-
beta-1 (Protein tyrosine phosphatase receptor type f
polypeptide-interacting protein-binding protein 1) (PTPRF-
interacting protein-binding protein 1) (hSGT2). [Xenopus
tropicalis]
FAKWTKEQVCLWLREQGLGGYVGSGKQWIVSGQTLLHASQQDLEKELGIKHPLHRKKLQLALQS
LGSE
605-672_UPI00006A1F7E    0.472
>smart|SAM-uniprot|UPI0000ECD2A9|UPI0000ECD2A9/552-619 PTPRF
interacting protein binding protein 1 [Gallus gallus]
FAKWTKEQVCNWLQDQGLGSYISNGRHWILSGQTLLQASQQDLEKELGIKHPLHRKKLQLALQA
LGSE
552-619_UPI0000ECD2A9    0.475
>smart|SAM-uniprot|Q92YL0|Q92YL0_RHIME/25-91 Putative adenylate
cyclase [Sinorhizobium meliloti]
VDGEGAMDIGAWLRDQGLGQYEGTFRQNDIDPEVLRHLTAEDLIGVGVASVGHRRKLLAAIAAL
REV
25-91_Q92YL0   0.476
>smart|SAM-uniprot|UPI0000E2294D|UPI0000E2294D/555-622
PREDICTED: PTPRF interacting protein, binding protein 2 (liprin
beta 2) [Pan troglodytes]
FAQWSTERVCAWLEDFGLAQYVIFARQWVSSGHTLLTATPQDMEKELGIKHPLHRKKLVLAVKA
INTK
555-622_UPI0000E2294D    0.484
>smart|SAM_PNT-uniprot|A0NC87|A0NC87_ANOGA/18-103
ENSANGP00000029784 [Anopheles gambiae]
DSDGEEECEQTYLPNDPHQWNAEHVSTWISWVSKNFDIFPPLEPARFPQAGSELAPFTKADFWV
CAGSAAGGNTLAKHFA
18-103_A0NC87  0.485
>smart|SAM-uniprot|UPI0000DC21AA|UPI0000DC21AA/555-622 Liprin
beta 2 [Rattus norvegicus]
FAQWSTERVCTWLQDFGLAQYVIFARQWVASGHTLLTATPQDMEKELGIKHPLHRKKLVLAVKA
INTK
555-622_UPI0000DC21AA    0.486
>smart|SAM-uniprot|Q4SVD1|Q4SVD1_TETNG/583-650 Chromosome
undetermined SCAF13765, whole genome shotgun sequence [Tetraodon
nigroviridis]
FAKWSTEQVCEWLEEIGLGQYVIVARHWVTSGQTLLAATPQDLERELVLKNPLHRKKLQLAIQT
ITSK
```

```
583-650_Q4SVD1 0.487
>smart|SAM-uniprot|UPI0000EBD637|UPI0000EBD637/553-620
PREDICTED: similar to PTPRF interacting protein, binding protein
2 (liprin beta 2) [Bos taurus]
FAQWSTERVCTWLEDFGLAQYVIFARQWVTSGHTLLTATPQDMEKELGIKHPLHRKKLVLAVKA
INTK
553-620_UPI0000EBD637    0.488
>smart|SAM-uniprot|UPI0000ECBA73|UPI0000ECBA73/556-623 Liprin-
beta-2 (Protein tyrosine phosphatase receptor type f
polypeptide-interacting protein-binding protein 2) (PTPRF-
interacting protein-binding protein 2). [Gallus gallus]
FAQWSTERVCNWLEDFGLGQYVIFARQWVTSGHTLLTATPQDMEKEMGIKHPLHRKKLVLAIKS
ISTK
556-623_UPI0000ECBA73    0.491
>smart|SAM-uniprot|UPI0000F2D7BF|UPI0000F2D7BF/552-619
PREDICTED: similar to Liprin-beta-2 (Protein tyrosine
phosphatase receptor type f polypeptide-interacting protein-
binding protein 2) (PTPRF-interacting protein-binding protein 2)
[Monodelphis domestica]
FAQWNTEQVCSWLEDFGLGQYVIFARQWVSSGHTLLTATPQDMEKELGIKHPLHRKKLVLAVKS
INTK
552-619_UPI0000F2D7BF    0.493
>smart|SAM-uniprot|UPI0000EB1088|UPI0000EB1088/581-648 Liprin-
beta-2 (Protein tyrosine phosphatase receptor type f
polypeptide-interacting protein-binding protein 2) (PTPRF-
interacting protein-binding protein 2). [Canis lupus familiaris]
FAQWSTERVCAWLEDFGLGQYVIFARQWVTSGHTLLSATPQDMEKELGIKQPLHRKKLILAVKA
INTK
581-648_UPI0000EB1088    0.494
>smart|SAM-uniprot|UPI000155EA28|UPI000155EA28/554-621
PREDICTED: similar to Liprin-beta-2 (Protein tyrosine
phosphatase receptor type f polypeptide-interacting protein-
binding protein 2) (PTPRF-interacting protein-binding protein 2)
[Equus caballus]
FAQWSTERVCTWLEDFGLGQYVIFARQWVTSGHTLLTATPQDMEKELGIKHPLHRKKLVLAVKA
INTK
554-621_UPI000155EA28    0.497
>smart|SAM-uniprot|Q969H4|CNKR1_HUMAN/4-70 Connector enhancer of
kinase suppressor of ras 1 [Homo sapiens]
VETWTPGKVATWLRGLDDSLQDYPFEDWQLPGKNLLQLCPQSLEALAVRSLGHQELILGGVEQL
QAL
4-70_Q969H4    0.499
```

```
>smart|SAM-uniprot|UPI0000D68174|UPI0000D68174/459-528
PREDICTED: similar to sex comb on midleg-like 2 (Drosophila)
[Mus musculus]
PVNWSVNDVITFLNRLDPPLADQLYPTIKKHDIDGKAMLLISSDTMIKYMGVKVGVAMKFENYI
RTLKEK
459-528_UPI0000D68174    0.501
>smart|SAM-uniprot|Q68M96|Q68M96_RHIME/2-61 RtsD [Sinorhizobium
meliloti]
DIAAWLRSLGLEEYASAFRDNDIDAQLLLHLKAEDLKELGVASIGHRRKLIDAIADLRDE
2-61_Q68M96    0.507
>smart|SAM-uniprot|Q7JQG5|Q7JQG5_DROME/253-318 LD09801p
[Drosophila melanogaster]
RQRNDELIIDWLLEMKHEEYAQLFIAAGYDLPTIARMTPEDLTAIGIKNPHHRERIKQRIDKLQ
VL
253-318_Q7JQG5 0.520
>smart|SAM-uniprot|UPI0000E1E745|UPI0000E1E745/4-70 PREDICTED:
connector enhancer of kinase suppressor of Ras 1 [Pan
troglodytes]
VETWTPGKVATWLRGLDDSLQDYPFEDWQLPGKNLLQLCPRSLEALAVRSLGHQELILGGVEQL
QAL
4-70_UPI0000E1E745  0.524
>smart|SAM-uniprot|UPI0000F1D271|UPI0000F1D271/480-546
PREDICTED: similar to cask-interacting protein 1 [Danio rerio]
AESKSSEAVVEWLTSAQLQFYTTNFLTAGYDLQTISRMTPEDLTAIGVTKPGHRKKMLSEISKM
NIP
480-546_UPI0000F1D271    0.529
>smart|SAM-uniprot|UPI0000F1DA4B|UPI0000F1DA4B/465-531
PREDICTED: similar to cask-interacting protein 1 [Danio rerio]
PEGKSSEAVIEWLSEFQLQVYAPNFINAGYDIPTISRMTPEDLTAIGVTKPGHRKKITSEINKI
SVN
465-531_UPI0000F1DA4B    0.534
>smart|SAM-uniprot|Q6DD51|CSKI2_XENLA/465-531 Caskin-2 [Xenopus
laevis]
LQGKDAEQIFCWLRGFQMETYVGNFISAGYDLPTIMRVTPEDLTAIGVTKPGHRKMISTEIGKL
IVA
465-531_Q6DD51 0.535
>smart|SAM-uniprot|UPI00006A0F24|UPI00006A0F24/467-533 Caskin-2.
[Xenopus tropicalis]
LQGKDAEQIYCWLRSFQMESYVENFISAGYDLPTIMRVTPEDLTAIGVTKPGHRKKISTEIGKL
IVA
467-533_UPI00006A0F24    0.536
```

```
>smart|SAM-uniprot|UPI0000E2401C|UPI0000E2401C/610-676
PREDICTED: CASK interacting protein 1 [Pan troglodytes]
SEGKSSEAVSQWLTAFQLQLYAPNFISAGYDLPTISRMTPEDLTAIGVTKPGHRKKIAAEISGL
SIP
610-676_UPI0000E2401C    0.537
>smart|SAM-uniprot|UPI000069DF96|UPI000069DF96/445-511 RNA-
binding protein with serine-rich domain 1 (SR-related protein
LDC2). [Xenopus tropicalis]
LPCQSSEAVFQWLCKFQLQLYAPNFINSGYDIPTISRMTPEDLTAIGVTKPGHRKKIFSEINNL
NIG
445-511_UPI000069DF96    0.539
>smart|SAM-uniprot|Q4TAH3|Q4TAH3_TETNG/319-385 Chromosome
undetermined SCAF7326, whole genome shotgun sequence [Tetraodon
nigroviridis]
MKGEDSEAIYQWLSEFQLEQYSSNFIKAGYDLPTISRITPEDLTAIGVTKPGHRKKISMEIARL
SIP
319-385_Q4TAH3 0.540
>smart|SAM-uniprot|UPI0000660E31|UPI0000660E31/1-63 Homolog of
Homo sapiens &quot;Caskin-2 [Takifugu rubripes]
DADAIGQWLSDFQLEQYTGNFVSAGYDVPTISRMTPEDLTAIGVTKPGHRKKISMEINNLNIP
1-63_UPI0000660E31  0.544
>smart|SAM-uniprot|UPI0000D55AC9|UPI0000D55AC9/303-369
PREDICTED: similar to CG12424-PA, isoform A [Tribolium
castaneum]
IHQGDQEIIHAWLVDLQYDEYFPLFVSAGYDLPTIGRMTPEDLTAIGIKKPNHRKRLKAEIAQL
NLP
303-369_UPI0000D55AC9    0.545
>smart|SAM-uniprot|UPI0000E8130C|UPI0000E8130C/511-577
PREDICTED: similar to cask-interacting protein 2 [Gallus gallus]
LEGKDAEAIYNWLSEFQLESYTVNFLSAGYDVPTISRMTPEDLTAIGVTKPGHRKKISTEIGQL
SIA
511-577_UPI0000E8130C    0.546
>smart|SAM-uniprot|UPI000065E323|UPI000065E323/3-70 Homolog of
Homo sapiens &quot;PARX protein [Takifugu rubripes]
EPGESIQEVVEWLLTLRLSQYTSAFLGAGYRSLEDCGELTEERLLELKILPTGHRRRLLRSLEA
LGEK
3-70_UPI000065E323  0.549
>smart|SAM-uniprot|UPI0001554BD0|UPI0001554BD0/1182-1248
PREDICTED: similar to diacylglycerol kinase eta2
[Ornithorhynchus anatinus]
VQKWGTDEVAAWLDLLSLGEYKETFIRHDIRGAELLHLERRDLKDLGITKVGHMKRILQGIKEL
GKG
```

```
1182-1248_UPI0001554BD0  0.552
>smart|SAM-uniprot|UPI0000F2DF5D|UPI0000F2DF5D/1151-1217
PREDICTED: similar to diacylglycerol kinase eta2 isoform 1
[Monodelphis domestica]
VQKWGTDEVAAWLDQLNLGEYTEIFIRHDIRGAELLHLERRDLKDLGITKVGHMKRILQGIKEL
GKS
1151-1217_UPI0000F2DF5D  0.553
>smart|SAM-uniprot|A6QPL5|A6QPL5_BOVIN/1143-1209 DGKH protein
[Bos taurus]
VHLWGTEEVAAWLEHLSLCEYKDIFTRHDIRGSELLHLERRDLKDLGVTKVGHMKRILCGIKEL
SRG
1143-1209_A6QPL5    0.554
>smart|SAM-uniprot|UPI00006A04B0|UPI00006A04B0/1132-1198
Diacylglycerol kinase eta (EC 2.7.1.107) (Diglyceride kinase
eta) (DGK-eta) (DAG kinase eta). [Xenopus tropicalis]
VQKWGTDEVAVWLDLLSLGEYREIFIRHDIRGSELLHLERRDLKDLGILKVGHMKRILQGIKEL
SKN
1132-1198_UPI00006A04B0  0.559
>smart|SAM-uniprot|A6NFX7|A6NFX7_HUMAN/1150-1216 Uncharacterized
protein DGKH [Homo sapiens]
VQKWGTEEVAAWLDLLNLGEYKDIFIRHDIRGAELLHLERRDLKDLGIPKVGHVKRILQGIKEL
GRS
1150-1216_A6NFX7    0.560
>smart|SAM-uniprot|UPI00006A0B49|UPI00006A0B49/1105-1169
Diacylglycerol kinase kappa (EC 2.7.1.107) (Diglyceride kinase
kappa) (DGK-kappa) (DAG kinase kappa) (142 kDa diacylglycerol
kinase). [Xenopus tropicalis]
VDKWGTEEVAAWLDVLGLGEYKDIFIRHDIQGSELVLLERRDLKDLGIAKVGHMKRILQGIKDI
T
1105-1169_UPI00006A0B49  0.562
>smart|SAM-uniprot|UPI0000F2C3EA|UPI0000F2C3EA/836-902
PREDICTED: similar to diacylglycerol kinase, delta 130kDa,
partial [Monodelphis domestica]
AHLWGTEEVAAWLEHLSLCEYKEIFVRHDIRGPELLQLERRDLKDLGVTKVGHMKRILHGIRKL
SRS
836-902_UPI0000F2C3EA    0.563
>smart|SAM-uniprot|Q6A0B7|Q6A0B7_MOUSE/967-1033 MKIAA0145
protein [Mus musculus]
VHLWGTEEVAAWLEHLSLCEYKDIFTRHDIRGSELLHLERRDLKDLGVTKVGHMKRILCGIKEL
SRS
967-1033_Q6A0B7     0.564
```

```
>smart|SAM-uniprot|UPI00005A4151|UPI00005A4151/1110-1176
PREDICTED: similar to diacylglycerol kinase, eta isoform 2
[Canis lupus familiaris]
VQKWGTEEVAAWLDLLNLGEYKEIFIRHDIRGAELLHLERRDLKDLGITKVGHMKRILQGIKEF
GRS
1110-1176_UPI00005A4151  0.565
>smart|SAM-uniprot|UPI0000EBD509|UPI0000EBD509/1053-1119
PREDICTED: similar to diacylglycerol kinase eta2 [Bos taurus]
VQKWGTEEVAAWLDLLNLGEYKEIFIRHDIRGAELLHLERRDLKDLGIPKVGHVKRILQGIKEL
ERS
1053-1119_UPI0000EBD509  0.566
>smart|SAM-uniprot|A0JP53|A0JP53_MOUSE/1007-1073 Dgkh protein
[Mus musculus]
VQNWGTEEVAAWLDLLNLGEYKEIFIRHDVRGAELLHLERRDLKDLGIPKVGHMKRILQGIKEL
ERN
1007-1073_A0JP53    0.568
>smart|SAM-uniprot|UPI00006A0FDA|UPI00006A0FDA/4-72 Connector
enhancer of kinase suppressor of ras 1 (Connector enhancer of
KSR1) (hCNK1) (Connector enhancer of KSR-like) (CNK homolog
protein 1). [Xenopus tropicalis]
VSSWSPETVRNYLRALGLDPAVQVYTFQEWNVTGKDLLCLSSQQLDALGVRCIGHQEIILEAIE
QLCAL
4-72_UPI00006A0FDA  0.569
>smart|SAM-uniprot|Q17FK9|Q17FK9_AEDAE/909-975 Tankyrase [Aedes
aegypti]
QVVAIESSVSAFLTSLQLEHLIELLEREQITMDILAEMGHEDLKQVGVSAYGFRHKILKGIATL
RAT
909-975_Q17FK9 0.573
>smart|SAM-uniprot|UPI0000EBEBAB|UPI0000EBEBAB/107-173
PREDICTED: similar to AIDA-1a [Bos taurus]
HDGYHPTSVAEWLDSIELGDYTKAFLINGYTSMDLLKKIWEVELINVADSSRQEQGITATITII
RVL
107-173_UPI0000EBEBAB    0.579
>smart|SAM-uniprot|UPI0000E46580|UPI0000E46580/1074-1140
PREDICTED: similar to diacylglycerol kinase eta
[Strongylocentrotus purpuratus]
VKLWTVEEVGAWLEALLMGEYKDQFIRNDIRGSELLSLERRDLKDIGVTKVGHIKRIQQAIKDL
DKR
1074-1140_UPI0000E46580  0.590
>smart|SAM-uniprot|UPI0000D55FB2|UPI0000D55FB2/1410-1476
PREDICTED: similar to CG31187-PA [Tribolium castaneum]
```

```
VASWGTQEVATWLETLQLSEYIDSFVKNDIRGRELLTLARRDLKDLGVTKVGHVKRILQAIKDL
TQG
1410-1476_UPI0000D55FB2  0.592
>smart|SAM_PNT-uniprot|UPI000155B956|UPI000155B956/44-126
PREDICTED: similar to B-cell lineage specific activator/TEL
oncogene fusion protein, partial [Ornithorhynchus anatinus]
APSAHRPVAPLAGGLQPVSWSRDDVGQWLKWAEKEFSLRPIDNNTFEMNGKALLLLTKEDFRYR
CPHSGDVLYELLQHIL
44-126_UPI000155B956     0.593
>smart|SAM_PNT-uniprot|UPI000065E96A|UPI000065E96A/1-87 Homolog
of Homo sapiens &quot;Splice Isoform G of Transcription factor
ETV7 [Takifugu rubripes]
VGSPPTLEDLWRLPGRLRINPSLWDKEDVALWLHWAQKEYSLRRPEKGRFEMNGRALCLLTKED
FRHRCPSSGDVLFEIL
1-87_UPI000065E96A  0.594
>smart|SAM_PNT-uniprot|Q5R3L3|Q5R3L3_HUMAN/31-117 Ets variant
gene 7 (TEL2 oncogene) (Ets variant gene 7 (TEL2 oncogene),
isoform CRA_f) [Homo sapiens]
QINLLGEGGICKLPGRLRIQPALWSREDVLHWLRWAEQEYSLPCTAEHGFEMNGRALCILTKDD
FRHRAPSSGDVLYELL
31-117_Q5R3L3  0.595
>smart|SAM_PNT-uniprot|UPI0000E816B8|UPI0000E816B8/20-103
PREDICTED: similar to Ets transcription factor TEL-2b [Gallus
gallus]
PRGRHTPADEGHSLALPGRLRIQPSLWSKDDVIHWLRWAEREYSLQQTDESKFEMNGKALCILT
KDDFRFRAPGSGMDGG
20-103_UPI0000E816B8     0.596
>smart|SAM-uniprot|UPI0000E4781C|UPI0000E4781C/1217-1284
PREDICTED: similar to mitogen-activated kinase kinase kinase 5
[Strongylocentrotus purpuratus]
GVEAKQEELSSWLQAVGLDEDSIRRFIEEDLTLPLVLNSMTRDDLRSLKIRVGAQVRIWDAVQQ
HRMN
1217-1284_UPI0000E4781C  0.598
>smart|SAM_PNT-uniprot|P97360|ETV6_MOUSE/39-125 Transcription
factor ETV6 [Mus musculus]
PRALRMEEDSIHLPTHLRLQPIYWSRDDVAQWLKWAENEFSLRPIESNKFEMNGKALLLLTKED
FRYRSPHSGDVLYELL
39-125_P97360  0.599
>smart|SAM_PNT-uniprot|Q0VC65|ETV6_BOVIN/38-124 Transcription
factor ETV6 [Bos taurus]
PRALRMEEDSIRLPAHLRLQPMFWSRDDVAQWLKWAENEFSLRPIDSNTFEMNGKALLLLTKED
FRYRSPHSGDVLYELL
```

