

# Is the HIV-TAR Binding Protein, TRP-185, an RNA Methyltransferase?

Todd M. Smith

Geospiza, Inc. 2442 NW Market St., #344 Seattle, WA 98107

[www.geospiza.com](http://www.geospiza.com)

The wealth of data from genome-scale DNA sequencing projects has created a new opportunity in biology. Public sequence repositories allow unexpected discoveries to be made by comparing sequences from diverse organisms as illustrated in the following example.

A BLASTP (1) search of the non-redundant database of protein sequences (National Center for Biotechnology Information [NCBI]) with a ribosomal RNA methyltransferase protein sequence, encoded by the thiostrepton resistance gene from the bacterium *Streptomyces laurentii*, revealed similarities to a family of protein sequences that are conserved in all species of life. Interestingly, one of these sequences, encoding a human protein TRP-185 (2), was initially discovered because it binds to a stem loop structure in HIV-1 RNA known as TAR (3). A detailed sequence analysis of the TRP-185 protein sequence raises the possibility that it is a methyltransferase, and methylation of HIV RNA may be important in regulating its life cycle.

The strongest evidence for TRP-185 RNA methylation comes from the regions of homology that it shares with three proteins (thiostrepton resistance protein (TSR), SPOU, and PET56) whose enzymatic activities have been well characterized.

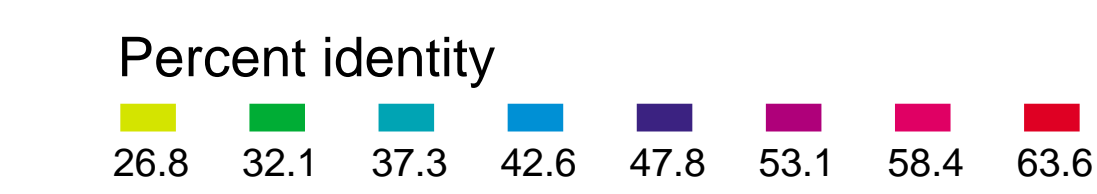
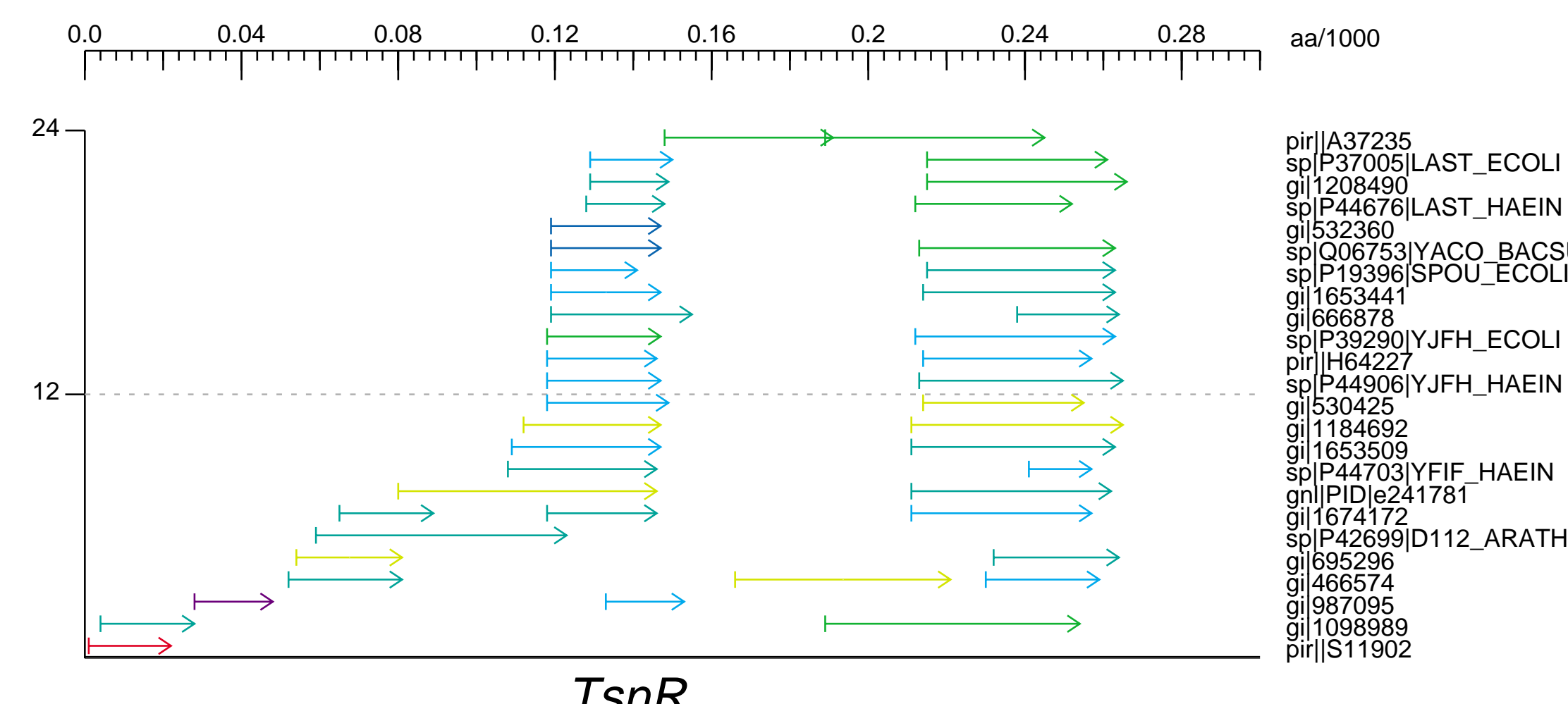
All three of these enzymes methylate RNA in a sequence dependent manner. The thiostrepton resistance protein (TSR) confers resistance to the antibiotic thiostrepton by methylating a 2'-OH of a specific adenosine on the 23S ribosomal RNA (4). In one case, a TSR from *Streptomyces azureus* (5), the methylation activity was shown to be highly dependent on the sequence of a stem loop structure in the 23S rRNA. Like TSR, TRP-185 binds to HIV RNA at a specific sequence in the stem loop structure of the TAR region (6). PET56 and SPOU methylate 2'-OHs of specific guanosine residues in RNA. PET56 methylates the 21 rRNA in yeast mitochondria (7, 8) and SPOU methylates *E. coli* transfer RNAs (9).

