Nucleic Acids and Molecular Biology 16

Homing Endonucleases and Inteins

Bearbeitet von
Marlene Belfort, Barry L. Stoddard, David W. Wood, Vicky Derbyshire

1. Auflage 2008. Taschenbuch. xix, 377 S. Paperback
ISBN 978 3 540 85235 3
Format (B x L): 15,5 x 23,5 cm
Gewicht: 611 g

Back to Basics: Structure, Function, Evolution and Application of Homing Endonucleases and Inteins

Marlene Belfort

It is a profound and necessary truth that the deep things in science are not found because they are useful; they are found because it was possible to find them.

Robert Oppenheimer 1904–1967

1 Introduction: Back to Basics

Oppenheimer's words resonate with this book's theme, which is how both applied and theoretical science can emanate from answers to basic questions – whether they are being asked in model organisms or in test tubes – about molecular structure and mechanism. Thus, fundamental work on DNA, RNA and proteins, which encode or constitute homing endonucleases and inteins, is leading to refined theories of evolution in prokaryotes and eukaryotes on the one hand, and the development of laboratory tools and health-care reagents on the other.

Homing endonucleases and inteins, sometimes referred to as "protein introns", are linked at many levels. First, homing endonucleases are frequently encoded by introns that self-splice at the RNA level, in analogy to inteins that self-splice at the protein level. Second, homing endonucleases similar to those encoded by introns are often found embedded within and co-translated with inteins. Third, both types of intervening sequence are mobile elements, capable of movement from genome to genome. Fourth, the endonuclease component of both introns and inteins imparts their mobility. Fifth, each of these mobile intervening sequences is thought to have originated from invasion of the gene encoding the self-splicing element, the intron or intein, by

M. Belfort (e-mail: belfort@wadsworth.org)
Wadsworth Center, New York State Department of Health, Center for Medical Sciences, 150 New Scotland Avenue, Albany, New York 12208, USA

Nucleic Acids and Molecular Biology, Vol. 16
Marlene Belfort et al. (Eds.)
Homing Endonucleases and Inteins

an endonuclease gene, the primordial mobile element (Fig. 1). The final unifying theme is the exploitation of introns, inteins and homing endonucleases by chemists, geneticists, structural biologists and engineers, to generate reagents and tools that are useful in basic research, biotechnology and medicine. This chapter serves as an introduction to the volume entitled *Homing Endonucleases and Inteins*, which provides a wonderful illustration of the point that fundamental studies of structure and mechanism fuel evolutionary theory and technology development alike.

2 What Is a Homing Endonuclease?

Homing endonucleases are rare-cutting enzymes that are most often encoded by introns or inteins, but they can also be free-standing, occurring between genes. The genesis of the homing endonuclease field dates back to 1970, with the observation, in genetic crosses between yeast mitochondria, of a significant polarity of recombination for markers of an rRNA gene (Dujon, this Vol.). In 1985, this phenomenon became attributable to an intron-encoded homing endonuclease that initiated recombination within the rRNA gene. Minimally, homing endonucleases are protein enzymes that make a site-specific double-strand break (DSB) at the "homing" site in intron-less or intein-less alleles, thereby initiating a gene conversion event through which the intron or intein is copied into the break site (Fig. 2A, B; reviewed by Chevalier and Stoddard 2001; Belfort et al. 2002; Dujon, this Vol.). For the group I and archaeal intron endonucleases and inteins, the recombinogenic ends created at the DSB engage in a strictly DNA-dependent recombination process that duplicates

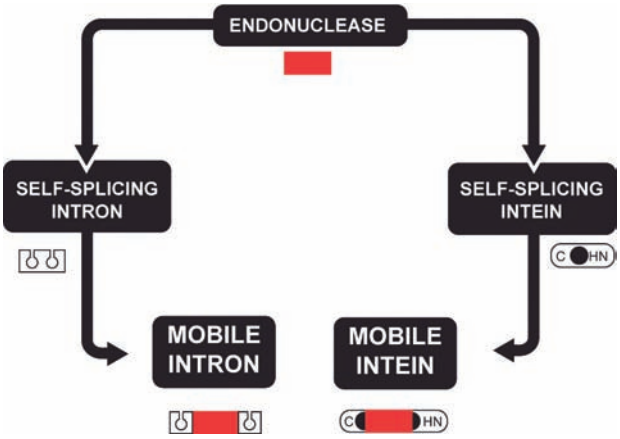

Fig. 1. Evolution of mobile introns and inteins. Endonuclease genes (*red*) are proposed to have invaded DNA encoding self-splicing introns or inteins, to generate mobile genetic elements. (Dassa and Pietrokovski, this Vol.)