```
38-124_Q0VC65  0.600
>smart|SAM_PNT-uniprot|UPI000155BF36|UPI000155BF36/75-161
PREDICTED: similar to Ets transcription factor TEL-2b
[Ornithorhynchus anatinus]
SFPLITEEEICRLPGQLRIQPSLWSKEDVIHWLRWAEKEYSLRRIVENKFEMNGKALCILTKDD
FRYRAPSSGDVLYELL
75-161_UPI000155BF36     0.601
>smart|SAM_PNT-uniprot|Q01842|POK_DROME/31-117 Ets DNA-binding
protein pokkuri [Drosophila melanogaster]
APSSQLAELKTQLPPSLPSDPRLWSREDVLVFLRFCVREFDLPKLDFDLFQMNGKALCLLTRAD
FGHRCPGAGDVLHNVL
31-117_Q01842  0.602
>smart|SAM_PNT-uniprot|O96416|O96416_DROVI/31-117 ETS DNA
binding protein Yan [Drosophila virilis]
PPSSQLAELKTQLPPSLPSDPRLWSRDDVLDFLRFCVREFDLPKLDFDLFQMNGKALCLLTRAD
FGHRCPGAGDVLHNVL
31-117_O96416  0.608
>smart|SAM-uniprot|Q294L8|Q294L8_DROPS/1733-1799 GA16074-PA
[Drosophila pseudoobscura]
ANNWSVNEVVTWLETMQLSEYVDSFLKNDIRGKELLTLGRRDLKDLGVVKVGHVKRILQAIKDL
SEN
1733-1799_Q294L8    0.609
>smart|SAM-uniprot|Q7PTK0|Q7PTK0_ANOGA/1658-1724
ENSANGP00000013418 [Anopheles gambiae]
VSNWGVGEVVTWLEAMQLAEYVDSFIKNDIRGKELLTLARRDLKDLGVTKVGHVKRILQAIKDL
GAG
1658-1724_Q7PTK0    0.612
>smart|SAM_PNT-uniprot|Q16SU2|Q16SU2_AEDAE/95-181 Ets [Aedes
aegypti]
APPSPLADLKTQLPPQLNTDPRIWGREEVAVFLRFCEREFDLPKFDLDLFQMNGKALCVLTKND
LAERSPGAGDVLHNVL
95-181_Q16SU2  0.613
>smart|SAM-uniprot|UPI0000DB7AC8|UPI0000DB7AC8/1293-1358
PREDICTED: similar to CG31187-PA [Apis mellifera]
VTTWGVQEVCCWLENLQLGEYTEKFISHDIRGRELLSLARRDLKELGIIKVGHVKRILQAINDL
NN
1293-1358_UPI0000DB7AC8  0.614
>smart|SAM-uniprot|Q17HC2|Q17HC2_AEDAE/1646-1712 Diacylglycerol
kinase, delta, kappa, eta [Aedes aegypti]
VNNWGVGEVVTWLEAMQMTEYVDSFIKNDIRGKELLTLARRDLKDLGVTKVGHVKRILQAIKDL
GST
1646-1712_Q17HC2    0.615
```

```
>smart|SAM_PNT-uniprot|UPI0000D55CB4|UPI0000D55CB4/47-133
PREDICTED: similar to Ets DNA-binding protein pokkuri (Protein
yan) (Protein anterior open) [Tribolium castaneum]
PPSSPLLDYKSQLPSNLASDPRVWSREDVATFLRWAEREFDLQPIDMDMFQMNGKAICLLTRTD
LAERAPGSGDVLYNVL
47-133_UPI0000D55CB4     0.617
>smart|SAM_PNT-uniprot|A4PIG5|A4PIG5_CIOIN/15-97 Ets
transcription factor Elf [Ciona intestinalis]
AKEDVAASDLAQLFATSPWEWSEDGVFAWMSHVVNQFDLDASNLKNLHINGKELQMFSQDEFEK
KVPYGNVLWAHLQFLS
15-97_A4PIG5   0.636
>smart|SAM_PNT-uniprot|Q66PI1|Q66PI1_9BIVA/69-154 ETS-family
transcription factor [Chlamys farreri]
PCPFDEKNMKSWAEKHPEHWNNNEVLDWIYFVAAENHLDVARLRGENFQTITGQKLCQMTQQDF
IDKDQEYGSYYFELFH
69-154_Q66PI1  0.647
>smart|SAM_PNT-uniprot|Q4H3K1|Q4H3K1_CIOIN/408-495 Transcription
factor protein [Ciona intestinalis]
PPNVTLNPDKHKRSTLPQDPSTWTRNHVKEWLVNSSKEYNLQELDLKKFASTDGYKLCQMTMRD
LCRITSKANAEVILNK
408-495_Q4H3K1 0.658
>smart|SAM-uniprot|Q6R5A4|Q6R5A4_DANRE/743-809 Bicaudal-C [Danio
rerio]
PSPAHADDLIELLAQLGLEKYIDIFRQQEIDYQTFLTLSDEDLKEVGVSTFGARRKMLLAIADL
SKK
743-809_Q6R5A4 0.677
>smart|SAM-uniprot|UPI000065F979|UPI000065F979/1-59 Homolog of
Brachydanio rerio &quot;Bicaudal-C. [Takifugu rubripes]
LPELLSQLGLIKYIDIFEQQEIDYPTFLTLSDEDLKEVGVFTFGARRKMLLAIADLSKS
1-59_UPI000065F979  0.681
>smart|SAM-uniprot|Q758Y4|VTS1_ASHGO/415-479 Protein VTS1
[Eremothecium gossypii]
CDPKLLKNIPAWLKSLRLHKYSASLNGKSWQELIDLDDAILEDMGVSALGARRKLLKAFAIVKE
C
415-479_Q758Y4 0.682
>smart|SAM-uniprot|Q6C9X3|Q6C9X3_YARLI/154-221 Similar to
sp|P39969 Saccharomyces cerevisiae YER114c BOI2 budding protein
[Yarrowia lipolytica]
VMSWTPIQVARWFENRGFEPAIPKLFVEHKISGAILTELELQHLKELDITSFGTRFEVYKEIEK
LRGD
154-221_Q6C9X3 0.683
```

```
>smart|SAM-uniprot|UPI000155D868|UPI000155D868/810-875
PREDICTED: similar to Protein bicaudal C homolog 1 (Bic-C)
[Equus caballus]
SSFKGSDLPELFSKLGLGKYTDVFQQQEIDLQTFLTLTDQDLKELGITTFGARRKMLLAISELN
KN
810-875_UPI000155D868    0.684
>smart|SAM-uniprot|Q99MQ1|BICC1_MOUSE/872-938 Protein bicaudal C
homolog 1 [Mus musculus]
NSCFKGSDLPELFSKLGLGKYTDVFQQQEIDLQTFLTLTDQDLKELGITTFGARRKMLLAISEL
SKN
872-938_Q99MQ1 0.685
>smart|SAM-uniprot|UPI0000E80747|UPI0000E80747/1087-1153
PREDICTED: similar to bicaudal-C [Gallus gallus]
NSLFKGSDLPELFSKLGLGKYTDVFQQQEIDLQTFLTLTDQDLKELGITTFGARRKMLLAISEL
NKN
1087-1153_UPI0000E80747  0.687
>smart|SAM-uniprot|Q08831|VTS1_YEAST/448-512 Protein VTS1
[Saccharomyces cerevisiae]
TDPKLLKNIPMWLKSLRLHKYSDALSGTPWIELIYLDDETLEKKGVLALGARRKLLKAFGIVID
Y
448-512_Q08831 0.688
>smart|SAM-uniprot|Q6CY29|VTS1_KLULA/384-448 Protein VTS1
[Kluyveromyces lactis]
CDPKLLKNVPAWLKSLRLHKYSEALGSKPWFELIYLDDEALENMGVSALGARRKLLKAFSIVRE
Y
384-448_Q6CY29 0.689
>smart|SAM-uniprot|UPI0000F2AE6A|UPI0000F2AE6A/869-935
PREDICTED: similar to serologically defined colon cancer antigen
3 [Monodelphis domestica]
NSSLKGSDLPELFSKLGLGKYTDVFQQQEIDLQTFLTLTDQDLKELGITTFGARRKMLLAISEL
NKN
869-935_UPI0000F2AE6A    0.693
>smart|SAM-uniprot|Q6BSL1|VTS1_DEBHA/512-576 Protein VTS1
[Debaryomyces hansenii]
SNPDLLNNIPAWLKLLRLHKYTDCLKDIYWKDLIEYDDLLLEKRGVKALGARRKLLKAFDAVKE
S
512-576_Q6BSL1 0.695
>smart|SAM-uniprot|A3LX24|A3LX24_PICST/566-630 Predicted protein
[Pichia stipitis]
ANIELLNNIPAWLKLLRLHKYTDCLKDTPWKELIELDNEQLELKGVAALGARRKLLKAFDAVKL
T
566-630_A3LX24 0.696
```

```
>smart|SAM-uniprot|Q6P0K1|Q6P0K1_DANRE/4-70 Connector enhancer
of kinase suppressor of Ras 1 [Danio rerio]
VASWSTERVTEWLKGLDAPLQQYPFSEWQLSGSDLLQLSSSRLENIGVHKIGHQELILEAVEKL
CAL
4-70_Q6P0K1    0.699
>smart|SAM-uniprot|UPI000065EE82|UPI000065EE82/1126-1192 Homolog
of Brachydanio rerio &quot;Bicaudal-C. [Takifugu rubripes]
NSSLKGSDLPELFSNLGLGKYTDVFQQQEIDLQTFLTLTDQDLKELGITTFGARRKMLLAISVV
PAA
1126-1192_UPI000065EE82  0.700
>smart|SAM-uniprot|Q54XX5|Q54XX5_DICDI/376-441 Ankyrin repeat-
containing protein [Dictyostelium discoideum]
FLKNVQDIFNWLKSIDLEQYWLNFVKEEIFMDLLLDIDERTLDSLGITYSGHRLKIIRNCRILR
DQ
376-441_Q54XX5 0.702
>smart|SAM-uniprot|A0D9Q7|A0D9Q7_PARTE/264-331 Chromosome
undetermined scaffold_42, whole genome shotgun sequence
[Paramecium tetraurelia]
MQEWQLEDVCNWLDSLHLSEYKDEFIKNQMTGKTLYALTDNDLKQDLGISVLGHRKQILQSIEE
YKKY
264-331_A0D9Q7 0.704
>smart|SAM-uniprot|UPI0000E45E6F|UPI0000E45E6F/1894-1960
PREDICTED: similar to KIAA1617 protein [Strongylocentrotus
purpuratus]
PLHWSVDRVVKFIKKTDCASLAKVFKDQEIDGQALLLLTLPTVQECMELKLGPAIKLCHHIERV
KIA
1894-1960_UPI0000E45E6F  0.714
>smart|SAM-uniprot|Q0CUP3|Q0CUP3_ASPTN/540-604 Protein VTS1
[Aspergillus terreus]
TDPNLLKDIPSWLRSLRLHKYTENLKDLKWTELVELNDKALEERGVNALGARNKMLKVFEQVRE
A
540-604_Q0CUP3 0.724
>smart|SAM-uniprot|Q2UKF0|Q2UKF0_ASPOR/542-606 Predicted protein
[Aspergillus oryzae]
TDPALLKDIPSWLRSLRLHKYTENLKDLKWTELVELDDKALEERGVNALGARNKMLKVFEQVRE
A
542-606_Q2UKF0 0.728
>smart|SAM-uniprot|A1D4C6|A1D4C6_NEOFI/538-602 SAM domain
protein [Neosartorya fischeri]
TDPNLLKDIPSWLRSLRLHKYTDNLKDLKWTELIELDDKALEDRGVNALGARNKMLKVFEQVKE
A
538-602_A1D4C6 0.729
```

```
>smart|SAM-uniprot|Q4IBN1|VTS1_GIBZE/546-610 Protein VTS1
[Gibberella zeae]
TDPTLLQDIPSWLRSLRLHKYTDNLKDMKWTDLIELDDKALEERGVNALGARRKMLKVFEQVKE
A
546-610_Q4IBN1 0.731
>smart|SAM_PNT-uniprot|UPI0000D99E09|UPI0000D99E09/48-132
PREDICTED: similar to E74-like factor 3 (ets domain
transcription factor, epithelial-specific ) [Macaca mulatta]
QMSLEGTEKTSWSGEQPQFWSKTQVLDWISYQVEKNKYDASAIDFSRCDMDGATLCNCALEELR
LVFGPLGDQLHAQLRD
48-132_UPI0000D99E09     0.739
>smart|SAM-uniprot|Q5BGC4|VTS1_EMENI/536-600 Protein vts1
[Emericella nidulans]
TDPNLLKDIPSWLRSLRLHKYTDNLKDLKWTELIELNDKQLEERGVNALGARNKMLKVFEQVKE
A
536-600_Q5BGC4 0.742
>smart|SAM-uniprot|Q6CME5|Q6CME5_KLULA/245-312 Similar to
sp|P39969 Saccharomyces cerevisiae YER114c BOI2 budding protein
[Kluyveromyces lactis]
VKNWVPEQVTAYLISLGFDVESASRFQKHKISGTILLELELAHLKELEINSFGTRFEIFKEIEA
LKEI
245-312_Q6CME5 0.832
>smart|SAM-uniprot|Q751J9|Q751J9_ASHGO/223-290 AGL293Cp
[Eremothecium gossypii]
AELWTPEQVTAYLLSTGFDEESSNKFQEHKISGAILLELELAHLKELEINSFGTRFEIFKEIEA
IKET
223-290_Q751J9 0.837
>smart|SAM-uniprot|Q6BKM8|Q6BKM8_DEBHA/264-332 Similar to
CA5428|CaBOI2 Candida albicans CaBOI2 budding protein
[Debaryomyces hansenii]
ADTWTPKQVSSYFAIILGFDMDVAGKFARHKITGAILFELDLTYLKELDIDSFGTRFEVYKEIE
KLKQL
264-332_Q6BKM8 0.839
>smart|SAM-uniprot|A3M0F7|A3M0F7_PICST/186-254 RHO protein
signal transduction [Pichia stipitis]
ASSWTPQQVSSYFAFVLGFDMDVAGKFSRHKITGPILFELDLGHLKELDIDSFGTRFEVYKEIE
KLKEL
186-254_A3M0F7 0.840
>smart|SAM-uniprot|Q6FTC8|Q6FTC8_CANGA/232-299 Candida glabrata
strain CBS138 chromosome G complete sequence [Candida glabrata]
VENWTPDEVTAYFIGHGFDIKSASSFQQHKISGKILLELQLEHLKELDISSFGTRFEMYKQIEM
IREV
```

```
232-299_Q6FTC8 0.841
>smart|SAM-uniprot|Q4SVU7|Q4SVU7_TETNG/3-70 Chromosome
undetermined SCAF13721, whole genome shotgun sequence [Tetraodon
nigroviridis]
EPPECVREVVGWLHTLHLSQYTPSFLGAGYRTLDDCRELTEGRLLELDVLPRGHRRRMLRSLEA
LGVT
3-70_Q4SVU7    0.842
>smart|SAM-uniprot|Q7R022|Q7R022_GIALA/14-80 GLP_456_60559_61104
[Giardia intestinalis]
VEFWSKHDVSVWLSSIDMKKYNQQLRKHDVCGRSLMLLTEDDLKQMGIDSLGDRKALLYEIKRL
DRW
14-80_Q7R022   0.869
>smart|SAM-uniprot|A0BV34|A0BV34_PARTE/226-293 Chromosome
undetermined scaffold_13, whole genome shotgun sequence
[Paramecium tetraurelia]
MQDWNTNDVCVWLECLGLSQYKENFQKNQMVGEILHNLTDKELKEELGIQILGHRKQILQQINM
HKKY
226-293_A0BV34 0.903
>smart|SAM-uniprot|Q299K5|Q299K5_DROPS/2-66 GA12577-PA
[Drosophila pseudoobscura]
PNHNIVCEWLRTIGLAQYGESFLENGYDELEICKQIGEIDLDAIGVDNLSHRGKLLKSVRMLRE
K
2-66_Q299K5    0.915
>smart|SAM-uniprot|UPI0000DB7366|UPI0000DB7366/1-67 PREDICTED:
similar to CG31163-PB, isoform B [Apis mellifera]
MTMASNIVVEWLRSLHLGQYSESFIDNGYDDLEICKQIGDPDLDAIGVFNQTHRARLLQSVKTL
REE
1-67_UPI0000DB7366  0.926
>smart|SAM-uniprot|Q9VD37|Q9VD37_DROME/2-66 CG13859-PA
[Drosophila melanogaster]
PHHNIVCEWLRTIGLAHYGESFLENGYDELEICKQIGEIDLDAIGVDNPSHRGKLLKSVRMLRE
K
2-66_Q9VD37    0.928
>smart|SAM-uniprot|UPI0000F2C2C7|UPI0000F2C2C7/1-65 PREDICTED:
similar to SAM domain containing 1 [Monodelphis domestica]
MCTNIVYEWLKALQLSQYAESFVDNGYDDLEVCKQIGDPDLDAIGVLAPQHRRRIHDAVRRLRE
Q
1-65_UPI0000F2C2C7  0.936
>smart|SAM-uniprot|A5D8S3|A5D8S3_DANRE/1-68 LOC100002699 protein
[Danio rerio]
MMTNGPNIVYEWLKTLQLCQYVEAFVDNGYDDLEVCKQIGDPDLDAIGVFIPHHRQRIHDAVQR
LKDD
```