The homologous regions shared by TRP-185, TSR and PET56 were studied further by submitting those sequences and sequences from seven other organisms as a training set to the MEME motif detection program (10) (<http://www.sdsc.edu/MEME>). Three motifs identified by MEME were subsequently used to search the non-redundant protein database using the MAST program (10) (<http://www.sdsc.edu/MEME>) to identify additional protein sequences. This search revealed a common structure shared by at least 44 sequences from 23 organisms. All but four of these amino acid sequences are encoded by bacteria.

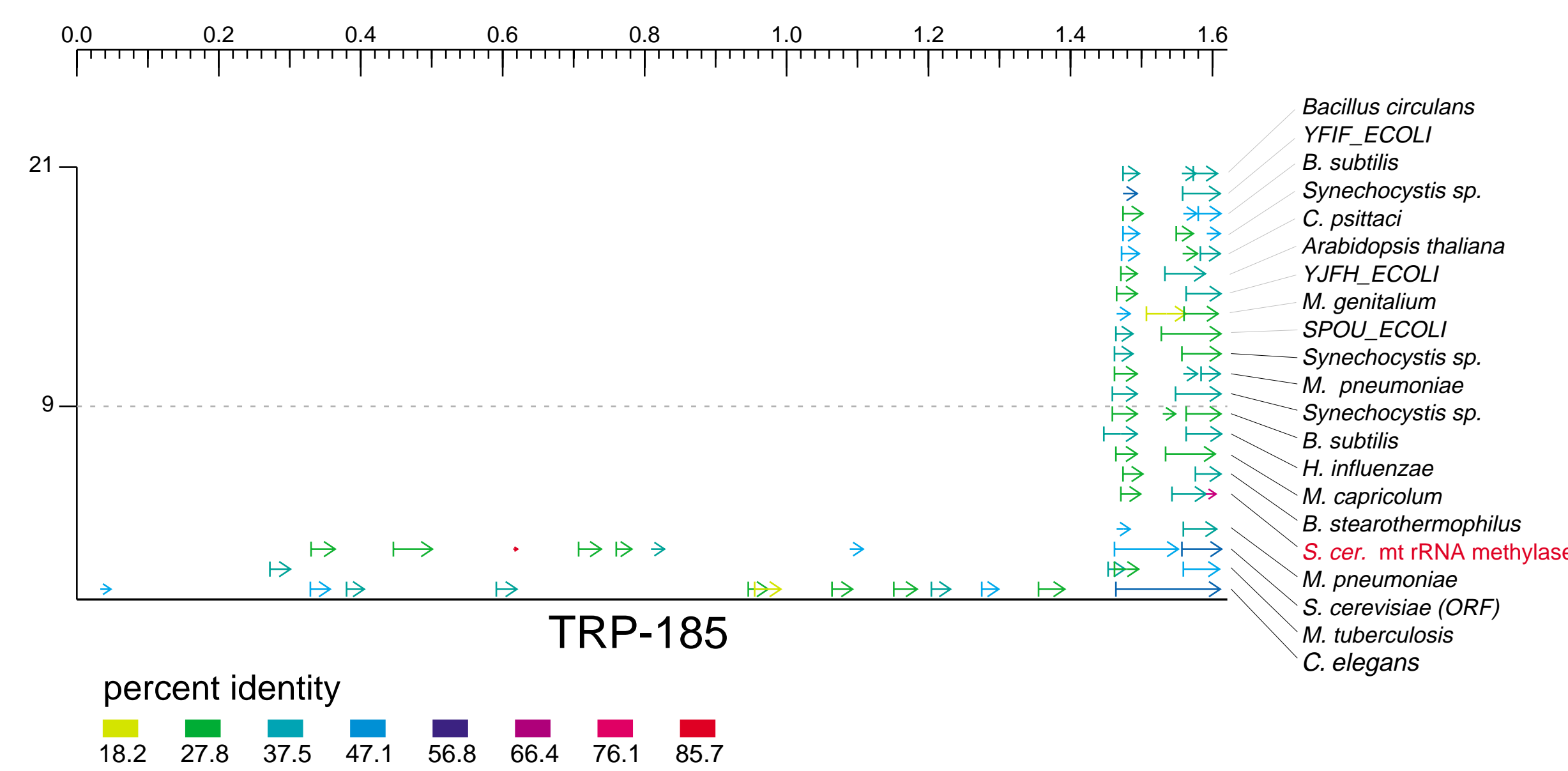
The discovery of a large family of putative and functionally characterized RNA methyltransferases raises a number of new and interesting questions. Previous analyses on a limited set of thiostrepton resistance proteins and other bacterial proteins suggested that a binding site for S-adenosyl-methionine (ado-met) is located within a conserved region (11). MEME places that site in the middle motif. When TRP-185 and PET56 are included in a multiple alignment of these proteins, it becomes apparent that TRP-185 contains the putative ado-met consensus sequence, raising the possibility that TRP-185 is able to methylate HIV-1 RNA in the TAR region. While no data exist to prove that TRP-185 is a methyltransferase, the features that it shares with TSR, PET56, and SPOU, such as the ability to bind specific RNA sequences and structures, combined with its high degree of local sequence similarity to these and other putative methyltransferases indicate that it is likely to have methyltransferase activity.

