We analyzed the proteome of a crenarchaeon, *Aeropyrum pernix* K1, by using the following four methods: (i) two-dimensional PAGE followed by MALDI-TOF MS, (ii) one-dimensional SDS-PAGE in combination with two-dimensional LC-MS/MS, (iii) multidimensional LC-MS/MS, and (iv) two-dimensional PAGE followed by amino-terminal amino acid sequencing. These methods were found to be complementary to each other, and biases in the data obtained in one method could largely be compensated by the data obtained in the other methods. Consequently a total of 704 proteins were successfully identified, 134 of which were unique to *A. pernix* K1, and 19 were not described previously in the genomic annotation. We found that the original annotation of the genomic data of this archaeon was not adequate in particular with respect to proteins of 10–20 kDa in size, many of which were described as hypothetical. Furthermore the amino-terminal amino acid sequence analysis indicated that surprisingly the translation of 52% of their genes starts with TTG in contrast to ATG (28%) and GTG (20%). Thus, *A. pernix* K1 is the first example of an organism in which TTG is the most predominant translational initiation codon.

The rapid accumulation of genomic sequence data has produced a huge set of new protein sequences, most of which have been derived from the translation of predicted ORFs with no accompanying experimental evidence for their expression and function. Functional prediction is often made by comparison with sequences already in the public databases.

Many microorganisms, in particular Archaea, live in extreme habitats such as high temperature, acidic pH, high salt concentration, etc., and many of them possess unusual cellular and molecular properties. They are collectively termed “extremophiles” and could potentially serve as valuable resources for novel biotechnological applications. Nonetheless there are few existing industrial applications in which either archaeal biomass or archaeal enzymes are used. This is partly due to the lack of data for the expression of individual genes predicted from genome analysis. Such will be best achieved by proteome analysis.

*Aeropyrum pernix* K1 is an aerobic hyperthermophilic crenarchaeon isolated in 1993 from a coastal solfataric thermal vent in Kodakara-jima Island of Kagoshima, Japan. It grows optimally at 90–95 °C (1). Many of the thermostable enzymes of this archaeon are expected to be useful for a variety of industrial applications. The complete genomic sequence of *A. pernix* K1 was established in 1999, and ~2,700 ORFs were predicted from the sequence of nearly 1.67 Mb in size. The data were made available to the public through DDBJ/EMBL/GenBank™ as well as the “Database of the Genomes Analyzed at NITE” (DOGAN) (2). About 1,600 of the predicted 2,700 ORFs were hypothetical (2). Moreover the number of predicted ORFs is much larger than those of other Archaea and bacteria with similar genome sizes, casting doubt over the authenticity of the predicted ORFs. Natale et al. (3) reannotated the *A. pernix* K1 genome using the Clusters of Orthologous Groups of Proteins database and reported the total number of its protein-coding genes to be 1,871. Similarly the current RefSeq contains an annotation reported by Pruitt et al. (4) in which 1,841 proteins were predicted in the *A. pernix* genome, and Guo et al. (5) re-evaluated the *A. pernix* K1 annotation and inferred a total of 1,610 ORFs as potential protein-coding genes. The confusion concerning the annotation of the *A. pernix* K1 genome is one of the factors that might have hindered wide spread utilization of *A. pernix* K1 enzymes, many of which are expected to possess excellent thermostability.

There is an additional problem: from the genomic and proteomic analyses performed to date, ATG is the most common initiation codon, and GTG and TTG are used in less than 10% of bacterial genes (6, 7). In contrast, however, of some 2,700
ORFs predicted in the genome of A. pernix K1, 43% were deduced to be initiated with ATG and 57% were deduced to be initiated with TTG, which differs greatly from other species. Furthermore in A. pernix K1, genes initiated with TTG were reported (8) despite that TTG has not been reported as an initiation codon in other organisms.

The problems described above can only be experimentally clarified by performing proteome analysis. For this purpose, we adopted four methods to maximize the number of detected proteins. Consequently we were able to identify 704 proteins, including 19 that were derived from the genomic regions in which no ORFs were predicted previously (2). The results suggest at the same time that the number of predicted ORFs in the current version of DOGAN is largely overestimated due to the inclusion of ORFs for non-conserved hypothetical proteins with molecular mass of 10–20 kDa. Furthermore amino-terminal amino acid sequences of 134 proteins were determined from which we were able to establish that surprisingly TTG is the most predominant initiation codon in A. pernix K1.

**EXPERIMENTAL PROCEDURES**

**Strain and Culture**

A. pernix strain K1 deposited at NITE Biological Resource Center (NBRC 100138) was cultured at 90 °C in 400 ml of jamarine-yeast extract-trypticase peptone medium (1) for 20 h, cooled down on ice, and harvested by centrifugation at 5,000 × g for 10 min. Cellular pellets were resuspended in 3.5% NaCl and recentrifuged.

**Protein Preparation**

Two-dimensional (2D)-PAGE and 1D-SDS-PAGE-LC-MS/MS—A. pernix K1 cells were suspended in an extraction buffer (67% acetic acid containing 33 mM MgCl2) and disrupted by sonication at 4 °C. Cell debris were removed by centrifugation and 4 volumes of 20 mM Tris-HCl (pH 6.8), 7 M urea, 2 M thiourea, 4% CHAPS, 50 mM DTT, and 40 mM lodoacetamide in 25 mM NH4HCO3 at room temperature for 35 min. The dried gel pieces were rehydrated and incubated in 25 mM NH4HCO3 containing modified trypsin at 37 °C for 4 h. 3% formic acid was added to stop the enzymatic reaction, and the resultant peptides were concentrated, desalted by passage through a µ-C18 ZipTip (Millipore, Billerica, MA), mixed with a matrix solution of 50% acetonitrile saturated with α-cyano-4-hydroxycinnamic acid (Sigma), and air-dried on the target plate.

**Mass Spectrometry of 2D-PAGE-MALDI-TOF MS**—The resulting peptide mixture was subjected to analysis on an Auto-Flex instrument (Bruker Daltonics, Bremen, Germany) with α-cyano-4-hydroxycinnamic acid as the matrix and operated in the reflector mode. Calibration was performed in the external mode using a peptide calibration standard kit (Bruker Daltonics). For peptide assignment the mass spectrum data were analyzed using the Mascot database search program (Matrix Science Ltd., London, UK) in the peptide mass fingerprinting mode against the database of putative proteins of A. pernix K1 containing the data for 2,694 ORFs as well as against the translation of the entire genomic sequence in all phases.

**1D-SDS-PAGE-LC-MS/MS Analysis**

Protein Separation by 1D-SDS-PAGE—Protein samples were dissolved in a lysis buffer containing 7 M urea, 2 M thiourea, 4% CHAPS, 50 mM DTT, and 40 mM Tris and incubated at room temperature for 1 h. Subsequently a 6× concentrated electrophoresis loading buffer containing 0.35 M Tris–HCl (pH 6.8), 10% SDS, 30% glycerol, and 9.3% DTT was added. 1D-SDS-PAGE was performed on 10% separating polyacrylamide gels with a Tris-Tricine running buffer containing 0.1 M Tris, 0.1 M Tricine, and 0.1% SDS.

