Home > Tools & Resources > Feature Articles > Everything You Ever Wanted to Know About Type II Restriction Enzymes

Everything You Ever Wanted to Know About Type II Restriction Enzymes

Type II Restriction Enzymes: Subtypes, naming conventions, and properties

What are Type II Restriction Enzymes used for?

Type II restriction enzymes are the familiar ones used for everyday molecular biology applications such as gene cloning and DNA fragmentation and analysis. These enzymes cleave DNA at fixed positions with respect to their recognition sequence, creating reproducible fragments and distinct gel electrophoresis patterns. Over 3,500 Type II enzymes have been discovered and characterized, recognizing some 350 different DNA sequences. Thousands more 'putative' Type II enzymes have been identified by analysis of sequenced bacterial and archaeal genomes, but remain uncharacterized.

How are Type II Restriction Enzymes named?

Restriction enzymes are named according to the micro-organism in which they were discovered. The restriction enzyme 'HindIII', for example, is the third of several endonuclease activities found in the bacterium Haemophilus influenzae serotype d. The prefix 'R.' is added sometimes to distinguish restriction enzymes from the modification enzymes with which they partner in vivo. Thus, 'R.HindIII' refers specifically to the restriction enzyme, and 'M.HindIII' to the modification enzyme. When there is no ambiguity, the prefix 'R.' is omitted.

Type II Restriction Enzymes Properties

Type II restriction enzymes are very diverse in terms of amino acid sequence, size, domain organization, subunit composition, co-factor requirements and modes of action. They are loosely classified into a dozen or so sub-types according to their enzymatic behavior. This is a practical classification that reflects their properties rather than their phylogeny. It does not necessarily reflect evolutionary or structural relationships, and the subtypes are not mutually exclusive. An enzyme can belong to several subtypes if it exhibits each of their defining characteristics. We discuss these subtypes in their order of importance; the four principal ones are Type IIP, IIS, IIC, and IIT.

Type IIP ('Palindromic' specificity; one domain)

- **Type IIP is the most important subtype, accounting for over 90% of the enzymes used in molecular biology.** Type IIP enzymes recognize symmetric (or 'palindromic') DNA sequences 4 to 8 base pairs in length and generally cleave within that sequence. They are the simplest and smallest of all restriction enzymes, typically 250-350 amino acids in length. Type IIP enzymes specific for 6-8 bp sequences mainly act as homodimers, composed of two identical protein chains that associate with each other in opposite orientations (Examples: EcoRI, HindIII, BamHI, NotI, PacI.) Each protein subunit binds roughly one-half of the recognition sequence and cleaves one DNA strand. Since the two subunits are identical, the enzyme is symmetric, and so the overall recognition sequence, and the positions of cleavage, are also symmetric. Usually, these enzymes cleave both DNA strands at once, each catalytic site acting independently of the other.

- Type IIP enzymes that recognize shorter, 4-bp, sequences often act as monomers composed of a single protein chain. (Examples:

MspI, HinP1I, BstNI, NciI.) These have only one catalytic site, and upon binding, cleave only one DNA strand. However, because they recognize sequences that are symmetric, they can bind in either orientation and ultimately cleave both DNA strands, first one and then the other. The switch in enzyme orientation that takes place is usually very fast, with little accumulation of 'nicked' intermediate molecules cleaved in only the first strand.

- Other Type IIP enzymes (Examples: SfiI, NgoMIV) act as complex homotetramers—dimers of homodimers—or higher order oligomers that bind to and cleave two or more recognition sequences at once.

- Depending on how close the subunits of Type IIP homodimers are to each other, the sequence recognized can be continuous (e.g., EcoRI: GAATTC), or discontinuous, with one unspecified internal bp (HinfI: GANTC), two (Cac8I: GCNNGC); three (AlwNI: CAGNNNCTG), four (PshAI: GACNNNNGTC), five (BglI: GCCNNNNNGGC), or more unspecified bp, up to a record nine (XcmI: CCANNNNNNNNNTGG).

- Type IIP enzymes cleave their recognition sequences at a variety of positions, depending on where the catalytic site is positioned in the protein relative to the sequence-recognition residues. Some generate 5'-overhangs ('staggered ends') of four bases (e.g., HindIII: A'AGCTT) or of two bases (NdeI: CA'TATG). Others generate 3'-overhangs of four (SacI: GAGCT'C) or two bases (PvuI: CGAT'CG). And yet others produce 'flush' (or 'blunt') ends (e.g., EcoRV: GAT'ATC). Enzymes with ambiguous base pairs in their recognition sequences can generate ends with an odd number of bases, including one base (NciI: CC'SGG), three bases (TseI: G'CWGC), five (PspGI: 'CCNGG), or more.