```
1-68_A5D8S3    0.944
>smart|SAM-uniprot|UPI000155472E|UPI000155472E/1-65 PREDICTED:
similar to N-acylaminoacyl-peptide hydrolase [Ornithorhynchus
anatinus]
MCAHVVYEWLKTLRLPQYAESFVDNGYDDLEVCKEIGDPDLDAIGVSAPRHRRRIQQAVRRLRL
D
1-65_UPI000155472E  0.945
>smart|SAM-uniprot|Q6FWY7|Q6FWY7_CANGA/201-268 Candida glabrata
strain CBS138 chromosome C complete sequence [Candida glabrata]
IIAWAPKDVAEYMLSRGFDFTTASKFQKHQITGIILLEMQTNMLKEIEISSFGIRFELEKEINH
LRSV
201-268_Q6FWY7 0.956
>smart|SAM-uniprot|UPI0000E21813|UPI0000E21813/911-977
PREDICTED: ephrin receptor EphA1 isoform 1 [Pan troglodytes]
PNFSAGCCVSEWLESIRMKRYILHFHSAGLDTMECVLELTAEDLTQMGITLPGHQKRILCSIQG
FKD
911-977_UPI0000E21813    1.030
>smart|SAM-uniprot|Q5LL22|Q5LL22_SILPO/1-66 Adenylate/guanylate
cyclase [Silicibacter pomeroyi]
MSRTIGSVTSWLADIGLEVYAERFEQADIDLEVLEHLSDQDLTDLGVTSLGHRRKILARIADMR
DA
1-66_Q5LL22    1.125
>smart|SAM-uniprot|UPI0000E4894C|UPI0000E4894C/1397-1459
PREDICTED: similar to conserved hypothetical protein
[Strongylocentrotus purpuratus]
FGEEPPLLSIFLKDLGYEKYDTNFLKAGISMIEMPYLSEEKLENIGIPIGPRLRILQEAQVMI
1397-1459_UPI0000E4894C  1.141
>smart|SAM-uniprot|A0CL90|A0CL90_PARTE/236-303 Chromosome
undetermined scaffold_20, whole genome shotgun sequence
[Paramecium tetraurelia]
MQDWNIEEVCIWLDCLGLSQYKENFIKNHMIGDTLHDLTDVELKEELGIEILGHRKLILQQINM
HKKY
236-303_A0CL90 1.171
>smart|SAM-uniprot|UPI0000ECAB95|UPI0000ECAB95/3-70 Centaurin-
delta 3 (Cnt-d3) (Arf-GAP, Rho-GAP, ankyrin repeat and
pleckstrin homology domain-containing protein 3). [Gallus
gallus]
SPCTPDSDIADWLATIHLERYRDLFKQHGFHLARDVAALDNKQLQQLGITATGHRKRILNLAEK
TRLL
3-70_UPI0000ECAB95  1.209
>smart|SAM_PNT-uniprot|Q4H3F7|Q4H3F7_CIOIN/129-211 GA repeat
binding protein alpha homolog [Ciona intestinalis]
```

```
PKVSQPEKVAELEIPGDPLLWNKTQVYQWMVWVAKEFNLDTSIIMDPDLDGIELNRMSQTEFVS
TFAYGNVLWSHHALLK
129-211_Q4H3F7 1.210
>smart|SAM-uniprot|Q96X32|Q96X32_USTMA/14-80 MAP kinase pathway-
interacting Ubc2 [Ustilago maydis]
VNKWSEQQVVDWLSSVGLSKYARDFKSNGITGDVLVLLDDEALRDIGVVTIGQRLALLAAIYRL
KQQ
14-80_Q96X32   1.242
>smart|SAM-uniprot|Q675P2|Q675P2_OIKDI/520-586 SH3 and multiple
ankyrin repeat domains 3-like protein [Oikopleura dioica]
AACWTTKEVVTWLEANNFEQYSSEFQENDISGENLLDLTKEELTEMGVTKIGHRKTLLTLIQRL
KEP
520-586_Q675P2 1.285
>smart|SAM-uniprot|UPI0000F21B52|UPI0000F21B52/1039-1105
PREDICTED: similar to SH3 and multiple ankyrin repeat domains 2
[Danio rerio]
IPLWSKHDVADWLDTLNLAEHKKAFLENDIEGSHLPNLQKEDLVDLGVTRVGHRMNIEKALKLL
MDR
1039-1105_UPI0000F21B52  1.292
>smart|SAM-uniprot|Q9JLU4|SHAN3_RAT/1749-1815 SH3 and multiple
ankyrin repeat domains protein 3 [Rattus norvegicus]
LQLWSKFDVGDWLESIHLGEHRDRFEDHEIEGAHLPALTKEDFVELGVTRVGHRMNIERALRQL
DGS
1749-1815_Q9JLU4    1.306
>smart|SAM-uniprot|UPI0000250F45|UPI0000250F45/2101-2167 SH3 and
multiple ankyrin repeat domains protein 1 (Shank1) (GKAP/SAPAP-
interacting protein) (SPANK-1) (Synamon) (Somatostatin receptor-
interacting protein) (SSTR-interacting protein) (SSTRIP).
[Rattus norvegicus]
LGFWTKFDVADWLEWLGLSEHRAQFLDHEIDGSHLPALTKEDYVDLGVTRVGHRMNIDRALKFF
LER
2101-2167_UPI0000250F45  1.307
>smart|SAM-uniprot|UPI00004D61B3|UPI00004D61B3/1152-1218
Synaptotagmin-3 (Synaptotagmin III) (SytIII). [Xenopus
tropicalis]
VHFWTKYDVADWLDYLNLGEHREHFLDNEIDGSHLPSLTKEDYIDLGVTRVGHRMNIERALKFF
IER
1152-1218_UPI00004D61B3  1.309
>smart|SAM-uniprot|Q5RGE7|Q5RGE7_DANRE/1535-1601 Novel protein
similar to vertebrate SH3 and multiple ankyrin repeat domains
family protein [Danio rerio]
```

```
LALWSKYDVGEWLESVGLGEHRARFLEHEIEGAHLPALTKDDLAELGVTRVGHRMNIERALKQL
LES
1535-1601_Q5RGE7    1.313
>smart|SAM-uniprot|UPI0000EBE749|UPI0000EBE749/1908-1974
PREDICTED: similar to Shank2E [Bos taurus]
VHLWTKPDVADWLESLNLGEHKEAFMDNEIDGSHLPNLQKEDLIDLGVTRVGHRMNIERALKQL
LDR
1908-1974_UPI0000EBE749  1.314
>smart|SAM-uniprot|A0E8G4|A0E8G4_PARTE/210-277 Chromosome
undetermined scaffold_82, whole genome shotgun sequence
[Paramecium tetraurelia]
MQEWNIDEVCTWLDYLGLSQYQEKFIKNHMIGEILHDLTDVELKDELGIDISAHRNLILSKQIY
IRNI
210-277_A0E8G4 1.315
>smart|SAM-uniprot|Q52KW0|SHAN2_XENLA/1226-1292 SH3 and multiple
ankyrin repeat domains protein 2 [Xenopus laevis]
IHMWTKQDVAEWLESLHLGEHREMFMDNEIDGTHLPNLQKEDLIDLGVTRVGHRMNIERALKQL
LDR
1226-1292_Q52KW0    1.316
>smart|SAM-uniprot|Q4TAT2|Q4TAT2_TETNG/1380-1446 Chromosome
undetermined SCAF7261, whole genome shotgun sequence [Tetraodon
nigroviridis]
LTLWNKYDVGDWLESVGLAEHRQRFQEHEIEGSHLPALTKEDYLELGVTRLGHRINIERALRGL
LDG
1380-1446_Q4TAT2    1.317
>smart|SAM-uniprot|UPI00005A206C|UPI00005A206C/1475-1541
PREDICTED: similar to SH3 and multiple ankyrin repeat domains 3
(Shank3) (Proline-rich synapse associated protein 2) (ProSAP2)
(SPANK-2) [Canis lupus familiaris]
LQFWSKFDVGDWLESIHLGEHRDRFEDHEIEGAHLPALTKDDLVELGVTRVGHRMNIERALRQL
DGS
1475-1541_UPI00005A206C  1.320
>smart|SAM-uniprot|UPI0000584890|UPI0000584890/972-1038
PREDICTED: similar to SH3 and multiple ankyrin repeat domains
protein 2 (Shank2) (Proline-rich synapse-associated protein 1)
(ProSAP1) (Cortactin-binding protein 1) (CortBP1) (GKAP/SAPAP-
interacting protein) (SPANK-3) [Strongylocentrotus purpuratus]
VSEWTVEDVSDWLEELNLGEYKESFTDNAISGEHLTSLGKEDLSELGVKRLGHRLTIIKALQKL
QNQ
972-1038_UPI0000584890   1.343
>smart|SAM-uniprot|Q876M8|Q876M8_ASPFU/64-130 Ste50p
[Aspergillus fumigatus]
```

```
ITEWTAEECANWLAALGLRQYCATFLENEIVGEALIALKHEELKEMGITSVGHRLTILKSVYET
KVK
64-130_Q876M8  1.414
>smart|SAM-uniprot|Q82Y00|Q82Y00_NITEU/3-63 Guanylate
cyclase:TPR repeat:SAM domain [Nitrosomonas europaea]
NDRVSAWLDSLGLTVYHESFEHNAITWDVLPELNEDDLEALGVLLGHRKILLRAIAQLSQN
3-63_Q82Y00    1.419
>smart|SAM_PNT-uniprot|O97145|O97145_STRPU/108-190 Ets4
transcription factor [Strongylocentrotus purpuratus]
IPDILEDCHKLNLIPNVYCWTPEDVQKWLLCVANRFELGELEMGHFYINGPTLATLQDVDFRHR
APKCGDILYSVVCLLK
108-190_O97145 1.423
>smart|SAM-uniprot|UPI000023E42A|UPI000023E42A/66-132
hypothetical protein FG04101.1 [Gibberella zeae]
ITEWTVEECADFIGTIGLPQYADRFMENEIVGEALVALQHDDLKSMGIASVGHRLTILKSVYDV
KKA
66-132_UPI000023E42A     1.426
>smart|SAM-uniprot|Q0MQZ4|Q0MQZ4_PENMA/64-126 STE50-like protein
[Penicillium marneffei]
ISDWTAEECADFVVGLGLRQYRTAFIVGEALIALKHEELKEMGITSAGHRLTILKSVYETKVK
64-126_Q0MQZ4  1.427
>smart|SAM-uniprot|UPI000051A5A7|UPI000051A5A7/349-415
PREDICTED: similar to CG10743-PA, isoform A [Apis mellifera]
SDKMDNTAVLRWLDDIGLPQHKEAFQNGKVDGRMLHRLTTEDLLNLGVTAQLHAASLRRGIQIL
REL
349-415_UPI000051A5A7    1.456
>smart|SAM-uniprot|P36622|STE4_SCHPO/8-73 Sexual differentiation
protein ste4 [Schizosaccharomyces pombe]
YWNWNNEAVCNWIEQLGFPHKEAFEDYHILGKDIDLLSSNDLRDMGIESVGHRIDILSAIQSMK
KQ
8-73_P36622    1.510
>smart|SAM-uniprot|UPI000155C33B|UPI000155C33B/257-323
PREDICTED: similar to PTPRF interacting protein, binding protein
1 (liprin beta 1) [Ornithorhynchus anatinus]
HGKLDFNWVTRWLDDIGLPQYKTQFDEARVDGRVLHYLTLDDLLSLKVVSVLHHLSIKRAIQVL
RIN
257-323_UPI000155C33B    1.533
>smart|SAM-uniprot|Q29HR7|Q29HR7_DROPS/358-423 GA15414-PA
[Drosophila pseudoobscura]
VSTWTVEQVVKYLARFYPDEAEAFKQQDVDGASLLLLTREDVINGFGFKLGPALRVFQIILGLQ
SH
358-423_Q29HR7 1.543
```

```
>smart|SAM-uniprot|UPI0000F2D7BF|UPI0000F2D7BF/624-690
PREDICTED: similar to Liprin-beta-2 (Protein tyrosine
phosphatase receptor type f polypeptide-interacting protein-
binding protein 2) (PTPRF-interacting protein-binding protein 2)
[Monodelphis domestica]
SAQLDHIWVTRWLDDIGLPQYKDQFHESRVDGRMLQYLTVNDLLFLKVTSQLHHLSIKCAIHVL
HVN
624-690_UPI0000F2D7BF    1.544
>smart|SAM-uniprot|UPI00006090D0|UPI00006090D0/710-776
PREDICTED: similar to mKIAA1230 protein isoform 1 [Mus musculus]
YGKLDFNWVTRWLDDIGLPQYKTQFDEGRVDGRMLHYMTVDDLLSLKVVSVLHHLSIKRAIQVL
RIN
710-776_UPI00006090D0    1.547
>smart|SAM-uniprot|UPI0000ECD2A9|UPI0000ECD2A9/624-690 PTPRF
interacting protein binding protein 1 [Gallus gallus]
HGKLDYHWVTRWLDDIGLPQYKTQFDEGKVDGRMLHYMTVDDLLSLKVVSVLHHLSIKRAIQVL
RIN
624-690_UPI0000ECD2A9    1.549
>smart|SAM-uniprot|UPI00006A1F7E|UPI00006A1F7E/677-743 Liprin-
beta-1 (Protein tyrosine phosphatase receptor type f
polypeptide-interacting protein-binding protein 1) (PTPRF-
interacting protein-binding protein 1) (hSGT2). [Xenopus
tropicalis]
YGKLDYRWVTRWLDDIGLPQYKTQFDDAKIDGRMLHYLAVDDLLSLKVVSVLHHLSIKRAIQVL
RIN
677-743_UPI00006A1F7E    1.550
>smart|SAM-uniprot|UPI000155E6CB|UPI000155E6CB/676-742
PREDICTED: similar to PTPRF interacting protein binding protein
1 isoform 3 [Equus caballus]
HGKLDFNWVTRWLDDIGLPQYKTQFDEGRVDGRMLHYMTVDDLLSLKVVSVLHHLSIKRAIQVL
RIN
676-742_UPI000155E6CB    1.552
>smart|SAM-uniprot|Q6DD79|Q6DD79_XENLA/677-743 LOC398525 protein
[Xenopus laevis]
YGRLDYRWVTRWLDDIGLPQYKTQFDDSKIDGRMLHYLAVEDLLSLKVVSVLHHLSIKRAIQVL
RIN
677-743_Q6DD79 1.553
>smart|SAM-uniprot|UPI0000DC21AA|UPI0000DC21AA/627-693 Liprin
beta 2 [Rattus norvegicus]
SALLDHVWVTRWLDDIGLPQYKDQFYESRVDGRMLQYLTVNDLLFLKVTSQLHHLSIKCAIHVL
HVN
627-693_UPI0000DC21AA    1.554
```

```
>smart|SAM-uniprot|UPI00001C1EF8|UPI00001C1EF8/627-693 Liprin-
beta-2 (Protein tyrosine phosphatase receptor type f
polypeptide-interacting protein-binding protein 2) (PTPRF-
interacting protein-binding protein 2). [Homo sapiens]
SALLDHIWVTRWLDDIGLPQYKDQFHESRVDRRMLQYLTVNDLLFLKVTSQLHHLSIKCAIHVL
HVN
627-693_UPI00001C1EF8    1.555
>smart|SAM-uniprot|Q4S5D4|Q4S5D4_TETNG/631-697 Chromosome 19
SCAF14731, whole genome shotgun sequence [Tetraodon
nigroviridis]
RGKLDHNWVTRWLDDIGLPQYKSHFEEARVDGRVLHHMTVEDLLSLKVGSVLHHLSIKRAIQVL
RIN
631-697_Q4S5D4 1.559
>smart|SAM-uniprot|UPI000155460B|UPI000155460B/653-719
PREDICTED: similar to coiled-coil like protein 1
[Ornithorhynchus anatinus]
SAQLDHIWVTRWLDDIGLPQYKDQFHESRVDGRVLQYLTVNDLLFLKVTSQLHHLSIKCAIHVL
HTH
653-719_UPI000155460B    1.566
>smart|SAM-uniprot|Q6DCK5|Q6DCK5_XENLA/670-736 Ppfibp2-prov
protein [Xenopus laevis]
YGRLDYRWVTRWLDDIGLPQYKTQFDDAKLDGRMLHYLAVEDLLSLKVVSVLHHLSIKRAIQLL
RIN
670-736_Q6DCK5 1.569
>smart|SAM-uniprot|UPI0000E4628B|UPI0000E4628B/78-144 PREDICTED:
similar to PTPRF interacting protein binding protein 1, partial
[Strongylocentrotus purpuratus]
QGDLDYHWVTRWLDDVGLPQYKDTFLEAKVDGRMLHYMTVDDLLLLKVTSAFHHVSIMRGIQCL
RLH
78-144_UPI0000E4628B     1.577
>smart|SAM-uniprot|Q86AT8|Q86AT8_DICDI/237-303 Similar to
Dictyostelium discoideum (Slime mold). Ankyrin repeat containing
protein [Dictyostelium discoideum]
ARIKKYKDLFDWLQKHGFEQYKDAFLKEEMFLDELGEMSEDILNKMGITSTGTRLRILKETSNL
ANE
237-303_Q86AT8 1.616
>smart|SAM-uniprot|UPI0000E80BD7|UPI0000E80BD7/242-309
PREDICTED: hypothetical protein [Gallus gallus]
IRKWTVDDVYNFIMSLPGCSDYAQTFKDHAIDGETLPLLTEEHLLDTMGLKLGPALKIRSQVSR
RLGN
242-309_UPI0000E80BD7    1.663
```