**Enzymatic Digestion for 1D-SDS-PAGE-LC-MS/MS**—After CBB staining, the gels were sliced into 5-mm-thick pieces from the top band (∼116 kDa) to the bottom line (4.4 kDa). In-gel digestion with modified trypsin was performed using Investigator ProPrep for 6 h and stopped with 3% formic acid. The resulting peptide mixtures were eluted from the gel and dried by evaporation. The peptides were diluted with 0.2% formic acid containing 0.005% heptfluoroisobutyric acid (HFBA) and 2% acetonitrile.

**Mass Spectrometry of 1D-SDS-PAGE-LC-MS/MS**—To analyze peptides, 2D-LC was combined with nano-ESI-MS/MS. The analysis was performed with a Finnigan LCQ DECA XP Plus ion trap mass spectrometer (ThermoElectron, San Jose, CA) coupled with an LC
were reduced by incubating in 5 mM DTT at 60 °C for 30 min, alkylated
12) were obtained.
min until 60 min (12 fractions). In this way a total of 96 fractions (8
ml/min, and fractions were collected every 3.5 min starting from 18
min, and digested with modified trypsin at 37 °C for 1 h. The samples
were adjusted to pH 4 with trifluoroacetic acid, desalted with a C18
aminopropyl-bonded silica gel (BONDENSIL-SAX, 40 μm, Varian, Palo
Alto, CA) was used to fill the column. Buffers used were 20 mM
Tris-HCl, pH 7.0 (Buffer C), and 20 mM Tris-HCl, pH 7.0, with 1 M NaCl
(Buffer D). Proteins were eluted with Buffer C from 0 to 5 min with a
linear gradient of Buffer C to Buffer D from 6 to 25 min and with Buffer
D from 26 to 32 min. The flow rate was 2 ml/min, and the eluate was
collected into eight 8-ml fractions. The fractions were concentrated to
D from 26 to 32 min. The flow rate was 2 ml/min, and the eluate was
linear gradient of Buffer C to Buffer D from 6 to 25 min and with Buffer
D from 26 to 32 min. The flow rate was 2 ml/min, and the eluate was
collected into eight 8-ml fractions. The fractions were concentrated to
linear gradient of Buffer C to Buffer D from 6 to 25 min and with Buffer
D from 26 to 32 min. The flow rate was 2 ml/min, and the eluate was
collected into eight 8-ml fractions. The fractions were concentrated to
94.9% water; desalted with a C18 reverse phase column; and evap-
orated. The digests were dissolved in 0.02% formic acid with 0.005%
HFBA, 2% acetonitrile, and 98% water prior to MS analysis.

Amino-terminal Amino Acid Sequence Analysis

Protein spots on 2D-PAGE were electroblotted onto a PVDF mem-
brane (Sequoi-Blot PVDF membrane, Bio-Rad) with a semidry blotting
apparatus (Bio Craft, Tokyo, Japan). The blotted membrane was
stained with CBB. Singly stained spots were excised from the PVDF
membrane and applied to a protein sequencer (model Procise 491cLC, Applied Biosystems, Foster City, CA) if the staining intensity
appeared to be strong enough for sequencing. For weakly stained
spots, two to six excised spots were combined by repeating 2D-
PAGE and then applied to the protein sequencer. Edman reactions
were performed according to the manufacturer’s instructions. To
identify each protein, the amino acid sequences obtained were com-
pared with the predicted amino acid sequence data translated from
the genomic sequence of A. pernix K1.

Miscellaneous

The genomic sequence data of A. pernix K1 along with the data for
2,694 annotated ORFs were downloaded from DOGAN (www.bio-
ite.go.jp/dogan/Top). Additional A. pernix K1 data for 1,610 anno-
tated ORFs were downloaded from the home page of Tianjin University
Bioinformatics Centre (TUBIC) (tubic.tju.edu.cn/Aper/).

RESULTS AND DISCUSSION

2D-PAGE Followed by TOF MS and Amino-terminal Analysis—About 500 protein spots were well separated from others
by 2D-PAGE with IEF in the first dimension (Fig. 1). Also about
70 basic protein spots, many of which were ribosomal pro-

Molecular & Cellular Proteomics 5.5 813
### Proteome Analysis of a Crenarchaeon, *A. pernix* K1

**Proteins whose amino-terminal amino acid sequence was determined**

Difference column shows the difference between the previously assigned start position (shown in DOGAN) and the newly identified start position. A blank indicates that no difference was detected, and N.D. indicates that the start position could not be assigned. — indicates no corresponding data because it is a newly identified ORF that was not predicted in DOGAN. ORFs newly identified were named by adding “‘a’” to the upstream ORF names of Ref. 2.