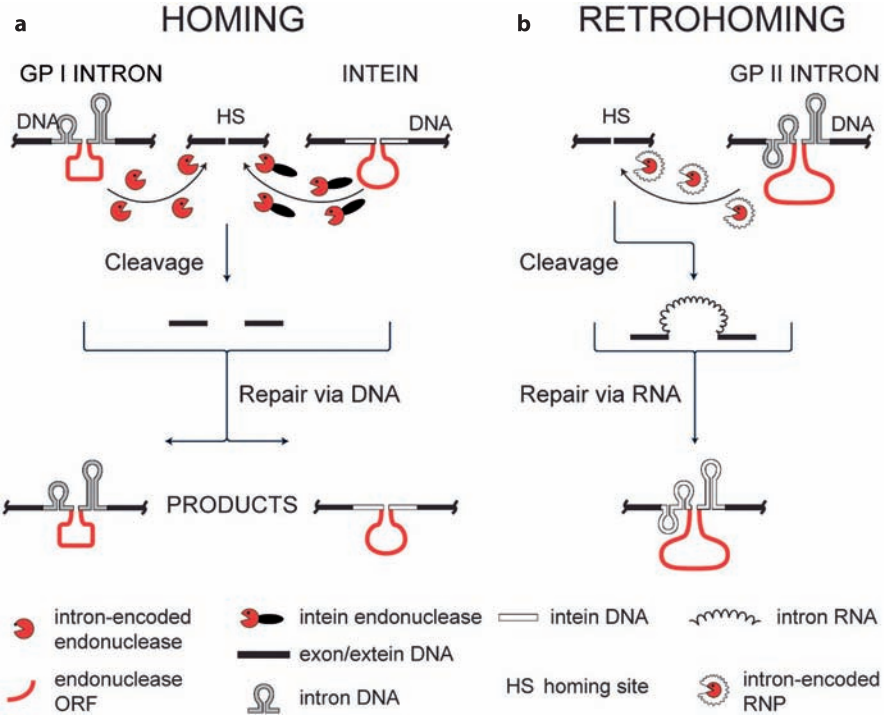

Fig. 2. Mobility of introns and inteins. **a** DNA-based homing of group I introns and inteins. The intron or intein endonuclease cleaves the homing site of a cognate intron- or intein-less allele. Gene conversion repairs the break to generate intron- or intein-containing products (Dujon, this Vol.). **b** Retrohoming of a group II intron. In this case, the homing site DNA is invaded by intron RNA, and the opposite strand is cleaved by an intron-encoded protein, which is part of an RNP complex. The intron is copied into cDNA to generate the intron-containing product (Lambowitz et al., this Vol.).

the intron or intein (Fig. 2A). The group II intron-encoded proteins are more complex, forming a ribonucleoprotein (RNP) particle with the intron RNA (Fig. 2B). The intron invades the DNA sense strand (mRNA-like strand) by reverse-splicing, whereas the endonuclease domain of the protein nicks the antisense strand. The intron acquisition event is completed with a cDNA copy of the intron, in a process termed retrohoming (Lambowitz et al., this Vol.).

The homing endonucleases fall within four families, characterized by the sequence motifs LAGLIDADG (Caprara and Waring, this Vol.; Chevalier et al., this Vol.; Dujon, this Vol.; Haber and Wolfe, this Vol), GIY-YIG (Edgell, this Vol.; Van Roey and Derbyshire, this Vol.), His-Cys box (Galburt and Jurica, this Vol.; Keeble et al., this Vol.) and HNH (Keeble et al., this Vol.). However, recent structural data support the hypothesis that the His-Cys box and HNH