This analysis revealed a distinct family of methyltransferases ubiquitous to all kingdoms of life. Based on primary sequence information, this family of proteins can be divided into three subfamilies. One subfamily, which includes TRP-185, a *C. elegans* protein, and a yeast homologue, has only been identified in eukaryotes and is characterized by large (1200 - 1600 AA) polypeptides containing conserved C-terminal RNA stem-loop binding motifs and large (>1000 aa) diverse N-terminal regions. The second, and largest, subfamily is characterized by small (300 aa) polypeptides observed primarily in bacteria. Members of the third subfamily lack the first motif (Table I), but do retain the putative ado-met binding site. Given their abundance in nature it is highly likely that all of these proteins play important roles in gene expression through RNA secondary structure recognition and modification.

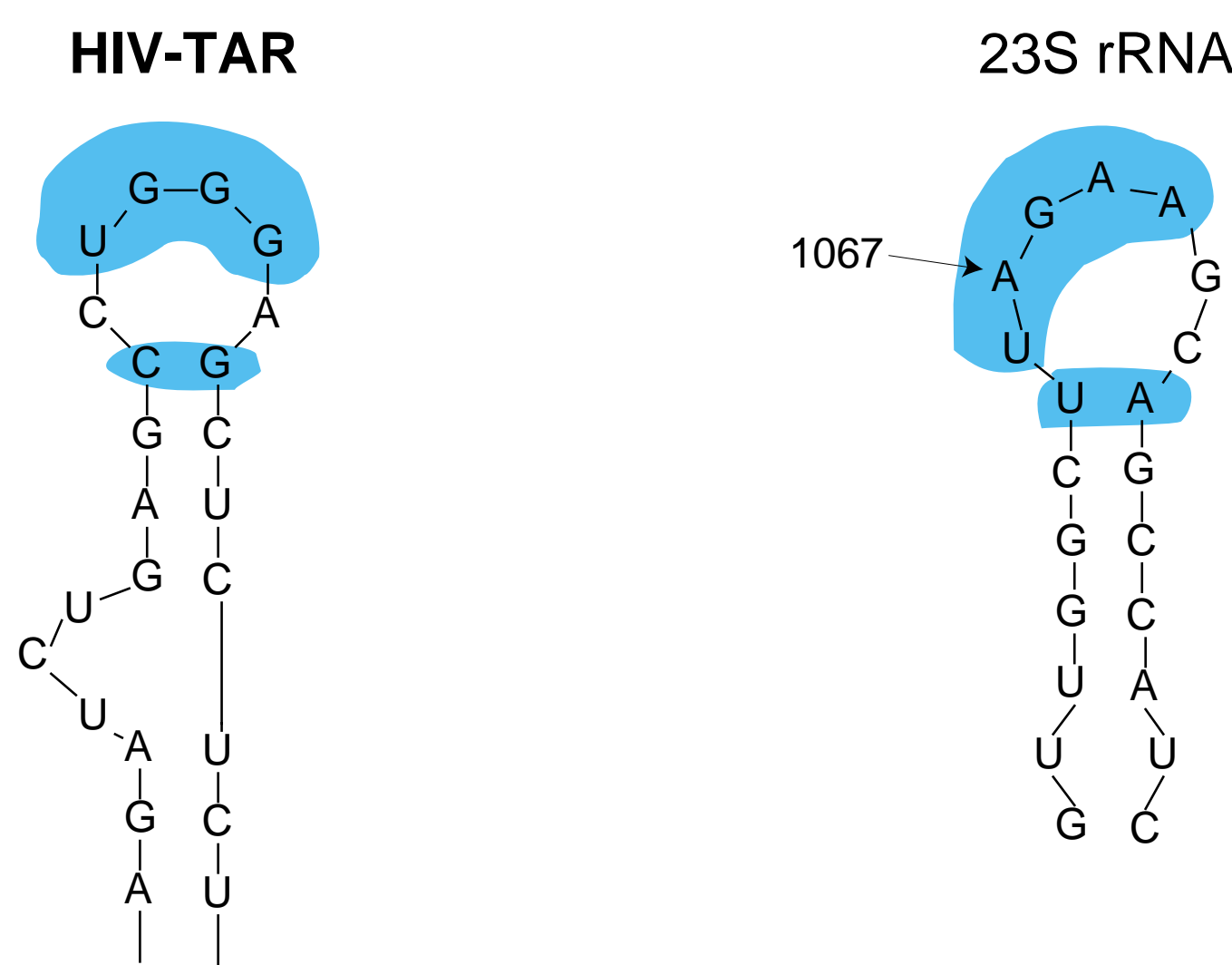
## Protein sequences matching *S. laurentii* TsnR