| ORF   | Direction | Start  | Stop   | Difference | Initiation codon | Amino-terminal Met<sup>a</sup> | Amino-terminal sequence<sup>b</sup> |
|-------|-----------|--------|--------|------------|-----------------|-------------------------------|-----------------------------------|
| APE0053 | <         | 32898  | 32017  | 21         | TTG             | ADMVVGAVDTPLPLMGY             |
| APE0079 | >         | 54129  | 54527  | N.D.       | GAG<sup>d</sup>  | (A)HLALHTWPEYDVTDFDYT        |
| APE0086 | >         | 59910  | 60401  | 105        | ATG             | PGVGEQAPDFGIAHTGKEI          |
| APE0087 | >         | 60825  | 60406  | 138        | ATG             | GLKHTWE(S)(G)GTTIEGQR         |
| APE0092 | >         | 66333  | 65371  | -3         | TTG             | VLVIVHEDFKLHS                 |
| APE0115 | >         | 86209  | 85004  | 153        | TTG             | GVNRLSEPLPEREREVELRA         |
| APE0119 | >         | 88998  | 89957  | 9          | ATG             | GEDKREIKDITDLPGVPPT             |
| APE0134 | >         | 99329  | 100084 | -45        | TTG             | PDIAVLLKASLNPDMARIQP           |
| APE0135 | >         | 100104 | 101030 |            | GTG             | MKVLAVAYDPDRFDGELATGV        |
| APE0140 | >         | 104906 | 104827 | -135       | TTG             | AKVLEHILPDLSEVS               |
| APE0145 | >         | 107854 | 108584 |            | ATG             | AYKVDTIFYVIGSEVDRYE           |
| APE0169 | >         | 127303 | 126080 | 12         | ATG             | VYDYLRSFEITRRFRASDV          |
| APE0171 | >         | 128428 | 127388 | 9          | TTG             | SVKVQGSLKFGTFRVVALA          |
| APE0176 | >         | 130047 | 130643 | -63        | TTG             | PVTVEW(Y)HYAILKAKGDL          |
| APE0182 | >         | 134821 | 135162 | 15         | GTG             | PNVNLKDASEVGLNVRITI           |
| APE0236 | >         | 172094 | 171336 | 9          | TTG             | AR(C)RGKVIVGGG                 |
| APE0243 | >         | 175883 | 175437 | -51        | TTG             | + MRIGVLQGGVQXEV               |
| APE0246 | >         | 176917 | 176357 | -48        | TTG             | + MIPKYSFSDSDVRLVYEFN         |
| APE0257 | >         | 184505 | 183945 |            | ATG             | + MYYVEVYEEVIGMPESI           |
| APE0319 | >         | 232765 | 232223 |            | ATG             | PRALISSADGFEDVellyPY          |
| APE0333 | >         | 239958 | 240503 | 9          | TTG             | + MILSRRDILALLAIDGLHVE        |
| APE0347 | >         | 252960 | 252316 |            | ATG             | GRGPRYVRVPLRRRE               |
| APE0363 | >         | 259013 | 258273 |            | GTG             | PTKVKKHFLNGQ                  |
| APE0381 | >         | 267488 | 267823 | -93        | TTG             | + MKRLSAKVDKGRVTIPQ           |
| APE0383 | >         | 268766 | 269209 | -33        | TTG             | GLEEYDVLLDVRDMSPPP            |
| APE0396 | >         | 276343 | 274664 | 12         | TTG             | ALPEKTPWPRGYSLT               |
| APE0409 | >         | 282320 | 281976 | -258       | TTG             | AKVQGDPDRPFAFTLKE             |
| APE0443 | >         | 310223 | 309815 | 9          | TTG             | + MERLLRIRSESGVYQVY            |
| APE0457 | >         | 317638 | 318996 | 9          | GTG             | AVDDPRIVVEPPG                 |
| APE0488 | >         | 340209 | 340775 | 9          | TTG             | VSRGAVSSLRVLDDRGFRF            |
| APE0490 | >         | 342822 | 342145 | -9         | TTG             | ARVPGKLVDLSVDSIWEVLG           |
| APE0521 | >         | 357894 | 358571 | 9          | TTG             | N.D. GGG<sup>d</sup>         |
| APE0526 | >         | 362154 | 363203 | -72        | TTG             | TTTVGILASDHVILAADDK            |
| APE0591 | >         | 397515 | 396676 | 39         | TTG             | + MEARLYDLRRKKLLAAEAG              |
| APE0595 | >         | 392413 | 400603 | 9          | TTG             | AGNSLYKVRTPDPLPSLWVF          |
| APE0632 | >         | 431129 | 429927 | -57        | TTG             | + MRRIVLGLSGSGGQLAARY          |
| APE0672 | >         | 449885 | 448944 | 15         | TTG             | + MTLGAGKGVGATAVLMM            |
| APE0676 | >         | 454639 | 453815 |            | ATG             | PGDKVTARRILKMKGPSRIV          |
| APE0681 | >         | 457052 | 456708 | -42        | TTG             | + MKAISIGPRLOVLKMLAHN        |
| APE0682 | >         | 457635 | 457084 | 30         | TTG             | + MLVGVISDHTNKLARLARAG        |
| APE0712 | >         | 476314 | 475118 | 69         | TTG             | + MEPVRLVFEWVM                 |
| APE0741 | >         | 493787 | 493143 |            | ATG             | VSFKRYELPPLPYNYNLEP            |
| APE0754 | >         | 499144 | 499716 | -15        | TTG             | QEOQVKVPRIPYKVKE              |
| APE0756 | >         | 499786 | 500649 | -12        | TTG             | VERAVVAGRM                    |
| APE0767 | >         | 505469 | 506389 | -24        | TTG             | AGGSEKSVFLKWSFWEV              |
| APE0788 | >         | 520508 | 521128 | 27         | GTG             | AVTAVGTSN                     |
| APE0813 | >         | 539396 | 540478 | 6          | ATG             | + MEGLVAEMRES                  |
| APE0890 | >         | 592644 | 593039 | 9          | GTG             | TSSAGAEQEVRSVTSIVYV            |
| APE0912 | >         | 597205 | 597877 |            | GTG             | + MDLQIOQKGLAVTQGGSSGLL         |
| APE0919 | >         | 602220 | 602984 |            | ATG             | AGDIKIDLVEPLKRGFQIVA          |
| APE0952 | >         | 623518 | 624243 | 3          | TTG             | TQHSVPGRLGSSL                 |
| APE0976 | >         | 640223 | 641706 | 15         | ATG             | AKVVEGRVCKTRGREAG              |
| APE0981 | >         | 645352 | 644738 | 6          | GTG             | + MRHPFKVWVTGVPVGKTT           |
| APE0996 | >         | 653298 | 652096 | -144       | TTG             | PIDTGEPEQORLISLREAV            |

<sup>a</sup> Initiation codon.<br/><sup>b</sup> Amino-terminal sequence.<br/><sup>c</sup> Amino-terminal sequence.<br/><sup>d</sup> N.D. indicates that the start position could not be assigned.
### Table I—continued

