Introns in protein-coding genes in Archaea

Yoh-ichi Watanabe[a,*], Shin-ichi Yokobori[b], Toshiro Inaba[a], Akihiko Yamagishi[b], Tairo Oshima[b], Yutaka Kawarabayasi[c,d], Hisasi Kikuchi[c], Kiyoshi Kita[a]

[a] *Department of Biomedical Chemistry, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan*
[b] *Department of Molecular Biology, School of Life Science, Tokyo University of Pharmacy and Life Science, 1432-1 Horinouchi, Hachioji, Tokyo 1092-0392, Japan*
[c] *Biotechnology Center, National Institute of Technology and Evaluation, 2-49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan*
[d] *National Institute of Advanced Industrial Science and Technology, Tsukuba, Ibaraki 305-8566, Japan*

Received 6 November 2001; accepted 20 November 2001

First published online 4 December 2001

Edited by Lev Kisselev

Abstract Introns in protein-coding genes are ubiquitous in eukaryotic cells, but pre-mRNA splicing has yet to be reported in archaeal and its viral genomes. We present evidence of introns in genes encoding a homolog of eukaryotic Cbf5p (centromere-binding factor 5; a subunit of a small nucleolar ribonucleoprotein) in three Archaea; *Aeropyrum pernix*, *Sulfolobus solfataricus* and *Sulfolobus tokodaii*. Splicing of pre-mRNAs in vivo was demonstrated by reverse transcriptase-mediated polymerase chain reaction. The exon–intron boundaries of these genes are predicted to be folded into a structure similar to the bulge–helix–bulge motif, suggesting that splicing of these pre-mRNAs probably depends on the splicing system elucidated for archaeal pre-tRNAs and rRNAs.

Key words: Bulge–helix–bulge motif; Pre-mRNA splicing; Cbf5p; *Aeropyrum pernix*; *Sulfolobus solfataricus*; *Sulfolobus tokodaii*

1. Introduction

Introns in protein-coding genes are ubiquitous in eukaryotic cells but are rarely found in prokaryotic cells. Although introns in protein-coding genes have been found in Bacteria (reviewed in [1]), pre-mRNA splicing in Archaea and its virus has yet to be reported. In archaeal precursor tRNAs [2] or rRNAs [3] that contain intervening sequences, the intervening sequence and the flanking region is folded into a bulge–helix–bulge motif [4–7] (Fig. 2D), which is cleaved by splicing/processing endonuclease followed by ligation (reviewed in [8]). A similar type of cleavage without ligation proceeds in some archaeal pre-rRNA processing (reviewed in [9]). In vivo and in vitro experiments of splicing of artificial RNA has demonstrated that the cleavage is independent of the mature RNA structures [10,11].

CBF5 (centromere-binding factor 5) is a gene encoding a subunit of a small nucleolar ribonucleoprotein involved in eukaryotic rRNA biogenesis, particularly in pseudouridylation [12]. Here, we present evidence of introns in genes encoding a Cbf5p homolog and their pre-mRNA splicing in some archaeal species.

2. Materials and methods

2.1. Archaeal cells

Aeropyrum pernix strain K1 (JCM 9820) was grown as described [13]. *Sulfolobus solfataricus* strain P2 (JCM 11322) and *Sulfolobus tokodaii* strain 7 (JCM 10545) were grown at 80°C as described [14] but without trace metal supplements.

2.2. RNA and DNA

RNAs from the archaeal cells were prepared with Trizol (Invitrogen, USA), and then treated with DNase I (Invitrogen, USA) before reverse transcription. DNA of *A. pernix* was prepared as described [13]. DNAs of *Sulfolobus* strains were prepared as described [15].

