

Supporting Information

Englert et al. 10.1073/pnas.1018307108

A

Score	Expectation	Protein ID	Protein Name	MW	% Coverage	Peptides	Comment
1262	6.3E-120	gi 20095118	hypothetical protein MK1682 [Methanopyrus kandleri AV19]	110919	26	view	
1053	5.3E-99	gi 11935049	keratin 1 [Homo sapiens] indistinguishable	66027	24.8	view	probable contaminant
778	1.5E-71	gi 14667176	PREDICTED: similar to Keratin, type I cytoskeletal 14 (Cytokeratin-14) (CK-14) (Keratin-14) (K14) [107386	16.4	view	probable contaminant
535	3.2E-47	gi 28317	unnamed protein product [Homo sapiens] indistinguishable	59492	17.4	view	probable contaminant
391	7.8E-33	gi 3318722	Chain E, Leech-Derived Trypsin InhibitorTRYPSIN COMPLEX	23457	26	view	enzyme used for digestion
370	8.4E-31	gi 181402	epidermal cytokeratin 2 [Homo sapiens] indistinguishable	65825	9.8	view	probable contaminant
340	9.4E-28	gi 84402	glutathione transferase (EC 2.5.1.18) - fluke (Schistosoma japonicum) (fragment) indistinguishable	25606	20.5	view	
115	0.00003	gi 20093608	HSP70 class molecular chaperones involved in cell morphogenesis [Methanopyrus kandleri AV19]	42733	5.3	view	tentative, only 1 peptide
65	3.2	gi 28590	unnamed protein product [Homo sapiens] indistinguishable	69250	2	view	tentative, only 1 peptide

B

```

1  MAPKSMKHI  RNNVVWELPE  DYKGCMKVP  G  RIYATEKLI  D  GMEKGVFDQV
51  ANVACLPGI  Y  GYSIALPDA  H  YGYGFFIG  V  AAFDVEEG  V  V  SPGGVGYDIN
101  CLAPGTKIL  T  EHGCVWKV  E  D  LPKMLTDQ  K  L  K  VYDVDEGR  E  DDSEIKFVME
151  RGIEEDERA  V  VLVTESGL  TI  EGSEDPV  L  T  PEGYVELGE  I  E  E  E  E  G  D  L  V  V  V  V  Y  P
201  FEGVEYEE  K  E  GTILDES  D  F  E  DVDPQVLR  Y  L  E  E  R  D  L  I  P  L  R  W  S  D  P  K  V  G  T  L  A  R
251  ILGFAMGD  G  H  LGEQAG  R  L  T  L  SFYGDERT  L  R  ELKRDLES  L  G  V  K  A  N  L  H  V  R  K  R
301  RYEIETAS  G  R  YEGEAT  S  V  E  L  RVASRSF  A  L  L  MEKLGMP  R  G  R  K  V  E  T  P  Y  K  V  P  D