enzymes share features at their active sites, and should be considered a single family, called ββα-Me (Keeble et al., this Vol.). All of the homing endonucleases recognize lengthy asymmetric or pseudosymmetric DNA sequences, ranging from a 14-bp homing site for I-DmoI, a member of the LAGLIDADG family, to a 40-bp site for I-TevI, a member of the GIY-YIG family (described in Chevalier et al., this Vol.; Van Roey and Derbyshire, this Vol., respectively). In addition, the enzymes exhibit varying degrees of sequence tolerance, with I-TevI again being exceptional, in this case in its promiscuity (Van Roey and Derbyshire, this Vol.). The conserved sequences and substrate-recognition characteristics stand in contrast to the properties of the restriction endonucleases, which usually recognize short palindromic DNA sequences with absolute sequence specificity. Thus, while both types of endonuclease cleave DNA, they have evolved independently. A growing understanding of the structures and mechanisms of some of these enzymes is facilitating their engineering for genomic applications (Dujon, this Vol.; Gimble, this Vol.).

3 What Is an Intein?

The discovery of inteins and protein splicing represented a breakthrough in our concept of the catalytic repertoire of proteins and of post-translational modification (Perler, this Vol.). Conserved residues at the intein–extein junctions facilitate splicing (Mills and Paulus, this Vol.; Perler, this Vol.). Several inteins are bifunctional proteins that not only catalyze protein splicing, but also function as endonucleases, to initiate homing of the intein gene (Figs. 1 and 2A). Additionally, several inteins have motifs suggesting an evolutionary relationship to intron-encoded homing endonucleases. The endonuclease and the protein-splicing component are genetically, structurally, and functionally separable (Dassa and Pietrokovski, this Vol.; Moure and Quiocho, this Vol.; Perler, this Vol.), supporting the hypothesis that the endonuclease genes invaded the genes of these self-splicing elements, which provided safe havens, while themselves acquiring mobile properties (Fig. 1).

4 Inteins and Homing Endonucleases as Molecular Mosaics

The invasion of self-splicing introns and inteins by endonuclease genes appears to have occurred multiple times, given that these elements encode endonucleases of different families and that some endonuclease genes of the same family emanate from different positions of group I introns. It appears that some of these endonucleases then adapted to function in other process-

es, e.g., repression of transcription (Van Roey and Derbyshire, this Vol.), and promotion of splicing through acquisition of maturase activity (Caprara and Waring, this Vol.).

Inteins share an ancestry with metazoan hedgehog proteins, which undergo a self-cleavage reaction. The common Hint (*hedgehog/intein*) domain is a structural unit with a mechanistic identity (Dassa and Pietrokovski, this Vol.). It is apparent that composite elements like mobile introns, inteins and hedgehog proteins have interchanged functional domains in the course of evolution.

Endonucleases themselves have evolved specificity by fusion of a catalytic domain, containing the conserved motif, with variant DNA-binding domains, e.g., the GIY-YIG and HNH endonucleases (Van Roey and Derbyshire, this Vol.). The modular nature of these enzymes is further illustrated for I-TevI, in which the DNA-binding domain is itself an assembly of small DNA-binding units, some of which are present in other homing endonucleases. These enzymes have evolved a broad range of binding specificities, through the shuffling of catalytic cartridges with DNA-binding cassettes. We can only speculate as to how such molecular mosaics are formed. The most popular view is that proposed for hybrid bacteriophage genomes, in which "illegitimate recombination takes place quasi-randomly along the recombining genomes, generating an unholy mélange of recombinant types" (Pedulla et al. 2003). This sloppy, non-homologous recombination would generate a mound of genetic junk, with only a miniscule number of recombinants being selected, on the basis of their function and/or viability.

A lingering question is whether homing endonucleases and their genes are maintained specifically to promote their own selfish lifestyles and that of their host elements (introns and inteins), or whether they additionally serve some useful function for the organism. While their invasiveness and success as selfish intruders are undisputed, a potential advantage to the host organism has been observed, in experiments with phage T4 and its relative T2 (Edgell, this Vol.). Here, GIY-YIG homing endonucleases act to promote the spread of genes from their host organism to its relatives. This is a satisfying observation, considering that 8% of the phage T4 genome comprises endonuclease genes. Another "useful" homing enzyme is HO endonuclease, the first member of the LAGLIDADG family to be discovered, and the first shown to make a DSB. This intriguing enzyme catalyzes mating-type switching in yeast (Haber and Wolfe, this Vol.).