



Title	Crystal structure of the putative RNA methyltransferase PH1948 from <i>Pyrococcus horikoshii</i> , in complex with the copurified S-adenosyl-L-homocysteine
Author(s)	Gao, Yong-Gui; Yao, Min; 姚, 閔 et al.
Citation	Proteins: Structure, Function, and Bioinformatics, 61(4), 1141-1145 https://doi.org/10.1002/prot.20678
Issue Date	2005-10-21
Doc URL	https://hdl.handle.net/2115/895
Rights	Copyright © 2005 Wiley-Liss, Inc.
Type	journal article
File Information	PSFB61(4).pdf



Crystal structure of the putative RNA methyltransferase PH1948 from *Pyrococcus horikoshii*, in complex with the co-purified S-adenosyl-L-homocysteine

Yong-Gui Gao^{1,3}, Min Yao^{1,2*}, Zhou Yong¹, and Isao Tanaka¹

¹Division of Biological Sciences, Graduate School of Science, Hokkaido University, Sapporo 060-0810, Japan

²RIKEN Harima Institute/SPRING-8, Hyogo 679-5148, Japan

³School of Life Sciences, Zhejiang University, Hangzhou 310029, China

Introduction. Methyltransferases (MTases) constitute a large family of enzymes that methylate the carbon, oxygen, or nitrogen atoms of their substrates, including DNA, RNA, proteins, phospholipids, and small molecules, thereby exerting important functions^{1,2}. These functions include biosynthesis, detoxification, signal transduction, protein sorting and repair, chromatin regulation, and gene silencing³. The structures of more than fifty MTases have been solved by X-ray crystallography and NMR. Of these structures, seven belong to the RNA MTases⁴ that act on RNA, and thus are directly related to gene transcription and expression. For example, methylation of the 5'-terminal cap plays an important role in mRNA export from the nucleus, efficient translation, and protection of the integrity of mRNAs⁵.

The classical biochemical approach to identify enzymes is the fractionation of cell extracts and purification of enzymes based on activity. This approach is clearly limited to enzymes with rather high specific activity and known substrates. For RNA

Grant sponsor: National Project on Protein Structural and Functional Analyses from the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

* To whom correspondence should be addressed, E-mail: yao@castor.sci.hokudai.ac.jp

MTases, the lack of appropriate substrates, instability, or low levels of activity of the enzymes, and possible redundancy of their functions *in vivo* have hampered their identification and functional characterization. Fourteen methylations of 23S rRNA are known in *E.coli*, but few of the methylating activities have been described to the level of the partial purification, and to date only one has been well characterized⁶. Structural analysis of RNA MTases either free or in a complex with the target RNA will assist in identifying the biochemical and cellular functions of the unknown RNA MTases³.

Pyrococcus horikoshi protein PH1948 is a putative RNA MTase. It was classified as belonging to Cluster of Orthologous Groups 2263 (COGs2263), the 15 members of which are found only in Archaea and cover all the Archaea species in the COG database. An NCBI conserved domain search indicated that PH1948 contains a highly conserved domain found in large sets of MTases, such as cytosine-C5 DNA MTases, ribosomal RNA adenine dimethylases, ribosomal protein L11 MTases, O-MTases, *etc.* The substrate and methyl donor for the methylation of COGs2263 members have not been characterized. Here, we report the crystal structure of PH1948 in complex with the co-purified binding factor, S-adenosyl-L-homocysteine (SAH), at 2.2 Å resolution. In addition, a structural comparison with RNA MTases ErmC' is also discussed.

Materials and Methods. Recombinant PH1948 protein was expressed in *Escherichia coli* strain B834 (DE3) by addition of 1 mM IPTG to LB broth (310 K) at OD₆₀₀ of about 0.6. After induction for 5 h, the cells were harvested and disrupted with a French Press. Highly pure PH1948 protein was prepared in 3 steps: the target protein

was captured with a HiTrap SP column (Amersham Biosciences Inc., Piscataway, NJ), then purified further with HiLoad 26/60 Superdex 75pg (Amersham Biosciences) chromatography, and finally passed through a Resource S column (Amersham Biosciences). The target peak of the last chromatography was collected and dialyzed overnight against 10 mM Tris-HCl buffer (pH 9.0), then concentrated to 5 mg ml⁻¹. PH1948 protein was confirmed by MALDI-TOF mass spectrometry. For production of Se-Met-substituted PH1948 (Se-Met PH1948), cells were cultured in minimal medium containing Se-Met. The procedure for purification of Se-Met PH1948 was the same as that of the native protein.