- Most Type IIP enzymes recognize DNA sequences that are unique, in which only one specific base pair can be present at each position (e.g. BglII: AGATCT), but some recognize 'degenerate' (ambiguous) sequences in which alternative bases can be present. The commonest alternatives are Y (pyrimidine, C or T) and R (purine, A or G), e.g., ApoI: RAATTY. Others include M (modifiable base, A or C) and K (not modifiable, G or T), e.g., AccI: GTMKAC; W (weak hydrogen bonding, A or T) e.g., BstNI: CCWGG; and S (strong hydrogen bonding, C or G), e.g., NciI: CCSGG. The atomic structure of the enzyme's binding site determines which base pair(s) can be recognized at each position. At unique binding sites, only the one base pair fits with respect to physical shape and hydrogen bonding. At ambiguous binding sites, either of the alternatives fit satisfactorily.

Type IIS ('Shifted cleavage'; two domains)

- In Type IIP restriction enzymes, the amino acids that catalyze cleavage and those that recognize the DNA are integrated into a single protein domain that cannot be effectively sub-divided. In Type IIS enzymes, in contrast, they are partitioned into separate domains linked by a short polypeptide connector. As a result, Type IIS proteins are larger than Type IIP proteins, typically 400-600 amino acids in length. When Type IIS enzymes bind to DNA, the catalytic domain is positioned to one side of, and several bases away from, the sequence bound by the recognition domain, and so cleavage is 'shifted' to one side of the sequence.

- Type IIS enzymes generally bind to DNA as monomers and recognize asymmetric DNA sequences. They cleave outside of this sequence, within one to two turns of the DNA. By convention, the recognition sequence is written in the orientation in which cleavage occurs downstream, to the right of the sequence. Cleavage often produces staggered ends of two or four bases. The exact positions of cleavage are indicated by the number of bases away from the recognition sequence in each strand. For example, the Type IIS enzyme FokI recognizes the asymmetric sequence GGATG in duplex DNA and cleaves this ('top') strand 9 bases to the right, and the complementary ('bottom') strand four bases further down, producing 4-base 5'-overhanging ends. The specificity of FokI is written: GGATG 9/13 or GGATG(9/13).

- The 'reach' of Type IIS enzymes, the separation between the recognition and cleavage sites, depends on physical parameters such as the structures of the two domains and the connector, and the helical twist of the bound DNA, rather than the actual number of base pairs in between. As a result, cleavage positions can vary somewhat, usually by ±1 base, and the longer the reach, the greater the possible variability. FokI cleaves mainly 9/13, for example, but occasionally cleaves 8/12 or 10/14 instead, depending on the site and the conditions of digestion.

- Type IIS cleavage domains have no inherent sequence-specificity, and so the sequence of the overhang they generate varies from one recognition site to another. Fragments produced by Type IIS-digestion of natural DNA molecules generally have different overhangs, therefore, and will not anneal to one another. However, if the sequence of the overhang is predetermined, by designing it into a PCR primer, for example, then it can be made to complement another and to be directional. This feature is used to great advantage in 'Golden Gate' assembly where multiple fragments can be stitched together in the correct order and orientation in a single ligation. The Type IIS enzymes, BsaI (GGTCTC 1/5), and BsmBI (CGTCTC 1/5), are very popular for this application. The advantage of using Type IIS enzymes for assembly is that the recognition sequence can be placed in the primer on either side of cleavage site. If placed 'inside', 3' to the cleaved end, it will be retained in the construct and can be re-used subsequently. If placed outside, 5' to the cleaved end, it will be lost, leading to a 'scar-less' assembly.

- The C-terminal cleavage domain (CD) of FokI (180 amino acids) can be separated from the N-terminal sequence-recognition domain, and grafted onto other sequence-specific proteins to convert these into 'engineered nucleases'. By grafting it to transcription factors that recognize infrequent sequences, and can be altered by mutagenesis, customized nucleases can be constructed that cleave eukaryotic genomes, ideally, at single sites of choice in vivo. Such 'gene targeting' reagents, termed zinc-finger nucleases (ZFNs), TALENs, and more recently dCas9 nucleases, are revolutionizing the genetic manipulation of higher organisms, and hold great promise for gene therapy and disease intervention in human medicine. The FokI CD has proved universally popular for these applications, although other Type IIS CDs might work as well or even better under certain circumstances.