## Protein sequences matching TRP-185



## Partial structures of the target RNA molecules



Partial structures of the RNA stem loop regions that TRP-185 (HIV-TAR) and TSR (23SRNA) bind to. The shaded regions denote the areas in which nucleotide substitutions decrease (or abolish) binding affinity.

## Motifs detected by MEME

name	Motif 1	Motif 3	Motif 2
Bsu,	PHNLGSMRRTA	LVIQSEKGM	LPMAGKVTSLNASVAAGLLMYE
Cel,	PNNLGGICRTS	YIFGAESKGL	IKQVGHTRSLNVHVTALMIK
Eco,	PHNVSAIIRTA	ILMGQEKTI	IPMIGMVQSLNVSVASALILYE
Hin,	PHNLGACLRRTA	LVMGAEQGM	IPMAGSVSSLNVSVATGVCLF
Mpn,	PHNFGAILRTC	LLVGNEDKGL	IPMNPKLNSLNVSVAVGILFG
PET56,	PHNIGAIIRSA	LVVGNESQGV	RAPEPIVDSLNVSVATALLIDN
Sce,	PPNLGGICRCL	ILLGTEAFGI	IQQFGVIRSMNIQTATAVIVHS
Syn,	PGNLGTILRRTA	VIFGSEGGQL	IPQAPQVESLNVAVGVMLYE
TRP-185,	PTNLGGLCRCT	LLLGNEREGL	IPQQGIRSLNVHVSAGALLIWE
TSR,	VGNIGAIVRTS	LLFGSEKGGP	IPMMSQTESLNVSVSLGIALHE
Multilevel consensus sequence	PHNLGAILRRTA G C	LVLGxExEGL LM I F ***** ado-met	IPMAGxVRSLSLNVSVAVGLLILYE Q T STAILI V

## (Top) Alignments of the motifs identified by MEME.

Ten amino acid sequences were submitted to the MEME server as a training set to detect conserved motifs. For clarity only the alignments of the motif regions are shown. The numbers preceding Motif 1, between the motifs and following Motif 2 are the number of amino acids. Below the motif alignments are consensus sequences predicted by MEME. The putative ado-met site previously predicted (11) is shown below Motif 3. Abbreviations: Bsu, *Bacillus subtilis*; Cel, *C. elegans*; Eco, *Escherichia coli*; Hin, *Haemophilus influenzae*; Mpn, *Mycoplasma pneumoniae*; PET56, *S. cerevisiae* mitochondrial rRNA methyltransferase; Sce, *S. cerevisiae*; Syn, *Synechocystis* sp.; TRP-185, *Homo sapiens* HIV TAR binding protein; TSR, *Streptomyces laurentii* thiostrepton resistance protein.