2.3. Reverse transcription and polymerase chain reaction (PCR)

Reverse transcription of RNA (2–3 μg) was carried out with random hexamer and ReverTra Ace (Toyobo, Japan) according to the manufacturer's instructions with slight modifications, then treated with RNase H (Takara Shuzo, Japan). PCR reaction mixtures (25 μl) contained: 1× reaction buffer (ExTaq, Takara Shuzo, Japan), 0.2 mM dNTPs each, 200 nM of each primers, 0.05 U/μl of ExTaq (Takara Shuzo, Japan) and template DNA (1.7–5% of the synthesized cDNA as above, or 40 ng of DNA). PCR conditions using a programmable incubator (PC700, Astec, Japan) were as follows: for *A. pernix*, 35 cycles at 94°C for 30 s, 68°C for 30 s and 72°C for 30 s; for *S. solfataricus*, 10 cycles at 94°C for 30 s, 68°C for 30 s and 72°C for 15 s, then 35 cycles at 94°C for 30 s, 62°C for 30 s and 72°C for 15 s; for *S. tokodaii*, 35 cycles at 94°C for 30 s, 60°C for 30 s and 72°C for 1 min. Oligodeoxynucleotides used as PCR primers (Japan Bio-Service Inc., Japan) were as follows: for *A. pernix*, P-517 (forward), 5′-CCTACCCCATGAGAGGCCGTTGGA-3′ (corresponding to 280567–280590 in GenBank accession number AP000060), P-518 (reverse), 5′-GGCCTATGGAGCTGCATCACGCAA-3′ (complementary to 280839–280862 in AP000060); for *S. solfataricus*, P-508 (forward), 5′-CGAGCCATGAAGTAGCGTATTGGGTT-3′ (complementary to 1846–1871 in accession number AE006673), P-509 (reverse), 5′-CTCTATTCCTATTGGTAACACACCGGTA-3′ (corresponding to 1740–1767 in AE006673); for *S. tokodaii*, P-504 (forward), 5′-AACGCTGATAAACCACCTGGACCTACA-3′ (corresponding to 151661–151687 in accession number AP000982), P-505 (reverse), 5′-CACTGAGGATCTTACGGGTGGTTTCTGA-3′ (complementary to 151955–151982 in AP000982). PCR products were purified by agarose gel electrophoresis, then cloned into pGEM-T vector (Promega, USA) with *Escherichia coli* strain DH5α as a host. The recombinant plasmids were extracted and then sequenced with DYEnamic ET Terminator Cycle Sequencing Kit (Amersham Pharmacia Biotech, USA) using vector-specific primers on an automated DNA sequencer (373A, Applied Biosystems, USA).

*Corresponding author. Fax: (81)-3-58413444.
E-mail address: ywatanab@m.u-tokyo.ac.jp (Y.-i. Watanabe).

Abbreviations: PCR, polymerase chain reaction; ORF, open reading frame; CBF5, centromere-binding factor 5

0014-5793/02/$22.00
PII: S0014-5793(01)03219-7

3. Results and discussion

Computational analysis of the sequenced genomes of *A. pernix* K1 [16], *S. solfataricus* P2 [17] and *S. tokodaii* strain 7 [18] predicted that the Cbf5p homologs from the archaeal kingdom Crenarchaeota are composed of two fragmented peptides encoded by two overlapping or adjacent open reading frames (ORFs) (Fig. 1), whereas other sequenced Euryarchaea have a single continuous *CBF5* gene [19–25]. The predicted amino-terminal Cbf5p peptides of these Crenarchaea stop at positions within or around conserved motif II [26], including a putative catalytic aspartate residue for pseudouridylation [12]. This conserved aspartate residue was missing at its corresponding position [27], in the amino-terminal peptides of *A. pernix*, *S. solfataricus* and *S. tokodaii* (Fig. 1, ApeN, SsoN and StoN). Furthermore, analysis of the *S. tokodaii* genome sequence revealed that part of motif II is missing in this species (Fig. 1C, StoN and StoC).

We noticed putative introns in the genes of the Cbf5p homologs in these crenarchaeal species at the position between the first and second letters of the codon corresponding to the conserved aspartate residue (Fig. 1). RNA secondary structure prediction [28,29] in addition to a comparison of the putative exon–intron boundaries of the crenarchaeal Cbf5p homologs suggest the possible folding of the corresponding region into a structure similar to a bulge–helix–bulge motif (Fig. 2A–C).

