JOURNAL OF BACTERIOLOGY, Sept. 2004, p. 6316–6319 Vol. 186, No. 18
0021-9193/04/$08.00+0 DOI: 10.1128/JB.186.18.6316–6319.2004

Bacterial Thymidylate Synthase with Intein, Group II Intron, and Distinctive ThyX Motifs

Xiang-Qin Liu* and Jing Yang

Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia, Canada

Received 8 April 2004/Accepted 25 May 2004

The ThyX class of thymidylate synthases was previously characterized by a common ThyX motif, $RHRX_7S$. We report bacterial ThyX sequences having distinctive ThyX motifs, suggesting a more general ThyX motif, $R/THRX_{7-8}S$. One ThyX sequence has an intein in its ThyX motif that was shown to do protein splicing and a group II intron in its gene, suggesting a hot spot for these self-splicing mobile elements.

The recently discovered ThyX class of thymidylate synthases is not similar to ThyA in its structure (16) or in its reductive mechanism (10, 19), revealing complexity in the evolution of thymidine production in DNA synthesis. ThyX exists in organisms lacking ThyA and shows a sporadic phylogenetic distribution indicating lateral gene transfers (19). Its presence in many pathogenic bacteria and absence in humans make ThyX an attractive target for potential antibacterial drugs. To realize this potential, knowledge is needed regarding important structural elements including the common ThyX motif found in previously known ThyX sequences.

Inteins and introns are rare and have not been found previously with ThyX. An intein is a protein intervening sequence that can self-excise and concomitantly splice together its flanking extein sequences (21). Many inteins also harbor an endonuclease domain that initiates intein homing, which confers genetic mobility on the intein (3, 5) and explains its sporadic phylogenetic distribution (20). Group II introns are another type of self-splicing mobile element, and most bacterial group II introns encode a reverse transcriptase-like (RTL) protein (23) that assists intron splicing and mobility, including retrotransposition (2, 11, 12). Group II introns are believed to be evolutionary ancestors of nuclear spliceosomal introns (4), but they are strongly excluded from conserved protein-coding genes in bacteria (6–8), although some bacteriophage genes encode both inteins and group I introns (9, 13, 14). Surprisingly, a bacterial ribonucleotide reductase (RIR)-encoding gene was found recently to encode multiple inteins and group II introns (18). To explore and understand this new phenomenon, we searched for similar inteins and introns in related genes.

An intein- and intron-encoding *thyX* gene was found during a BLAST search (1) of the GenBank database. This gene is from the oceanic N_2-fixing cyanobacterium *Trichodesmium erythraeum*. As illustrated in Fig. 1, the three exon and extein coding sequences are 196, 80, and 444 bp long, respectively, and they together predicted a 240-amino-acid ThyX sequence that is very similar to known ThyX sequences. The intron was identified by its strong sequence similarity to known group II introns in this organism, which included the *T.er*.I4 intron in an RIR-encoding gene (18) and the *Tr.e*.I1 intron in an intergenic sequence (6). It is more than 80% identical to the other introns in the ~680-nucleotide folded region (data not shown), although it lacks the RTL coding sequence present in domain IV of the other introns.

The intein was recognized by its intein sequence motifs (Fig. 2), and its boundaries were readily identified through comparisons with inteinless ThyX sequences. The intein clearly has sequence motifs (A, B, F, and G) for a splicing domain, but it either lacks or has incomplete sequence motifs (C, D, E, and H) for a homing endonuclease domain. It showed less than 15% sequence identity and no insertion site similarity to other known inteins. For example, it showed 14% sequence identity and 28% sequence similarity to the *Synechocystis* sp. strain PCC6803 DnaB intein. Nevertheless, it showed efficient protein splicing in a recombinant protein in *Escherichia coli* (Fig. 3).