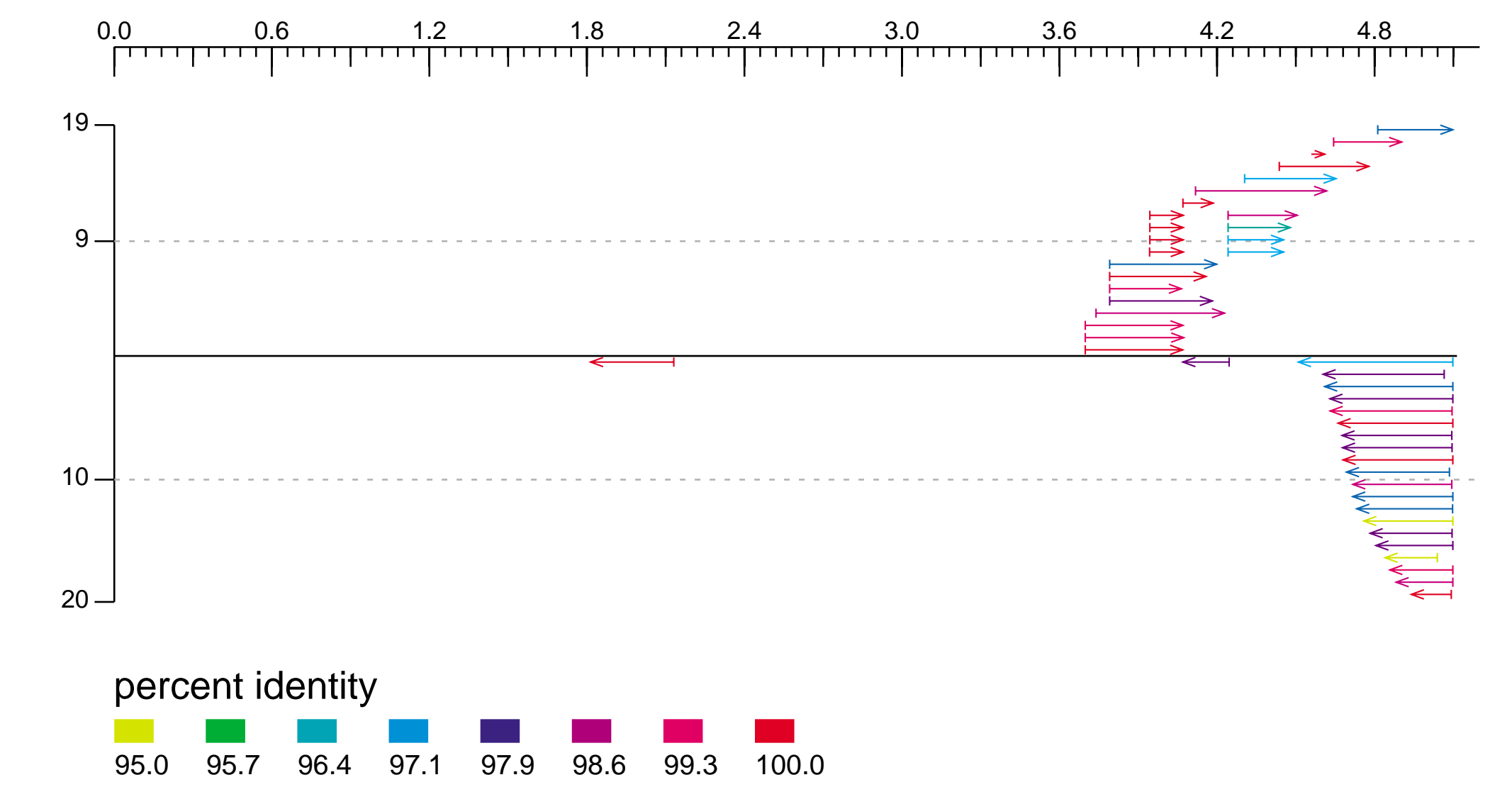
## (Right) Motif diagrams, determined by MAST, for the putative methyltransferases.