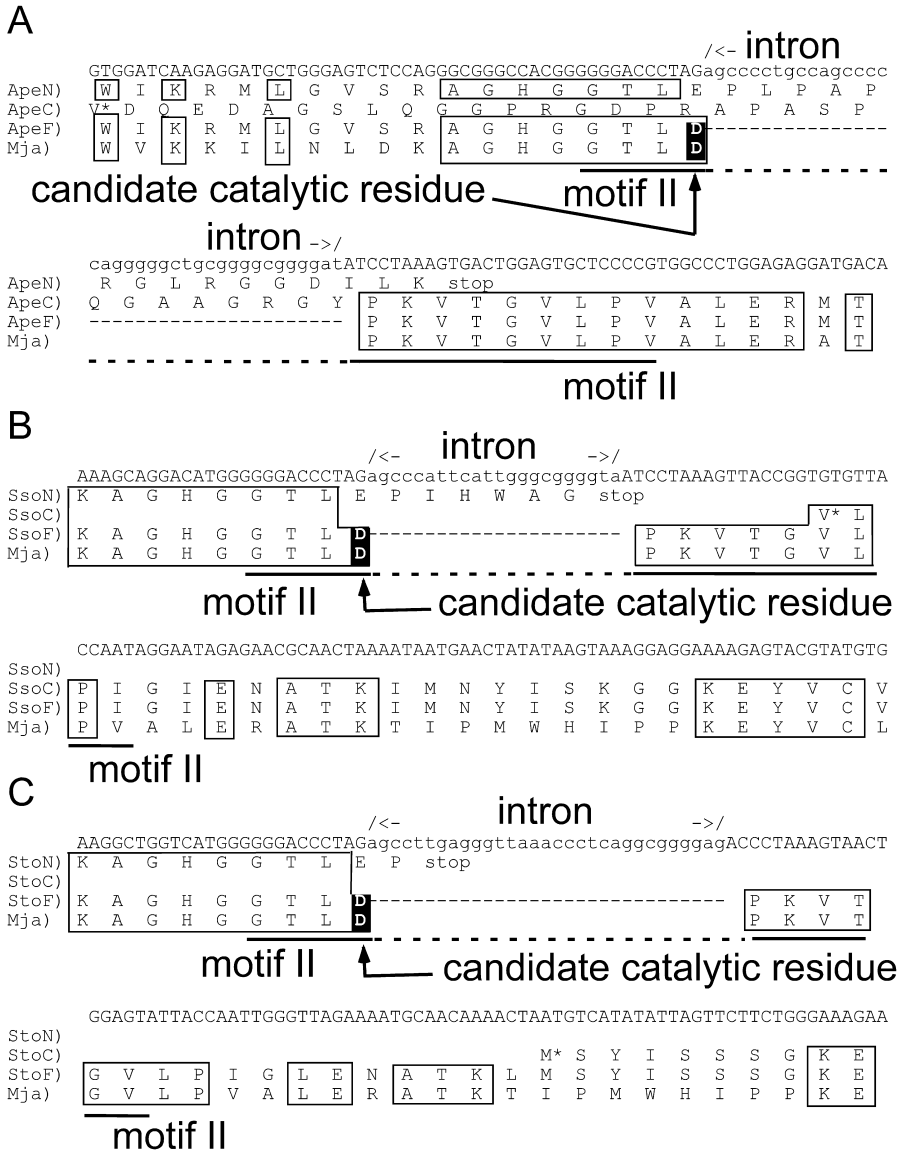

Fig. 1. Parts of the *CBF5* genes of the Crenarchaea (A) *A. pernix* (GenBank accession number AP000060), (B) *S. solfataricus* (accession number AE006673) and (C) *S. tokodaii* (accession number AP000982) along with an alignment of part of the Cbf5p homologs of these Crenarchaea without (amino-terminal fragments: ApeN, ORF APE0989, accession number BAA79973.1; SsoN, ORF SSO5761, accession number AAK40721.1; and StoN, ORF STS055, accession number BAB65384.1, and carboxyl-terminal fragments: ApeC, ORF APE0990, accession number BAA79974.1; SsoC, ORF SSO0393, accession number AAK40720.1; and StoC, ORF ST0402, accession number BAB65385.1) and with (ApeF, SsoF, and StoF) pre-mRNA splicing, and the Euryarchaeon *Methanococcus jannaschii* (Mja, accession number AAB98132). The intron is shown in lower case. Conserved amino acid residues between the crenarchaeal and *M. jannaschii* sequences are boxed. The amino-termini of carboxyl-terminal peptides predicted in previous studies are indicated with an asterisk. The candidate catalytic aspartate residues are shown in a black background.

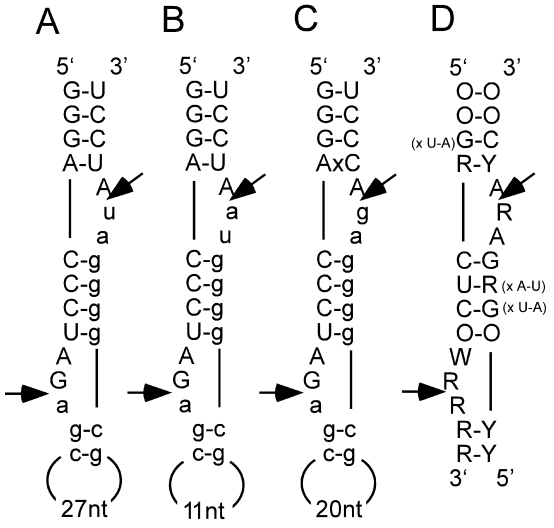

Fig. 2. Predicted folding of putative exon–intron boundaries of *CBF5* pre-mRNAs of *A. pernix* (A), *S. solfataricus* (B) and *S. tokodaii* (C) and bulge–helix–bulge motif (D; adopted from [32]). The intron is shown in lower case. The spliced sites are indicated by arrows.

Cleavage at the predicted sites followed by ligation was expected to restore the ORF and the putative catalytic aspartate residues at the conserved position in the predicted peptide (Fig. 1, ApeF, SsoF and StoF). Reverse transcriptase-mediated PCR (see Section 2) of *S. solfataricus* and *S. tokodaii* total RNAs produced cDNAs from spliced CBF5 mRNA and its precursor (Fig. 3B,C, lane 2), while total RNA of *A. pernix* gave only cDNA from spliced *CBF5* mRNA (Fig. 3A, lane 2). Ligation between the predicted splice sites was confirmed by sequencing analysis of the cDNAs. These results suggest that splicing of the *CBF5* pre-mRNA occurs at the RNA level in these crenarchaeal species.

Since archaeal Cbf5p homologs have a putative function in biogenesis of stable RNAs for translation [27], the expression of the Cbf5p homologs, tRNA, and rRNA (in *A. pernix* [30]) may be co-regulated by the same RNA splicing/processing machinery utilized during the post-transcriptional step in these archaeal species.

Conservation of the intron position and sequence similarity (not shown) suggests a possible origin of the introns identified in this study to be a common ancestor of *A. pernix*, *S. solfataricus* and *S. tokodaii*. Further study of *CBF5* genes in other crenarchaeal species may elucidate more detailed information pertaining to the origin of these introns.

Several full-genome sequences of Archaea are now available [16–25], and recombinant proteins expressed from cloned thermophilic archaeal protein-coding genes are highly sought after by protein biochemists and structural biologists. Although no other introns have been identified in archaeal protein-coding genes, our results suggest this possibility in other protein-coding genes of Archaea. Therefore, we suggest that the confirmation of the native protein or cDNA sequences of such genes, in particular the highly useful thermophilic archaeal genes, should perhaps be performed before functional and/or structural analysis of these recombinant proteins proceeds (e.g. [31]).

