

## ACKNOWLEDGEMENTS

We thank D. Tautz for his constant support. Furthermore, we thank all members of the laboratory for their help in optimizing the cloning protocol. This work was supported by the EC program 'Conservation of genetic resources III.2 – Development of universal techniques for the rapid screening of genetic variation'; Contract No. BIO2-CT92-0476 and SFB 190.

## REFERENCES

- 1 Bauer, J. *et al.* (1992) *Strategies* 5, 62–64
- 2 Deininger, P.L. (1983) *Anal. Biochem.* 129, 216–223
- 3 Rassmann, K., Schlötterer, C. and Tautz, D. (1991) *Electrophoresis* 12, 113–118
- 4 Yanish-Perron, C., Viera, J. and Messing, J. (1985) *Gene* 33, 103–119

Contributed by Christian Schlötterer (schlotc@hp01.boku.ac.at) Institut für Tierzucht und Genetik, Veterinärmedizinische Universität Wien, 1210 Wien, Austria; and Christian Wolff (chwolff@zi.biologie.uni-muenchen.de), Zoologisches Institut, Luisenstrasse, 14, D-80333 München, Germany.

## LETTERS

### A new intein in cyanobacteria and its significance for the spread of inteins

Inteins are protein 'introns' encoded inside the polypeptide sequence of other proteins. The inteins splice out post-translationally by a proteolytic cleavage and ligation process. Inteins appear to autocatalyze their own excision and some are site-specific endonucleases<sup>1</sup>. Inteins are mobile genetic elements and at least one can home, that is, insert a copy of its DNA into its integration site in an intein-less allele<sup>2</sup>. Fifteen inteins have been found in various organisms, including mycobacteria<sup>3–6</sup>, thermophilic archaeobacteria<sup>7,8</sup>, yeast<sup>9,10</sup> and chloroplast of red alga<sup>11</sup>. Inteins are not very similar to one another<sup>5</sup>, but homologous sites in archaeobacterial DNA polymerases and in mycobacterial gyrase-A proteins<sup>6</sup> contain homologous inteins. However, the mycobacterial RecA proteins<sup>4</sup> and DNA polymerases also contain different inteins in different integration sites.

Searching for inteins in sequence databases<sup>5</sup> I have identified a 429 amino acid region of an open reading frame (ORF) from the thermophilic cyanobacterium *Synechocystis* sp. (Ssp)<sup>12</sup> (GenBank Accession No. D64003, ORF slr0833 positions 381–809). The region contains all of the known intein motifs<sup>5</sup> and can be confidently identified as an intein (expectant value<sup>13</sup> of  $7.8 \times 10^{-13}$  for finding the multiple conserved regions). This is

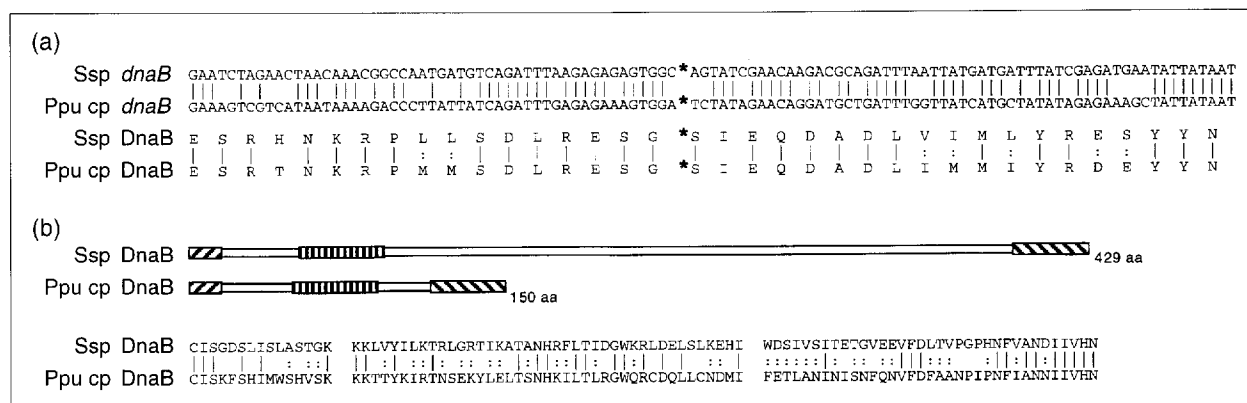
the first intein reported in cyanobacteria and the first eubacterial intein outside mycobacteria. The two regions flanking this intein have significant sequence similarity to prokaryotic replicative DnaB helicases (BLAST search<sup>14</sup> *P*-values of  $2.2 \times 10^{-22}$  to  $6.5 \times 10^{-130}$ ).