- In general, the cleavage domains of Type IIS enzymes, including FokI, contain only one catalytic site. In order to cleave duplex DNA, these enzymes form 'transient homodimers', the CD of a bound enzyme molecule combining with the CD of a second molecule to assemble the two catalytic sites needed for cleavage of both DNA strands. As a rule, Type IIS CDs cannot cleave DNA on their own, only when dimerized, and so individual enzyme molecules do not 'nick' DNA. In some cases, the second molecule of the dimer can be unbound, but in other cases it, too, must be bound to a recognition site, the intervening DNA between the two enzymes looping out. The latter enzymes cleave DNA efficiently only when multiple recognition sites are present. If only one site is present, cleavage can sometimes be improved by the addition of short, double-stranded 'helper' oligonucleotides that contain the recognition sequence and to which enzyme molecules can attach specifically.

- Because the FokI CD is only active when dimerized, in order to use it for gene targeting, ZFN or TALEN reagents are constructed in pairs designed to recognize opposed genomic sequences a few base pairs apart. This positions the two CDs, one attached to each reagent, close enough together to dimerize, and thence to cleave the DNA between the two binding sites. The need to use two reagents, rather than only one, improves the accuracy of gene targeting and reduces the likelihood of undesirable, 'off-target' cleavage.

Type IIC (Combined 'restriction-and-modification' enzymes; three domains)

- Restriction enzymes are encoded for the most part by bacteria and archaea. They are potentially toxic to the host cell, and for each restriction enzyme a protective 'antidote' is also made in the form of one or more DNA-methyltransferases (MTases). The enzymes recognize the same sequence as the restriction enzyme and chemically alter each of the sites in the cell's own DNA, to prevent them from becoming cleaved. This DNA-'modification' involves transfer of a methyl group to one base in each strand of the recognition sequence. The methyl groups protrude into the major groove of the DNA and create obstructions that, through steric hindrance, prevent the restriction enzyme from binding to that site.

- Invariably, the MTases that partner with Type IIP and Type IIS enzymes are separate proteins encoded by separate genes. Although both kinds of enzymes recognize the same DNA sequence, they act independently of one other and share no structural or amino acid sequence similarities.

- In contrast, in Type IIC enzymes, restriction and modification activities are combined into a single, composite, enzyme. Whereas Type IIS enzymes comprise two domains, recognition and cleavage. Type IIC enzymes comprise three domains: one for cleavage,

one for methylation, and another for sequence-recognition that is shared by both enzyme activities. The additional domain makes Type IIC enzymes larger than Type IIS enzymes, typically 800-1200 amino acids in length. Some bind as monomers, others as homodimers, and yet others assemble into complex oligomers with molecular masses exceeding 500 kDa.

- Type IIC enzymes can catalyze two competing reactions at once. The co-factor S-adenosylmethionine (SAM) is universally required for the methyltransferase reaction. Some Type IIC enzymes also require SAM for cleavage, others are merely stimulated by SAM, and yet others require no SAM at all. If SAM is present, methylation can proceed alongside cleavage and prevent complete digestion. Due in part to their complexity and size, Type IIC enzymes are not used a great deal in molecular biology. They are very interesting in terms of biochemistry and enzymology, however, and so we discuss them in some detail here.

- The cleavage domain of Type IIC enzymes forms the N-terminal 200 amino acids of the protein. A connector joins this to an adenine-specific DNA-methyltransferase domain of around 400 amino acids. The sequence motifs within this domain places it the 'gamma'-class of methyltransferases, and so Type IIC enzymes are alternatively referred to as 'Type IIG'. The MTase domain is followed by a DNA-binding domain comprising one, or sometimes two, 'target-recognition domains' (TRDs), of approximately 200 amino acids each, that either form the C-terminus of the protein, or a separate protein chain. Type IIC enzymes typically recognize asymmetric sequences. Those with single TRDs recognize short, continuous sequences (e.g., MmeI: TCCRAC; BseRI: GAGGAG). Those with two TRDs recognize longer 'bipartite' (discontinuous) sequences (e.g., BcgI: CGANNNNNNTGC; CspCI: CCACNNNNNTTG).

- Because their recognition and cleavage domains are separate, Type IIC enzyme also cleave outside of their recognition sequences. Their 'reach' tends to be slightly longer than Type IIS enzymes, between one turn of the DNA helix and two, and with most enzymes, cleavage results in 2-base 3'-overhangs (e.g., MmeI: TCCRAC 20/18; EciI: GGCGGA 11/9). Type IIC catalytic domains contain only one catalytic site, and so transient pairing between the CDs of neighboring enzyme molecules is assumed to take place prior to cleavage. Some Type IIC enzymes cleave DNA containing single recognition sequences poorly, and can be stimulated by the addition of oligonucleotides containing additional sites, suggesting that these enzymes must bind to recognition sequences before they can pair effectively.