| ORF   | Directiona | Start   | Stop    | Difference | Initiation codon | Amino-terminal Metb | Amino-terminal sequencec |
|-------|------------|---------|---------|------------|-------------------|----------------------|--------------------------|
| APE1043 | <         | 672699  | 671698  | -144       | TTG               | +                    | MKKIVLGGPGGLVAAYNL       |
| APE1065 | >         | 683529  | 684746  | 12         | TTG               | +                    | MKLYFEAKIEMYGVETP        |
| APE1072 | >         | 684753  | 685646  | 15         | ATG               |                      | AVLVDSETRVLQGITGREG      |
| APE1109 | >         | 700527  | 699271  | 93         | TTG               |                      | SEGRTTTVSKIKADIGSAL      |
| APE1117 | >         | 702122  | 702940  |            | ATG               |                      | ARRPLYGHAAFLQAEGRKI       |
| APE1197 | >         | 750622  | 751791  | 18         | TTG               |                      | VSRPHG                   |
| APE1335 | >         | 845237  | 846256  | -6         | TTG               | +                    | SVLETRELVFTPGPGSFED       |
| APE1386 | >         | 882894  | 884165  | 15         | TTG               | +                    | MQPPTDLLEARQL            |
| APE1472 | >         | 936176  | 935247  |            | ATG               |                      | ARSWLWDYKTAWQVCPGC       |
| APE1473 | >         | 938016  | 936186  | -78        | TTG               |                      | VDISILIGPQGGG            |
| APE1505 | >         | 953060  | 953815  | 9          | TTG               |                      | VDRKKGITPMDILGKGIP       |
| APE1509 | >         | 955147  | 955632  | 24         | GTP               |                      | ARRLPSSPRGIVKIDEGE       |
| APE1520 | >         | 960942  | 960538  | -6         | TTG               | +                    | MKVETGSTFYKTERRLQVLVD    |
| APE1538 | >         | 970516  | 971163  | -24        | TTG               | +                    | MKPVLAVNMTYSSAFEGAGA     |
| APE1568 | >         | 990854  | 990525  | -3         | TTG               | +                    | MKVLKIVSSDVKALTGM         |
| APE1596 | >         | 1012109 | 1013329 |            | GTG               |                      | ARRIVESYPPVRDELOVE        |
| APE1644 | >         | 1042380 | 1042817 | 12         | ATG               |                      | DSESQQSEGAPAPEEVKEV      |
| APE1660 | >         | 1051099 | 1050593 | 60         | GTP               |                      | MERFDLVWVGGPYPGPAV       |
| APE1669 | >         | 1055599 | 1054196 | 9          | TTG               | +                    | SGSRVLSVLNEREVETPET       |
| APE1688 | >         | 1063850 | 1064722 | 48         | ATG               |                      | TGCILGIPGEADPVNNVIII      |
| APE1692 | >         | 1066818 | 1076441 | 93         | ATG               |                      | GLEVGKTLATKEKVLNKV       |
| APE1702 | >         | 1072392 | 1073153 |            | ATG               |                      | MLPTRIDKKXMIPQDA          |
| APE1758 | >         | 1108961 | 110412  | 6          | ATG               |                      | +                      |
| APE1764 | >         | 1112675 | 1111662 | 15         | TTG               |                      | +                      |
| APE1771 | >         | 1116234 | 1115389 |            | GTG               | +                    | +                      |
| APE1775 | >         | 1119301 | 1117685 | 114        | GTG               |                      | +                      |
| APE1786 | >         | 1125447 | 1128990 | -75        | TTG               |                      | +                      |
| APE1818 | >         | 1146015 | 1146398 |            | ATG               |                      | +                      |
| APE1838 | >         | 1162125 | 1161796 | -78        | TTG               | +                    | +                      |
| APE1844 | >         | 1165746 | 1164433 |            | ATG               |                      | +                      |
| APE1872 | >         | 1185763 | 1184540 | 90         | TTG               |                      | +                      |
| APE1937 | >         | 1226227 | 1225514 | -75        | TTG               | +                    | +                      |
| APE1952 | >         | 1232929 | 1234314 | -6         | TTG               | +                    | AVEKYYKDVWVGPAGSSL        |
| APE1961 | >         | 1240727 | 1239909 |            | ATG               |                      | +                      |
| APE1992 | >         | 1240800 | 1242128 |            | TTG               | +                    | KELDLSDPLSMVLKIAETT      |
| APE2050 | >         | 1289059 | 1288040 | 9          | ATG               | +                    | MIIGQVGGYPRG             |
| APE2072 | >         | 1306352 | 1308019 | N.D.       | GACd              | +                    | PIEVLKLRSARHEPKNW         |
| APE2085 | >         | 1316082 | 1316528 | 27         | ATG               | +                    | +                      |
| APE2089 | >         | 1319044 | 1319442 |            | GTP               | +                    | +                      |
| APE2125 | >         | 1342056 | 1342550 |            | TTG               |                      | +                      |
| APE2138 | >         | 1351675 | 1351307 |            | ATG               |                      | +                      |
| APE2147 | >         | 1360415 | 1358976 | 63         | TTG               |                      | GETVAKAMARWHVSLRFSDI     |
| APE2149 | >         | 1360726 | 1361550 |            | ATG               |                      | +                      |
| APE2154 | >         | 1367465 | 1366569 | -18        | TTG               |                      | VEIKLDDTPEFTGKMGSY       |
| APE2158 | >         | 1369691 | 1368480 |            | ATG               |                      | +                      |
| APE2174 | >         | 1380449 | 1380123 |            | ATG               |                      | +                      |
| APE2196 | >         | 1394092 | 1393391 |            | ATG               |                      | +                      |
| APE2211 | >         | 1403501 | 1403055 | 15         | ATG               | +                    | +                      |
| APE2219 | >         | 1407574 | 1408452 |            | GTP               | +                    | +                      |
| APE2225 | >         | 1411301 | 1411771 | 27         | TTG               | +                    | +                      |
| APE2229 | >         | 1419485 | 1419406 | 45         | TTG               | +                    | MKAARLHYNKPLICEDVDY      |
| APE2278 | >         | 1453439 | 1454191 |            | ATG               |                      | +                      |
| APE2289 | >         | 1462461 | 1463036 | 87         | TTG               |                      | +                      |
| APE2291 | >         | 1465851 | 1465132 |            | ATG               |                      | +                      |
| APE2300 | >         | 1473432 | 1472434 | 63         | TTG               | +                    | +                      |
| APE2308 | >         | 1476819 | 1477670 |            | ATG               | +                    | +                      |
| APE2342 | >         | 1497585 | 1496803 | -12        | TTG               | +                    | +                      |
| APE2349 | >         | 1496265 | 1500161 | -24        | TTG               | +                    | +                      |
| APE2354 | >         | 1502075 | 1503292 |            | ATG               | +                    | +                      |

**Proteome Analysis of a Crenarchaeon, A. pernix K1**
Proteome Analysis of a Crenarchaeon, *A. pernix* K1

### Table I—continued

| ORF     | Direction^a | Start  | Stop   | Difference | Initiation codon | Amino-terminal Met^b | Amino-terminal sequence^c |
|---------|--------------|--------|--------|------------|------------------|-----------------------|--------------------------|
| APE2362 | <            | 1507857| 1506425| 33         | TTG              | VLPQPVRMASEDPAISTEDM  |
| APE2384 | >            | 1517045| 1518229|            | GTG              | VSRPNIHVKSBP          |
| APE2390 | >            | 1518794| 1519267| 9          | GTG              | MINAQWSKPNVYKFPGRME  |
| APE2423 | >            | 1534230| 1535378| -126       | ATG              | ARMLMIPGTPVDEVLLAM    |
| APE2437 | <            | 1543770| 1544351| -126       | TTG              | PSARIDLQQGLDRGSPEDL   |
| APE2454 | <            | 1558082| 1557762| 24         | TTG              | ARRKKQDRPYP(C)PS(C)GA |
| APE2458 | >            | 1559754| 1561052|            | GTG              | GNVDFAIERWGVALDAG     |
| APE2484 | >            | 1577489| 1577911| -45        | TTG              | TLIYXFRHGKAEKPMPGVDD  |
| APE2489 | >            | 1579351| 1579803| 6          | ATG              | VVFIRKRIPVRASDIMIT    |
| APE2493 | <            | 1583501| 1581849| 12         | GTG              | GADYPGPLVSLIEL       |
| APE2495 | <            | 1584068| 1583607| 12         | GTG              | +                     | ME(C)IF(C)RIKGEAEAYT   |
| APE2595 | <            | 1651372| 1650716| 12         | GTG              | GADYPGPLVSLIEL       |
| APE2595 | >            | 1651372| 1650716| 12         | GTG              | GADYPGPLVSLIEL       |
| APE2604 | >            | 1658128| 1658400|            | GTG              | +                     | MKF(C)PK(C)GGVMPRKDL   |
| APE2606 | >            | 1659684| 1660222| 6          | ATG              | +                     | MKF(C)PK(C)GGVMPRKDL   |
| APE2606 | >            | 1662330| 1662875| 12         | ATG              | TAGIKLYVYSTYSDVHR     |
| APE2606 | >            | 1664900| 1665287| 33         | ATG              | RGRGRHERRGEMPLPSEDE  |
| APE2606 | >            | 1667572| 1667916| 33         | ATG              | +                     | MRLAVASTGPDAGRGAG      |
| APE2606 | >            | 1669222| 1669668| 33         | ATG              | +                     | MRLAVASTGPDAGRGAG      |
| APE2606 | >            | 1670872| 1671216| 33         | ATG              | +                     | MRLAVASTGPDAGRGAG      |
| APE2606 | >            | 1672528| 1672872| 33         | ATG              | +                     | MRLAVASTGPDAGRGAG      |
| APE2606 | >            | 1674184| 1674528| 33         | ATG              | +                     | MRLAVASTGPDAGRGAG      |
| APE2606 | >            | 1675840| 1676184| 33         | ATG              | +                     | MRLAVASTGPDAGRGAG      |
| APE2606 | >            | 1677500| 1677844| 33         | ATG              | +                     | MRLAVASTGPDAGRGAG      |
| APE2606 | >            | 1679156| 1679498| 33         | ATG              | +                     | MRLAVASTGPDAGRGAG      |
| APE2606 | >            | 1680812| 1681156| 33         | ATG              | +                     | MRLAVASTGPDAGRGAG      |
| APE2606 | >            | 1682468| 1682811| 33         | ATG              | +                     | MRLAVASTGPDAGRGAG      |
| APE2606 | >            | 1684124| 1684467| 33         | ATG              | +                     | MRLAVASTGPDAGRGAG      |
| APE2606 | >            | 1685776| 1686119| 33         | ATG              | +                     | MRLAVASTGPDAGRGAG      |
| APE2606 | >            | 1687432| 1687775| 33         | ATG              | +                     | MRLAVASTGPDAGRGAG      |
| APE2606 | >            | 1689088| 1689430| 33         | ATG              | +                     | MRLAVASTGPDAGRGAG      |
| APE2606 | >            | 1690744| 1691087| 33         | ATG              | +                     | MRLAVASTGPDAGRGAG      |
| APE2606 | >            | 1692400| 1692743| 33         | ATG              | +                     | MRLAVASTGPDAGRGAG      |
| APE2606 | >            | 1694056| 1694399| 33         | ATG              | +                     | MRLAVASTGPDAGRGAG      |
| APE2606 | >            | 1695712| 1696055| 33         | ATG              | +                     | MRLAVASTGPDAGRGAG      |
| APE2606 | >            | 1697368| 1697711| 33         | ATG              | +                     | MRLAVASTGPDAGRGAG      |
| APE2606 | >            | 1699024| 1699367| 33         | ATG              | +                     | MRLAVASTGPDAGRGAG      |