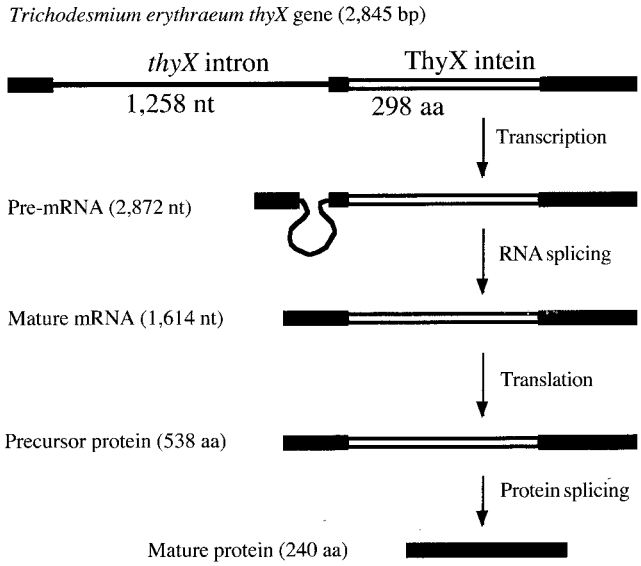

FIG. 1. Illustration of the *T. erythraeum thyX* gene and predicted expression products. Black boxes represent the three exons and exteins. aa, amino acids; nt, nucleotides.

* Corresponding author. Mailing address: Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia B3H 4H7, Canada. Phone: (902) 494-1208. Fax: (902) 494-1355. E-mail: pxqliu@dal.ca.

```
ThyX    CLSGNTKVRFRYSSSSQEAKYYEETIEKLANLWHYGSKNQYTSKDAKCMQENISSRNIFTLDTQTNQIVSSKITNIYINGEKETYTIKTVSGKEI
DnaB    CISGDS--LISLASTGKRVSIKDLLDEKDFEIWAINEQTMKLESAKVSRVFCTGKKLVYILKTR-------------------------LGRTI
            A

ThyX    RATLEHQFWTNQGWKRLKDFNNSTQL-----------------------------------------------------------------
DnaB    KATANHRFLTIDGWKRLDELSLKEHIALPRKLESSSLQLMSDEELGLLGHLIGDGCTLPRHAIQYTSNKIELAEKVVELAKAVFGDQINPRISQE
            B                                   C

ThyX    ---------CEVQLAGNKVTPQEVKFLEKEMFN----EKWIPVRNYDG---------YKIYSLDILNSYLIQKENK--------EHSRSKSKNCL
DnaB    RQWYQVYIPASYRLTHNKKNPITKWLENLDVFGLRSYEKFVPNQVFEQPQRAIAIFLRHLWSTDGCVKLIVEKSSRPVAYYATSSEKLAKDVQSL
                                            D                   E                           H

ThyX    LPNLNYGIRYDRLSEVNINR--------------LVMENLKLLGEYKELEVRHLN-------ENSFNNKPKKFAWNS------------------
DnaB    LLKLGINARLSKISQNGKGRDNYHVTITGQADLQIFVDQIGAVDKDKQASVEEIKTHIAQHQANTNRDVIPKQIWKTYVLPQIQIKGITTRDLQM

ThyX    ------------------SKSDQIDNINNN----SFSDNSGVFVEIESIEKFGKEITYDLEVEHPEHNFIANGLVVHN
DnaB    RLGNAYCGTALYKHNLSRERAAKIATITQSPEIEKLSQSDIYWDSIVSITETGVEEVFDLTVPGP-HNFVANDIIVHN
                                                                F           G
```

FIG. 2. Intein sequence comparison. The *T. erythraeum* ThyX intein sequence is aligned with the sequence of the previously identified *Synechocystis* sp. strain PCC6803 DnaB intein, and putative intein sequence motifs (A through H) are underlined. Dashes represent gaps introduced to optimize the alignment.

The *T. erythraeum thyX* gene is only the second bacterial gene (except for phage genes) known to encode inteins and introns, following an RIR-encoding gene (18). It is not certain why these two genes appear to be hot spots for intein and intron insertions, which most likely involve intein homing and group II intron retrotransposition. The ThyX intein likely had a homing endonuclease domain and later lost it, on the basis of observation of putative remnants of this domain in the intein. The ThyX intron and the *T.er*.I4 intron of the RIR-encoding gene, showing strong sequence identity, are likely related through recent retrotransposition, and it is tempting to speculate that the RTL-less ThyX intron is assisted in retrotransposition by the RTL protein encoded in the *T.er*.I4 intron. It is interesting that the *thyX* gene and the RIR-encoding gene both encode proteins involved in nucleic acid metabolism, because such genes have been recognized as favored homes of inteins and introns for various reasons (8, 17). We further noticed that the ThyX intein is inserted in the conserved ThyX motif, which prompted further analysis of ThyX motifs.