In conclusion, this is the first report of introns in protein-coding genes in Archaea.

Acknowledgements: We thank Dr. Y. Sako (Kyoto University, Japan) and Dr. N. Kurosawa (Soka University, Japan) for kindly providing the *A. pernix* and *S. solfataricus* cells, respectively.

References

[1] Edgell, D.R., Belfort, M. and Shub, D.A. (2000) J. Bacteriol. 182, 5281–5289.
[2] Daniels, C.J., Gupta, R. and Doolittle, W.F. (1985) J. Biol. Chem. 260, 3132–3134.
[3] Kjerms, J. and Garrett, R.A. (1985) Nature 318, 675–677.
[4] Thompson, L.D. and Daniels, C.J. (1988) J. Biol. Chem. 263, 17951–17959.
[5] Kjems, J., Jensen, J., Olesen, T. and Garrett, R.A. (1989) Can. J. Microbiol. 35, 210–214.

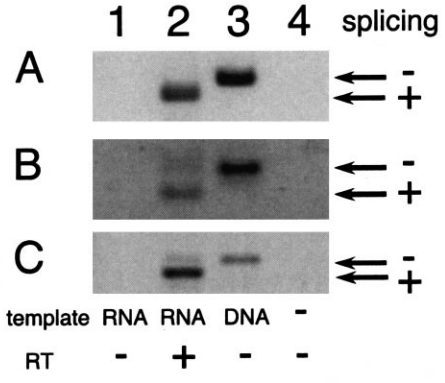

Fig. 3. Products of reverse transcriptase (RT)-mediated and genomic PCR (lanes 2 and 3, respectively) amplified with primers specific to the flanking region of the predicted introns of the *CBF5* genes of (A) *A. pernix*, (B) *S. solfataricus* and (C) *S. tokodaii*, run through a 2% agarose gel (A,C) and a 3% Nusieve 3:1 agarose gel (B). Lane 1, same as lane 2 but without the treatment with reverse transcriptase; lane 4, same as lane 3 but without template DNA.