^a Direction indicates the direction of each ORF.

^b + indicates that Met was detected at the amino terminus.

^c Most cysteine residues and one alanine could not experimentally be detected most likely due to degradation, and hence they are placed in parentheses.

^d Indicates that the amino terminus of the corresponding protein products of APE0079, APE0521, APE2072, and APE2493 might be modified.

Proteins, were observed on RFHR gels. They were individually cut out and digested with trypsin as described under “Experimental Procedures,” and the resultant peptide mixtures were analyzed by MALDI-TOF MS with 400–3000 m/z ranges. The mass spectra obtained were then examined by using the MASCOT database search program. In addition, the protein spots were transferred to a PVDF membrane and subjected to amino-terminal amino acid sequencing.

Consequently a total of 300 proteins were identified (Supplemental Table 1). Of them, 187 (62.3%) corresponded to those annotated in other organisms, 80 (26.7%) corresponded to conserved hypothetical proteins, and 33 (11.0%) corresponded to non-conserved hypothetical proteins. However, six of the identified proteins matched the genomic regions of *A. pernix* K1 in which no ORF had been assigned previously. Of these 300 proteins, 134 proteins were found to have unblocked amino termini, and their 6–21 amino-terminal amino acid residues were successfully sequenced enabling determination of the corresponding genomic regions encoding them. These included the six proteins mentioned above that were derived from the regions without assigned ORFs (Table I). Of the 134 proteins, 73 (54.5%) corresponded to annotated, 45 (33.6%) corresponded to conserved hypothetical, and 16 (11.9%) corresponded to non-conserved hypothetical proteins.

**1D-SDS-PAGE-LC-MS/MS Analysis**—In this method, proteins prepared from *A. pernix* K1 cells were first resolved by SDS-PAGE, and the gel was sliced and analyzed as described under “Experimental Procedures” (Fig. 2). A total of 630 proteins were identified accordingly (Supplemental Table 1). Of them, 357 (56.7%) corresponded to annotated, 161 (25.6%) corresponded to conserved hypothetical, and 112 (17.8%) corresponded to non-conserved hypothetical proteins. Also 14 proteins matched the genomic regions without previously assigned ORFs.

**MD-LC-MS/MS Analysis**—In this method, the whole cell lysate was applied to a SAX column for the first dimensional separation and then to a GPC column for the second dimensional separation. In the first dimension, proteins were separated according to their pI values into eight fractions, whereas in the second dimension, proteins in each of the SAX fractions were separated by their molecular mass into 12 fractions. As a consequence, a total of 96 fractions were obtained as exemplified in Fig. 3. Proteins in the resultant fractions were then digested with trypsin and analyzed by 2D-LC-MS/MS. In addition, proteins retained on the SAX column were treated similarly. A total of 404 proteins were thus identified (Supplemental Table 1), 235 (58.2%) of which corresponded to annotated, 94 (23.3%) of which corresponded to conserved hypothetical, and 75 (18.6%) of which corresponded to non-conserved hypothetical proteins. 10 proteins corresponded to the genomic regions without previously assigned ORFs.

By combining the results obtained with the four methods mentioned above, a total of 704 proteins were successfully identified in *A. pernix* K1. The proteins identified by each method are listed in Supplemental Table 1. Of them, 382
Several features of the identified proteins such as molecular mass, pI, hydropathy, protein class, and codon usage were then compared with those of the ORFs predicted in the genome of *A. pernix* K1. If the statistical distribution of these values is not similar between the observed and predicted proteins, then the annotation of the genomic data needs to be appropriately corrected.

**Molecular Mass Distribution**—The molecular mass distribution of proteins predicted from the genomic sequence of *A. pernix* K1 and those experimentally identified by 2D-PAGE, 1D-SDS-PAGE-LC-MS/MS, and MD-LC-MS/MS was compared in groups of 10 kDa up to and higher than 120 kDa (Fig. 5A). Although 49.4% of the proteins were predicted in the molecular mass range of 10–20 kDa in the genome analysis, only 22.9% were actually observed in the same range. Consequently the number of predicted proteins in this molecular mass range in the current version of DOGAN appears to be overrepresented. The average molecular mass of the proteins identified by 2D-PAGE was 32.4 kDa, whereas it was 34.8 kDa by 1D-SDS-PAGE-LC-MS/MS or MD-LC-MS/MS. With the latter two methods, it is possible to identify proteins harboring a larger molecular mass value, whereas such is not the case with 2D-PAGE as the separation of larger proteins becomes poorer. In any event, it is obvious that about half of the proteins predicted from the genomic data in the 10–20-kDa range appear to be incorrectly assigned.

This could be due in part to incorrect prediction of ORFs in the genome analysis of *A. pernix* K1 because of the poorer quality computer software used at that time to assign ORFs. The percentages of ORFs identified by our proteome analysis were 35, 47, 51, 57, and 55%, respectively, in the molecular mass ranges of 20–40, 40–60, 60–80, 80–100, and >100 kDa. Therefore, proteins of larger sizes could in general be more frequently identified in our analysis. Of the proteins of 80 kDa or higher, 14% of those identified by our analysis were predicted to possess a transmembrane domain, whereas the value was much higher for those that could not have been identified in our analysis in which case 61% were predicted to possess a transmembrane domain. Likewise of the proteins smaller than 80 kDa, 9% of identified and 32% of unidentified were predicted to possess a transmembrane domain. From these data, it appears that many of the proteins not identified in our proteome analysis, in particular those with higher molecular mass, are likely to be membrane proteins.