351  WIK  E  A  P  L  W  V  K  R  N  F  L  A  G  L  F  A  A  D  G  S  V  V  K  F  K  R  Y  T  P  L  P  I  N  L  T  Q  A  K  V  E  E  L  E  E  N  L  R
401  E  F  M  N  D  V  A  K  L  L  R  E  F  G  I  E  T  T  L  Y  E  V  K  S  K  N  V  V  Y  K  L  A  I  V  G  E  E  N  I  K  R  F  L  G  K  V  G  Y  E
451  Y  D  P  E  K  K  V  E  G  L  A  A  Y  A  Y  L  K  L  K  E  R  V  K  K  D  R  K  E  A  A  E  T  A  A  E  V  Y  E  E  T  G  S  I  T  K  A  H  E  A  V
501  A  D  V  V  N  R  R  F  V  E  R  V  V  Y  D  G  G  I  S  S  V  R  V  P  E  D  F  P  T  F  E  R  F  K  E  E  R  V  L  A  G  G  F  V  I  E  E  V  V  E
551  V  K  G  V  E  P  E  Y  D  R  F  Y  D  I  G  V  C  H  G  A  H  N  F  I  A  D  G  V  V  V  H  N  C  G  V  R  V  M  K  T  D  L  T  E  D  D  V  R  P  K
601  L  R  E  L  L  E  T  I  F  R  N  V  P  A  G  L  G  S  R  H  R  R  V  R  L  S  T  Q  E  L  R  Q  V  M  L  Y  G  A  E  W  A  V  E  E  G  F  G  F  D  E
651  D  L  D  H  I  E  S  R  G  N  M  T  H  A  Y  E  T  I  G  W  E  E  Y  G  P  R  D  D  V  A  S  K  R  A  I  E  R  G  R  P  Q  L  G  T  L  G  S  G  N  H
701  F  L  E  V  Q  V  V  D  E  I  Y  D  K  E  A  A  E  K  M  G  I  R  E  E  G  Q  V  T  I  M  V  H  T  G  S  R  G  F  G  H  Q  V  C  S  D  H  L  R  I  M
751  E  R  S  M  R  D  V  E  R  R  F  G  V  R  I  P  D  R  Q  L  A  C  A  A  M  G  T  D  E  A  K  R  Y  F  N  A  M  N  A  A  A  N  Y  A  F  A  N  R  Q  M
801  I  S  H  W  T  R  E  S  F  V  E  V  F  G  D  E  Y  G  D  A  D  D  M  G  I  E  V  I  Y  D  I  A  H  N  M  A  K  I  E  K  H  P  V  D  G  E  E  R  W  L
851  V  V  H  R  K  G  A  T  R  A  F  S  E  E  A  L  K  K  H  G  E  P  V  P  F  E  G  L  P  Q  P  V  L  I  P  G  D  M  G  T  G  S  Y  I  L  I  G  T  E  K
901  A  M  E  E  T  W  G  S  T  C  H  G  A  G  R  T  M  S  R  A  A  A  K  R  K  F  W  G  E  D  V  A  R  E  L  E  R  Q  G  I  L  V  K  A  A  S  M  P  V  V
951  A  E  E  A  P  P  A  Y  K  D  V  D  E  V  V  R  A  V  A  E  A  G  I  S  D  P  V  V  R  L  R  P  I  G  V  V  K  G
  