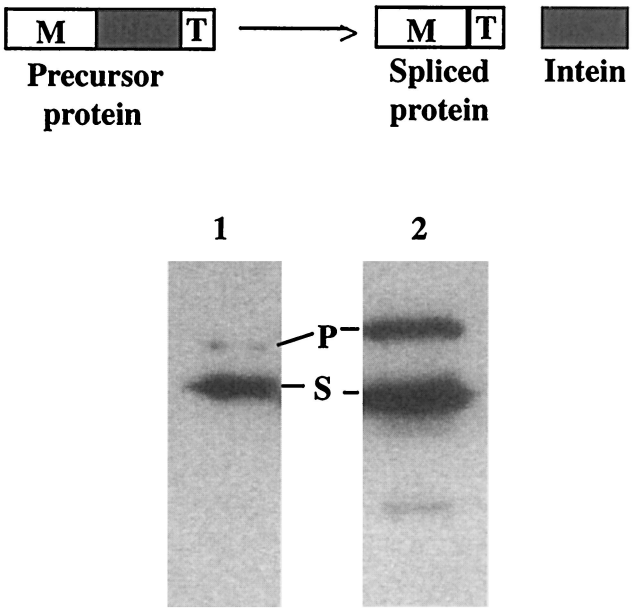

FIG. 3. Protein splicing of *T. erythraeum* ThyX intein. (Top) Schematic illustration of the fusion protein construct consisting of the maltose binding protein sequence (M), the intein sequence (gray box), and the thioredoxin sequence (T). (Bottom) Observation of protein splicing. Protein production and splicing were carried out with *E. coli*, and the resulting products were visualized by Western blotting with anti-thioredoxin antibody as previously described (22). Lanes: 1, protein splicing of the *Synechocystis* sp. strain PCC6803 DnaB mini-intein as a known standard; 2, protein splicing of the *T. erythraeum* ThyX intein. The letter P marks the position of the precursor protein, which matches closely the predicted sizes of 74 and 91 kDa in lanes 1 and 2, respectively. The letter S marks the position of the spliced protein, which matches closely the predicted sizes of 57 and 56 kDa in lanes 1 and 2, respectively.

New ThyX sequences having distinctive ThyX motifs were found through BLAST searches of the GenBank database. Although not having inteins or introns, they showed ThyX motifs that are similar to but distinct from the previously defined ThyX motif $RHRX_7S$ (Table 1 and Fig. 4). Most of the new ThyX sequences (*T. erythraeum*, *Nostoc punctiforme*, *Nostoc* sp. strain PCC7120, *Synechococcus* sp. strain WH8102, mycobacteriophage Bxz1, *Streptomyces coelicolor*, and *Streptomyces avermitilis* MA-4680) are readily identified because they

TABLE 1. Distinctive ThyX motifs in ThyX sequences of various organisms

ThyX motif	Organism	GenBank accession number
$THRX_8S$	*Trichodesmium erythraeum*	NZ_AABK02000059
	Nostoc punctiforme	ZP_00106512
	Nostoc sp. strain PCC7120	BAB72473
	Synechococcus sp. strain WH 8102	ZP_00114730
$THRX_7S$	*Thermosynechococcus elongatus*	NP_681946
	Thermophilic bacteriophage RM 378	NP_835722
$RHRX_8S$	Mycobacteriophage Bxz1	NP_818164
	Streptomyces coelicolor	086840
	Streptomyces avermitilis MA-4680	NP_823693
$RHRX_7S$	*Synechocystis* sp. strain PCC6803	NP_440394
	Helicobacter pylori J99	NP_224139
	Thermotoga maritima	NP_228259