The crystallization conditions for PH1948 consisted of 0.1 M HEPES-Na, pH 7.2, 28% PEG400 and 0.1 M CaCl₂. Crystals of Se-Met PH1948 with dimensions of 0.35×0.25×0.1 mm were obtained under the same conditions at 20°C within 2 weeks. Two diffraction datasets (SAD data and refinement data) were collected under 100 K at BL44B2 of SPring-8 (Hyogo, Japan) and BL5A of PF (Tsukuba, Japan), respectively. All data were processed using the program HKL2000⁷. The structure was determined at a resolution of 2.7 Å by the selenomethionyl single-wavelength anomalous diffraction (SAD) method⁸. The sites of 8 Se atoms were found, and used for phase calculation by SOLVE⁹. After phase improvement with NCS (non-crystallographic symmetry)-averaging by DM¹⁰, the initial model was built automatically to 85.5% by ARP/wARP¹¹. The refinement at 2.2 Å resolution was carried out semi-automatically by LAFIRE¹² with CNS¹³, and NCS-restraint was applied for the main chain throughout refinement.

Table 1. Data collection and refinement statistics

Crystal	Se-Met	Native
Beamline	BL44B2 (SPring-8)	NW12 (PF)
Wavelength (Å)	0.9793	1.00
Space group	C2	C2
Unit cell parameters (Å, °)	a=205.5, b=43.3, c=118.5 β=92.2	a=207.0, b=43.1, c=118.2 β=92.1
Resolution (Å)	39.86-2.7(2.8-2.7)	39.86-2.20 (2.28-2.20)
Unique reflections	28286	53514
Completeness (%)	96.9 (78.4)	99.3 (94.9)
Redundancy	6.5 (5.1)	5.0 (4.1)
I/σ (I)	13.1 (2.5)	25.2 (5.6)
R _{merge} ^a (%)	8.6 (34.9)	4.6 (21.8)
Refinement		
R/R-free (%)		24.2/26.8
RMSD bond length/angle (Å/°)		0.011/1.4
B factor (Å ²)		
protein molecules		45.1
water molecules		37.6
others		44.4
Ramachandran plot (%)		
most favored regions		91.2
additionally allowed regions		8.7
generously allowed regions		0.1
disallowed regions		0

Values in parentheses refer to the highest resolution shell.

^a $R_{merge} = \frac{\sum_h \sum_j | \langle I \rangle_h - I_{hj} |}{\sum_h \sum_j I_{hj}}$, where $\langle I \rangle_h$ is the mean intensity of symmetry – equivalent reflections.

Results and discussion. The final model of PH1948 includes 799 of 828 residues constituting four protein molecules (one molecule contains 207 residues), four cofactors, and 312 water molecules in an asymmetric unit (PDB code, 1WY7). Due to poor electron density, N-terminal residues 1-3, loop residues 181-185, and C-terminal residues 205-207 of molecule A, residues 1, loop residues 183-187, and C-terminal residues 205-207 of molecule B, loop residues 181-186 of molecule C, and C-terminal residues 205-207 of molecule D could not be built. Except for the N-, C-terminus (1-3,

205-207 residues) and flexible loop (179-190), the structures of four copies in an asymmetric unit are very similar with averaging root means square diversity of 0.105 Å.

Fig. 1a shows a monomer (Molecule D) of the protein PH1948-cofactor complex. PH1948 has a single domain composed of seven α -helices and seven β -strands (a MTase fold and an N-terminal α -helix α N). The seven-stranded β -sheet is assembled in the order β 3- β 2- β 1- β 4- β 5- β 7- β 6 with β 7 anti-parallel to the other six parallel strands. With the β -strands flanked by three α -helices on one side (α Z, α A, α B) and three α -helices on the other (α C- α E), the core structure is formed with the typical features of the SAM-dependent MTase fold².