The Ssp DnaB intein has weak sequence similarity to known inteins. The sequence can be aligned end-to-end with the mycobacterial GyrA inteins and the *Mycobacterium tuberculosis* RecA intein with 22–25% amino acid identity, but is only similar to the conserved regions of other inteins. The Ssp DnaB intein and the DnaB intein from the chloroplast of the red alga *Porphyra purpurea* (Ppu) share only local sequence similarities and have distinctly different lengths. Nevertheless, both are found in the same point in the *dnaB* gene. The integration points are in the center of a region of 15 identical amino acids. However, nine of these have silent substitutions, including the two amino acids flanking the integration point (Fig. 1). The two DnaB inteins could be the result of a single integration event into a common ancestor of the cyanobacterial and chloroplast *dnaB* genes. Such an event was suggested to explain the presence of similar group I introns at an identical position in the leucine

tRNA gene of diverse chloroplasts and cyanobacteria<sup>15,16</sup>. Red alga chloroplasts probably originated from cyanobacteria<sup>17</sup> about 1.25 to 2.1 billion years ago<sup>18,19</sup>. This scenario implies that (1) inteins are extremely ancient and (2) the DnaB inteins survived in their hosts a remarkably long time. The different conservation of the DnaB hosts and the DnaB inteins (Fig. 1) can be the result of different selection pressures. DnaB replicative helicase is a multifunctional protein closely interacting with DNA and other proteins<sup>20</sup>. Selection on the DnaB helicases is probably linked to changes in the proteins with which they interact and to the evolution of the whole DNA replication process.

Conversely, inteins might be molecular parasites that merely need to ensure the transmission of their genes<sup>21</sup>. Rapid and efficient protein splicing is essential for minimal interference in the function of the protein host. Inteins can also increase their survival chances by homing into intein-less alleles, spreading in the population. The Ppu DnaB intein is the shortest intein (150 compared with 360–538 amino acids) and is missing the endonuclease-like motifs found in all other inteins<sup>5</sup>, while the Ssp DnaB intein has a typical length and all the characteristic motifs. It seems that over their long separate evolution the two inteins have opted for different survival strategies.

Alternatively, separate integration events might have led to the presence of different inteins in



**FIGURE 1.** Comparison of the two DnaB inteins. (a) The *Synechocystis* sp. (Ssp) and *Porphyra* (Ppu) inteins integration sites in the DnaB proteins. Bars mark identities, colons conserved amino acid substitutions and asterisks the integration sites. The two DnaB proteins can be aligned end-to-end (415 amino acids) with only three gaps, 37% amino acid identities and 79% amino acid identities and conserved substitutions. (b) Sequence similarity of the two inteins. The similar blocks in the two DnaB intein sequences are hatched and shown below the diagram. No other significantly similar regions were found in the two sequences using the Macaw multiple alignment program<sup>25</sup> and the Smith-Waterman optimal alignment algorithm<sup>26</sup>. The *P*-values for the block alignments are  $9.5 \times 10^{-1}$ ,  $1.1 \times 10^{-8}$  and  $6.8 \times 10^{-7}$  (Ref. 25). The GenBank sequence accession codes are D64003 for Ssp and U38804 for Ppu. Abbreviations: aa, amino acid; cp, chloroplast.

homologous hosts. Such events occurred in the RecA proteins and DNA polymerases, but at different integration sites. Separate integration at the same sites in the *dnaB* genes implies a particular susceptibility of the sites (such as cleavage by the same endonuclease). The difference in nucleotide sequences around the integration sites does not completely rule out this idea. Restriction endonucleases can cleave ambiguous targets and this was shown for a homing intein endonuclease<sup>2</sup>. It might also be that it is difficult to 'dislodge' the inteins from their particular integration site or that being in this position somehow assists the protein splicing.