```
>smart|SAM-uniprot|Q9VKL0|Q9VKL0_DROME/795-862 CG31868-PA
[Drosophila melanogaster]
VNSWSVDDVCGFVGGIDICAEYVQSFRDQSIDGTGLPLLTEDHLVNSLGMKLGPALKLRSILAK
KLGG
795-862_Q9VKL0 1.672
>smart|SAM-uniprot|UPI0000DA1F03|UPI0000DA1F03/316-383
PREDICTED: similar to sterile alpha motif domain containing 7
[Rattus norvegicus]
IHKWTVDDVYNFIRSLPGCSDYAQVFKDHAIDGETLPLLTEQHLRGTMGLKLGPALKIQSQVSQ
HVGN
316-383_UPI0000DA1F03    1.673
>smart|SAM-uniprot|Q4SVD1|Q4SVD1_TETNG/678-744 Chromosome
undetermined SCAF13765, whole genome shotgun sequence [Tetraodon
nigroviridis]
ASDGRLFSFAGWLDDIGLPQYKDQFNEGRVDGQMLQYLTVNDLLYLKVTSQLHHLSIKCAIHVL
HVN
678-744_Q4SVD1 1.680
>smart|SAM-uniprot|P78365|P78365_HUMAN/366-433 Polyhomeotic 2
homolog [Homo sapiens]
SHQVNVEDVYEFIRSLPGCQEIAEEFRAQEIDGQALLLLKEDHLMSVMNIKLGPALKIYARISM
LKDS
366-433_P78365 1.763
>smart|SAM-uniprot|UPI00005A45C5|UPI00005A45C5/760-827
PREDICTED: similar to l(3)mbt-like isoform II [Canis lupus
familiaris]
VAKWTIDEVFSFVQTLTGCEDQARLFKDEMIDGEAFLLLTQADIVKIMSVKLGPALKIYNAILM
FKNA
760-827_UPI00005A45C5    1.769
>smart|SAM-uniprot|Q5RGR0|Q5RGR0_DANRE/664-731 Novel protein
similar to human and mouse l(3)mbt-like 3 [Danio rerio]
VASWSIDEVIQFIQGLPGCKEQVRTFRDEQIDGEAFLLLTQVDLVKILSIKLGPALKIYNSILM
FKTT
664-731_Q5RGR0 1.772
>smart|SAM-uniprot|UPI0000D9E7C6|UPI0000D9E7C6/540-607
PREDICTED: similar to l(3)mbt-like 3 isoform 2 [Macaca mulatta]
VARWTVDEVAEFVQSLLGCEEHAKCFKKEQIDGKAFLLLTQIDIVKVMKIKLGPALKIYNSILM
FRNS
540-607_UPI0000D9E7C6    1.773
>smart|SAM-uniprot|UPI0000E813CE|UPI0000E813CE/756-823
PREDICTED: similar to mKIAA0681 protein [Gallus gallus]
VAKWTIDEVFSFVQTLTGCEDQAKLFKDEMIDGEAFLLLTQADIVKIMSVKLGPALKIYNAILM
FKNA
```

```
756-823_UPI0000E813CE    1.774
>smart|SAM-uniprot|Q32NI4|Q32NI4_XENLA/598-665 MGC131153 protein
[Xenopus laevis]
VGKWSADEVAQFVHQLPGCKDQSSVFKDEQIDGEAFLLLTQSDLVKILGLKLGPALKIYNSILM
FHTL
598-665_Q32NI4 1.775
>smart|SAM-uniprot|UPI0000E24C72|UPI0000E24C72/540-607
PREDICTED: hypothetical protein isoform 5 [Pan troglodytes]
VARWTVDEVAEFVQSLLGCEEHAKCFKKEQIDGKAFLLLTQTDIVKVMKIKLGPALKIYNSILM
FRHS
540-607_UPI0000E24C72    1.776
>smart|SAM-uniprot|UPI00006D58F6|UPI00006D58F6/324-391
PREDICTED: similar to sterile alpha motif domain containing 7
isoform 2 [Macaca mulatta]
IQKWTVDDVHSFISSLPGCSDYAQVFKDHAIDGETLPLLTEEHLRGTMGLKLGPALKIQSQVSQ
HVGS
324-391_UPI00006D58F6    1.777
>smart|SAM-uniprot|UPI00001CF8F1|UPI00001CF8F1/404-471
PREDICTED: similar to sterile alpha motif domain containing 11
[Rattus norvegicus]
INKWTVDDVCNFVGGLSGCGEYARVFREQGIDGETLPLLTEEHLLNTMGLKLGPALKIRAQVAK
RLGR
404-471_UPI00001CF8F1    1.778
>smart|SAM-uniprot|UPI0000E2569F|UPI0000E2569F/768-835
PREDICTED: similar to mKIAA0681 protein isoform 3 [Pan
troglodytes]
VAKWTIDEVFGFVQTLTGCEDQARLFKDEMIDGEAFLLLTQADIVKIMSVKLGPALKIYNAILM
FKNA
768-835_UPI0000E2569F    1.779
>smart|SAM-uniprot|UPI0000E1F217|UPI0000E1F217/531-598
PREDICTED: sterile alpha motif domain containing 11, partial
[Pan troglodytes]
VTKWTVDDVCSFVGGLFGCGEYTRVFREQGINGETLPLLTEEHLLSTMGLKLGPALKIRAQVAR
RLGR
531-598_UPI0000E1F217    1.780
>smart|SAM-uniprot|UPI000065D5A4|UPI000065D5A4/347-414 Homolog
of Homo sapiens &quot;SAMD11 protein [Takifugu rubripes]
ISKWSVEDVCGFISSLAGCAEYTQVFREQAIDGETLPLLTEEHLLSTMGLKLGPALKIRSQVAR
RVGR
347-414_UPI000065D5A4    1.781
>smart|SAM-uniprot|UPI0000F1D97F|UPI0000F1D97F/389-456
PREDICTED: hypothetical protein [Danio rerio]
```

```
IRKWTVDDVYSFISEIPSCAEYAQTFKEHMIDGETLPLLTEDHLLDTLGLKLGPALKIRSQLSR
RLGN
389-456_UPI0000F1D97F    1.782
>smart|SAM-uniprot|UPI000065E8F8|UPI000065E8F8/241-308 Homolog
of Brachydanio rerio &quot;Polyhomeotic Ph2beta homolog (Phc2
protein). [Takifugu rubripes]
IRKWTVNDVYNFINSIPACSEYAQTFKDHMIDGETLPLLSEEHLLDTLGLKLGPALKIRSQVES
RFIH
241-308_UPI000065E8F8    1.783
>smart|SAM-uniprot|Q4S7Z0|Q4S7Z0_TETNG/1222-1289 Chromosome 9
SCAF14710, whole genome shotgun sequence [Tetraodon
nigroviridis]
VVAWSVDEVFRFVQNLIGCEEQARVFKEEMIDGEAFLLLTQTDIVKIMSIKLGPALKISNAILM
FKST
1222-1289_Q4S7Z0    1.787
>smart|SAM-uniprot|Q4SK78|Q4SK78_TETNG/238-305 Chromosome 15
SCAF14568, whole genome shotgun sequence [Tetraodon
nigroviridis]
VRKWTVNDVYSFISSIPTCSEYAQTFKDHMIDGETLPLLSEEHLLDTLGLKLGPALKIRSQVSQ
QIQT
238-305_Q4SK78 1.792
>smart|SAM-uniprot|UPI00004D9966|UPI00004D9966/801-865
Lethal(3)malignant brain tumor-like 3 protein (L(3)mbt-like 3
protein) (H-l(3)mbt-like protein). [Xenopus tropicalis]
VGKWSADEVAQFVHQLPGCTEQSSVFRDEQIDGEAFLLLTQNDLVKILGLKLGPALKIYNSILM
F
801-865_UPI00004D9966    1.793
>smart|SAM-uniprot|A3EXT2|A3EXT2_9HEMI/161-228 Polyhomeotic-like
protein 2-like protein [Maconellicoccus hirsutus]
PWKWTVTDVCEFIKKLPEASDYVEEFSTHEIDGQALMLLKENHLISVMNMKLGLALKIVNKINA
LREP
161-228_A3EXT2 1.794
>smart|SAM-uniprot|UPI0000F31BF7|UPI0000F31BF7/525-592 Sterile
alpha motif domain-containing protein 11. [Bos taurus]
VSKWTVDDVCSFVGGLSGCGEYAPVFREQGIDGETLPLLTEEHLLTTMGLKLGPALKIRAQVAK
RLGR
525-592_UPI0000F31BF7    1.796
>smart|SAM-uniprot|UPI0000F2E1F2|UPI0000F2E1F2/325-392
PREDICTED: hypothetical protein [Monodelphis domestica]
IRKWAVDDVFDFITGLPGCSDYAQIFRDHAIDGETLPLLTEEHLLDTMGLKLGPVLKIRAQTPT
MMII
325-392_UPI0000F2E1F2    1.797
```

```
>smart|SAM-uniprot|UPI0000E45C59|UPI0000E45C59/790-857
PREDICTED: similar to polyhomeotic Ph2alpha homolog
[Strongylocentrotus purpuratus]
PSSWSVENVAMFIRSLPGCAGYADEFQSQEIDGQALMLLKEDHLMTALSMKLGPALKIINKINT
LKDK
790-857_UPI0000E45C59    1.798
>smart|SAM-uniprot|Q4RVQ7|Q4RVQ7_TETNG/418-485 Chromosome 9
SCAF14991, whole genome shotgun sequence [Tetraodon
nigroviridis]
ISKWSVEDVCGFISSLAGCAEYTQVFREQAIDGETLPLLSEKHLLNTMGLKLGPALKIRSQVAR
RVGR
418-485_Q4RVQ7 1.800
>smart|SAM-uniprot|Q4TBZ6|Q4TBZ6_TETNG/213-279 Chromosome
undetermined SCAF7068, whole genome shotgun sequence [Tetraodon
nigroviridis]
PSQWSVEEVTAFITTLPGCSDVAEAFRLQEIDGQALLLLTEDHLMTSMNIKLGPALKICAHINA
LKH
213-279_Q4TBZ6 1.809
>smart|SAM-uniprot|Q16MU8|Q16MU8_AEDAE/1474-1541
Lethal(3)malignant brain tumor [Aedes aegypti]
PLCWSVEDVASYVERFPGCSMVGNQIREEQVNGMAFLSMTQEDLIKYLDVKLGPAIKLYNRIIH
LRLQ
1474-1541_Q16MU8    1.812
>smart|SAM-uniprot|UPI000065FB40|UPI000065FB40/212-279 Homolog
of Brachydanio rerio &quot;Polyhomeotic Ph2beta homolog (Phc2
protein). [Takifugu rubripes]
PGQWNIEDVYEFISALPGCLEIAEEFRSQEIDGQALLLLKEDHLMATMNIKLGPALKIFAQINS
LKDS
212-279_UPI000065FB40    1.814
>smart|SAM-uniprot|UPI0000437DA3|UPI0000437DA3/760-827
Polyhomeotic-like protein 2. [Danio rerio]
PTKWNVEDVYEFICSLPGCHEIAEEFRSQEIDGQALMLLKEDHLMSTMNIKLGPALKIFARISM
LKDS
760-827_UPI0000437DA3    1.818
>smart|SAM-uniprot|UPI00006607F9|UPI00006607F9/636-702 Homolog
of Brachydanio rerio &quot;Polyhomeotic Ph1 homolog. [Takifugu
rubripes]
PAQWSVEEVCRFISSLQGCEELAAQFLSQEIDGQALLLLREEHLISTMNIKLGPALKICASINT
LRE
636-702_UPI00006607F9    1.819
```

```
>smart|SAM-uniprot|Q4SY21|Q4SY21_TETNG/235-301 Chromosome
undetermined SCAF12233, whole genome shotgun sequence [Tetraodon
nigroviridis]
PTEWNTEEVSQFIASLQGCKELASQFLSQEIDGQALLLLKEEHLMSTMNIKLGPALKICAHINS
LRT
235-301_Q4SY21 1.820
>smart|SAM-uniprot|UPI000051A220|UPI000051A220/1150-1217
PREDICTED: similar to l(3)mbt-like 3 isoform b [Apis mellifera]
PRRWSNEEVIKFIQSVPNCKEIGNIFRQHNIDGEAFLMLTQEDLVSLLGLRLGPAIKLYNSIVL
LRRR
1150-1217_UPI000051A220  1.821
>smart|SAM-uniprot|A4QNV8|A4QNV8_DANRE/786-853 Zgc:56685 protein
[Danio rerio]
PTKWNVQEVFEFIRSLPGCQEIADEFRAQEIDGQALLLLKEDHLMSAMNIKLGPALKIFARINM
LKDS
786-853_A4QNV8 1.822
>smart|SAM-uniprot|UPI00005A5775|UPI00005A5775/325-392
PREDICTED: similar to sterile alpha motif domain containing 7
[Canis lupus familiaris]
IQKWTVNDVHNFISGLPGCSDYAQVFKDHAIDGETLPLLTEEHLRSTLGLKLGPALKIQSQVSQ
HVES
325-392_UPI00005A5775    1.823
>smart|SAM-uniprot|UPI0000660A3A|UPI0000660A3A/212-278 Homolog
of Homo sapiens &quot;Homolog of polyhomeotic 3 [Takifugu
rubripes]
PSQWSVEEVTAFITTLPGCSDVAEAFRLQEIDGQALLLLTEDHLMTSMNIKLGPALKICAHINS
LKN
212-278_UPI0000660A3A    1.824
>smart|SAM-uniprot|UPI0000F204D5|UPI0000F204D5/492-559
PREDICTED: hypothetical protein [Danio rerio]
PVDWTVADVASYFTAAGFPEQAIAFRTQEIDGKSLLLMQRNDVLTGLSIRLGPALKIYERHVKV
LQKT
492-559_UPI0000F204D5    1.825
>smart|SAM-uniprot|UPI00005EB781|UPI00005EB781/914-981
PREDICTED: similar to polyhomeotic like 3 (Drosophila), isoform
1 [Monodelphis domestica]
PSIWTVDEVWAFIHSLPGCQDIADEFRAQEIDGQALLLLKEDHLMSAMNIKLGPALKICARINS
LKES
914-981_UPI00005EB781    1.826
>smart|SAM-uniprot|A0JMR9|A0JMR9_XENLA/776-843 MGC154340 protein
[Xenopus laevis]
```

```
PSHWSVEEVYEFISSLQGCQDLAEDFRSQEIDGQALLLLKEEHLMSALNIKLGPALKICAKINL
LKET
776-843_A0JMR9 1.827
>smart|SAM-uniprot|UPI000065DD1C|UPI000065DD1C/23-90 Homolog of
Homo sapiens &quot;Dnaj-like protein [Takifugu rubripes]
PANWGVADVVNYFKATGFEEQATAFQDQEIDGKSLLLMTRNDVLTGLSIKLGPALKIYEYHVKP
LQTQ
23-90_UPI000065DD1C 1.829
>smart|SAM-uniprot|Q4V7W5|PHC2_XENLA/277-344 Polyhomeotic-like
protein 2 [Xenopus laevis]
PTKWNVEDVYDFVRSLPGCQEISEEFRAQEIDGQALLLLKEDHLMSAMNIKLGPALKLYARISM
LKDS
277-344_Q4V7W5 1.830
>smart|SAM-uniprot|UPI0000E80BD6|UPI0000E80BD6/971-1038
PREDICTED: similar to early development regulator 3 [Gallus
gallus]
PSVWTVDEVWAFIHSLPGCQDIADEFRAQEIDGQALLLLKEDHLMSAMNIKLGPALKICARINS
LKES
971-1038_UPI0000E80BD6   1.831
>smart|SAM-uniprot|UPI00006A0796|UPI00006A0796/412-479
Polyhomeotic-like protein 1 (hPH1) (Early development regulatory
protein 1). [Xenopus tropicalis]
PSRWSVEEVYEFISSLQGCQDLAEDFRSQEIDGQALLLLKEEHLMSALNIKLGPALKICAKINL
LKET
412-479_UPI00006A0796    1.834
>smart|SAM-uniprot|UPI0000F1D982|UPI0000F1D982/822-889
PREDICTED: similar to polyhomeotic like 3 (Drosophila), [Danio
rerio]
PSQWSVEQVCSFISTLPGCHDVAAEFGCQEIDGQALLLLTEDHLMSAMNLKLGPALKICAHINT
LKQT
822-889_UPI0000F1D982    1.835
>smart|SAM-uniprot|Q4RIF1|Q4RIF1_TETNG/605-672 Chromosome 11
SCAF15043, whole genome shotgun sequence [Tetraodon
nigroviridis]
PTKWNVEEVYDFIRSLPGCQEIADEFRSQEIDGQALLLLKEDHLMSTMNIKLGPALKIFARINM
LKDS
605-672_Q4RIF1 1.838
>smart|SAM-uniprot|P39769|PHP_DROME/1510-1577 Polyhomeotic-
proximal chromatin protein [Drosophila melanogaster]
ISSWSVDDVSNFIRELPGCQDYVDDFIQQEIDGQALLLLKEKHLVNAMGMKLGPALKIVAKVES
IKEV
1510-1577_P39769    1.853
```

```
>smart|SAM-uniprot|UPI0000DBF086|UPI0000DBF086/240-309 sex comb
on midleg-like 2 (Drosophila) [Rattus norvegicus]
PSTWSVDEVIQFMKHTDPHISGPLADLFRQHEIDGKALLLLKSELIMKYMGLKLGPALKLCYYI
EKLKER
240-309_UPI0000DBF086    1.861
>smart|SAM-uniprot|UPI0000E25C7E|UPI0000E25C7E/255-324
PREDICTED: sex comb on midleg-like 1 [Pan troglodytes]
PSTWSVEAVVLFLKQTDPVALCPLVDLFRSHEVDGKALLLLTSDVLLKHFGVKLGTAVKLCYYI
DRLKQG
255-324_UPI0000E25C7E    1.863
>smart|SAM-uniprot|Q1JQ83|Q1JQ83_XENLA/630-699 MGC130971 protein
[Xenopus laevis]
PAMWSVDDVMRFVKEADPQSLAPHAELFRRHEIDGTALLLLKSDMIMKYMGLKLGPALKLCYHI
ERLKQG
630-699_Q1JQ83 1.865
>smart|SAM-uniprot|Q4R8H0|Q4R8H0_MACFA/590-658 Testis cDNA
clone: QtsA-12509, similar to human sex comb on midleg homolog 1
(Drosophila) (SCMH1), [Macaca fascicularis]
PSSWTVEDVMQFVQEADPQLGPHADLFRKHEIDGKALLLLRSGMMMKYMGLKLGPALKLSYHID
RLKQG
590-658_Q4R8H0 1.867
>smart|SAM-uniprot|UPI0000F32863|UPI0000F32863/223-292 Sex comb
on midleg-like protein 1. [Bos taurus]
PSIWSVDEVTMFLQDVDPQTFDPLVDLFRDHGIDGKALLLLRSDMMIKYMGLKVGTALKICHYI
DRLKRE
223-292_UPI0000F32863    1.868
>smart|SAM-uniprot|UPI0000DA230B|UPI0000DA230B/803-870
PREDICTED: similar to Lethal(3)malignant brain tumor-like
protein (L(3)mbt-like) (L(3)mbt protein homolog) (H-l(3)mbt
protein) (H-L(3)MBT) (L3MBTL1) [Rattus norvegicus]
VSKWTIEEVFGFVQTLTGSEDQARLFKEEMIDGEAFLLLTQADIVKIMSVKLGPALKIYNAILM
FKNN
803-870_UPI0000DA230B    1.869
>smart|SAM-uniprot|Q4SV88|Q4SV88_TETNG/618-687 Chromosome 1
SCAF13775, whole genome shotgun sequence [Tetraodon
nigroviridis]
PSSWSIEEVMQFVRDADPTALAPHAELFRKHEIDGKALMLLRSDMIMKYMGLKLGPALKLCHHI
ERLKQG
618-687_Q4SV88 1.871
>smart|SAM-uniprot|UPI0000D61EC3|UPI0000D61EC3/363-430
Polyhomeotic-like protein 2 (hPH2) (Early development regulatory
protein 2). [Homo sapiens]
```