As the concentrations of the predominant methyl donor, S-adenosyl-L-methionine (SAM), in *E.coli* were reported to be 300-500 μ M, co-purification and crystallization of bound SAM or SAH have been found in some MTases^{2,14,15}. In the present structure of PH1948, the electron density map of $F_o - F_c$ indicated clearly that each of the four molecules in the asymmetric unit contained one SAH molecule, although no cofactors were added for purification and crystallization (Fig. 1b). The presence of SAH was also confirmed by refining PH1948 with SAH and SAM, respectively. The bound SAH was presumably derived from methyl transfer of SAM during the growth of the *E. coli* cells, and the MTase PH1948 must have retained the cofactor throughout the experimental procedures.

The interactions of PH1948 with SAH are shown in Fig. 2a. Motif 2, a glycine-rich region between β 1 and α A, Motif 3, an acid loop between β 2 and α B with the

sequence EVD, and Motif 4 (DV) prior to α C interact directly with SAH in a mode common to the SAM-dependent MTases^{2,3} (common motifs I-III in MTases). Motif 1, located in the loop between α N and α Z, is highly conserved in the members of COGs2263. The O atom of Glu 26 and NE of Gln 27 interact with sulfur and the O2' hydroxyl of SAH. Motif 5, with the highly conserved sequence NPPF following the strand β 4, is the catalytic site for members of the PH1948 family. Apart from Pro121 making van der Waals contact with the adenine ring, there is no direct interaction between Motif 5 and SAH. Remarkably, the binding of SAH was enhanced by van der Waals contacts with Leu25, Val80, Val106, and Phe133. Especially, the ring of the conserved residue, Phe133, is stacked to the adenine ring with a distance of about 3.4 Å. Phe18, Val62, and Leu63 are also involved in formation of the binding pocket. With these interactions, SAH is held deeply within the binding pocket of PH1948, which results in co-purification of the complex of PH1948 and SAH.

The results of DALI¹⁶ searches indicated that the 3D structure of PH1948 is very similar to the putative RNA MTase, Rv2118c (PDB¹⁷: 1I9G), the putative RNA (guanine-N2) MTase, Mj0882 (PDB¹⁸: 1DUS), and the rRNA (adenine-N6) MTase ErmC' (PDB¹⁹: 1QAN) with Z-scores of 18.4, 17.5, and 14.7, respectively. Among these RNA MTases, ErmC' is relatively well characterized although the structure with bound substrate has not been determined^{19,20}. Fig. 2b shows that the proteins both bind SAH in similar pockets, and the active site is superposed well with the consensus N[P/I]P[Y/F] except for the orientations of the aromatic rings of Phe122 and Tyr104. The aromatic ring can be rotated easily to make a face-to-face π -stacking interaction

with the target base upon substrate binding¹⁸. The motif [D/N/S][P/I]P[Y/F] is a very common mode of substrate binding employed by N-MTases and is not nucleotide-specific but is used for nitrogens conjugated to a planar system, *e.g.*, an amide moiety after nucleophilic attack⁴. In the active pocket of ErmC', the loop (158-171) connecting $\beta 6$ and $\beta 7$ plays an important role in binding and orienting the target adenine, and the N terminal loop (10-12) is involved in binding the cofactor and fitting the orientation of the cofactor and target base^{19,20}. In the case of PH1948, similar residues are present at the corresponding positions (residues 179-191 and 26-28; Fig. 2, bottom). The two loops of PH1948 may employ the same function as that of ErmC'. Especially, residues 26-28 are completely conserved among PH1948 family members, suggesting that this region is likely related to the substrate specificity, and the side chain of Tyr28 would experience a conformational change to fit the target upon binding. ErmC' has an additional C-terminal domain to bind RNA¹⁹. Unlike ErmC', the long anti-parallel $\beta 6$ and $\beta 7$ sheets in PH1948 protrude out the seven-stranded sheet, and the large loop 179-191 is twisted with the top deviating away from the loop 14-32 resulting in the formation of an active crevice. Combined with the N-terminal helix αN and αZ , a groove filled with positively charged residues is formed, which is predicted to bind RNA (electrostatic potential surface not shown). Upon RNA binding, the flexible and conserved loop 179-191 of PH1948 may undergo significant motion to bind and orient the target RNA base, similar to the motion of ErmC' loop 158-171²⁰. Our structural findings and comparison with the structure of rRNA (adenine-N6) MTase ErmC' demonstrated that PH1948 is a SAM-dependent

nitrogen-MTase, with a positively charged groove in its N-terminus as the binding site for the most likely RNA substrate.