Sequence ID gives the database and accession number (database|accession number) for each sequence. E-values are determined by the MAST program. The motif diagram gives structural information about each protein. The numbers with in brackets are the motif number determined from the training set in MEME, and the other numbers are the number of amino acids preceding or following each motif. Where known, the gene/protein name is given in parenthesis (red). Abbreviations: A, *Arabidopsis*; B, *Bacillus*; C, *Caenorhabditis*; Ch, *Chlamydia*; E, *Escherichia*; H, *Haemophilus*; Ho, *Homo*; He, *Helicobacter*; M, *Mycoplasma*; Me, *Methanococcus*; My, *Mycobacterium*; S, *Streptomyces*; Sa, *Saccharomyces*.

## Sequences identified by MAST

Organism	E-Value	Motif Diagram
<i>A. thaliana</i>	6.5e-08	91-[1]-86-[3]-20-[2]-31
<i>A. thaliana</i>	0.0026	87-[3]-14-[2]-77
<i>B. circulans</i>	3.6e-12	12-[1]-74-[3]-14-[2]-01
<i>B. stearothermophilus</i>	1.6e-09	12-[1]-77-[3]-14-[2]-11
<i>B. subtilis</i>	1.7e-23	107-[1]-76-[3]-13-[2]-10
<i>B. subtilis</i>	1.1e-13	114-[1]-75-[3]-13-[2]-03
<i>B. subtilis</i>	1.1e-10	12-[1]-77-[3]-14-[2]-11
<i>B. subtilis</i>	2.1e-07	57-[3]-14-[2]-11
<i>C. elegans</i>	1.3e-18	1049-[1]-78-[3]-13-[2]-10
<i>Ch. psittaci</i>	5.7e-13	10-[1]-74-[3]-14-[2]-15
<i>Ch. trachomatis</i>	1.9e-11	63-[3]-14-[2]-10
<i>E. coli</i>	3.3e-26	105-[1]-76-[3]-13-[2]-06
<i>E. coli</i>	1.9e-11	207-[1]-75-[3]-13-[2]-07
<i>E. coli</i>	3.1e-09	11-[1]-74-[3]-15-[2]-14
<i>E. coli</i>	6.6e-08	235-[1]-75-[3]-13-[2]-79
<i>E. coli</i>	0.081	109-[3]-13-[2]-74
<i>E. coli</i> (SpoU)	4.9e-25	29-[1]-75-[3]-13-[2]-69
<i>H. influenzae</i>	1.1e-07	11-[1]-74-[3]-15-[2]-17
<i>H. influenzae</i>	0.002	107-[3]-13-[2]-89
<i>H. influenzae</i>	3.0e-27	105-[1]-77-[3]-13-[2]-08
<i>Ho. sapiens</i> (TRP-185)	3.8e-20	1473-[1]-78-[3]-13-[2]-14
<i>He. pylori</i>	7.8e-10	95-[1]-74-[3]-13-[2]-02
<i>M. capricolum</i>	1.5e-07	81-[1]-77-[3]-13-[2]-09
<i>M. genitalium</i>	1.1e-17	105-[1]-78-[3]-13-[2]-03
<i>M. genitalium</i>	4.0e-08	15-[1]-79-[3]-14-[2]-15
<i>M. pneumoniae</i>	2.4e-23	105-[1]-78-[3]-13-[2]-03
<i>M. pneumoniae</i>	2.9e-08	15-[1]-79-[3]-14-[2]-15
<i>M. tuberculosis</i>	5.2e-14	170-[1]-76-[3]-13-[2]-20
<i>Me. jannaschii</i>	7.6e-05	14-[1]-86-[3]-13-[2]-74
<i>My. leprae</i>	6.7e-09	26-[1]-74-[3]-15-[2]-11
<i>My. tuberculosis</i>	1.2e-10	140-[1]-80-[3]-13-[2]-12
<i>My. tuberculosis</i>	5.1e-10	118-[1]-78-[3]-13-[2]-08
<i>S. laurentii</i> (TsnR)	3.9e-14	126-[1]-77-[3]-13-[2]-11
<i>S. actuosus</i> (NshR)	5.5e-18	126-[1]-77-[3]-13-[2]-15
<i>S. aureus</i> (TSR)	1.0e-18	104-[1]-77-[3]-13-[2]-10
<i>Sa. cerevisiae</i>	5.6e-14	1295-[1]-79-[3]-13-[2]-06
<i>Sa. cerevisiae</i> mt (PET56)	1.0e-16	251-[1]-90-[3]-23-[2]-05
<i>Synechocystis</i> sp.	1.1e-07	11-[1]-76-[3]-14-[2]-09
<i>Synechocystis</i> sp.	0.0013	5-[1]-84-[3]-13-[2]-95
<i>Synechocystis</i> sp.1	1.9e-22	130-[1]-76-[3]-13-[2]-12
<i>Synechocystis</i> sp.1	9.0e-20	227-[1]-76-[3]-13-[2]-25
unspecified	0.0003	14-[1]-85-[3]-13-[2]-27
unspecified	0.0074	14-[1]-85-[3]-13-[2]-91
unspecified	0.86	76-[3]-13-[2]-75