```
VSKWNVEDVYEFIRSLPGCQEIAEEFRAQEIDGQALLLLKEDHLMSAMNIKLGPALKIYARISM
LKDS
363-430_UPI0000D61EC3    1.872
>smart|SAM-uniprot|UPI00005A15E8|UPI00005A15E8/702-769
PREDICTED: similar to l(3)mbt-like 3 [Canis lupus familiaris]
VARWTVEEVAEFVQSLLGCEEHAKCFKKEQIDGKAFLLLTQTDIVKVMKIKLGPALKIYNSILM
FRNS
702-769_UPI00005A15E8    1.873
>smart|SAM-uniprot|UPI00005A2CEB|UPI00005A2CEB/553-621
PREDICTED: similar to Polycomb protein SCMH1 (Sex comb on midleg
homolog 1) isoform 1 [Canis lupus familiaris]
PSSWTVEDVMQFVREADPQLGPHADLFRKHEIDGKALLLLRSDMMMKYMGLKLGPALKLSFHID
RLKQG
553-621_UPI00005A2CEB    1.874
>smart|SAM-uniprot|UPI0000F2C803|UPI0000F2C803/474-541
PREDICTED: similar to l(3)mbt-like 4 (Drosophila), [Monodelphis
domestica]
VAKWTVDEVANFVQSLPGCEDHSRYFKKEQIDGKAFLLLTQTDIVKVMRIKLGPALKIYNSILM
FRNA
474-541_UPI0000F2C803    1.876
>smart|SAM-uniprot|Q9UQR0|SCML2_HUMAN/628-697 Sex comb on
midleg-like protein 2 [Homo sapiens]
PSTWSVDEVIQFMKHTDPQISGPLADLFRQHEIDGKALFLLKSDVMMKYMGLKLGPALKLCYYI
EKLKEG
628-697_Q9UQR0 1.877
>smart|SAM-uniprot|UPI0000F2147C|UPI0000F2147C/553-620
PREDICTED: hypothetical protein [Danio rerio]
VTHWSVEEVADFIHSLPGCEEQAKQFREEQIDGKAFLLLTQRDIVKIMSVKLGPALKIYNSILM
FKHA
553-620_UPI0000F2147C    1.878
>smart|SAM-uniprot|UPI0000ECA191|UPI0000ECA191/622-690 Polycomb
protein SCMH1 (Sex comb on midleg homolog 1). [Gallus gallus]
PSCWTVEEVMQFIREADPQLGPHADLFRKHEIDGKALLLLRSDMMMKYMGLKLGPALKLTYHID
KLKQG
622-690_UPI0000ECA191    1.889
>smart|SAM-uniprot|Q96GD3|SCMH1_HUMAN/590-658 Polycomb protein
SCMH1 [Homo sapiens]
PSSWTVEDVMQFVREADPQLGPHADLFRKHEIDGKALLLLRSDMMMKYMGLKLGPALKLSYHID
RLKQG
590-658_Q96GD3 1.890
```

```
>smart|SAM-uniprot|UPI0000660F04|UPI0000660F04/161-228 Homolog
of Homo sapiens &quot;Splice Isoform 4 of Lethal [Takifugu
rubripes]
KPDASVLQVFRFVQNLIGCEEQARVFKDEMIDGEAFLLLTQTDIVKIMSIKLGPALKISNAILM
FKST
161-228_UPI0000660F04    1.908
>smart|SAM-uniprot|UPI0000D9C6B0|UPI0000D9C6B0/1-67 PREDICTED:
similar to l(3)mbt-like 3 [Macaca mulatta]
MSSLVSKVFGFVQTLTGCEDQARLFKDEMIDGEAFLLLTQADIVKIMSVKLGPALKIYNAILMF
KNA
1-67_UPI0000D9C6B0  1.929
>smart|SAM-uniprot|UPI0000E8015B|UPI0000E8015B/684-745
PREDICTED: similar to KIAA1798 protein isoform 2 [Gallus gallus]
VSKWSTDEVDEHQSLPLCVFLLKQIDGEAFLLMTQSDIVKIMSIKLGPALKIFNSILMFKAA
684-745_UPI0000E8015B    1.950
>smart|SAM-uniprot|UPI0000EB4740|UPI0000EB4740/815-883
Lethal(3)malignant brain tumor-like 3 protein (L(3)mbt-like 3
protein) (H-l(3)mbt-like protein). [Canis lupus familiaris]
AGAACMKRVDRFVISLEKCEAEGKCVFSFSQIDGEAFLLMTQTDIVKIMSIKLGPALKIFNSIL
MFKAA
815-883_UPI0000EB4740    1.952
>smart|SAM-uniprot|Q1RLH6|Q1RLH6_CIOIN/446-513 Zinc finger
protein [Ciona intestinalis]
VKTWSIEKVAEFINVLTGKTECGTFFIDHEIDGESLLLLTQEDILKILKIKLGPAVKIYNAILM
FKTV
446-513_Q1RLH6 1.965
>smart|SAM-uniprot|Q5KJP5|Q5KJP5_CRYNE/17-83 Protein kinase
regulator, putative [Filobasidiella neoformans]
VLQWDEDAVVSYLSSIGLGQYEAAIHEHGIAGDVLVALDHETLQDMGMSSVGHRLMLLRAVYEL
KME
17-83_Q5KJP5   2.022
>smart|SAM-uniprot|Q4H2U6|Q4H2U6_CIOIN/636-704 Sex comb on
midleg like protein [Ciona intestinalis]
PTKWSVNDVVSYIGEREPALQPHLHLFNKHEIDGAALLLLNNNNIVKFMSLKLGPTLKLCSLVE
GLKEK
636-704_Q4H2U6 2.322
>smart|SAM-uniprot|UPI0000E7FAC4|UPI0000E7FAC4/635-704
PREDICTED: hypothetical protein [Gallus gallus]
PSTWSIEEVMCFVKEADPQALAPHAELFRRHEIDGKALLLLRSDMIMKYMGLKLGPALKLCYHI
ERLRQG
635-704_UPI0000E7FAC4    2.338
```

```
>smart|SAM-uniprot|UPI0000EBC71D|UPI0000EBC71D/583-651
PREDICTED: similar to sex comb on midleg homolog 1 (Drosophila)
isoform 7 [Bos taurus]
PSSWTVEDVMQFVREADPQFGPHADLFRKHEIDGKALLLLRSDTMMKYMGLKLGPALKLCYHID
RLKQG
583-651_UPI0000EBC71D    2.341
>smart|SAM-uniprot|UPI0000F2C14C|UPI0000F2C14C/348-417
PREDICTED: similar to sex comb on midleg-like 4 (Drosophila),
isoform 2 [Monodelphis domestica]
PSTWTVEDVVWFVKDADPQALGPHVELFRKHEIDGNALLLLKSDMIMKYLGLKLGPALKLCYHI
DKLKQS
348-417_UPI0000F2C14C    2.343
>smart|SAM-uniprot|UPI0000F2032D|UPI0000F2032D/540-608
PREDICTED: similar to sex comb on midleg homolog 1 (Drosophila),
[Danio rerio]
PNLWTVEDVMQFIRDIDPQLGPHADLFRKHEIDGKALLLLRSDVMMKYMGLKLGPALKLSFHID
RLKEG
540-608_UPI0000F2032D    2.347
>smart|SAM-uniprot|Q0IHT6|Q0IHT6_XENTR/635-704 Sex comb on
midleg-like 2 [Xenopus tropicalis]
PATWSVDDVMRFVKEADPQALAPHAELFRRHEIDGTALLLLRSDMIMKYMGLKLGPALKLCYHI
ERLKQG
635-704_Q0IHT6 2.351
>smart|SAM-uniprot|UPI00004BBD7D|UPI00004BBD7D/343-412
PREDICTED: similar to sex comb on midleg-like 4 [Canis lupus
familiaris]
PSTWTVEDVVRFVKDADPQALGPHVELFRKHEIDGNALLLLKSDMIMKYLGLKLGPALKLCYHI
DKLKQA
343-412_UPI00004BBD7D    2.353
>smart|SAM-uniprot|UPI00006A0C5F|UPI00006A0C5F/605-673 Sex comb
on midleg homolog 1. [Xenopus tropicalis]
PSLWNTEDVMQFVRDADPQLGFHAELFRKHEIDGKALLLLRSDVMMKYMGLKLGPALKLSHHID
KLKQG
605-673_UPI00006A0C5F    2.355
>smart|SAM-uniprot|UPI000155C8CF|UPI000155C8CF/243-312
PREDICTED: similar to sex comb on midleg-like 4 (Drosophila)
[Ornithorhynchus anatinus]
PSAWTVDDVVCFVEGVDPRAFGPHAALFRKHEIDGSALLLLKSDMVMKYLGLKLGPALKLCYHI
DKLKQT
243-312_UPI000155C8CF    2.357
>smart|SAM-uniprot|UPI0000F209EB|UPI0000F209EB/338-407
PREDICTED: hypothetical protein [Danio rerio]
```

```
PSRWSVDEVVWFIKDADPQALGPHVELFRKHEIDGDALLLLKSDMIMKYLGLKLGPALKLCYHI
DKLKQN
338-407_UPI0000F209EB    2.358
>smart|SAM-uniprot|Q9VHA0|SCM_DROME/803-871 Polycomb protein Scm
[Drosophila melanogaster]
PIDWTIEEVIQYIESNDNSLAVHGDLFRKHEIDGKALLLLNSEMMMKYMGLKLGPALKICNLVN
KVNGR
803-871_Q9VHA0 2.361
>smart|SAM-uniprot|Q172L2|Q172L2_AEDAE/808-876
Lethal(3)malignant brain tumor [Aedes aegypti]
TSDWTIEDVIRFIAVQDPALAVHADLFRKHEIDGKALLLLNSDMMMKYMGLKLGPALKICNLVS
RAKGR
808-876_Q172L2 2.362
>smart|SAM-ensembl|ENSPTRP00000049763|ENSPTRP00000049763/101-170
sex comb on midleg-like 4 [Source:RefSeq_peptide;Acc:NP_932347]
[Pan troglodytes]
PSAWTVEDVVWFVKDADPQALGPHVELFRKHEIDGNALLLLKSDMVMKYLGLKLGPALKLCYHI
DKLKQA
101-170_ENSPTRP00000049763    2.363
>smart|SAM-uniprot|Q4RRA4|Q4RRA4_TETNG/186-255 Chromosome 14
SCAF15003, whole genome shotgun sequence [Tetraodon
nigroviridis]
PSSWGVEEVVSFIKDADPQALGPHADTFRRHEIDGDALLLLKSEMMMKYLGLKLGPALKLSYHI
DNLKQS
186-255_Q4RRA4 2.365
>smart|SAM-uniprot|Q4T6D4|Q4T6D4_TETNG/11-78 Chromosome
undetermined SCAF8815, whole genome shotgun sequence [Tetraodon
nigroviridis]
PNLWTVDNVMQYIRDIDPLLAPHADLFRKHEIDGKALLLLRSDTMMKYMGLKLGPALKLTFHID
KLKR
11-78_Q4T6D4   2.366
>smart|SAM-uniprot|UPI000065D460|UPI000065D460/182-251 Homolog
of Homo sapiens &quot;OTTHUMP00000040508 [Takifugu rubripes]
PSSWGVEEVVSFIKDADPQALGPHTDAFRKHEIDGDALLLLKSEMMMKYLGLKLGPALKLSYHI
DKLKQS
182-251_UPI000065D460    2.373
>smart|SAM-uniprot|UPI0000D572FF|UPI0000D572FF/527-604
PREDICTED: similar to CG9495-PA [Tribolium castaneum]
TQEWGIEEVIQFIESADSCLGVHADLFRKHSLVNWELLQEIDGKALLLLNSDMMMKYMGLKLGP
ALKICNLVSRLKGR
527-604_UPI0000D572FF    2.401
```