Acknowledgments. We thank Ms. A. Morita and R. Ogawa for help in protein expression and purification. We also thank the staffs of beamline BL44B2, SPring-8 and beamline NW12, Photon Factory, Japan, for their kindly help with data collection.

REFERENCES

1. Armengaud J, Urbonavicius J, Fernandez B, Chaussinand G, Bujnicki JM, Grosjean H. N2-methylation of guanosine at position 10 in tRNA is catalyzed by a THUMP domain-containing, S-adenosylmethionine-dependent methyltransferase, conserved in Archaea and Eukaryota. *J Biol Chem* 2004; 279: 37142-37152.
2. Cheng XD, Roberts RJ. AdoMet-dependent methylation, DNA methyltransferases and base flipping. *Nucleic Acids Res* 2001; 29: 3784-3795.
3. Martin JL, McMillan FM. SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. *Curr Opin Struc Biol* 2002; 12, 783-793.
4. Schubert HL, Blumenthal RM, Cheng XD, Many paths to methyltransfer: a chronicle of convergence. *Trends Biochem Scie* 2003; 28: 329-335.
5. Chiang, P. K., Gordon, R. K., Tal, J., Zeng, G. C., Doctor, B. P., Pardhasaradhi, K. & McCann, P. P. S-Adenosylmethionine and methylation. *FASEB J.* 1996; 10, 471-480.
6. Bügl, H., Fauman, E. B., Staker, B. L., Zheng, F., Kushner, S. R., Saper, M. A., Bardwell, J. C. A. & Jakob, U. RNA methylation under heat shock control. *Molecular Cell.* 2002; 6: 349-360
7. Otwinowski, Z., Minor, W. Processing of X-ray diffraction data collected in

- oscillation mode. *Methods Enzymol.* 1997; 276: 307-326
8. Hendrickson WA. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* 1991; 254: 51-58.
 9. Terwilliger TC, Berendzen J. Automated MAD and MIR structure solution. *Acta Crystallogr D Biol Crystallogr* 1999; 55: 849-861.
 10. Cowtan KD, Main P. Phase combination and cross validation in iterated density-modification calculations. *Acta Cryst D Biol Crystallogr.* 1996; 52: 43-48.
 11. Lamzin VS, Wilson KS. Automated refinement of protein models. *Acta Cryst D Biol Crystallogr* 1993; 49: 129-147.
 12. Yao M, Zhou Y, Tanaka I. LAFIRE: a building and fitting for automatic refinement. Manuscript in preparation.
 13. Brünger AT, Adams PD, Clore GM, Delano WL, Gros P, Grosse-Kunstleve RW, Jiang J-S, Kuszewski J, Nilges M, Pannas NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Cryst D Biol Crystallogr* 1998; 54: 905-921
 14. Miller DJ, Ouellette N, Evdokimova E, Savchenko A, Edwards A, Anderson WF. Crystal complexes of a predicted S-adenosylmethionine-dependent methyltransferase reveal a typical AdoMet binding domain and a substrate recognition domain. *Protein Sci* 2003; 12:1432-1442.
 15. Tanaka Y, Tsumoto K, Yasutake Y, Umetsu M, Yao M, Fukada H, Tanaka I, Kumagai I. How oligomerization contributes to the thermostability of an archaeon protein. Protein L-isoaspartyl-O-methyltransferase from *Sulfolobus tokodaii*. *J Biol Chem*, 2004; 279 (31): 32957-32967.
 16. Holm L, Sander C. Touring protein fold space with Dali/FSSP. *Nucleic Acids Res*