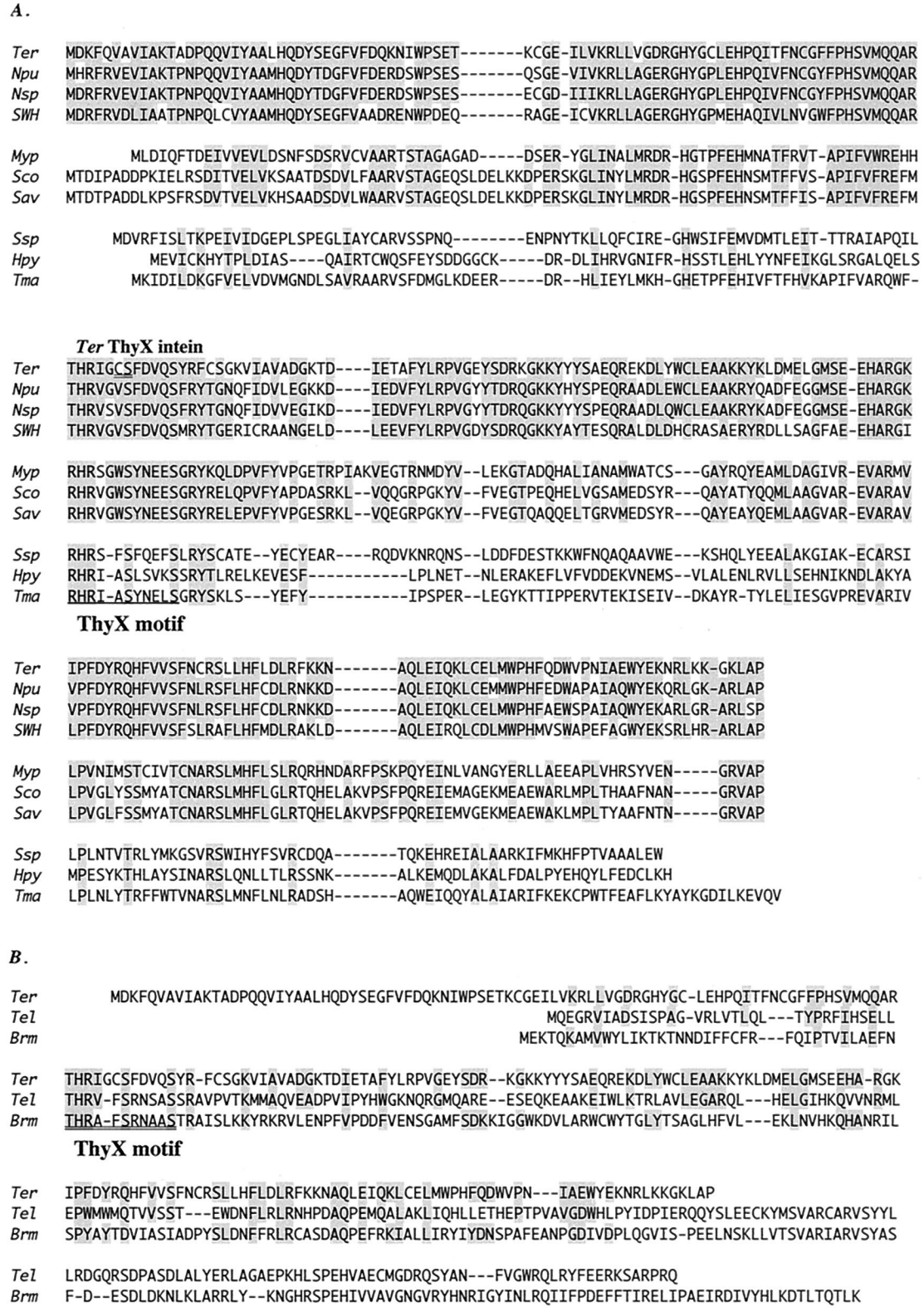

FIG. 4. ThyX protein sequence comparisons. In panel A, sequences are grouped according to ThyX motifs including $THRX_8S$ (*T. erythraeum* [*Ter*], *N. punctiforme* [*Npu*], *Nostoc* sp. strain PCC7120 [*Nsp*], and *Synechococcus* sp. strain WH8102 [*SWH*]), $RHRX_8S$ (*Nostoc* sp. strain PCC7120 [*Nsp*], *S. coelicolor* [*Sco*], and *S. avermitilis* MA-4680 [*Sav*]), and $RHRX_7S$ (*Synechocystis* sp. strain PCC6803 [*Ssp*], *H. pylori* J99 [*Hpy*], and *T. maritima* [*Tma*]). Positions conserved within each group are highlighted in gray. In panel B, positions conserved between the *T. erythraeum* sequence and the *T. elongatus* or thermophilic bacteriophage RM378 sequences are highlighted in gray. The intein insertion site in the *T. erythraeum* sequence is indicated by underlining of the two flanking residues (CS). The ThyX motif is also underlined, with the conserved R/THR and S residues in bold.