```
>smart|SAM-uniprot|Q6C5C5|Q6C5C5_YARLI/3-69 Similar to
tr|AAM08676 Aspergillus fumigatus AAM08676 Ste50p [Yarrowia
lipolytica]
SQAWNTDQVGQWISSLGFPQYSKSFIDNNITGDVLVHLDHDDISDIGVTKIGHRVLILKSIYQL
KEK
3-69_Q6C5C5    2.724
```

Chapter 2

TMKink: a method to predict transmembrane helix kinks

TMKink: A method to predict transmembrane helix kinks

Alejandro D. Meruelo,[1] Ilan Samish,[2] and James U. Bowie[3]*

[1]Medical Scientist Training Program, UCLA-DOE Institute for Genomics and Proteomics, Molecular Biology Institute, UCLA, Los Angeles, California 90095-1570
[2]Department of Structural Biology, Weizmann Institute of Science, Rehovot 76100, Israel
[3]Department of Chemistry and Biochemistry, UCLA-DOE Institute for Genomics and Proteomics, Molecular Biology Institute, UCLA, Los Angeles, California 90095-1570

Received 28 February 2011; Accepted 22 April 2011
DOI: 10.1002/pro.653

Abstract: A hallmark of membrane protein structure is the large number of distorted transmembrane helices. Because of the prevalence of bends, it is important to not only understand how they are generated but also to learn how to predict their occurrence. Here, we find that there are local sequence preferences in kinked helices, most notably a higher abundance of proline, which can be exploited to identify bends from local sequence information. A neural network predictor identifies over two-thirds of all bends (sensitivity 0.70) with high reliability (specificity 0.89). It is likely that more structural data will allow for better helix distortion predictors with increased coverage in the future. The kink predictor, TMKink, is available at http://tmkinkpredictor.mbi.ucla.edu/.

Keywords: membrane protein; protein structure; structure prediction; protein folding

Introduction

Roughly half of all transmembrane helices contain bends or other deviations from ideality.[1,2] Distortions in helix geometry can facilitate conformational changes required for protein function by providing sites of flexibility[3,4] and can be important for positioning key residues precisely in the structure.[5] Kinks that open the polar backbone to alternative hydrogen bonds are often wedged by water, thereby providing a polar region within the hydrophobic core.[6] Proline kinks can also prevent off-pathway events during the folding of membrane proteins, thereby serving as a negative design feature.[7]

Because of the common occurrence of helix breaks in membrane proteins, predicting where they occur could be an important tool for membrane protein structure prediction.[8,9] Existing structure prediction efforts have typically started with the prediction of transmembrane helices that are then packed together in a separate step.[8,10–13] Clearly, knowing where helix deviations are likely to occur would be useful information for packing together transmembrane segments, but this requires that local sequence at least partially encodes the distortion. An early indication that local sequence can provide predictive information about helix deviations was the work of Rigoutsos *et al.*, who found predictive patterns in transmembrane sequences, although the database at the time was too small to perform rigorous cross validation.[14] Langelaan *et al.* developed a kink prediction method exploiting the recent dramatic improvement in database size, but the performance is hard to assess because the database did not exclude homologous proteins.[2]

One sequence signature that is clearly a powerful indicator of helix kinking is the presence of a proline, an amino acid that is incompatible with a

---

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

Grant sponsor: NIH; Grant number: RO1 GM063919; Grant sponsors: Ruth L. Kirschstein NRSA Predoctoral Fellowship Award to Promote Diversity in Health-Related Research, Molecular Biology Whitecome Stipend.

**Correspondence to:* James U. Bowie, Department of Chemistry and Biochemistry, UCLA-DOE Institute for Genomics and Proteomics, Molecular Biology Institute, UCLA, Los Angeles, CA 90095-1570. E-mail: bowie@mbi.ucla.edu

helix.[1,2,5,15,16] Yohannan *et al*. found that kinks can often be identified by looking for prolines in the aligned sequences of homologs.[5] For 36 of 39 kinks examined, Pro occurred near the kink in at least 10% of homologous family members. More recent work with a larger database suggests that the Pro signature may much be less common than originally seen,[1] however (also see below). Nevertheless, it is clear that Pro in either the protein itself or in a homolog provides strong predictive information.

There are indications that nonproline residues can also provide information about kink formation, although the picture is much less clear. Hall *et al*. found that Ser, Thr, and Gly are common in kinked helices.[1] Ser and Thr may collaborate with Pro to modulate bend angle.[17] Langelaan *et al*. did not observe the enhancement in Ser, Thr, and Gly frequencies but did find changes in the prevalence of other polar residues.[2] Clearly as our database expands, our understanding of residue preferences is evolving.

Here, we examine kinked helices and find that there are distinct residue preferences in kinked versus nonkinked helices in a nonredundant database. We exploit these differences and residue conservation to predict kinked helices using a neural network algorithm.

Results and Discussion

Kinked helix search space

To identify possible sequence differences between kinked and nonkinked helices, we first constructed a library of kinked and nonkinked regions from a database of 41 nonhomologous, high-resolution membrane protein structures. Although there are now many more unique structures available, we believe it is important to reduce biases as much as possible by only using unrelated proteins.

We examined nine-residue segments of transmembrane helices and identified 323 kinked and 567 nonkinked segments, defined by strict bend angle criteria as described in Methods. Using only bend angle criteria has the disadvantage of lumping together many different helix anomalies (π-helix, 3_{10} helix, etc.).[16] Moreover, some helix distortions do not lead to a change in bend angle and are ignored by this criterion. Nevertheless, we are restricted by the small data set of nonhomologous high-resolution membrane protein structures currently available, so we opted not to refine the kink type categories further.

We bolstered the limited sequence data available by adding information from homologous sequences to each segment. We counted the amino acids found at each position in the kinked and nonkinked segments, but we reduced counting biases by (1) weighting the counts by sequence divergence so as not to overcount close homologs to the protein of known structures and (2) using only 100 randomly chosen homologs per segment so that all segments were roughly equally weighted (see Methods). We were unable to find 100 homologs for 75 of the kinked segments, and these were eliminated from the residue preference analysis.

Amino acid preferences in kinked helices

The observed amino acid abundance ratios for kinked versus nonkinked helices are shown in Figure 1 (histograms corresponding to this data can be found in Supporting Information Fig. S1); the kink center was defined as position 5. As expected from prior work,[1,2,5] Pro is highly overrepresented at positions 5 through 9 of the kinks. The bias of Pro toward the C-terminus of kinks makes sense because the loss of the hydrogen bond and steric clashes occur at residues preceding the proline. The spread of Pro over many kink positions at least partly reflects the difficulty of defining the center of a kink as well as diversity of kink structures. The overall occurrence of Pro in kinks is lower than we had observed previously.[5] In particular, we found that in a smaller, less diverse database, ~ 90% of kinks contained Pro in 10% of homologs. In the current database, the percentage decreased to 56%. Although Pro had the most pronounced change, other residues also exhibited significant biases. Other than Pro, the residues that were at least twofold overrepresented in kinked helices were as follows: Trp at position 1, Asn at position 4, Trp at position 6, and Glu at position 8. Residues at least twofold underrepresented were Glu at position 1, Asn at position 2, Gln at position 3, Thr, Lys, and Arg at position 4, Asn at position 6, His at postion 7, Gln at postion 8, and Arg at postion 9. The relative dearth of strongly polar residues in kinked helical regions was also observed by Langelaan *et al*.[2] This result is perhaps surprising as polar side chains might be expected to help satisfy any broken backbone hydrogen bonds[1] or support hydrogen bonding to water molecules that often wedge kinks.[6] We do not see the preferences for Gly, Ser, and Thr observed by Hall *et al*.[1] nor the dramatic enhancement of Asp noted by Langelaan *et al*.,[2] perhaps because of differences in our database construction. Because of the variety of kinks and the likely variety of kinking mechanisms, however, understanding the reason for the residue preferences in kinks is not straightforward. Nevertheless, the results indicate that there are differences in amino acid composition in kink positions that could be exploited for kink prediction.

A kink predictor

We developed a neural network analogous to secondary structure prediction algorithms.[19–21] A feedforward network consisting of an input, hidden, and

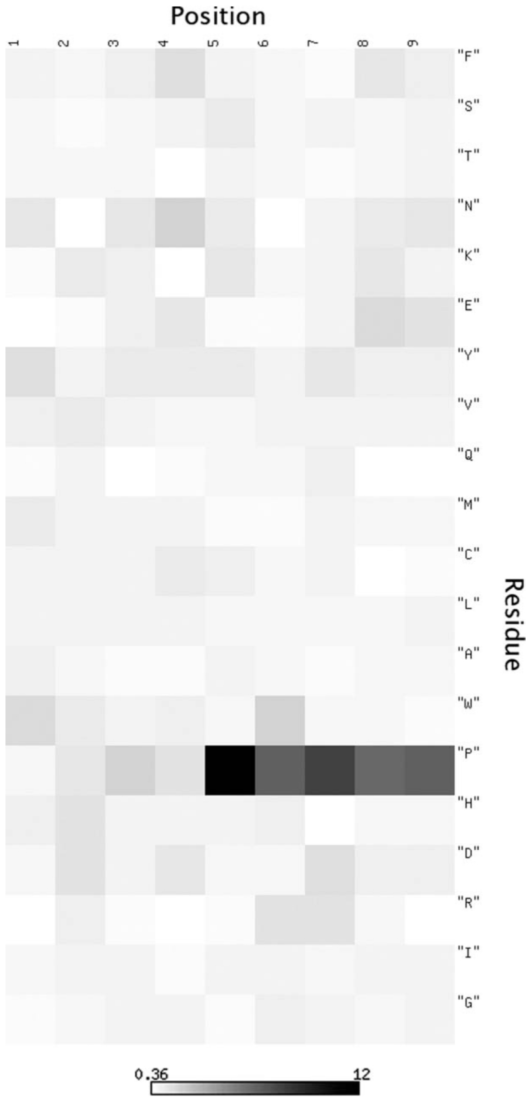


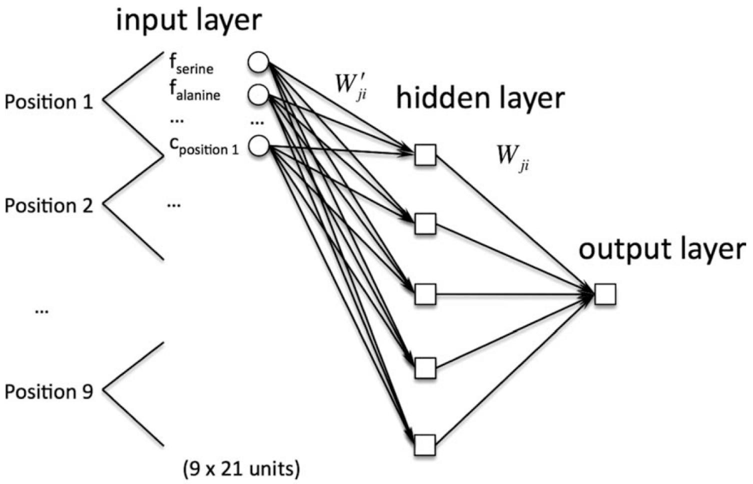
Figure 1. Amino acid composition differences in kinked and nonkinked transmembrane segments. The ratio of residue frequencies in kinked versus nonkinked nine-residue segments from the strict database is shown. Darker regions signify residues that are overrepresented when comparing kinked to nonkinked structures. The position in the segment is labeled along the horizontal axis, and residue type is labeled along the vertical axis. A numerical scale has been provided in the bottom half of the figure. A more quantitative view of this data is provided by histograms in the Supporting Information. This heatmap was generated using Matrix2png.[18]

output layer was constructed as shown in Figure 2. The network inputs were amino acid sequence composition at each position and a measure of sequence conservation as described under Methods. The weights were adjusted using back propagation,[21] and we employed early stopping to prevent overtraining of the network.[22] Network performance was assessed using the leave-one-out method. We found that five hidden nodes provided the best performance on the strict set of training kinks and nonkinks. We also tested performance with 7, 9, 15, and 19 residue window sizes and found that a nine-residue window resulted in the best network performance.

The performance of the network can be seen in the receiver–operator characteristic plot shown in Figure 3. We defined a network output threshold that maximized the Matthews correlation coefficient for the strict and relaxed kink databases. The threshold point for the strict database is shown in Figure 3 and yields a Matthews correlation coefficient of 0.40. At this threshold, the sensitivity is 0.46 and the specificity is 0.99. Thus, about half of all kinks are predicted and those that are predicted are almost always correct.

We next tested the kink predictor with more realistic criteria. The training data set employed strict criteria to rigorously separate kinks and nonkinks, but it omits visually obvious kinks. We therefore tested our kink predictor using more relaxed criteria for kink identification that more accurately reflect what is seen by eye (see Methods). We also assessed prediction performance more accurately in practice by performing predictions across all possible protein windows rather than just selected nine-residue windows in our strict database. Using the relaxed criteria, the sensitivity increased to 0.70 and the specificity diminished somewhat to 0.89. The statistics described in Table I indicate that the majority of kinks are predicted, and when a kink is predicted, it is almost always a correct prediction. As there were no known freely available TM kink prediction algorithms with which to compare our method, we compared our prediction algorithm to a well-known secondary structure prediction algorithm developed for soluble proteins PSIPRED.[23] We defined predicted kinks as a predicted coil or strand and nonkinks as predicted helices. The results using the relaxed kink criteria are shown in Table I. Considering the very different physicochemical basis of helix formation in soluble and membrane proteins, PSIPRED does surprisingly well. Nevertheless, our method clearly outperforms PSIPRED.

The performance for our neural network predictions is illustrated for two known structures in Figure 4. These examples were chosen to simply highlight the types of correct and incorrect predictions possible. To give a fuller picture, predictions for all structures in the database are given in the supplement. The predicted segments were always excluded during network training (see Methods). For the structure 1OTS [Fig. 4(A)], seven kinks were correctly identified (highlighted in red; true positives), four were missed (dark green; false negatives), three



$$X_j = \sum_i W_{ji} I_i \qquad Y_j = \frac{1}{1+e^{-(Xj-\sigma)}}, \sigma = 0.5$$

Figure 2. Neural network design. The neural network consisted of three layers: input, hidden, and output. The inputs are the weighted fractions of each residue type and a conservation score for each position in a set of homologous sequences (see Methods). The total number of inputs was 189 because of 9 positions × 20 residues for each candidate, and nine conservation score inputs. Two sets of weight matrices were evolved through learning: one between the input and hidden layers and the other between the hidden and output layers. For the output layer, the corresponding expressions are highlighted below and have been left general so as to also apply to the hidden layer. I_i denotes input (i.e., output from hidden node *i*) and W_{ji} its corresponding weight. *i* denotes the hidden node number and *j* the output node number into which it goes. Y_j denotes the output from a given output node. Sigma is the threshold of a neuron. The expressions are similar for the hidden layer, with weights W'_{ji} used in place of the previous weights.

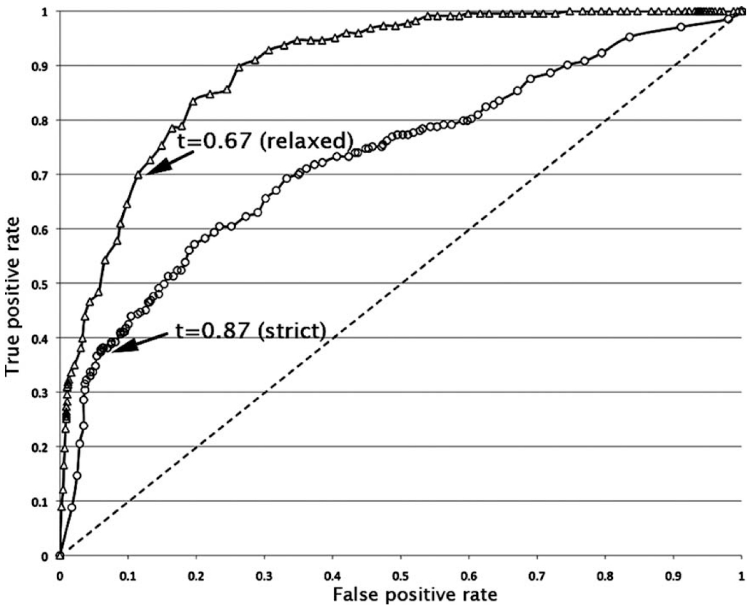


Figure 3. Receiver–operator characteristic plots for prediction performance. The black solid lines represent the receiver–operator curves for leave-one-out validation on the strict and relaxed databases. Statistics are provided in Table I.

Table I. *Statistical Summary of Prediction Performance*

Structures	Sensitivity	Specificity	Jack-knife correlation coefficient	Training correlation coefficient	Kink prediction	Nonkink prediction
Strict	0.46	0.99	0.40	0.57	96%	78%
Relaxed	0.70	0.89	0.56	—	62%	92%
PSIPRED relaxed	0.64	0.81	0.42	—	54%	86%

Prediction performance was assessed using the leave-one-out method for the strict set of kink and nonkink structures as well as for the relaxed set of kink and nonkink structures. The PSIPRED results are considered equivalent to a jack-knifed procedure because none of the proteins used in the current database were included in the original soluble protein training database.

nonkinked helices were correctly predicted as nonkinked (true negatives), and there were no false positives. Looking at the four kinks that were missed, it appears from Figure 4(A) that these are relatively subtle bends and may have less distinctive sequence signatures. For the structure 1H2S [Fig. 4(B)], two kinks were correctly identified, one was missed, and the algorithm predicts kinks at five positions that are not kinked. The seventh helix illustrates an error in kink identification. We predict a kink at a point where the helix clearly breaks, but the bend angle does not change. Thus, we would argue that the algorithm actually predicted this deviation correctly, but our method of identifying true kinks in

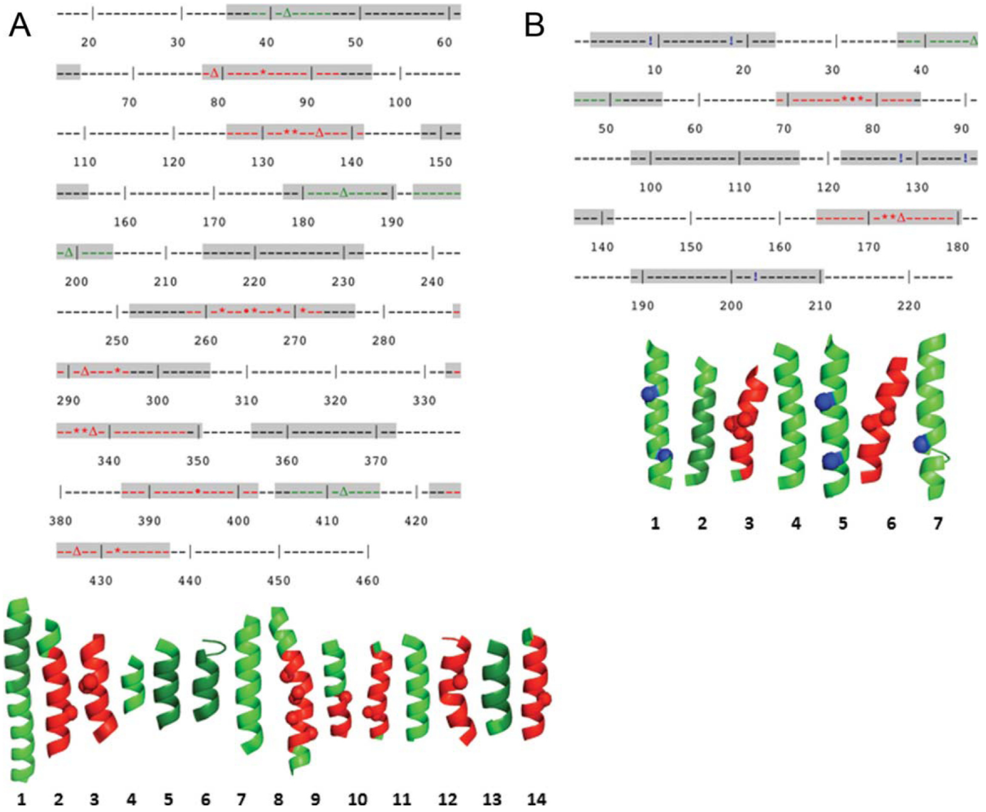


Figure 4. Examples of predictions. Illustration of kink prediction results using the relaxed database. (A) PDB code 1OTS, chain A and (B) PDB code 1H2S, chain A. The upper section of each figure illustrates the prediction results as a one-dimensional sequence. Transmembrane regions have been highlighted in gray. Regions highlighted in red correspond to correctly predicted regions (true positives), and blue exclamation points to incorrectly predicted positions (false positives). The maximum bend angle in each kinked region has been indicated by a Δ. The center of each kink prediction has been marked with a *, and overlap of a prediction center and the maximum bend angle has been marked with a ●. Kinked regions not predicted have been highlighted in green (false negatives). The bottom of the figure shows the structures of the individual transmembrane helices. The coloring scheme is the same as for the upper figure. The correctly predicted kinks are highlighted in red, and the center of each kink prediction is indicated by a red ball at the Cα position. False negatives are highlighted in dark green. False-positive predictions are indicated by a blue ball at the Cα position.

structures is not foolproof. The one missed kink and five mispredicted kinks that are not in fact kinked may be attributed to the fact that 1H2S chain A had only 118 homologs compared to other structures that often had thousands of homologs.

Conclusion

Our results indicate that there are significant sequence differences between kinked and nonkinked transmembrane helices and that these differences can be exploited to predict kinks. This does not mean that kinks are generated only by local sequence, however, as we cannot necessarily link cause and effect. For example, certain residue types may provide stronger long-range contacts that can distort helices. Alternatively, if kink occurrence is more frequent at functional sites, the residue preferences could reflect the likelihood of being in functional sites rather than kinks. It is likely that prediction results could be greatly improved with more structures. Sequence preferences other than proline are much more subtle and seem to vary with database construction. Some of the more subtle sequence pattern information is likely masked because of the limited number of kink classifications we can use. Clearly not all helix distortions are the same, but because of the small dataset, we used a simple binary classification of kinked or nonkinked based solely on bend angle. Ideally, one would like to parse helix bends into distinct classes as the sequence signatures may be quite different. In spite of these limitations, we have been able to develop a prediction method that provides useful information that can be incorporated into helix packing algorithms.[8,11,24] Kink prediction will only improve in accuracy and refinement as more structures become available.

Methods

Structure database

A nonredundant database of membrane proteins was used as the source from which all kinks and nonkinks were identified. None of the sequences shared more than 30% sequence identity, and only structures solved at a resolution of 2.5 Å or better resolution were retained. This resulted in a total of 41 membrane protein structures: 1C3W, 1EHK, 1H2S, 1JB0, 1K4C, 1KB9, 1KQF, 1NKZ, 1OKC, 1OTS, 1Q16, 1SU4, 1U19, 1V54, 1XIO, 1XQF, 1Z98, 2A65, 2AHY, 2BHW, 2BL2, 2BS2, 2F2B, 2FBW, 2J58, 2J7A, 2J8C, 2NQ2, 2RH1, 2UUI, 2VPZ, 2W2E, 2WGM, 2Z73, 2ZXE, 3B45, 3B9W, 3BKD, 3C02, 3DDL, and 3KCU.

Kinked and nonkinked transmembrane helix identification

Kink and nonkink identification was performed differently for training and for validation purposes. To ensure that we only trained on the clearest, most well-defined kinks, we were particularly strict in our training set definitions. Subsequently, to more accurately assess our prediction performance, we relaxed the criteria to identify all genuine kinks and nonkinks. The strict and relaxed criteria are described below.