- 1998; 26: 316-319.
17. Gupta A, Kumar PH, Dineshkumar TK, Varshney U, Subramanya HS. Crystal structure of Rv2118c: an AdoMet-dependent methyltransferase from *Mycobacterium tuberculosis* H37Rv. *J Mol Biol.* 2001; 312(2): 381-391.
 18. Bujnicki JM, Rychlewski L, RNA: (guanine-N2) methyltransferases RsmC/RsmD and their homologs revisited –bioinformatics analysis and prediction of the active site based on the uncharacterized Mj0882protein structure. *BMC Bioinformatics* 2002; 3: 10-21.
 19. Schluckebier G, Zhong P, Stewart KD, Kavanaugh TJ, Abad-Zapatero C. The 2.2 Å structure of the rRNA methyltransferase ErmC' and its complexes with cofactor and cofactor analogs: implications for the reaction mechanism. *J Mol Biol.* 1999; 289(2):277-91.
 20. Maravic G, Feder M, Pongor S, Flogel M, Bujnicki JM, Mutational analysis defines the roles of conserved amino acid residues in the predicted catalytic pocket of the rRNA:m6A methyltransferase ErmC'. *J Mol Biol.* 2003; 332(1):99-109.

Figure captions

Fig. 1 PH1948 in complex with SAH and the omit map of SAH.

(a) The protein is shown as a ribbon representation, and SAH as a stick model. The secondary structural elements are labeled $\beta 1$ - $\beta 7$, αA - αE and αZ , consistent with standard nomenclature among MTases except for an additional helix, αN , in the N-terminus. (b) A ball-and-stick model of SAH superimposed with the omit map. The map was calculated with coefficients $F_o - F_c$ contoured at 2.5σ calculated without SAH.

Fig. 2 Protein PH1948-SAH complex and structural comparison with ErmC'.

(a) Stick model showing the interaction of PH1948 and SAH. SAH is shown in green and the important protein residues in gray labeled with the single-letter abbreviations. Oxygen atoms and water are shown in red, nitrogen atoms in blue, and sulfur atoms in brown. The interactions among protein, SAH, and water are indicated by dashed lines. (b) Structural comparison of PH1948 with the N-terminal catalytic domain of ErmC'. The two structures are shown as ribbons, with the similar fold in gray, and varying parts in red (PH1948) and blue (ErmC'). The residues in the catalytic site are shown as sticks, and the cofactor SAH as lines. The sequences of the peptide segments possibly involved in substrate binding are compared; *, :, and . indicate identity, strong similarity, and weak similarity, respectively.

