Nucleic Acids Research, 2002, Vol. 30, No. 1 383–384

InBase: the Intein Database

Francine B. Perler*

New England Biolabs Inc., 32 Tozer Road, Beverly, MA 01915, USA

Received August 31, 2001; Accepted September 18, 2001

ABSTRACT

Inteins are self-catalytic protein splicing elements. InBase (http://www.neb.com/neb/inteins.html), the Intein Database and Registry, is a curated compilation of published and unpublished information about protein splicing. It presents general information as well as detailed data for each intein, including tabulated comparisons and a comprehensive bibliography. An intein-specific BLAST server is now available to assist in identifying new inteins.

INTRODUCTION

Inteins are in-frame intervening sequences that disrupt a host gene and its protein product; the host gene and its protein product are called exteins (1). Inteins are post-translationally excised from a protein precursor by a self-catalytic protein splicing mechanism (2,3). Consequently, two or more stable proteins [the intein(s) and the extein] are produced from a single gene. Extein ligation forming a native peptide bond and the presence of conserved intein motifs differentiate intein-mediated protein splicing from other post-translational processing events. Most inteins are bifunctional proteins with separate structural domains responsible for protein splicing and homing endonuclease activities. Over 75% of currently registered inteins contain either a DOD or H-N-H class homing endonuclease domain (4). Homing endonucleases initiate mobilization of the intein gene into the same site in a homologous host gene lacking the intervening sequence (5).

There are >115 inteins registered in InBase from Archaea, Bacteria and Eukarya. InBase (http://www.neb.com/neb/inteins.html) compiles information about inteins, often submitted by researchers prior to publication. Several subsets of data are tabulated for easy comparison including organism data, motif sequences, proximal insertion site sequences, selected properties and intein alleles (inteins present at the same insertion site in homologous genes from different species).

NEW DEVELOPMENTS

InBase has expanded since the 2000 NAR Database Issue (6). The InBase home page has been reorganized to move background information into a separate section called Intein Basics. This section is suitable for students and contains PDF files that allow users to download protein splicing figures. An intein-specific BLAST server is now available and the amino acid sequence is present in individual intein pages. The Identification of Intein section reflects the growing list of intein polymorphisms. A new splicing mechanism for inteins lacking an N-terminal nucleophile has been included in the Splicing Mechanism section (7).

The InBase BLAST server was added in response to requests from researchers and genome sequencing groups, since intein identification is not always trivial. Searches of general sequence databases often yield hits with very low scores and low probability values due to (i) the small size of mini-inteins (as few as 134 amino acids), (ii) the low level of sequence similarity even in conserved motifs and (iii) the high degree of polymorphisms in conserved splice junction residues. By limiting the search process to intein sequences, significant scores and *P*-values can be obtained. Since inteins are predominantly found in extein active sites and cofactor or substrate binding pockets, identification of inteins can potentially help locate these elements in uncharacterized proteins.

Several important advances have been reported in the protein splicing field. Most notable was the discovery of a non-canonical protein splicing pathway for inteins beginning with Ala, instead of Ser, Thr or Cys (7). The first crystal structure of an intein precursor shed light onto the amino acids that assist catalysis and the need for conformational changes to align nucleophiles during the sequential steps in the protein splicing pathway (8). Numerous protein engineering applications take advantage of the C-terminal α-thioester formed on target proteins purified from intein vectors (2,3,9–11), such as the commercially available IMPACT™ system (NEB). Papers using intein vectors for protein purification are not included in the bibliography unless they add to our understanding of intein technologies. A green (APP) label in the bibliography section highlights application papers.

ORGANIZATION OF THE DATABASE

Since few textbooks cover protein splicing, InBase provides background material suitable for classroom use. At the same time, detailed information is presented in a layered format with general discussions and tables pointing to more specific data. The InBase home page lists the accessible sections in the database:

1. Intein Basics
2. The Mechanism of Protein Splicing
3. The Intein Registry
4. Intein Motifs
5. Identifying Inteins
   A. Conserved Intein Features
   B. BLAST the InBase Sequence Database
6. Online Submission of Intein Data
7. The Intein Bibliography
8. Intein Links

*Tel: +1 978 927 5054; Fax: +1 978 921 1350; Email: perler@neb.com

The Intein Registry (section 3A) lists all known inteins sorted by Domain of Life, genus and species of the host organism, while section 3B sorts inteins by extein insertion site. Individual intein records contain detailed information about each intein, including insertion site sequence data, comments on unusual properties, submitter contact information and a reference list for each intein. Section 5 describes the criteria for intein identification, including a description of conserved motifs and polymorphisms. Intein data can be submitted confidentially or for immediate release using the online submission form or by email. References throughout the database are linked to the Bibliography section. The bibliography includes annotations for reviews, application papers, related papers and recent papers. PubMed hot links allow the reader to retrieve abstracts from the National Library of Medicine.

DATABASE AVAILABILITY AND CITATION

InBase can be found by clicking the Technical Resource button on the New England Biolabs Home Page (http://www.neb.com) or directly at http://www.neb.com/neb/inteins.html. Users of InBase are requested to cite this article when referencing the database.

ACKNOWLEDGEMENTS

I am grateful to my co-workers at NEB, Ellen M. Zaglakas and Ching Lin for help in maintaining InBase, Janos Posfai and Tamas Vincze for developing and maintaining the InBase BLAST server, and to all the intein workers who have submitted their published and unpublished data, especially Shmuel Pietrokovski.

REFERENCES

1. Perler,F.B., Davis,E.O., Dean,G.E., Gimble,F.S., Jack,W.E., Neff,N., Noren,C.J., Thorner,J. and Belfort,M. (1994) Protein splicing elements: inteins and exteins—a definition of terms and recommended nomenclature. *Nucleic Acids Res.*, **22**, 1125–1127.
2. Noren,C.J., Wang,J. and Perler,F.B. (2000) Dissecting the chemistry of protein splicing and its applications. *Angew. Chem. Int. Ed.*, **39**, 450–466.
3. Paulus,H. (2001) Inteins as enzymes. *Bioorg. Chem.*, **29**, 119–129.
4. Belfort,M. and Roberts,R.J. (1997) Homing endonucleases: keeping the house in order. *Nucleic Acids Res.*, **25**, 3379–3388.
5. Gimble,F.S. and Thorner,J. (1992) Homing of a DNA endonuclease gene by meiotic gene conversion in *Saccharomyces cerevisiae*. *Nature*, **357**, 301–306.
6. Perler,F.B. (2000) InBase, the Intein Database. *Nucleic Acids Res.*, **28**, 344–345.
7. Southworth,M.W., Benner,J. and Perler,F.B. (2000) An alternative protein splicing mechanism for inteins lacking an N-terminal nucleophile. *EMBO J.*, **19**, 5019–5026.
8. Poland,B.W., Xu,M.Q. and Quiocho,F.A. (2000) Structural insights into the protein splicing mechanism of PI-*Sce*I. *J. Biol. Chem.*, **275**, 16408–16413.
9. Blaschke,U.K., Silberstein,J. and Muir,T.W. (2000) Protein engineering by expressed protein ligation. *Methods Enzymol.*, **328**, 478–496.
10. de Grey,A.D. (2000) Mitochondrial gene therapy: an arena for the biomedical use of inteins. *Trends Biotechnol.*, **18**, 394–399.
11. Perler,F.B. and Adam,E. (2000) Protein splicing and its applications. *Curr. Opin. Biotechnol.*, **11**, 377–383.