## ESTs matching TRP-185 mRNA



The identification of TRP-185 as a member of a family of methyltransferases has immediate implications. For example, inhibitors of TRP-185 could provide new treatments for HIV infections. Additionally, the bacterial homologues, which are small relative to TRP-185, may provide useful models for understanding TRP-185 activity. The question remains however, as to the endogenous function of TRP-185; does it bind and methylate RNA molecules in uninfected cells? A BLASTN search of the EST (expressed sequence tag) database (NCBI) with the TRP-185 cDNA sequence (GenBank accession U38847) identified 61 partial human cDNA sequences from a wide variety of tissues that are highly similar (more than 95% identity) to TRP-185. This indicates at least a modest level of expression of the TRP-185 gene in most tissues.

Finally, these observations demonstrate the necessity of biochemical data for interpreting sequence information since the detailed characterization of the *S. aureus* methyltransferase, PET56, and TRP-185 gave the only clues as to the function of this important family of proteins.

## Summary

- HIV TAR - stem loop region of the viral RNA involved in gene regulation
- TRP-185 - ubiquitous cellular protein that binds to TAR
- Carboxy-terminal end is homologous to numerous bacterial proteins, some of which are known to be rRNA methyltransferases.
- Biochemical recognition of the stem loop region is analogous to the *Streptomyces azureus* *tsr* gene product.
- Data indicate TRP-185 shares a domain with a family of proteins that play important roles in RNA metabolism and function.
- Methylation of the HIV TAR region is possible and inhibition of this reaction could lead to novel therapies.

## References

- Altschul, S., Gish, W., Miller, W., Myers, E. & Lipman, D. J. Mol. Biol. 215, 403-410 (1990).
- Smith, T.M., Jiang, Y.-F. & Floss, H.G. in Biotechnology of Industrial Antibiotics, Second Edition (ed. Strohl, W.R.) (Marcel Dekker, New York, 1997).
- Wu, F., Garcia, J., Sigman, D. & Gaynor, R. Genes Dev 5, 2128-40 (1991).
- Cundliffe, E. in Mechanism of Action of Antibacterial Agents (ed. Hahn, E.F.) 329-42 (Springer-Verlag, New York, 1979).
- Bechthold, A. & Floss, H.G. Eur J Biochem 224, 431-7 (1994).
- Wu-Baer, F., Lane, W.S. & Gaynor, R.B. Embo J 14, 5995-6009 (1995).
- Sirum-Connolly, K. & Mason, T.L. Science 262, 1886-9 (1993).
- Sirum-Connolly, K., Peltier, J.M., Crain, P.F., McCloskey, J.A. & Mason, T.L. Biochimie 77, 30-9 (1995).
- Persson, B.C., Jager, G. & Gustafsson, C. Nucleic Acids Res. 25, 4093-7 (1997).
- Bailey, T.L. & Elkan, C. in Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology 28-36 (AAAI Press, Menlo Park, CA, 1994).