[6] Thompson, L.D. and Daniels, C.J. (1990) J. Biol. Chem. 265, 18104–18111.
[7] Kjems, J. and Garrett, R.A. (1991) Proc. Natl. Acad. Sci. USA 88, 439–443.
[8] Lykke-Andersen, J., Aagaard, C., Semionenkov, M. and Garrett, R.A. (1997) Trends Biochem. Sci. 22, 326–331.
[9] Garrett, R.A., Dalgaard, J., Larsen, N., Kjems, J. and Mankin, A.S. (1991) Trends Biochem. Sci. 16, 22–26.
[10] Armbruster, D.W. and Daniels, C.J. (1997) J. Biol. Chem. 272, 19758–19762.
[11] Lykke-Andersen, J. and Garrett, R.A. (1997) EMBO J. 16, 6290–6300.
[12] Zebarjadian, Y., King, T., Fournier, M.J., Clarke, L. and Carbon, J. (1999) Mol. Cell. Biol. 19, 7461–7472.
[13] Sako, Y., Nomura, N., Uchida, A., Ishida, Y., Morii, H., Koga, Y., Hoaki, T. and Maruyama, T. (1996) Int. J. Syst. Bacteriol. 46, 1070–1077.
[14] Brock, T.D., Brock, K.M., Belly, R.T. and Weiss, R.L. (1972) Arch. Mikrobiol. 84, 54–68.
[15] Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (1987) Current Protocols in Molecular Biology, John Wiley and Sons, New York.
[16] Kawarabayasi, Y., Hino, Y., Horikawa, H., Yamazaki, S., Haikawa, Y., Jin-no, K., Takahashi, M., Sekine, M., Baba, S., Ankai, A., Kosugi, H., Hosoyama, A., Fukui, S., Nagai, Y., Nishijima, K., Nakazawa, H., Takamiya, M., Masuda, S., Funahashi, T., Tanaka, T., Kudoh, Y., Yamazaki, J., Kushida, N., Oguchi, A., Aoki, K., Kubota, K., Nakamura, Y., Nomura, N., Sako, Y. and Kikuchi, H. (1999) DNA Res. 6, 83–101.
[17] She, Q., Singh, R.K., Confalonieri, F., Zivanovic, Y., Allard, G., Awayez, M.J., Chan-Weiher, C.C.-Y., Clausen, I.G., Curtis, B.A., De Moors, A., Erauso, G., Fletcher, C., Gordon, P.M.K., Heikamp-de Jong, I., Jeffries, A.C., Kozera, C.J., Medina, N., Peng, X., Thi-Ngoc, H.P., Redder, P., Schenk, M.E., Theriault, C., Tolstrup, N., Charlebois, R.L., Doolittle, W.F., Duguet, M., Gaasterland, T., Garrett, R.A., Ragan, M.A., Sensen, C.W. and der Oost, J.V. (2001) Proc. Natl. Acad. Sci. USA 98, 7835–7840.
[18] Kawarabayasi, Y., Hino, Y., Horikawa, H., Jin-no, K., Takahashi, M., Sekine, M., Baba, S., Ankai, A., Kosugi, H., Hosoyama, A., Fukui, S., Nagai, Y., Nishijima, K., Otsuka, R., Nakazawa, H., Takamiya, M., Kato, Y., Yoshizawa, T., Tanaka, T., Kudoh, Y., Yamazaki, J., Kushida, N., Oguchi, A., Aoki, K., Masuda, S., Yanagii, M., Nishimura, M., Yamagishi, A., Oshima, T. and Kikuchi, H. (2001) DNA Res. 8, 123–140.
[19] Bult, C.J., White, O., Olsen, G.J., Zhou, L., Fleischmann, R.D., Sutton, G.G., Blake, J.A., FitzGerald, L.M., Clayton, R.A., Gocayne, J.D., Kerlavage, A.R., Dougherty, B.A., Tomb, J.-F., Adams, M.D., Reich, C.I., Overbeek, R., Kirkness, E.F., Weinstock, K.G., Merrick, J.M., Glodek, A., Scott, J.D., Geoghagen, N.S., Weidman, J.F., Fuhrmann, J.L., Nguyen, D.T., Utterback, T., Kelley, J.M., Peterson, J.D., Sadow, P.W., Hanna, M.C., Cotton, M.D., Hurst, M.A., Roberts, K.M., Kaine, B.B., Borodovsky, M., Klenk, H.P., Fraser, C.M., Smith, H.O., Woese, C.R. and Venter, J.C. (1996) Science 273, 1058–1073.
[20] Smith, D.R., Doucette-Stamm, L.A., Deloughery, C., Lee, H.-M., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., Harrison, D., Hoang, L., Keagle, P., Lumm, W., Pothier, B., Qiu, D., Spadafora, R., Vicare, R., Wang, Y., Wierzbowski, J., Gibson, R., Jiwani, N., Caruso, A., Bush, D., Safer, H., Patwell, D., Prabhakar, S., McDougall, S., Shimer, G., Goyal, A., Pietrovski, S., Church, G.M., Daniels, C.J., Mao, J.-I., Rice, P., Nolling, J. and Reeve, J.N. (1997) J. Bacteriol. 179, 7135–7155.
