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

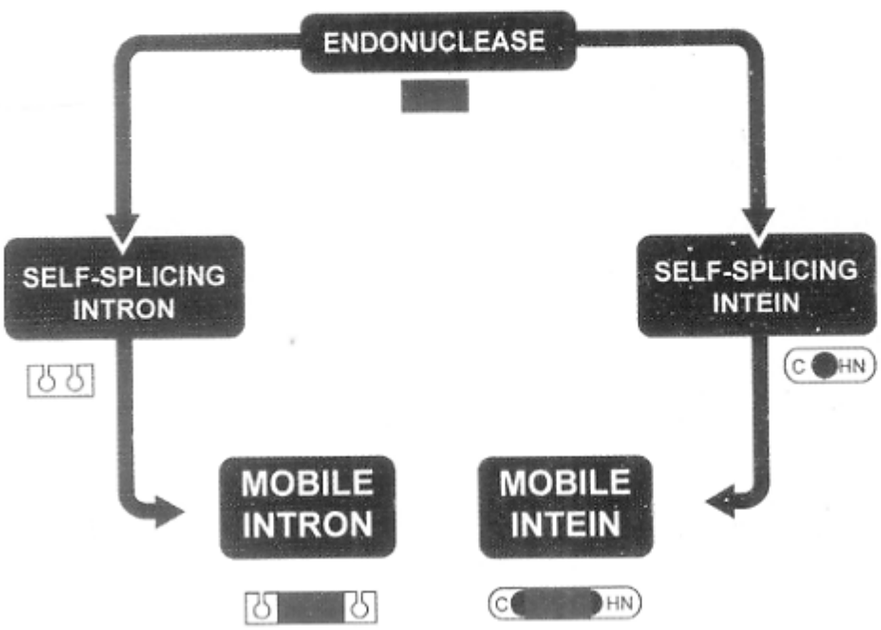

Fig. 1. Evolution of mobile introns and inteins. Endonuclease genes (*red*) are proposed to have invaded DNA encoding self-splicing introns or inteins, to generate mobile genetic elements. (Dassa and Pietrokovski, this Vol.)

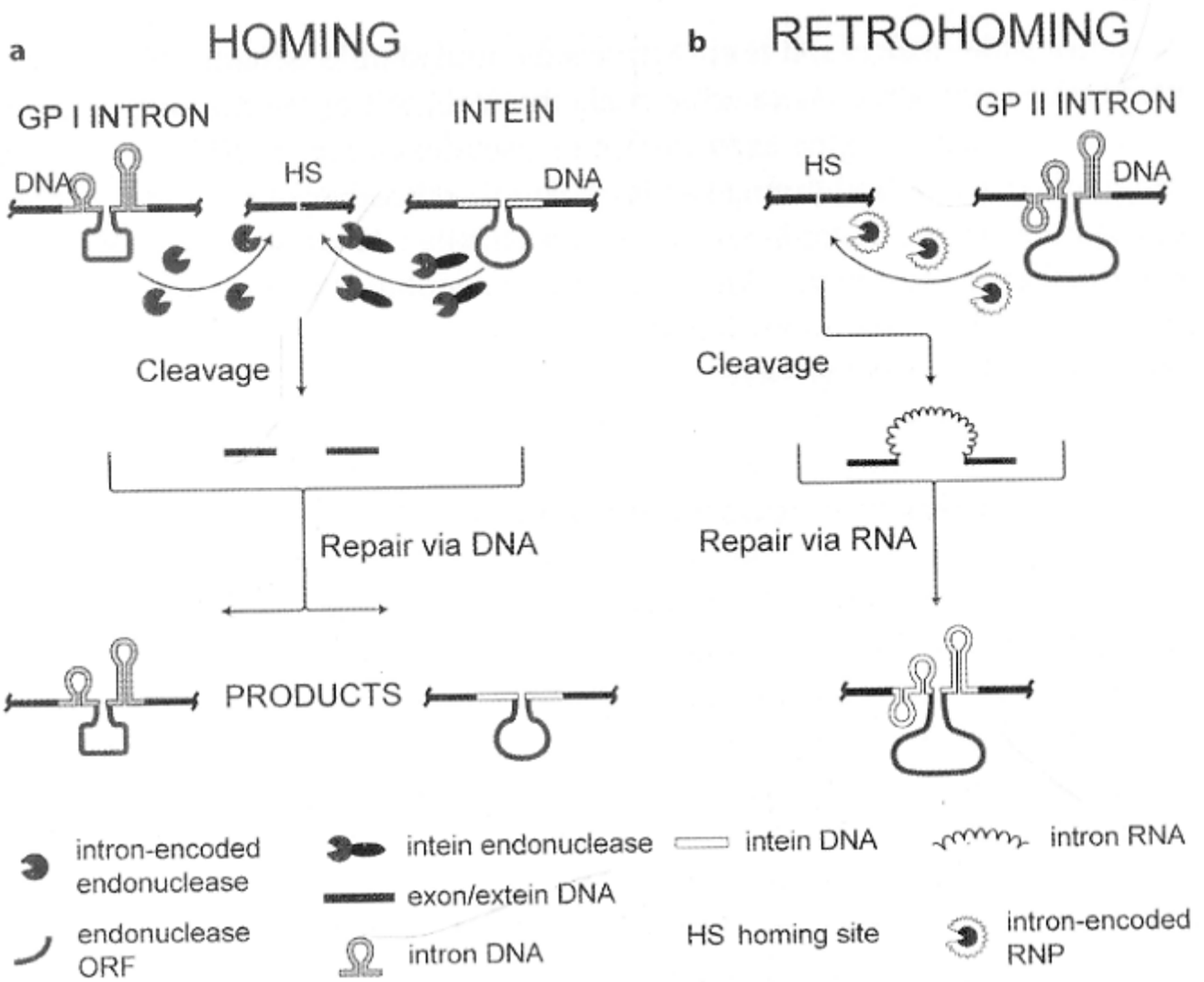

Fig. 2. Mobility of introns and inteins. **a** DNA-based homing of group I introns and inteins. The intron or intein endonuclease cleaves the homing site of a cognate intron- or intein-less allele. Gene conversion repairs the break to generate intron- or intein-containing products (Dujon, this Vol.). **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