are more than 20% identical and 30% similar to the functionally identified *Helicobacter pylori* J99 ThyX protein and to the structurally determined *Thermotoga maritima* ThyX protein. ThyX sequences from the two thermophilic sources (*Thermosynechococcus elongatus* and thermophilic bacteriophage RM378), however, were less easy to identify because of their striking differences from other ThyX sequences at the N and C termini (Fig. 4B). Nevertheless, they are approximately 18% identical and 34% similar to *T. erythraeum* ThyX over four-fifths of the *T. erythraeum* ThyX sequence, which is quite significant in light of the generally low levels of sequence conservation among ThyX proteins. For example, the ThyX sequences of two cyanobacterial species (*T. erythraeum* and *Synechocystis* sp. strain PCC6803) are only 17% identical and 33% similar. The shortened N-terminal sequence of the *T. elongatus* and thermophilic bacteriophage RM378 ThyX proteins, relative to that of other ThyX proteins, may be compensated for by the extended C-terminal sequence, although the extended C-terminal sequence does not show similarity to that of other ThyX sequences. *T. elongatus* ThyX is also the only recognizable thymidylate synthase (which is functionally essential) that could be predicted from the complete genome sequence of this organism.

These ThyX sequences, when grouped by their ThyX motifs as in Table 1, showed higher sequence identities within a group than between groups. In particular, the ThyX sequences of four cyanobacteria (*T. erythraeum*, *N. punctiforme*, *Nostoc* sp. strain PCC7120, and *Synechococcus* sp. strain WH8102), having the same ThyX motif, $THRX_8S$, have more than 60% sequence identity with each other. But they show less than 20% sequence identity with the ThyX sequences of other cyanobacteria (e.g., *Synechocystis* sp. strain PCC6803) that have a different ThyX motif. The previously identified ThyX motif $RHRX_7S$ has the widest distribution, encompassing bacteria, archaea, and eucarya; therefore, it likely is the original ThyX motif present in the common ancestor of ThyX proteins. The new ThyX motifs $THRX_8S$ and $THRX_7S$, found in some cyanobacterial species, likely represent later divergence from the presumed original ThyX motif, and they could also have been acquired through lateral gene transfer. Interestingly, three of the four distinctive ThyX motifs have been found with bacteriophage, which could have facilitated lateral transfer of *thyX* genes.

We suggest a more general ThyX motif, $R/THRX_{7-8}S$, in order to accommodate all four of the distinctive ThyX motifs (Table 1). While this report was under review, others reported the functional characterization of two residues of the ThyX motif $RHRX_{7-8}S$ located at the catalytic site (15). The absolutely conserved S residue (S84) was shown to function as a nucleophile, and the first R residue (R74) was shown to participate in flavin adenine dinucleotide and dUMP binding. But our findings show that the first R residue is not absolutely conserved. However, when a ThyX motif begins with T instead of an R, it either has an R immediately before a T (in *T. erythraeum*, *N. punctiforme*, *Nostoc* sp. strain PCC7120, and *Synechococcus* sp. strain WH8102) or has a second R internally (in *T. erythraeum* and thermophilic bacteriophage RM378), and this could suggest functional replacement of alternative R residues.

This work was supported by research grants from the National Science and Engineering Research Council of Canada.