The largest five of the protein-coding ORFs of *A. pernix* K1 are APE0620, APE0609, APE0057, APE1340, and APE1213. Of them, the products of APE1340 and APE0609 were identified in our proteome analysis. The former has homology to the reverse gyrase of *Pyrococcus furiosus* that was recently experimentally proven to be necessary for the growth of this bacterium at high temperature (11). The APE0609 protein is similar to a surface layer protein of *Staphylothermus marinus*. The surface layer protein of *S. marinus* forms a complex with a protease that is likely to play a role in taking up external proteins.
peptides and proteins. A protein similar to the protease of this complex is encoded by APE0607 that was identified in our analysis. This protein is likely to have a similar function in A. pernix.

The remaining proteins were not identified in our analysis most likely because they are membrane proteins as they appear to possess a transmembrane domain. APE1213 is a paralogue of APE0607, but its expression might be different from the latter. The function of APE0620 and APE0057 remains to be investigated.

Isoelectric Point Distribution—
The pI values of the identified and predicted proteins were compared with each other in the pI range from 3 to 13 (Fig. 5B). The average pI values of identified proteins were 7.25 (2D-PAGE), 7.55 (1D-SDS-PAGE-LC-MS/MS), and 7.70 (MD-LC-MS/MS), whereas the value for the predicted proteins was calculated to be 8.68. In the pI range between 5 and 7, the proteins predicted from the genomic data are much fewer than those identified by proteome analysis, whereas proteins in the high pI range (>10) show an opposite distribution pattern. Also proteins identified

![Fig. 3. Separation of proteins by MD-LC.](image)

Proteins were separated into S1–S8 fractions in the first dimension SAX column chromatography (A). Proteins were subsequently separated by gel permeation chromatography in the second dimension. In B, a chromatogram of proteins of fraction S3 obtained in A is shown that were subjected to second dimensional separation into G1 through G12.

![Fig. 4. The number of proteins identified by the three methods used.](image)

The number of proteins identified by each of the three methods, i.e. 2D-PAGE, 1D-SDS-PAGE-LC-MS/MS, and MD-LC-MS/MS, is schematically illustrated along with the number of proteins redundantly identified by two or by all three methods as indicated in the respective overlapping regions.
by 2D-PAGE were much more likely to be distributed in the pI range of 5–7 than those identified in other methods, although the reason for this is not clear.

Hydropathy Distribution—The GRAVY score indicates the hydrophilicity or hydrophobicity of a protein (12); it can be calculated as an arithmetic mean of the sum of the hydropathy index of each amino acid of a protein. About 70% of the predicted proteins concentrated in the neutral range (−0.4 to 0.4). On the other hand, 85% of experimentally identified proteins were found to be distributed in the same range regardless of the identification methods used. The results indicated, therefore, that a large portion of proteins of A. pernix are in the neutral GRAVY score range (Fig. 5C). The averages of the GRAVY score for identified proteins were −0.15 (2D-PAGE), −0.13 (1D-SDS-PAGE-LC-MS/MS), and −0.16 (MD-LC-MS/MS), whereas that of the predicted proteins was −0.02.

Protein Class Distribution—Proteins predicted from the genome of A. pernix K1 as well as those identified by 2D-PAGE, 1D-SDS-PAGE-LC-MS/MS, and MD-LC-MS/MS were grouped into six functional classes (Fig. 5D) and compared. More experimentally identified proteins were found to be categorized in “metabolism” and “genetic information processing” than those predicted from the genome analysis, whereas a distinctly large proportion of predicted proteins were categorized in “non-conserved hypothetical proteins.” With respect to the distribution pattern of proteins in the six protein classes, differences among the methods used were marginal.

Codon Usage Pattern—It is known that a characteristic bias in codon usage exists in each species of organisms (13). To examine whether and to what extent differences in codon usages exist between the experimentally identified ORFs and the ORFs predicted from the A. pernix K1 genomic sequence, codon usages in individual ORFs were plotted against the categories of proteins described above, namely molecular mass, pI, hydropathy, and protein class.

An example is shown in Fig. 6: in A, the codon usage patterns of proteins categorized by their molecular mass are shown, and in B, similar patterns of proteins categorized by their protein class are shown. As described above, a large proportion of predicted proteins were classified in the molecular mass range of 10–20 kDa. Indeed many of them were found to deviate from the average use of TCC, whereas predicted proteins larger than 40 kDa appear to match well with
those of experimentally confirmed proteins. Therefore, it seems that the usage patterns of various codons will serve as good tools to evaluate whether a particular ORF predicted from the genomic sequence is likely to be a true gene or not. Indeed this is one of the bases on which algorithms for the prediction of genes/ORFs in the genomic sequence data rely. A similar analysis was performed with respect to protein classes as shown in Fig. 6B. The patterns of experimentally identified versus predicted ORFs were found to be quite different when proteins categorized as "non-conserved hypothetical proteins" were analyzed. Interestingly such a clear difference shown in Fig. 6A and B, was not observed when a similar analysis was performed with proteins categorized by their pl and hydropathy values.

Complementarity of the Methods Used—For the proteome analysis of A. pernix K1, we adopted the high resolving power of 2D-PAGE (14) including RFHR-2D-PAGE (9) and combined it with MALDI-TOF MS for "peptide mass fingerprinting." However, 2D-PAGE-MALDI-TOF MS has limitations in the detection of less abundant or hydrophobic proteins as well as proteins of extremely large sizes. Introduction of improved chaotropes and development of novel zwitterionic detergents (15) were found to improve the situation to some extent.

An alternative approach was to omit the first dimensional separation and apply an enriched membrane protein fraction directly to 1D-SDS-PAGE-LC-MS/MS. Most of the bands on the 1D-SDS-PAGE gel consisted of multiple proteins, but the ability of HPLC in connection with ESI tandem mass spectrometry is powerful enough to analyze a mixture of derived peptides so that conventional tryptic digestion of proteins followed by mass spectrometric analysis led to the identification of each protein in the mixture. This method is called the "shotgun method" (16), and it gives a considerable advantage in the characterization of membrane proteins because separation was targeted at the peptides rather than proteins so that solubility problems that are often encountered with hydrophobic proteins could largely be alleviated. A disadvantage of the shotgun method is that intensive computational analysis of the entire data set is always required, and no information regarding the charge of the intact protein could be obtained.

MD-LC-MS/MS is a third alternative method we adopted in which proteins were separated by MD-LC (10, 17, 18), digested with a specific enzyme, and ionized with ESI, and then their mass spectra were measured. In regular MD-LC systems proteins and peptides are separated according to a variety of their properties, such as pl, relative molecular mass, and hydrophobicity (17). However, a disadvantage of the MD-LC-MS/MS method is that not all of the peptide fragments could be detected, and their quantity is low. Also the sensitivity of detection will progressively decrease as the number of fractions increases.

Comparison with the Results Obtained by Other Researchers—Guo et al. (5) examined the genomic data of A. pernix K1 and reported that 1,610 ORFs can be recognized as such (tubic.tju.edu.cn/Aper/). Therefore, their data were compared with the proteins experimentally identified. Of the 704 identified proteins, 692 were included as ORFs predicted by Guo et al. (5), but the remaining 12 were not. The molecular mass distribution of the proteins derived from TUBIC ORFs is very similar to the identified proteins as shown in Fig. 5A. However, their other characteristics slightly but significantly deviate from those of the proteins we experimentally characterized (Fig. 5, B–D).