Separate integration events seem to account for the presence of distinct group I introns at the same location in nuclear rDNA genes of different phyla<sup>22</sup>. It is not known if inteins confer any advantage to their hosts, but the separate integrations of inteins into RecA proteins were proposed as indication for selective advantage<sup>4</sup>. Such advantage might help to explain the very long persistence time of the DnaB inteins or the incentive for separate integrations. The hedgehog developmental regulatory proteins might be relevant to this issue. These proteins undergo autoproteolytic cleavage at a specific site<sup>23</sup> that is significantly similar to the N-terminal splice junction motif of inteins<sup>24</sup>. Experimental evidence suggests that

the autoprocessing of the *Drosophila* hedgehog protein regulate its range of action<sup>23</sup>. Similarly, some inteins might regulate the activity of their host proteins.

SHMUEL PIETROKOVSKI  
pietro@sparky.fhrc.org

Fred Hutchinson Cancer Research Center,  
1124 Columbia St, Seattle, WA 98104, USA.

#### Acknowledgement

The author is a Howard Hughes Medical Institute Fellow of the Life Sciences Research Foundation.

#### References

- Cooper, A.A. and Stevens, T.H. (1995) *Trends Biochem. Sci.* 20, 351–356
- Gimble, F.S. and Thorner, J. (1992) *Nature* 357, 301–306
- Davis, E.O. *et al.* (1992) *Cell* 71, 201–210
- Davis, E.O., Thangaraj, H.S., Brooks, P.C. and Colston, M.J. (1994) *EMBO J.* 13, 699–703
- Petrokovski, S. (1994) *Protein Sci.* 3, 2340–2350
- Fsihi, H., Vincent, V. and Cole, S.T. (1996) *Proc. Natl. Acad. Sci. U. S. A.* 93, 3410–3415
- Perler, F.B. *et al.* (1992) *Proc. Natl. Acad. Sci. U. S. A.* 89, 5577–5581
- Hodges, R.A., Perler, F.B., Noren, J.N. and Jack, W.E. (1992) *Nucleic Acids Res.* 20, 6153–6157
- Kane, P.M. *et al.* (1990) *Science* 250, 651–657
- Gu, H.H., Xu, J., Gallagher, M. and Dean, G.E. (1993) *J. Biol. Chem.* 268, 7372–7391
- Reith, M.E. and Munholland, J. (1995) *Plant Mol. Biol. Rep.* 13, 333–335

- Kaneko, T. *et al.* (1995) *DNA Res.* 2, 153–166
- Henikoff, S. and Henikoff, J.G. (1994) *Genomics* 19, 97–107
- Altschul, S.F. *et al.* (1990) *J. Mol. Biol.* 215, 403–410
- Kuhnel, M.G., Strickland, R. and Palmer, J.D. (1990) *Science* 250, 1570–1573
- Xu, M.Q. *et al.* (1990) *Science* 250, 1566–1570
- Gray, M.W. (1993) *Curr. Opin. Genet. Dev.* 3, 884–890
- Han, T-M. and Runnegar, B. (1992) *Science* 257, 232–235
- Doolittle, R.F. *et al.* (1996) *Science* 271, 470–477
- Kornberg, A. and Baker, T.A. (1992) *DNA Replication*, W.H. Freeman
- Hickey, D.A. (1994) *Trends Genet.* 10, 147–149
- Vader, A. *et al.* (1994) *Nucleic Acids Res.* 22, 4553–4559
- Porter, J.A. *et al.* (1995) *Nature* 374, 363–366
- Koonin, E.V. (1995) *Trends Biochem. Sci.* 20, 141–142
- Schuler, G.D., Altschul, S.F. and Lipman, D.J. (1991) *Protein Struct. Funct. Genet.* 9, 180–190
- Smith, T.F. and Waterman, M.S. (1981) *J. Mol. Biol.* 147, 195–197

**50% off all new student subscriptions!**

Did you know that as a student you are entitled to a **special discount** on a personal subscription to *Trends in Genetics*?

See the subscription order form for details.