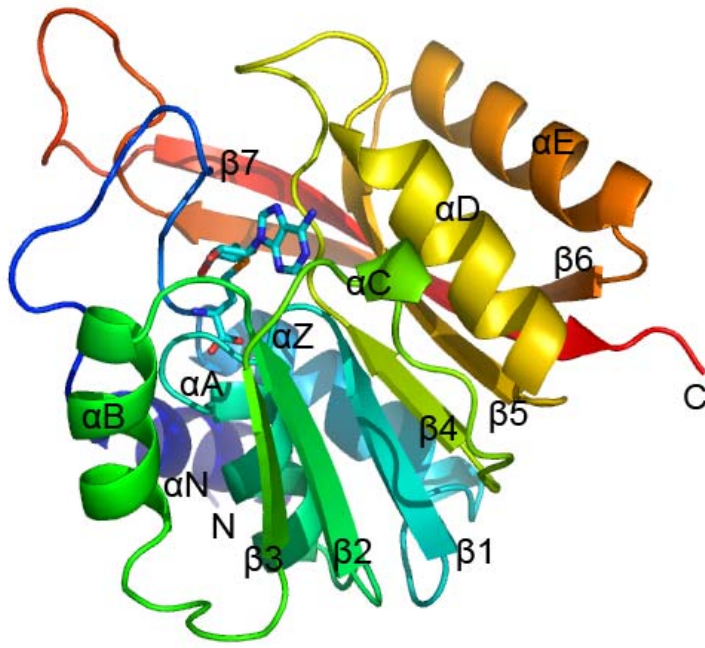


Fig. 1a

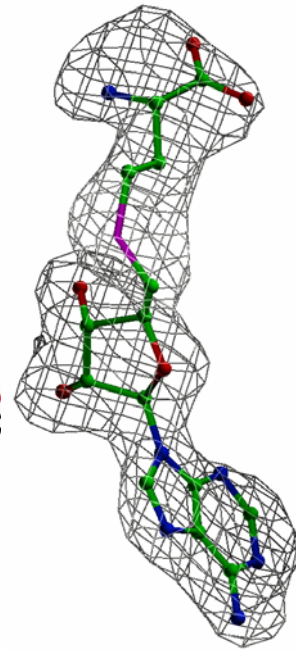


Fig. 1b

Fig.2a

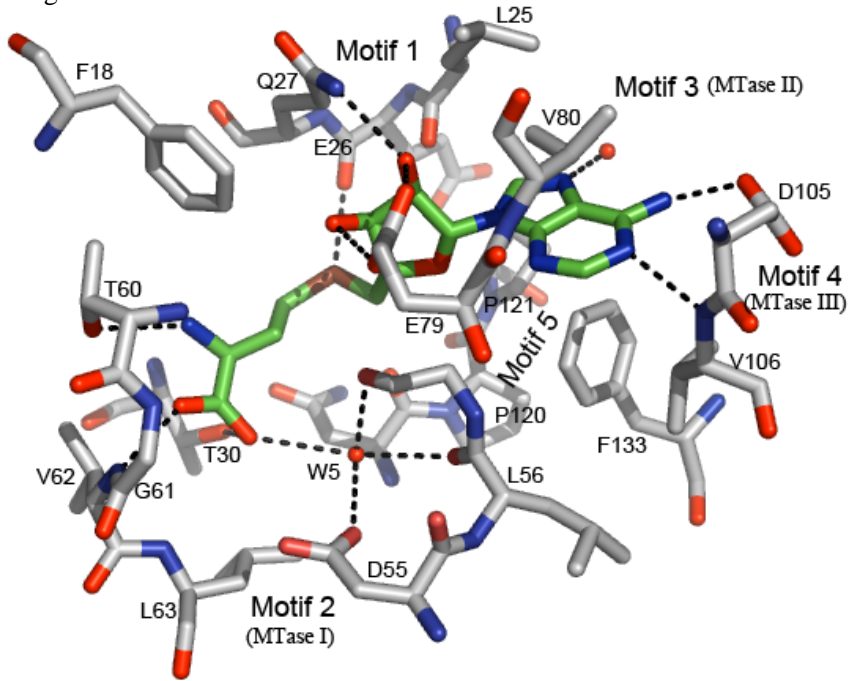
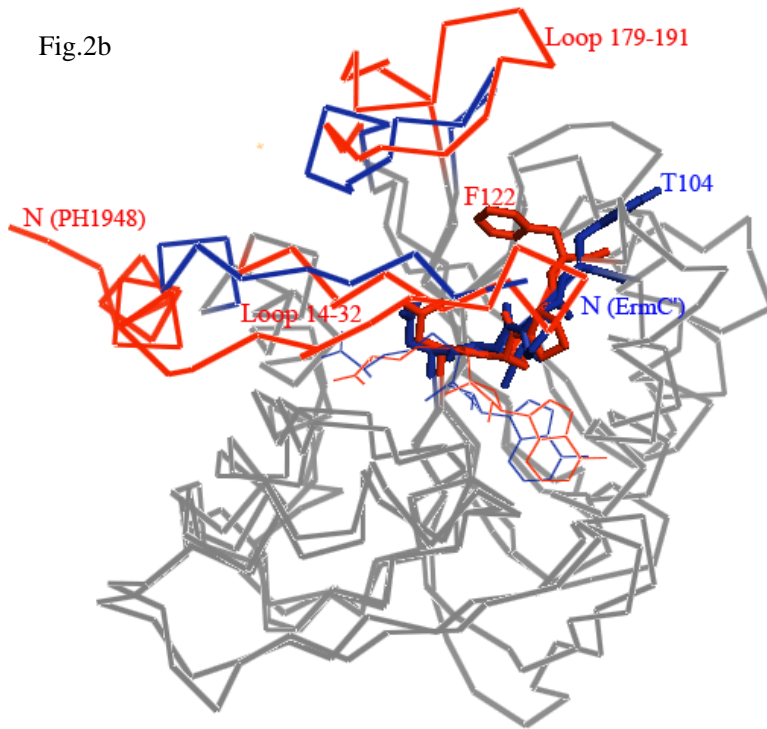


Fig.2b



PH1948:26-EQY	179-IPLQFFFHRKKLE
ErmC':10-QNF	158-VPREYFHPKPKVN
...	:* :*. ..*::