Kinked and straight helices were identified by measuring bend angles of nine-residue stretches of transmembrane helices. Only unique chains for each structure were used. For each chain, the transmembrane regions defined in the Orientations of Proteins in Membranes (OPM) database[25] were used as a starting point. The transmembrane region was then extended to the end of the helix if the region was defined as helical by DSSP.[26] These extended helices constituted the template from which the nine-residue segments were extracted, although only segments with a center residue within at least one residue of the transmembrane region (as defined by OPM) were retained. Kinks and nonkinks for transmembrane helical regions were identified using bend angles measured using the ProKink plug-in available for Simulaid.[27]

To extract kinked segments using strict criteria for training purposes, the nine-residue segments were filtered using several steps. First, the bend angle of all segments was sorted from largest to smallest. Beginning with the largest bend angle segment, nearby segments were eliminated if they overlapped. This process was repeated for every remaining structure in order of decreasing bend angle. Those segments that remained were then filtered using two criteria. The average bend angle over four neighboring sliding windows was required to be greater than (or equal to) 13°, or the bend angle for any individual window was required to be greater than 24°. If only three or fewer neighboring windows were available because they were at a helix end, then any candidate with a bend angle of 13° or less was thrown out. Ultimately, using the strict criteria, 323 kinked structures were obtained.

To identify clear nonkinked segments, we also employed strict criteria. First, the bend angle of all segments was sorted from smallest to largest. Any window exceeding 20° in bend angle was immediately discarded. Beginning with the smallest bend angle segment, nearby segments were eliminated if they overlapped. This process was repeated for every remaining structure in order of increasing bend angle. Those structures that remained were then filtered using two criteria: the average bend angle over four neighboring segments was required to be less than 8°, and the bend angle for every individual segment had to be less than 20°. In this manner, 567 nonkinked segments were obtained.

For prediction assessment we employed relaxed criteria that identified more visually obvious bends. A kink was identified as an authentic kink if the average bend angle over three neighboring windows was greater than (or equal to) 10° or the bend angle of an individual window was greater than 18°. A total of 2048 relaxed kink windows were extracted. This does not mean that there are 2048 kinks because overlapping windows can contain the same kink. We wanted to identify all kink windows for assessment of predictions because all possible windows are tested in kink prediction. Nonkink regions were identified as all regions other than the identified kink regions above.

Family building

Similar sequences in RefSeq's nonredundant protein sequence database[28] were identified using PSIBLAST with one iteration, and a P-value cutoff of $1e-10$ to ensure close similarity between the known structure and its sequence homologs. The number of sequences identified was limited to 30,000. Any aligned sequences with an alignment length less than 70% of the length of the original sequence were discarded. All nonnative gapped regions were discarded.

To count amino acid frequencies at each position, the counts were weighted by (1-% identity) to incorporate some information about the likelihood a residue might have changed in the homologous sequence (% identity refers to the whole protein). Thus, when a homolog sequence is 100% identical to the sequence of the known structure, the new amino acid is not counted. In this manner, close homologs do not overweight the counting statistics.

For input to the neural network, each residue at a given position of the original candidate sequence was added at full weight a fixed number of times per homolog. The number of times was optimized to maximize network performance; this value was 2.

This weighting scheme is described by the following equation for the 180 inputs corresponding to the 9 positions × 20 amino acids:

$$I(\mathrm{res_i}, \mathrm{pos}_j) = \frac{\sum\limits_{\text{all homologs}} [(1 - \%\ \text{identity})S(\mathrm{res}_i) + \lambda]}{\sum\limits_{1}^{N_{\mathrm{pos_j}}} [(1 - \%\ \text{identity}) + \lambda]}$$

$$S(\mathrm{res}_i) = \begin{cases} 1, \text{if } \mathrm{res}_i = \mathrm{res}_i \\ 0, \text{if } \mathrm{res}_i \neq \mathrm{res}_i \end{cases},$$

$$\lambda = \begin{cases} 2, \text{if } \mathrm{res}_i = \mathrm{res}_i \text{ in known structure} \\ 0, \text{if } \mathrm{res}_i \neq \mathrm{res}_i \text{ in known structure} \end{cases},$$

$$\mathrm{N}_{\mathrm{pos}_j} = \#\ \text{residues at position } j, \text{excluding gaps}.$$

Use of sequence conservation

Sequence conservation scores were used as inputs to the neural network in addition to the primary sequence. Conservation scores were calculated for the nine-residue segment belonging to each candidate sequence window. This was accomplished by using the PSIBLAST data produced by the family building step described above. From all pairwise alignments for each PDB/chain, 3000 were randomly selected and conservation scores calculated using SCORECONS with the Trident scoring method.[29] For this purpose, ClustalW-formatted pseudo-alignments were constructed.[30] Positions with excessive gaps were assigned a score of 0.

Neural network design

A feedforward neural network consisting of an input, hidden, and output layer was constructed. There were 9 (positions) × 20 (amino acids) = 180 inputs to the network related to the fractional occurrence of all possible residue/position combinations (as described by the input equation above). In addition, we used nine inputs corresponding to the conservation scores of the nine residues in the native sequence. Two sets of weight matrices were evolved through learning: one between the input and hidden layers and the other between the hidden and output layers. Weights were randomly assigned to values between −0.5 and 0.5 to start. The number of hidden nodes was optimized according to network performance, and the best performance was found to occur using five hidden nodes. The output layer is governed by the following equations, which have been left general so as to also apply to the hidden layer. I_i denotes the input (i.e., output from hidden node i) and W_{ji} its corresponding weight. i denotes the hidden node number (e.g., 3) and j the output node number into which it goes. Y_j denotes the output from a given output node. Sigma is the threshold of a neuron and was set to 0.5[21]:

$$X_j = \sum_i W_{ji} I_i$$

$$Y_j = \frac{1}{1 + e^{-(X_j - \sigma)}}.$$

The expressions are similar for the hidden layer, with weights W'_{ji} used in place of the previous weights. For each cycle, the weight matrices were adjusted once for each example in the training set. Training was performed over 1000 cycles. The order of training examples during weight adjustments was shuffled every cycle to reduce any noise due to ordering.

Weight adjustment was performed in the following fashion. The output layer consisted of a single node predicting whether a candidate sequence was a kink (maximum output value of 1) or a nonkink (minimum output value of 0). The error in this output value for each training example was computed as follows:

$$\delta = D - Y,$$

where D is the correct value and Y is the predicted value.[21] Adjustments to the weight matrix between the hidden and output layers were made according to the expression:

$$W_{ji}(t+1) = W_{ji}(t) + \eta\delta_j I_i,$$
$$\delta_j = \delta,$$

where η is the learning rate. The optimal learning rate was found to be 0.1 (according to how far training progressed). The weight matrix between the input and hidden layers was adjusted by first back propagating the error according to the expression[21]:

$$\delta'_i = \sum_j W_{ji}\delta.$$

The weight layer adjustments were then made using the W_{ji} $(t + 1)$ expression above but with weights W'_{ji} used in place of the previous weights, error δ_I' in place of the previous error, and inputs to the appropriate layer.

Early stopping was used to prevent overtraining. To find the point at which overtraining occurs, we evaluated the network every 10th cycle (of 1000 cycles) for its ability to predict structures in a test set that was not used for training. The test set consisted of 50 kinked and 50 nonkinked structures. The total error over all test set examples was computed as follows:

$$\delta_T = \sum_{\text{all examples in test set}} \delta.$$

The training network for which the lowest total error occurred on the test set was taken as the best network.[22]

Training and test set creation

To ensure no bias in the datasets used for training and early stopping, kink and nonkink selection was randomized. Test set kink selection was done by selecting 50 structures at random from the total number of identified strict kinks. Similarly, test set nonkink selection was done by selecting 50 structures at random from the total number of identified strict nonkinks. These two groups became the test set used for early stopping. The remaining strict structures (273 kinks and 517 nonkinks) were retained for training of the neural network.

Evaluation

Because of limited data for membrane proteins, we chose to use the jack-knife or leave-one-out method to evaluate our strict and relaxed predictions. The threshold for each database (strict or relaxed) was chosen using a receiver–operator plot where the Matthews correlation coefficient was maximized (see Fig. 3). Relaxed predictions were done across all possible protein windows in transmembrane segments. Relaxed evaluation was done using the jack-knife approach for trained examples from the strict database and the best network (i.e., one with lowest total error) for all other structures. Assessment of kinks and nonkinks using the relaxed criteria was done separately. Because it is difficult to unambiguously separate the transition region between kinked and nonkinked regions, a buffer region of four residues to the left and right of the original kinked region centers was created. We only evaluated nonkinked region centers that were more than four intervening residues from this unambiguously defined transition kink area. These residue separations were arrived at after visual evaluation of many structures.

Kinked regions that passed the relaxed criteria were merged together to form contiguous regions when separated by one residue or less. These merged regions defined the contiguous kink regions that would be checked to see if they contained predicted kinks or not. If a contiguous kinked region contained a single kink prediction, it was recorded as a single true positive. If a contiguous kinked region did not contain a single kink prediction, it was recorded as a single false negative.

Nonkinks were evaluated differently. When evaluating a nonkink region, every single nine-residue window contained within the contiguous region was evaluated individually (rather than as a continuous region). If a given window was correctly predicted as a nonkink, it was recorded as a single true negative. If it was incorrectly predicted as a kink, it was recorded as a single false positive. All of these true and false negatives and positives were summed over all predictions. Predictions using the secondary structure prediction algorithm PSIPRED[23] were evaluated in the same way except that predicted coils or strands were defined as kinks. As with kink prediction, a nine-residue window around the predicted coil or strand was defined as the kink region.

Website predictor

De novo prediction by our web application available online is done solely using the primary amino acid sequence of entire proteins. The transmembrane helical regions can be input manually or determined automatically by Proteus2.[31]

We ran three trials of our neural network using five hidden nodes. This resulted in jack-knife networks with correlation coefficients of 0.39, 0.38, and 0.40. We chose to use the network with the highest correlation coefficient. We looked at all networks making up this jack-knife network and retained the network with the lowest total error when early stopped using the test set. This network was installed for use in online predictions.

The network threshold for online predictions was chosen to be that determined by the receiver–operator curve analysis for the relaxed database ($t =$

0.67). This was done because the relaxed database most closely mirrors practical usage of the prediction software on entire proteins.

Acknowledgment

The authors thank Zheng Cao for helpful feedback on web program development, Duilio Cascio for helpful program testing suggestions, Luki Goldschmidt for assistance with the cluster, Tom Holton and Alex Lisker for help setting up the MBI webspace, Amit Oberai for helpful discussion about program development, and members of the Bowie lab for careful reading of the manuscript.

References

1. Hall SE, Roberts K, Vaidehi N (2009) Position of helical kinks in membrane protein crystal structures and the accuracy of computational prediction. J Mol Graph Model27:944–950.
2. Langelaan DN, Wieczorek M, Blouin C, Rainey JK (2010) Improved helix and kink characterization in membrane proteins allows evaluation of kink sequence predictors. J Chem Inf Model50:2213–2220.
3. Bright JN, Shrivastava IH, Cordes FS, Sansom MSP (2002) Conformational dynamics of helix S6 from Shaker potassium channel: simulation studies. Biopolymers64:303–313.
4. Shi L, Liapakis G, Xu R, Guarnieri F, Ballesteros JA, Javitch JA (2002) Beta2 adrenergic receptor activation. Modulation of the proline kink in transmembrane 6 by a rotamer toggle switch. J Biol Chem277:40989–40996.
5. Yohannan S, Faham S, Yang D, Whitelegge JP, Bowie JU (2004) The evolution of transmembrane helix kinks and the structural diversity of G protein-coupled receptors. Proc Natl Acad Sci USA101:959–963.
6. Miyano M, Ago H, Saino H, Hori T, Ida K (2010) Internally bridging water molecule in transmembrane alpha-helical kink. Curr Opin Struct Biol20:456–463.
7. Wigley WC, Corboy MJ, Cutler TD, Thibodeau PH, Oldan J, Lee MG, Rizo J, Hunt JF, Thomas PJ (2002) A protein sequence that can encode native structure by disfavoring alternate conformations. Nat Struct Biol9: 381–388.
8. Barth P, Wallner B, Baker D (2009) Prediction of membrane protein structures with complex topologies using limited constraints. Proc Natl Acad Sci USA106: 1409–1414.
9. Gimpelev M, Forrest LR, Murray D, Honig B (2004) Helical packing patterns in membrane and soluble proteins. Biophys J87:4075–4086.
10. Freddolino PL, Kalani MYS, Vaidehi N, Floriano WB, Hall SE, Trabanino RJ, Kam VWT, Goddard WA (2004) Predicted 3D structure for the human β2 adrenergic receptor and its binding site for agonists and antagonists. Proc Natl Acad Sci USA101:2736–2741.
11. Kim S, Chamberlain AK, Bowie JU (2003) A simple method for modeling transmembrane helix oligomers. J Mol Biol329:831–840.
12. Treutlein HR, Lemmon MA, Engelman DM, Brünger AT (1992) The glycophorin A transmembrane domain dimer: sequence-specific propensity for a right-handed supercoil of helices. Biochemistry31:12726–12732.
13. Nugent T, Jones DT (2010) Predicting transmembrane helix packing arrangements using residue contacts and a force-directed algorithm. PLoS Comput Biol 6: e1000714.
14. Rigoutsos I, Riek P, Graham RM, Novotny J (2003) Structural details (kinks and non-alpha conformations) in transmembrane helices are intrahelically determined and can be predicted by sequence pattern descriptors. Nucleic Acids Res31:4625–4631.
15. Cordes FS, Bright JN, Sansom MSP (2002) Proline-induced distortions of transmembrane helices. J Mol Biol323:951–960.
16. Riek R, Rigoutsos I, Novotny J, Graham RM (2001) Non-α-helical elements modulate polytopic membrane protein architecture. J Mol Biol306:349–362.
17. Deupi X, Olivella M, Govaerts C, Ballesteros JA, Campillo M, Pardo L (2004) Ser and Thr residues modulate the conformation of pro-kinked transmembrane alpha-helices. Biophys J86:105–115.
18. Pavlidis P, Noble WS (2003) Matrix2png: a utility for visualizing matrix data. Bioinformatics19:295–296.
19. Holley LH, Karplus M (1989) Protein secondary structure prediction with a neural network. Proc Natl Acad Sci USA86:152–156.
20. Rost B, Sander C (1993) Prediction of protein secondary structure at better than 70% accuracy. J Mol Biol232:584–599.
21. Sun Z, Rao X, Peng L, Xu D (1997) Prediction of protein supersecondary structures based on the artificial neural network method. Protein Eng10:763–769.
22. Prechelt L (1998) Automatic early stopping using cross validation: quantifying the criteria. Neural Netw11: 761–767.
23. McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. Bioinformatics16:404–405.
24. Yin H, Slusky JS, Berger BW, Walters RS, Vilaire G, Litvinov RI, Lear JD, Caputo GA, Bennett JS, DeGrado WF (2007) Computational design of peptides that target transmembrane helices. Science315: 1817–1822.
25. Lomize MA, Lomize AL, Pogozheva ID, Mosberg HI (2006) OPM: orientations of proteins in membranes database. Bioinformatics22:623–625.
26. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers22: 2577–2637.
27. Visiers I, Braunheim BB, Weinstein H (2000) Prokink: a protocol for numerical evaluation of helix distortions by proline. Protein Eng13:603–606.
28. Pruitt KD, Tatusova T, Maglott DR (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res 35:D61–D65.
29. Valdar WS (2002) Scoring residue conservation. Proteins48:227–241.
30. Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD (2003) Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res31:3497–3500.
31. Montgomerie S, Cruz JA, Shrivastava S, Arndt D, Berjanskii M, Wishart DS (2008) PROTEUS2: a web server for comprehensive protein structure prediction and structure-based annotation. Nucleic Acids Res 36: W202–W209.

Chapter 3

Structural differences between mesophilic and thermophilic membrane proteins

Introduction

Protein thermostability remains a focal point of interest for protein scientists[1]. Thermophiles, in contrast to mesophiles, evolved to survive at higher temperatures, ranging from 45 to 122° C[2]. As a result of their greater thermostability, thermophiles are of great interest to those studying how to improve protein thermostability. Knowledge gained can be applied towards protein engineering.

The differences in thermostability between mesophilic and thermophilic soluble proteins have been extensively studied. Properties studied include packing density, burial fraction, hydrogen bonding, ion pairs, and length of surface loops[1,3–7]. The only general rule has been that the number of ion pairs increases with rising, optimal growth temperature[8].

Several groups argue that protein thermostabilization is achieved by different mechanisms for different groups of proteins[8,9]. Berezovsky and Shakhnovich argue that for soluble proteins, there are two major physical mechanisms of thermostabilization: structure-based and sequence-based[9]. Some hyperthermophilic proteins are described as being more structurally compact than their mesophilic homologues. Other thermophilic proteins retain similar structure but differ in a small number of strong, sequence-based interactions. The strategy adopted by a given protein is thought to depend on its evolutionary origin. Organisms that began life in an extreme environment such as hyperthermophilic archaea have proteins that are structurally more compact.

Organisms that began as mesophiles but migrated into a warmer environment adopted a sequence-based evolutionary strategy based on a few strong interactions[9].

The vast majority of thermostability studies comparing mesophilic and thermophilic proteins have been conducted on soluble proteins. This has largely been because of an abundance of structural data for these proteins. While no comparable number of structures exist for membrane proteins, approximately 60% of approved drugs targets are membrane proteins[11]. Understanding the thermostability of membrane proteins is thus of critical importance to lowering their crystallization barrier and increasing the amount of structural data available for study[12].

Membrane proteins differ from soluble proteins in many ways. Soluble proteins exist in an aqueous environment whereas membrane proteins are surrounded by phospholipid bilayers. To interface with such an environment, membrane proteins have adapted a much greater abundance of hydrophobic residues and far fewer polar or charged residues. As a result of these differences, the same questions investigated to study thermostability in soluble proteins may lead to different answers when looking at membrane proteins.