- Type IIC enzymes with single TRDs cleave on only one side of their recognition sequence—by convention to the right of 'top' strand depicted as the recognition sequence (e.g., BpuEI: CTTGAG 16/14). Remarkably, those with two TRDs cleave on both sides, and in doing so excise a small fragment that contains the recognition sequence within it (e.g., BsaXI: 9/12 ACNNNNNCTCC 10/7). Because these enzymes cleave on both sides, they are also sometimes referred to as 'Type IIB' enzymes. Some are single chain proteins that likely act as homo-tetramers. Others comprise two protein chains, one ('RM') for catalysis and containing the cleavage and methyltransferase domains, the other for sequence recognition (specificity: 'S') containing the two TRDs. The latter form hetero-trimers of two RM subunits and one S subunit, which assemble into oligomers of up to four trimers in order to cleave DNA.

- Type IIC enzymes have diverged widely in the course of evolution, and unlike Type IIP and S enzymes, fall into distinct, close-knit, families. Members of these families are closely similar in amino acid sequence and predicted structure, yet recognize a variety of different DNA sequences. By correlating the sequences recognized with the amino acids at the 'contact' positions within the TRDs, an amino acid-to-base pair 'recognition code' is emerging that reveals how these proteins recognize DNA. This is enabling the specificities of Type IIC enzymes such as MmeI to be rationally changed, and might eventually allow 'designer' enzymes with specificities of choice to be constructed for individual customer-specific applications.

Type IIT (two different catalytic sites; heterodimers)

- Regardless of whether they act as monomers, homodimers or higher-order oligomers, all of the restriction enzymes discussed so far, belonging to the Type IIP, S, C, G and B subclasses, use one catalytic site for DNA cleavage. If this site is disrupted by mutation, the enzyme becomes inactive and cleaves neither strand. Type IIT enzymes, in contrast, use two different catalytic sites for cleavage, each of which is specific for one particular strand. Type IIT enzymes combine features of both Type IIP and Type IIS

enzymes, and so they are intermediate in size, between 350-450 amino acids. Disrupting either catalytic site of a Type IIT enzyme does not inactivate it, but rather turns it into a strand-specific 'nicking' enzyme. These cleave one DNA strand normally, but cannot cleave the other.

- Type IIT enzymes recognize asymmetric sequences. Some cleave within the sequence (e.g., BssSI: C'ACGAG); others cleave on the periphery, and appear to be Type IIS enzymes with a very short reach (e.g., GCAATG 2/0).

- Some Type IIT enzymes are heterodimers, composed of two different protein chains, each of which contains one catalytic site. In some, the two subunits are similar in size (e.g., BbvCI: CC'TCAGC; 275 and 285 aa). Both subunits are involved in DNA recognition in these enzymes, and so both are needed for activity. In other heterodimers, the two subunits are of different sizes (e.g., BtsI: GCAGTG 2/0; 328 and 164 amino acids). The large subunit of these is active on its own, recognizing the DNA and cleaving one strand, while the small subunit on its own is inactive.

- Other Type IIT enzymes are heterodimeric in function, but are joined into a single protein chain. Gene fusion is a common event in nature, and both fusion, and the reverse, gene separation, can be readily replicated in the laboratory. Some of these 'single-chain heterodimers' comprise joined subunits—now, domains— of similar size (e.g., BsrBI: CCG'CTC), while others clearly comprise one large and one small subunit (e.g., BsmI: GAATGC 1/-1).

- DNA-nicking enzymes ('nickases') derived from Type IIT restriction enzymes are used to study the biological effects of DNA-strand breaks in replication, recombination and transcription. They are also used in advanced technologies such as fluorescent bar-coding and optical mapping of individual DNA molecules, and in molecular diagnostic tests based on strand-displacement amplification (SDA). SDA is an isothermal alternative to PCR in which nicking enzymes are used to repetitively generate 3'-OH ends from which DNA polymerase then repetitively initiates polymerization. Versions of SDA offer a rapid way to screen for and identify infectious agents such as viruses at point-of-care locations, and under less-than-ideal, or non-laboratory, conditions. The technique is ideally suited for diagnosing neglected, but increasingly significant, tropical diseases and for routine monitoring of influenza, hepatitis, and others at home.

- Depending on which catalytic site of a Type IIT enzyme is disrupted, the resulting nicking enzyme will cleave either only the 'top' DNA strand (the one depicted as the recognition sequence), or only the 'bottom' DNA strand (the complement). These two activities are distinguished by the prefixes 'Nt.' and 'Nb.' For example, disrupting the catalytic site in one subunit of BbvCI generates 'Nt.BbvCI' (CC'TCAGC) which cleaves only the 'top' strand of the CCTCAGC recognition sequence, and disrupting the catalytic site in the other subunit generates 'Nb.BbvCI' (GC'TGAGG) which cleaves only the complementary, 'bottom', strand.