REFERENCES

1. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:**3389–3402.
2. **Belfort, M., and P. S. Perlman.** 1995. Mechanisms of intron mobility. J. Biol. Chem. **270:**30237–30240.
3. **Belfort, M., and R. J. Roberts.** 1997. Homing endonucleases: keeping the house in order. Nucleic Acids Res. **25:**3379–3388.
4. **Cavalier-Smith, T.** 1991. Intron phylogeny: a new hypothesis. Trends Genet. **7:**145–148.
5. **Cooper, A. A., and T. H. Stevens.** 1995. Protein splicing: self-splicing of genetically mobile elements at the protein level. Trends Biochem. Sci. **20:** 351–356.
6. **Dai, L., N. Toor, R. Olson, A. Keeping, and S. Zimmerly.** 2003. Database for mobile group II introns. Nucleic Acids Res. **31:**424–426.
7. **Dai, L., and S. Zimmerly.** 2002. Compilation and analysis of group II intron insertions in bacterial genomes: evidence for retroelement behavior. Nucleic Acids Res. **30:**1091–1102.
8. **Edgell, D. R., M. Belfort, and D. A. Shub.** 2000. Barriers to intron promiscuity in bacteria. J. Bacteriol. **182:**5281–5289.
9. **Ghim, S. Y., S. K. Choi, B. S. Shin, and S. H. Park.** 1998. An 8 kb nucleotide sequence at the 3′ flanking region of the sspC gene (184 degrees) on the Bacillus subtilis 168 chromosome containing an intein and an intron. DNA Res. **5:**121–126.
10. **Giladi, M., G. Bitan-Banin, M. Mevarech, and R. Ortenberg.** 2002. Genetic evidence for a novel thymidylate synthase in the halophilic archaeon Halobacterium salinarum and in Campylobacter jejuni. FEMS Microbiol. Lett. **216:**105–109.
11. **Ichiyanagi, K., A. Beauregard, S. Lawrence, D. Smith, B. Cousineau, and M. Belfort.** 2002. Retrotransposition of the Ll.LtrB group II intron proceeds predominantly via reverse splicing into DNA targets. Mol. Microbiol. **46:** 1259–1272.
12. **Lambowitz, A. M., M. G. Caprara, S. Zimmerly, and P. S. Perlman.** 1999. Group I and group II ribozymes as RNPs: clues to the past and guides to the future, p. 451–485. *In* R. F. Gesteland, T. R. Cech, and J. F. Atkins (ed.), The RNA world. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
13. **Lazarevic, V.** 2001. Ribonucleotide reductase genes of Bacillus prophages: a refuge to introns and intein coding sequences. Nucleic Acids Res. **29:**3212–3218.
14. **Lazarevic, V., B. Soldo, A. Dusterhoft, H. Hilbert, C. Mauel, and D. Karamata.** 1998. Introns and intein coding sequence in the ribonucleotide reductase genes of Bacillus subtilis temperate bacteriophage SPβ. Proc. Natl. Acad. Sci. USA **95:**1692–1697.
15. **Leduc, D., S. Graziani, G. Lipowski, C. Marchand, P. Le Marechal, U. Liebl, and H. Myllykallio.** 2004. Functional evidence for active site location of tetrameric thymidylate synthase X at the interphase of three monomers. Proc. Natl. Acad. Sci. USA **101:**7252–7257.
16. **Lesley, S. A., P. Kuhn, A. Godzik, A. M. Deacon, I. Mathews, A. Kreusch, G. Spraggon, H. E. Klock, D. McMullan, T. Shin, J. Vincent, A. Robb, L. S. Brinen, M. D. Miller, T. M. McPhillips, M. A. Miller, D. Scheibe, J. M. Canaves, C. Guda, L. Jaroszewski, T. L. Selby, M. A. Elsliger, J. Wooley, S. S. Taylor, K. O. Hodgson, I. A. Wilson, P. G. Schultz, and R. C. Stevens.** 2002. Structural genomics of the Thermotoga maritima proteome implemented in a high-throughput structure determination pipeline. Proc. Natl. Acad. Sci. USA **99:**11664–11669.
17. **Liu, X. Q.** 2000. Protein-splicing intein: genetic mobility, origin, and evolution. Annu. Rev. Genet. **34:**61–76.
18. **Liu, X. Q., J. Yang, and Q. Meng.** 2003. Four inteins and three group II introns encoded in a bacterial ribonucleotide reductase gene. J. Biol. Chem. **278:**46826–46831.
19. **Myllykallio, H., G. Lipowski, D. Leduc, J. Filee, P. Forterre, and U. Liebl.** 2002. An alternative flavin-dependent mechanism for thymidylate synthesis. Science **297:**105–107.
20. **Perler, F. B.** 2002. InBase: the intein database. Nucleic Acids Res. **30:**383–384.
21. **Perler, F. B., E. O. Davis, G. E. Dean, F. S. Gimble, W. E. Jack, N. Neff, C. J. Noren, J. Thorner, and M. Belfort.** 1994. Protein splicing elements: inteins and exteins—a definition of terms and recommended nomenclature. Nucleic Acids Res. **22:**1125–1127.
22. **Wu, H., M. Q. Xu, and X. Q. Liu.** 1998. Protein trans-splicing and functional mini-inteins of a cyanobacterial dnaB intein. Biochim. Biophys. Acta **1387:** 422–432.
23. **Zimmerly, S., G. Hausner, and X. Wu.** 2001. Phylogenetic relationships among group II intron ORFs. Nucleic Acids Res. **29:**1238–1250.