Assignment of the Codons for Translation Initiation—Of the 134 proteins whose amino-terminal sequences were experimentally determined, 50 were found to possess Met at their amino terminus. By comparing the nucleotide sequences corresponding to the amino-terminal Met, seven of them were found to possess ATG, and 14 others contained GTG. In
is likely to be that of a processed protein, although the nature of the processing remains to be clarified further. The genomic nucleotide sequence present in the immediate upstream of APE2493 is ATA. However, because ATA is not likely to serve as a translational initiator, it may be that the amino-terminal amino acid sequence corresponding to APE2493 was similarly processed, although the processing has not been elucidated yet.

To summarize the data for translational initiation codons corresponding to the 130 sequences other than the four proteins mentioned above, TTG was found to be most frequent (52%), whereas ATG and GTG, respectively, were found in 28 and 20% of the cases. Of the 130 ORFs, six proteins were found to be derived from the region in which no ORFs were previously assigned. Of the remaining 124 ORFs, the initiation codons of 89 (72%) were different from the positions that were assigned previously (2).

**Characteristics of the Region Upstream of the Putative Initiation Codons**—The mechanism of transcriptional initiation in Archaea has been speculated to be more closely related to that of eukaryotes (19, 20). However, three groups of transcription-associated proteins have been identified in Archaea: one group more similar to prokaryotes, another group more similar to eukaryotes, and a third group more similar to both prokaryotes and eukaryotes. Several homologues of bacterial transcriptional factors (21–24) have been identified in Archaea, and Tolstrup et al. (25) have shown that the translation process of internal genes of operons in Archaea was similar to that in bacteria.

To characterize the genomic regions likely to function in translational initiation in *A. pernix* K1, the nucleotide frequency of the sequences surrounding the ORFs for 130 proteins mentioned above were analyzed according to the method of Xiu-Feng et al. (26). The ORFs were categorized into two groups: in Group 1 the ORFs in question are not well separated from their immediate upstream neighbor, whereas in Group 2 they are more than 50 bp away from each other. In the region preceding the 130 ORFs, there is a G box at the position −10 upstream of the initiation codon with a typical sequence of GGTG regardless of the ORF category, whereas ORFs of Group 1 harbor in addition an AT box at the −42 position upstream of the initiation codon and a weak C box at the −35 position (Fig. 7, A and B).

**Mechanisms of Translational Initiation**—In view of the finding mentioned above, all the identified ORFs of *A. pernix* were reassigned by taking the presence of three initiation codons and of a G box into consideration. Consequently TTG was found to be the most predominant initiation codon (38% of all ORFs) followed by ATG (33%) and GTG (29%). After the reassignment, the frequency of occurrence of each nucleotide in the upstream region was plotted. There is a distinct G box at the region surrounding the −10 position upstream of the initiation codon harboring GGTG as a typical sequence as in the case of the 130 genes for the experimentally characterized
proteins mentioned above. In addition, Group 1 ORFs possess a weak AT box at the -42 position and a weak C box at the -35 position, whereas the weak C boxes were not so clear in Group 2 ORFs (data not shown).

For the experimental identification of translational initiation codons, a large scale amino-terminal sequencing of Synechocystis sp. strain PCC6803 was performed by Sazuka and Ohara (27, 28) in which amino-terminal amino acid analysis is shown. The results shown here, however, are different, and TTG is most predominant in A. pernix K1 or not. Archaea are known to possess a eukaryote-like positive regulator facilitating recruitment of the TATA-binding protein for transcriptional activation (34). Furthermore two types of translational initiation mechanisms have been reported in Archaea, namely leadered and leaderless translation. The former has been shown to occur in internal genes (25), which possess a G-rich region in their 5' flanking region that is likely to play a role in ribosomal binding, whereas the latter involves scanning for the first initiation codon along the transcripts (35). N-Formylmethionyl tRNA, deformylase, and methionyl aminopeptidase have been proven to play the roles in the translational initiation in E. coli in particular in A. pernix K1, although it is not clear whether it is the most efficient initiation codon in A. pernix K1 or not. Archaea are known to possess a eukaryote-like positive regulator facilitating recruitment of the TATA-binding protein for transcriptional activation (34). Furthermore two types of translational initiation mechanisms have been reported in Archaea, namely leadered and leaderless translation. The former has been shown to occur in internal genes (25), which possess a G-rich region in their 5' flanking region that is likely to play a role in ribosomal binding, whereas the latter involves scanning for the first initiation codon along the transcripts (35). N-Formylmethionyl tRNA, deformylase, and methionyl aminopeptidase have been proven to play the roles in the translational initiation in E. coli in particular in combination with ATG. A. pernix K1 possesses no homologues for N-formylmethionyl tRNA and deformylase, and the Met-tRNA genes are present in triplicate, although there is no sequence similarity between them, and two of them have an intron (2). These might possibly be related, at least to some extent, to the less frequent translational initiation at ATG in A. pernix K1.

Acknowledgment—We thank T. Sekigawa of NITE for valuable assistance in the computational analysis of the genomic data of A. pernix K1.

| Species                        | ATG | GTG | TTG | Other | G + C content |
|-------------------------------|-----|-----|-----|-------|---------------|
| A. pernix K1 (this study)*    | 28  | 20  | 52  | 0     | 56            |
| A. pernix K1 (old)            | 43  | 57  | 0   | 56    |               |
| A. aeolicus                   | 82  | 10  | 7   | <1    | 43            |
| A. fulgidus                   | 76  | 22  | 2   | 0     | 49            |
| B. subtilis                   | 78  | 9   | 13  | <1    | 44            |
| B. burgdorferi                | 69  | 9   | 22  | 0     | 29            |
| C. trachomatis                | 89  | 8   | 4   | 0     | 41            |
| C. efficiens                  | 54  | 32  | 14  | 0     | 63            |
| E. coli                      | 83  | 14  | 3   | <1    | 51            |
| H. influenzae                 | 89  | 11  | 0   | 0     | 38            |
| H. pylori                     | 82  | 10  | 8   | <1    | 39            |
| M. jannaschii                 | 67  | 13  | 18  | 0     | 31            |
| M. thermoautotrophicum       | 62  | 22  | 15  | 1     | 50            |
| M. tuberculosis              | 61  | 33  | 5   | 1     | 66            |
| M. genitalium                | 92  | 8   | 0   | 0     | 32            |
| M. pneumoniae                | 92  | 4   | 3   | <1    | 40            |
| P. horikoshii OT3            | 71  | 29  | 0   | 0     | 42            |
| R. prowazekii                | 83  | 17  | 0   | 0     | 29            |
| S. aureus N315               | 83  | 7   | 8   | 2     | 33            |
| S. avermitilis               | 59  | 37  | 4   | <1    | 71            |
| S. tokodaii                  | 77  | 23  | <1  | 0     | 33            |
| Synechocystis sp.             | 58  | 34  | 8   | <1    | 48            |
| T. pallidum                  | 82  | 10  | 7   | <1    | 53            |

* The percentage of start codons in A. pernix based on the amino-terminal amino acid analysis is shown.
REFERENCES

1. O'Farrell, P. H. (1975) High resolution two-dimensional electrophoresis of proteins. J. Chromatogr. B 877, 11–18

2. Kawakami, T., Anyoji, H., and Nishimura, T. (2002) Development of proteome analysis systems for efficient expression profiling: towards high throughput identification and quantification of disease target proteins. J. Mass Spectrom. Soc. Jpn. 3, 135–141