```

Fig. S1. Massspectrometric protein sequencing result. (A) The putative RNA ligase protein (shown in Fig. 1B) was cut out of the Coomassie stained protein gel and used for in gel tryptic digest and subsequent LC MS/MS analysis (1). The hypothetical protein MK1682 from *M. kandleri*, which is annotated as an intein-containing precursor [molecular weight (MW) of 110 kDa] was identified with the sequence of 30 tryptic fragments that cover of the precursor protein sequence with 26%. All other massspectrometric hits are either obvious contaminants or the identification is tentative as only one identified tryptic fragment matches the hit. (B) The amino acid sequence for MK1682 is shown. The intein region is shaded in gray, and the identified tryptic fragments matching MK1682 are highlighted in bold red. The tryptic fragments cover 51% of the MK1682 extein sequence.

1. Sauerwald A, et al. (2005) RNA-dependent cysteine biosynthesis in archaea. *Science* 307:1969–1972.

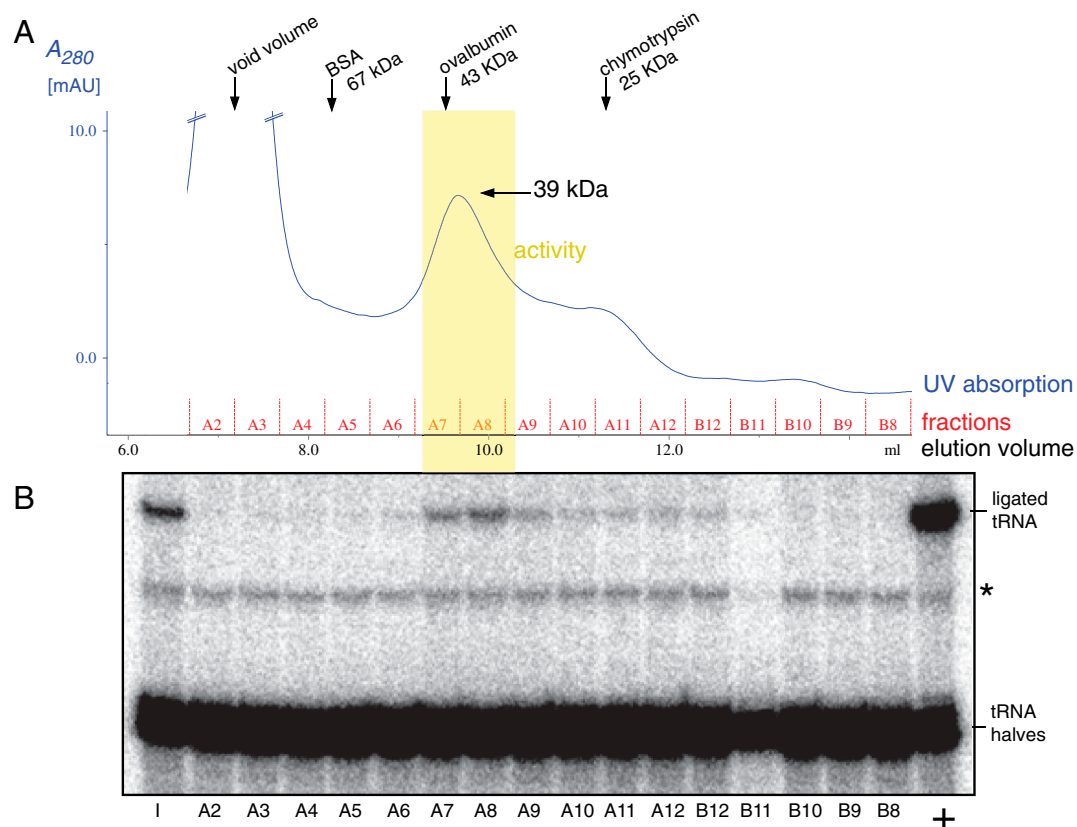


Fig. S2. Characterization of the Superdex 75 chromatographic step during the purification of the tRNA splicing ligase activity from *M. kandleri* extracts. (A) The elution profile for the Superdex 75 HR10/30 separation of the concentrated active Mono S elution fraction is shown. The UV absorption A_{280} is plotted against the elution volume. The collected fractions are indicated. The Superdex column has been calibrated with the marker proteins BSA (elution volume 8.25 mL, native MW of 67 kDa), ovalbumin (9.55 mL, 43 kDa), and chymotrypsin (11.3 mL, 25 kDa). (B) RNA ligase activity assay for the Superdex 75 fractions. Lane I: concentrated, active Mono S elution fraction that was loaded on the Superdex 75 column. Lane +: positive control by the action of T4 polynucleotide kinase/3'-phosphatase and T4 RNA ligase 1. The RNA ligase activity peak correlates to an elution volume of 9.9 mL. Hence, the native MW of the RNA ligase activity can be estimated to be 39 kDa. The calculated MW of MK1682 extein protein is 56 kDa. RtcB may be a very compact shaped protein—a phenomenon also observed for human RtcA—the 3'-terminal phosphate RNA cyclase, which has a native MW by gel filtration of 25 kDa (1) and a calculated MW of 39 kDa (2).

1. Filipowicz W, Strugala K, Konarska M, Shatkin AJ (1985) Cyclization of RNA 3'-terminal phosphate by cyclase from HeLa cells proceeds via formation of N(3')pp(5')A activated intermediate. *Proc Natl Acad Sci USA* 82:1316–1320.
2. Genschik P, Billy E, Swianiewicz M, Filipowicz W (1997) The human RNA 3'-terminal phosphate cyclase is a member of a new family of proteins conserved in Eucarya, Bacteria and Archaea. *EMBO J* 16:2955–2967.