[21] Klenk, H.P., Clayton, R.A., Tomb, J.-F., White, O., Nelson, K.E., Ketchum, K.A., Dodson, R.J., Gwinn, M., Hickey, E.K., Peterson, J.D., Richardson, D.L., Kerlavage, A.R., Graham, D.E., Kyrpides, N.C., Fleischmann, R.D., Quackenbush, J., Lee, N.H., Sutton, G.G., Gill, S., Kirkness, E.F., Dougherty, B.A., McKenney, K., Adams, M.D., Loftus, B., Peterson, S., Reich, C.I., McNeil, L.K., Badger, J.H., Glodek, A., Zhou, L., Overbeek, R., Gocayne, J.D., Weidman, J.F., McDonald, L., Utterback, T., Cotton, M.D., Spriggs, T., Artiach, P., Kaine, B.P., Sykes, S.M., Sadow, P.W., D'Andrea, K.P., Bowman, C., Fujii, C., Garland, S.A., Mason, T.M., Olsen, G.J., Fraser, C.M., Smith, H.O., Woese, C.R. and Venter, J.C. (1997) Nature 390, 364–370.
[22] Kawarabayasi, Y., Sawada, M., Horikawa, H., Haikawa, Y., Hino, Y., Yamamoto, S., Sekine, M., Baba, S., Kosugi, H., Hosoyama, A., Nagai, Y., Sakai, M., Ogura, K., Otuka, R., Nakazawa, H., Takamiya, M., Ohfuku, Y., Funahashi, T., Tanaka, T., Kudoh, Y., Yamazaki, J., Kushida, N., Oguchi, A., Aoki, K., Nakamura, Y., Robb, T.F., Horikoshi, K., Masuchi, Y., Shizuya, H. and Kikuchi, H. (1998) DNA Res. 5, 55–76.
[23] Kawashima, T., Yamamoto, Y., Aramaki, H., Nunoshiba, T., Kawamoto, T., Watanabe, K., Yamazaki, M., Kanehori, K., Amano, N., Ohya, Y., Makino, K. and Suzuki, M. (1999) Proc. Jpn. Acad. 75, 213–218.
[24] Ruepp, A., Graml, W., Santos-Martinez, M.L., Koretke, K.K., Volker, C., Mewes, H.W., Frishman, D., Stocker, S., Lupas, A.N. and Baumeister, W. (2000) Nature 407, 508–513.
[25] Ng, W.V., Kennedy, S.P., Mahairas, G.G., Berquist, B., Pan, M., Shukla, H.D., Lasky, S.R., Baliga, N.S., Thorsson, V., Sbrogna, J., Swartzell, S., Weir, D., Hall, J., Dahl, T.A., Welti, R., Goo, Y.A., Leithauser, B., Keller, K., Cruz, R., Danson, M.J., Hough, D.W., Maddocks, D.G., Jablonski, P.E., Krebs, M.P., Angevine, C.M., Dale, H., Isenbarger, T.A., Peck, R.F., Pohlschroder, M., Spudich, J.L., Jung, K.W., Alam, M., Freitas, T., Hou, S., Daniels, C.J., Dennis, P.P., Omer, A.D., Ebhardt, H., Lowe, T.M., Liang, P., Riley, M., Hood, L. and DasSarma, S. (2000) Proc. Natl. Acad. Sci. USA 97, 12176–12181.
[26] Koonin, E.V. (1996) Nucleic Acids Res. 24, 2411–2415.
[27] Watanabe, Y. and Gray, M.W. (2000) Nucleic Acids Res. 28, 2342–2352.
[28] Mathews, D.H., Sabina, J., Zuker, M. and Turner, D.H. (1999) J. Mol. Biol. 288, 911–940.
[29] Zuker, M., Mathews, D.H. and Turner, D.H. (1999) in: RNA Biochemistry and Biotechnology (Barciszewski, J. and Clark, B.F.C., Eds.), NATO ASI Series, pp. 11–43, Kluwer Academic Publishers, Dordrecht.
[30] Nomura, N., Sako, Y. and Uchida, A. (1998) J. Bacteriol. 180, 3635–3643.
[31] Christendat, D., Yee, A., Dharamsi, A., Kluger, Y., Savchenko, A., Cort, J.R., Booth, V., Mackereth, C.D., Saridakis, V., Ekiel, I., Kozlov, G., Maxwell, K.L., Wu, N., McIntosh, L.P., Gehring, K., Kennedy, M.A., Davidson, A.R., Pai, E.F., Gerstein, M., Edwards, A.M. and Arrowsmith, C.H. (2000) Nat. Struct. Biol. 7, 903–909.
[32] Kleman-Leyer, K., Armbruster, D.W. and Daniels, C.J. (1997) Cell 89, 839–847.