3. Soppa, J. (1999) Translation initiation in archaea: facts, factors and future aspects. Mol. Microbiol. 31, 1265–1266

4. Soppa, J. (1999) Normalized nucleotide frequencies allow the detection of archaeal promoter elements for different archaeal groups and reveal base-specific TFB contacts upstream of the TATA box. Mol. Microbiol. 31, 1589–1592

5. Kyripides, N. C., and Ouzounis, C. A. (1999) The eubacterial transcriptional activator Lrp is present in the archaeon Pyrococcus furiosus. Trends Biochem. Sci. 22, 140–141

6. Vierke, G., Engelmann, A., Hebbeln, C., and Thomm, M. (2003) A novel archaeal transcription regulator of heat shock response. J. Biol. Chem. 278, 18–26

7. Lee, S. J., Engelmann, A., Horiarcher, R. Q., Qu, Q., Vierke, G., Hebbeln, C., Thomm, M., and Boos, W. (2003) TrmB, a sugar-specific transcriptional regulator of the trehalose/maltose ABC transporter from the hyperthermophilic archaeon Thermococcus litoralis. J. Biol. Chem. 278, 983–990

8. Bell, S. D., Caims, S. S., Robison, R. L., and Jackson, S. P. (1999) Transcriptional regulation of an archaeal operon in vivo and in vitro. Mol. Cell 4, 971–982

9. Holstrup, N., Sensen, C. W., Garrett, R. A., and Clausen, I. G. (2000) Two different and highly organized mechanisms of translation initiation in the archaeon Sulfolobus solfataricus. Extremophiles 4, 175–179

10. Xu-Feng, W., Susan, M. B., and John, A. B. (2004) Revealing gene transcription and translation patterns in archaea, using interactive clustering model. Extremophiles 8, 291–299

11. Sazuka, T., and Ohara, O. (1996) Sequence features surrounding the translation initiation sites assigned on the genome sequence of Synechocystis sp. strain PCC6803 by amino-terminal protein sequencing. DNA Res. 3, 225–232

12. Sazuka, T., and Ohara, O. (1997) Towards a proteome project of cyanobacterium Synechocystis sp. strain PCC6803: linking 130 protein spots with their respective genes. Electrophoresis 18, 1252–1258

13. Rocha, E., Danchin, A., and Vial, A. (1999) Translation in Bacillus subtilis: roles and trends of initiation and termination, insights from a genome analysis. Nucleic Acids Res. 27, 3567–3576

14. Oikamoto, R., Macfarlane, J. L., and Wösteholme, D. R. (1990) Evidence for the frequent use of TTG as the translation initiation codon of mitochondrial protein genes in the nematodes, Ascaris suum and Caenorhabditis elegans. Nucleic Acids Res. 18, 6113–6118

15. Becerra, S. P., Koczot, F., Fabisch, P., and Rose, J. A. (1988) Synthesis of adeno-associated virus structural proteins requires both alternative mRNA splicing and alternative initiations from a single transcript. J. Virol. 62, 2745–2754

16. Curran, J., and Kolakofsky, D. (1988) Ribosomal initiation from an ACG codon in the Sendai virus P/C mRNA. EMBO J. 7, 245–251

17. Ringquist, S., Shniedling, S., Barrick, D., Green, L., Binkley, J., Stormo, G. D., and Gold, L. (1992) Translation initiation in Escherichia coli: sequences within the ribosome-binding site. Mol. Microbiol. 6, 1219–1229

18. Ouhoumouch, M., Dewhurst, R. E., Thomm, W., Thomm, M., and Geiduschek, E. P. (2003) Activation of archaeal transcription by recruitment of the trehalose/maltose ABC transporter from the hyperthermophilic archaeon Thermococcus litoralis. J. Biol. Chem. 278, 983–990

19. Bell, S. D., Caims, S. S., Robison, R. L., and Jackson, S. P. (1999) Transcriptional regulation of an archaeal operon in vivo and in vitro. Mol. Cell 4, 971–982

20. Holstrup, N., Sensen, C. W., Garrett, R. A., and Clausen, I. G. (2000) Two different and highly organized mechanisms of translation initiation in the archaeon Sulfolobus solfataricus. Extremophiles 4, 175–179

21. Xu-Feng, W., Susan, M. B., and John, A. B. (2004) Revealing gene transcription and translation patterns in archaea, using interactive clustering model. Extremophiles 8, 291–299

22. Sazuka, T., and Ohara, O. (1996) Sequence features surrounding the translation initiation sites assigned on the genome sequence of Synechocystis sp. strain PCC6803 by amino-terminal protein sequencing. DNA Res. 3, 225–232

23. Sazuka, T., and Ohara, O. (1997) Towards a proteome project of cyanobacterium Synechocystis sp. strain PCC6803: linking 130 protein spots with their respective genes. Electrophoresis 18, 1252–1258

24. Rocha, E., Danchin, A., and Vial, A. (1999) Translation in Bacillus subtilis: roles and trends of initiation and termination, insights from a genome analysis. Nucleic Acids Res. 27, 3567–3576

25. Oikamoto, R., Macfarlane, J. L., and Wösteholme, D. R. (1990) Evidence for the frequent use of TTG as the translation initiation codon of mitochondrial protein genes in the nematodes, Ascaris suum and Caenorhabditis elegans. Nucleic Acids Res. 18, 6113–6118

26. Becerra, S. P., Koczot, F., Fabisch, P., and Rose, J. A. (1988) Synthesis of adeno-associated virus structural proteins requires both alternative mRNA splicing and alternative initiations from a single transcript. J. Virol. 62, 2745–2754

27. Curran, J., and Kolakofsky, D. (1988) Ribosomal initiation from an ACG codon in the Sendai virus P/C mRNA. EMBO J. 7, 245–251

28. Ringquist, S., Shniedling, S., Barrick, D., Green, L., Binkley, J., Stormo, G. D., and Gold, L. (1992) Translation initiation in Escherichia coli: sequences within the ribosome-binding site. Mol. Microbiol. 6, 1219–1229

29. Ouhoumouch, M., Dewhurst, R. E., Hausner, W., Thomm, M., and Geiduschek, E. P. (2003) Activation of archaeal transcription by recruitment of the TATA-binding protein. Proc. Natl. Acad. Sci. U. S. A. 100, 5097–5102

30. Kozak, M. (1999) Initiation of translation in prokaryotes and eukaryotes. Gene (Amst.) 234, 187–208

31. Schultzman, L. H., and Howard, M. O. (1973) Recognition of altered E. coli formylmethionine transfer RNA by bacterial T factor. Biochem. Biophys. Res. Commun. 51, 275–282

32. Mazel, D., Pochet, S., and Marliere, P. (1994) Genetic characterization of polypeptide deformylase, a distinctive enzyme of eubacterial translation. EMBO J. 13, 914–923

33. Ben-Bassat, A., Bauer, K., Chang, S. Y., Myambo, K., Boosman, A., and Chang, S. (1987) Processing of the initiation methionine from proteins: properties of the Escherichia coli methionine aminopeptidase and its gene structure. J. Bacteriol. 169, 751–757