While there are many factors that may contribute to thermostability in membrane proteins, I investigated burial and packing differences between mesophiles and thermophiles. Packing differences may explain the thermostability difference between mesophilic and thermophilic membrane proteins. Large scale studies using non-redundant soluble proteins have shown that there is no difference in the packing of soluble proteins[6]. Membrane protein packing is different from soluble protein packing[13–16], so thermophilic adaptation may be different. Some groups have found there to be greater packing in membrane proteins than in soluble proteins[13,14]. Faham *et al.* has argued that packing is the dominant force driving membrane protein folding[16]. And Hildebrand *et al.* found lesser packing in membrane proteins than in soluble proteins[15]; Adamian and Liang have found more cavities in membrane proteins to explain this[17].

To study thermostability in membrane proteins, a non-redundant database was created consisting of alpha-helical membrane proteins to examine the differences between mesophiles and thermophiles. Structurally analogous pairs of mesophiles and thermophiles were identified using structure and sequence information. Using these pairs and the database, burial and packing density were evaluated to see if they helped explain the difference in thermostability between thermophiles and mesophiles.

Results and Discussion

Mesophile-thermophile database construction

To explore the structural differences between mesophiles and thermophiles that may contribute to thermostability of membrane proteins, I constructed a library of mesophiles and thermophiles from a database of 126 non-homologous, high-resolution alpha-helical membrane protein structures. Using optimal growth temperature data, these structures were divided into 101 mesophiles and 25 thermophiles (list provided in Methods). To reduce biases further in looking at properties of membrane proteins, only unique chains (as described in Methods) were used. Prosthetic groups were carefully included and author-recommended oligomer forms were used. Burial and packing were studied within the context of this database. In addition, a database of structurally similar pairs of mesophiles and thermophiles was created (as described in Methods). Differences found using the unpaired structure database may be due to variation in function among the proteins studied. Structurally similar pairs are more likely to share function. Thus, differences found using the paired structure database are more likely to explain differences in thermostability. A difference observed in both the unpaired and paired structure databases is more likely to explain a difference in thermostability. Lastly, those proteins with a greater number of residues were weighed more heavily as has been described in the Methods. The rationale was that a protein with more residues contributes more data to analyze than one with fewer, and thus should have greater weight.

Burial in mesophiles and thermophiles

Burial characterizes how exposed a residue to what is outside of a protein such as solvents and substrates. I divided analysis of burial values into analysis of soluble regions and analysis of transmembrane regions. Soluble regions of membrane proteins are the closest relative to regions of soluble proteins while transmembrane regions are a region that might be expected to differ due to the environment in which membrane proteins reside.

Burial values were calculated for mesophiles and thermophiles in the unpaired and paired structure databases, and tabulated in Table 3-1 below. Little or no difference in burial values was found in four comparisons:

(1) soluble regions of unpaired mesophiles and thermophiles
(2) transmembrane regions of unpaired mesophiles and thermophiles
(3) soluble regions of paired mesophiles and thermophiles
(4) transmembrane regions of paired mesophiles and thermophiles

The distributions of these values were all found to be approximately normal, with no outliers.

Table 3-1: Burial and packing in mesophiles and thermophiles

Measurement		Soluble	Transmembrane
Burial	Unpaired ratio	1.02±0.03[1]	1.1±0.1[1]
	Paired difference	0.003±0.004[2]	0.007±0.008[2]
Packing	Unpaired ratio	1.006±0.005[1]	1.008±0.005[1]
	Paired difference	0.004±0.001[2]	0.001±0.002[2]

[1] mean burial or packing ratio (mesophiles/thermophiles) with standard error
[2] mean burial or packing difference (mesophile-thermophile) with standard error

An unpaired ratio of 1 indicates no burial or packing preference when mesophiles are compared to thermophiles. A paired difference of 0 indicates no burial or packing difference between paired mesophiles and thermophiles.

Packing in mesophiles and thermophiles

Similar to the analysis for burial values, the analysis of packing values was divided into soluble region and transmembrane region analyses. Packing describes how many atoms fit into a fixed volume. Packing values were calculated for mesophiles and thermophiles in the unpaired and paired structure databases, and tabulated in Table 3-1 above. Little or no difference in packing values was found in four comparisons:

(1) soluble regions of unpaired mesophiles and thermophiles

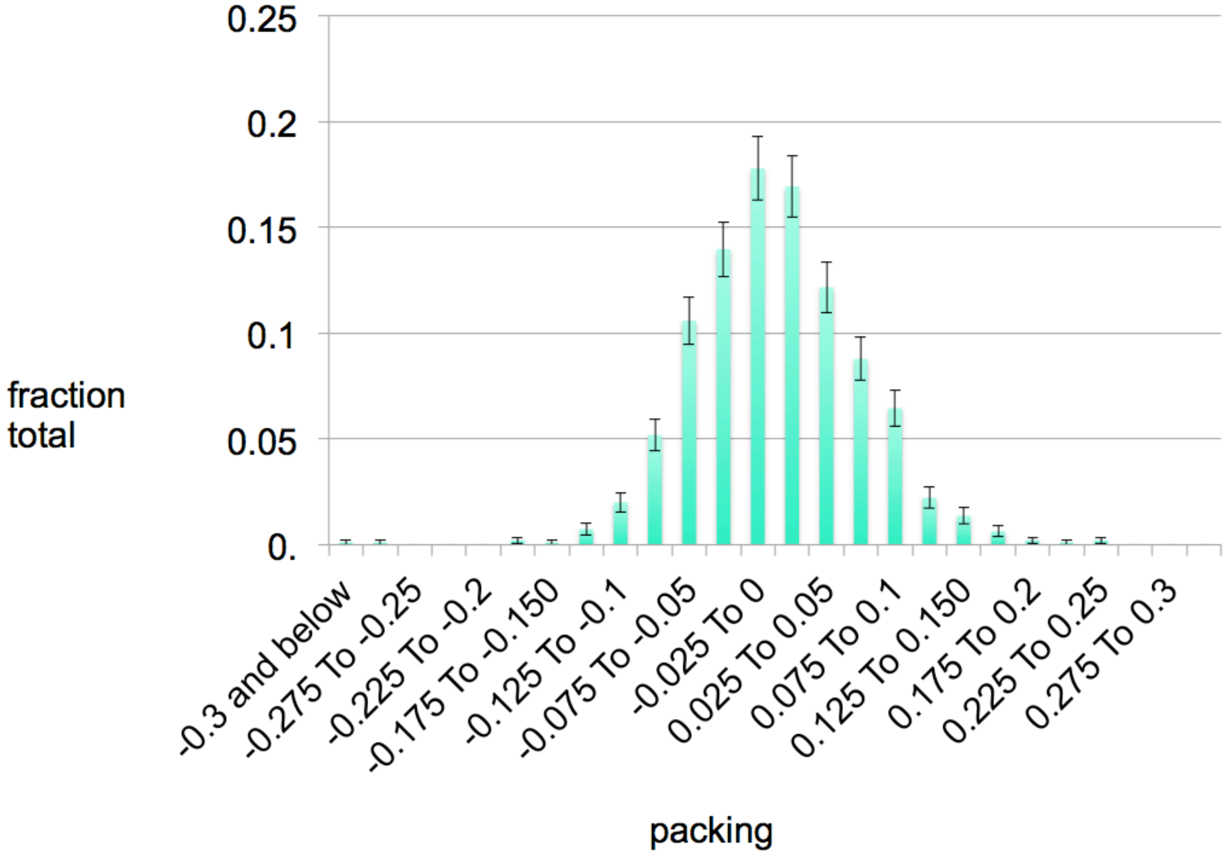
(2) transmembrane regions of unpaired mesophiles and thermophiles

(3) soluble regions of paired mesophiles and thermophiles

(4) transmembrane regions of paired mesophiles and thermophiles

The distributions of these values were all found to be approximately normal. The transmembrane pair packing distribution, in addition to having a normal distribution, also had 2 outliers (shown in Figure 3-1 below). These outliers were explained by too short alignment lengths (31 and 28 residues). With such short alignment lengths, the likelihood of alignment of one protein to any other protein becomes extremely high (e.g., one alpha helix from one protein to the alpha helix of another protein) and packing differences lose significance. As a result, these outliers were eliminated from analysis.

Figure 3-1: Average packing difference distribution for transmembrane regions of mesophile-thermophile pairs



The 'packing' axis describes the distribution of the difference in average packing density for every mesophile-thermophile protein pair. The 'fraction total' axis describes the fraction of protein pairs found in each packing value difference bin.

It is surprising that both membrane and soluble proteins have no packing or burial differences despite their vastly different environments[6] and the differing role of packing in membrane proteins[13–16]. Looking solely at burial and packing values, soluble and membrane proteins appear closer cousins than previously thought[13–16]. However, these results are in agreement with previous work by Joh *et al.*, who found little difference in packing energetics of membrane and water soluble proteins[18].

With no burial or packing differences, the question arises as to what the thermostability differences can be attributed to. The structure of membrane proteins is dominated by alpha helices and hydrogen bonding to stabilize such helices. As a result, one might expect differences in thermostability to arise from differences in interhelical hydrogen bonding or other stability adaptations such as kinking of helices. Ions pairs may be another contributor to thermostability as these have also been found in membrane proteins[19–21]. In addition, differences within soluble regions of membrane proteins such as surface loop length may be distinguishing features of thermophiles from mesophiles. Lastly, structural pairs may reveal sequence determinants of stability differences that have not previously been apparent from pure sequence analysis. These parameters have not been studied for mesophiles and thermophiles in membrane proteins and are an active area of exploration. As additional membrane protein structures are deposited, it may become clear that several different mechanisms may have evolved to explain differences in thermostability for different groups of membrane proteins as has been the case for soluble proteins[8,9]. For example, the thermostability of hyperthermophiles may be explained by structural differences between hyperthermophiles and mesophiles. And the thermostability of thermophiles may be explained by sequence differences between thermophiles and mesophiles.

Conclusion

Even though membrane proteins differ from soluble proteins in many fundamental ways such as an abundance of hydrophobic residues and the role of packing[13–16], burial and packing values appear to be shared between mesophiles and thermophiles in both types of proteins. The addition of new non-redundant structures to the existing membrane protein database in the future may modify this picture, but the present data unambiguously suggest this conclusion. Little or no difference in burial or packing values exist between mesophiles and thermophiles in membrane and soluble proteins. One question this observation raises is what can explain the thermostability differences between mesophiles and thermophiles in membrane proteins. The only general pattern for soluble proteins relates the number of ion pairs and rising growth temperature[8]. For membrane proteins, many properties such as surface loop length, interhelical hydrogen bonding, kinking of helices, ion pairs, and sequence determinants remain to be explored. Ultimately, as for soluble proteins, a combination of different mechanisms may have evolved and will have to be investigated in the future[8,9].

Methods

Structure database

A nonredundant database of mesophiles and thermophiles was created using alpha-helical membrane proteins from Stephen White's database (http://blanco.biomol.uci.edu/mpstruc/listAll/list)[10]. Growth temperatures were obtained from the American Type Culture Collection (http://www.atcc.org/ATCCAdvancedCatalogSearch/AllCollectionSearch/tabid/454/Default.aspx) and additional papers from the literature. Structures were divided into mesophiles and thermophiles based on a 50 °C cutoff; those greater than or equal to this cutoff were considered thermophiles.

Each group of structures (mesophiles, thermophiles) was made non-redundant separately. To reduce each set of structures to a non-redundant set, a random structure was chosen and its sequence was blasted against that of all other structures. If the random structure's sequence had a similarity expectation value of less than 1e-10, the structure was removed. This was repeated for randomly chosen structures until the number of member structures did not change for 20 iterations.

For each structure, the author-recommended oligomer was chosen as described by REMARK 350 in each Protein Data Bank (PDB) file. To further reduce the dataset to a meaningful one for burial and packing analysis, a number of additional requirements were enforced for member structures. Only unique chains (i.e., non-identical) were used for analysis. Nuclear magnetic resonance (NMR) and electron microscope (EM) structures were eliminated. For each oligomer, three or more transmembrane helices were required to retain it in the structure database; those with fewer transmembrane helices were thought to contain insufficient data to perform transmembrane packing and burial analysis. Structures lacking membrane protein topology when examined by eye were removed. The prosthetic groups for each structure were carefully examined and retained in the oligomeric structures so that they would be considered for burial and packing analysis. Lipids and other non-intrinsic components were removed from each structure for consistency in analysis. The members making up the structure database were as follows.

Mesophiles (101): 1AIJ, 1BCC, 1BGY, 1COW, 1EFR, 1FFT, 1KPL, 1KQF, 1L0V, 1LGH, 1NEK, 1NKZ, 1PW4, 1Q16, 1Q90, 1QO1, 1U7G, 1V54, 1YEW, 1ZCD, 1ZOY, 1ZRT, 2A79, 2ACZ, 2BHW, 2BL2, 2BS2, 2C3E, 2CK3, 2E74, 2EI4, 2EXW, 2GMH, 2GSM, 2H8A, 2I37, 2J58, 2JLN, 2NUU, 2O01, 2O7L, 2OAR, 2OAU, 2Q7M, 2QCU, 2QI9, 2QKS, 2QTS, 2RCR, 2RDD, 2V8N, 2VQG, 2W5J, 2WCD, 2WSS, 2WSW, 2X2V, 2XUT, 2YFY, 2ZUQ, 2ZW3, 2ZXE, 3B8E, 3B9W, 3B9Y, 3BF0, 3BKD, 3CHX, 3CK6, 3CX5, 3D31, 3DQB, 3EAR, 3EHZ, 3FH6, 3G5U, 3H90, 3HD6, 3JQO, 3K07, 3KCU, 3KG2, 3KP9, 3LIM, 3LRB, 3LUT, 3M71, 3MK7, 3MKT, 3NAF, 3NVO, 3O7Q, 3OAA, 3P0G, 3PJZ, 3QAK, 3QE7, 3QNQ, 3RLB, 3RVY, and 3S0X

Thermophiles (25): 1EHK, 1JB0, 1ORQ, 2A0L, 2B2F, 2F2B, 2HN2, 2ONK, 2QJU, 2VPZ, 2ZY9, 3A5C, 3AQP, 3ARC, 3B4R, 3BVD, 3DIN, 3GIA, 3HYW, 3K5B, 3KBC, 3KDS, 3M9S, 3MP7, and 3ORG

Transmembrane region assignment

Transmembrane regions used for analysis were determined using the unpublished work of Seong-Kyu Han described here: Transmembrane regions were obtained from a consensus database constructed from the Protein Data Bank of Transmembrane Proteins (PDBTM) and Orientations of Proteins in Membranes (OPM) database, or using TMDET to identify transmembrane regions for those proteins not present in these databases [22–24].

Mesophile-thermophile structural pair assignment

Mesophile-thermophile structural pairs were identified by Seong-Kyu Han as has been described here: Structurally homologous pairs were identified using the structural alignment tool TM-align[25]. Aligned residues were defined by those that showed greater than 1Å root-mean-square deviations of Cα atoms (Cα RMSD) since these were considered areas of interesting change between proteins with similar function. Pairs that did not match transmembrane (or soluble) regions between proteins were excluded from packing and burial analysis.

Burial calculation

Burial data was obtained by calculating relative solvent accessibilities per residue using a program called surfarea produced using the Ezprot library by Frank Pettit (http://www.doe-mbi.ucla.edu/local/software/ezprot); burial values resulted from subtracting the accessibility from 100. Results were cross-checked with another popular program for burial calculations called naccess (http://www.bioinf.manchester.ac.uk/naccess/). Burial values were calculated separately for soluble and transmembrane regions of mesophiles and thermophiles belonging to the database, as well as for the mesophile-thermophile structural pairs. For the database, values were averaged over all mesophiles and thermophiles separately. In computing the average, burial values were weighted by the number of residues in a given protein to obtain a weighted burial average over all proteins; this ensured that burial values from smaller proteins did not unnecessarily bias the results calculated over all proteins. For the structural pairs, difference values between each mesophile-thermophile pair were averaged over all mesophile-thermophile pairs.

Packing calculation

Packing values were calculated using Rother *et al.*'s Voronoia program for computing packing densities[26]. Surface residues were excluded from the calculation since they are known to pack far less efficiently than buried residues[15]. A 0.20 Å grid distance and the ProtOr radii were chosen as parameters based on similar ones chosen by Hildebrand *et al.*[15]; the ProtOr radii was chosen since I was interested in an experimental rather than molecular dynamics application.

Packing values were calculated separately for soluble and transmembrane regions of mesophiles and thermophiles belonging to the database, as well as for the mesophile-thermophile structural pairs (as was done above for burial values). For the database, values were averaged over all mesophiles and thermophiles separately. In computing the average, packing values were weighted by the number of residues in a given protein to obtain a weighted packing average over all proteins. For the structural pairs, difference values between each mesophile-thermophile pair were averaged over all mesophile-thermophile pairs; if either packing value for a pair was zero, this pair’s data was excluded from the packing analysis.

References

1. Jaenicke R, Böhm G (1998 [cited 2012]) The stability of proteins in extreme environments. Current Opinion in Structural Biology 8:738–748.

2. Madigan MT, Martinko JM, Brock TD Brock biology of microorganisms. Upper Saddle River, NJ: Pearson Prentice Hall; 2006.

3. Vogt G, Woell S, Argos P (1997 [cited 2012]) Protein thermal stability, hydrogen bonds, and ion pairs. J. Mol. Biol. 269:631–643.

4. Vogt G, Argos P (1997 [cited 2012]) Protein thermal stability: hydrogen bonds or internal packing? Fold Des 2:S40–46.

5. Vetriani C, Maeder DL, Tolliday N, Yip KS, Stillman TJ, Britton KL, Rice DW, Klump HH, Robb FT (1998 [cited 2012]) Protein thermostability above 100 degreesC: a key role for ionic interactions. Proc. Natl. Acad. Sci. U.S.A. 95:12300–12305.

6. Karshikoff A, Ladenstein R (1998 [cited 2012]) Proteins from thermophilic and mesophilic organisms essentially do not differ in packing. Protein Engineering 11:867–872.

7. Thompson MJ, Eisenberg D (1999 [cited 2012]) Transproteomic evidence of a loop-deletion mechanism for enhancing protein thermostability. J. Mol. Biol. 290:595–604.

8. Szilágyi A, Závodszky P (2000 [cited 2012]) Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. Structure 8:493–504.

9. Berezovsky IN, Shakhnovich EI (2005 [cited 2012]) Physics and evolution of thermophilic adaptation. Proc. Natl. Acad. Sci. U.S.A. 102:12742–12747.

10. White SH (2004 [cited 2012]) The progress of membrane protein structure determination. Protein Sci. 13:1948–1949.

11. Yildirim MA, Goh K-I, Cusick ME, Barabasi A-L, Vidal M (2007 [cited 2012]) Drug[mdash]target network. Nat Biotech 25:1119–1126.

12. Janda I, Devedjiev Y, Cooper D, Chruszcz M, Derewenda U, Gabrys A, Minor W, Joachimiak A, Derewenda ZS (2004 [cited 2012]) Harvesting the high-hanging fruit: the structure of the YdeN gene product from Bacillus subtilis at 1.8 angstroms resolution. Acta Crystallogr. D Biol. Crystallogr. 60:1101–1107.

13. Eilers M, Shekar SC, Shieh T, Smith SO, Fleming PJ (2000 [cited 2012]) Internal packing of helical membrane proteins. Proc. Natl. Acad. Sci. U.S.A. 97:5796–5801.

14. Gerstein M, Chothia C (1999 [cited 2012]) Perspectives: signal transduction. Proteins in motion. Science 285:1682–1683.

15. Hildebrand PW, Rother K, Goede A, Preissner R, Frömmel C (2005 [cited 2012]) Molecular packing and packing defects in helical membrane proteins. Biophys. J. 88:1970–1977.

16. Faham S, Yang D, Bare E, Yohannan S, Whitelegge JP, Bowie JU (2004 [cited 2012]) Side-chain Contributions to Membrane Protein Structure and Stability. Journal of Molecular Biology 335:297–305.

17. Adamian L, Liang J (2001 [cited 2012]) Helix-helix packing and interfacial pairwise interactions of residues in membrane proteins. J. Mol. Biol. 311:891–907.

18. Joh NH, Oberai A, Yang D, Whitelegge JP, Bowie JU (2009 [cited 2012]) Similar energetic contributions of packing in the core of membrane and water-soluble proteins. J. Am. Chem. Soc. 131:10846–10847.

19. Craven KB, Zagotta WN (2004 [cited 2012]) Salt bridges and gating in the COOH-terminal region of HCN2 and CNGA1 channels. J. Gen. Physiol. 124:663–677.

20. Kim J-M, Altenbach C, Kono M, Oprian DD, Hubbell WL, Khorana HG (2004 [cited 2012]) Structural origins of constitutive activation in rhodopsin: Role of the K296/E113 salt bridge. Proc. Natl. Acad. Sci. U.S.A. 101:12508–12513.

21. Law CJ, Almqvist J, Bernstein A, Goetz RM, Huang Y, Soudant C, Laaksonen A, Hovmöller S, Wang D-N (2008 [cited 2012]) Salt-bridge dynamics control substrate-induced conformational change in the membrane transporter GlpT. J. Mol. Biol. 378:828–839.

22. Lomize MA, Lomize AL, Pogozheva ID, Mosberg HI (2006) OPM: orientations of proteins in membranes database. Bioinformatics 22:623–5.

23. Tusnády GE, Dosztányi Z, Simon I (2005 [cited 2012]) PDB_TM: selection and membrane localization of transmembrane proteins in the protein data bank. Nucleic Acids Res. 33:D275–278.

24. Tusnády GE, Dosztányi Z, Simon I (2005 [cited 2012]) TMDET: web server for detecting transmembrane regions of proteins by using their 3D coordinates. Bioinformatics 21:1276–1277.

25. Zhang Y, Skolnick J (2005 [cited 2012]) TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Res. 33:2302–2309.

26. Rother K, Hildebrand PW, Goede A, Gruening B, Preissner R (2009 [cited 2012]) Voronoia: analyzing packing in protein structures. Nucleic Acids Res. 37:D393–395.