doi:10.1016/j.jmb.2008.02.058 *J. Mol. Biol.* (2008) **378**, 653–665

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Spatially Directed Assembly of a Heterotetrameric Cre-Lox Synapse Restricts Recombination Specificity

Kathy A. Gelato[1], Shelley S. Martin[2], Patty H. Liu[2], April A. Saunders[2] and Enoch P. Baldwin[2,3]*

[1]*Biochemistry and Molecular Biology Graduate Group, University of California, Davis, Davis, CA 95616, USA*

[2]*Department of Molecular and Cellular Biology, University of California, Davis, Davis, CA 95616, USA*

[3]*Chemistry Department, University of California, Davis, Davis, CA 95616, USA*

Received 29 November 2007; received in revised form 13 February 2008; accepted 25 February 2008
Available online
4 March 2008

Edited by J. Karn

The pseudo-fourfold homotetrameric synapse formed by Cre protein and target DNA restricts site-specific recombination to sequences containing dyad-symmetric Cre-binding repeats. Mixtures of engineered altered-specificity Cre monomers can form heterotetramers that recombine nonidentical asymmetric sequences, allowing greater flexibility for target site selection in the genome of interest. However, the variety of tetramers allowed by random subunit association increases the chances of unintended reactivity at nontarget sites. This problem can be circumvented by specifying a unique spatial arrangement of heterotetramer subunits. By reconfiguring intersubunit protein–protein contacts, we directed the assembly of two different Cre monomers, each having a distinct DNA sequence specificity, in an alternating (ABAB) configuration. This designed heterotetramer preferentially recombined a particular pair of asymmetric Lox sites over other pairs, whereas a mixture of freely associating subunits showed little bias. Alone, the engineered monomers had reduced reactivity towards both dyad-symmetric and asymmetric sites. Specificity arose because the organization of Cre-binding repeats of the preferred substrate matched the programmed arrangement of the subunits in the heterotetrameric synapse. When this "spatial matching" principle is applied, Cre-mediated recombination can be directed to asymmetric DNA sequences with greater fidelity.

Keywords: Cre-LoxP site-specific recombination; protein engineering; orthogonal protein–protein interface; library screening/selection; directed subunit assembly

**Corresponding author.* Department of Molecular and Cellular Biology and the Chemistry Department, University of California, Davis, Davis, CA 95616, USA. E-mail address: epbaldwin@ucdavis.edu.

Present addresses: K. A. Gelato, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany; P. H. Liu, Department of Neurological Surgery, University of California, San Francisco, 13 Parnassus Avenue, San Francisco, CA 94143, USA; A. A. Saunders, Department of Molecular and Integrative Neurosciences, The Scripps Research Institute, La Jolla, CA 92037, USA.

Abbreviations used: HJ, Holliday junction; CTH, C-terminal helix; CTD, C-terminal domain; YSSR, tyrosine site-specific recombinase; CreWT, wild-type Cre recombinase with a Met-His_6 N-terminal tag fused to Ser2; CreALSHG, CreWT with five substitutions, Ile174Ala, Thr258Leu, Arg259Ser, Glu262His, and Glu266Gly; CreAAF, CreWT with three substitutions, Met299Ala, Val304Ala, and Ala334Phe; CreAA, CreWT split-interface mutant with two substitutions, Met299Ala and Val304Ala; ALSHG-F, CreALSHG split-interface mutant with an additional substitution, Ala334Phe; LoxP, natural Cre recombinase recognition site; LoxM7, LoxP variant that is the preferred substrate of CreALSHG, containing T7C, C8T, and G9A left arm 13-bp repeat and the dyad-related C26T, G27A, and A28G right arm 13-bp repeat substitutions; LoxPM7, chimeric Lox site with a LoxP left arm 13-bp repeat and a LoxM7 right arm 13-bp repeat; LoxM7P, chimeric Lox site with a LoxM7 left arm 13-bp repeat and a LoxP right arm 13-bp repeat.

Introduction

Controlled protein oligomerization underlies specificity in many biological processes. Multisubunit complex assembly provides for precise targeting of binding or enzymatic activities as well as regulation of those activities through allosteric interactions and proximity effects. For sequence-specific binding proteins and enzymes, oligomerization of nucleic-acid-interacting domains increases recognition site size and target specificity. The larger sites reduce the chance occurrences of recognition sequences elsewhere in the host genome, thereby minimizing detrimental off-target effects. For homo-oligomeric proteins, this increased specificity requires no extra protein-coding capacity but limits recognition to repeated elements. Hetero-oligomeric assemblies can expand the range of possible target sequences by combinatorial mixing of differently specific partners, as is the case with homeodomain, bZIP coiled-coil, and bHLH transcription factors.[1,2] However, the increased number of potential recognition sequences necessarily reduces the uniqueness of binding interactions. Specificity is achieved by a variety of mechanisms including regulated protein expression and localization, or stabilization of individual subunit combinations.[3–6] Here, we enforced pairing of two variant tyrosine site-specific recombinase (YSSR) subunits in a defined spatial arrangement in order to direct recombination to specific asymmetric substrates, thus eliminating the requirement for repeated recognition sites while retaining a narrow specificity.

YSSRs efficiently induce crossovers between 20- and 40-bp DNA target sequences. In nature, the resulting integrations, excisions, inversions, and translocations are employed in chromosome and plasmid segregation, plasmid amplification, virus integration, and gene regulation.[7–10] In biotechnology, YSSRs have been harnessed to manipulate chromosome structure in living cells and organisms to generate controlled gene deletions, integrate transgenes, and excise viral DNA.[11–16] However, the high sequence specificity of the naturally occurring YSSRs and the relaxed specificity of engineered versions remain an obstacle for the widespread use in applications requiring both flexible recognition potential and extreme precision, such as gene therapy.[17] One critical limitation is the homotetrameric nature of recombination complexes, which enforces recombination to occurring between nearly identical, dyad-symmetric DNA sequences.

One well-studied YSSR, bacteriophage P1 Cre protein, carries out site-specific recombination at 34-bp LoxP sites.[18–21] Cre-LoxP recombination is widely utilized for genome manipulations because of its robust activity and simple requirements. Numerous biochemical, biophysical, and structural investigations have generated paradigms for YSSR recombination mechanism and function.[22,23]

Recombination begins by assembly of a synaptic complex containing four Cre monomer subunits and two LoxP sites (Fig. 1a).[24] As is typical for YSSRs, LoxP sites contain an inverted dyad of Cre-binding 13-bp repeats, separated by an asymmetric intervening sequence, the 8-bp spacer (Fig. 1b).[19] In the complex, "crossing-over" recombination results from two pairs of single-strand exchanges, effected by reversible cleavages and strand swaps within the 8-bp spacers (Fig. 1a and b). The first exchange forms a Holliday junction (HJ) intermediate and the second exchange forms the products.[21] The order and progression of strand exchanges are directed by interplay between 8-bp spacer asymmetry, an associated DNA bend, and differentiation of Cre monomers into cleaving and noncleaving conformations. This aspect of Cre function has been extensively studied and reviewed elsewhere.[27–30]

The two domains of a Cre monomer comprise a clamp that surrounds the 13-bp repeat, creating an extensive interface of protein–DNA contacts.[24] Monomers are recruited to each LoxP 13-bp repeat and the active tetrameric recombination complex is subsequently assembled through a pseudo-fourfold cyclic arrangement of Cre–Cre interactions (Fig. 1c). A crucial intersubunit contact is a domain swap in which the C-terminal helix (CTH, residues 333–340) packs against the C-terminal domain (CTD, residues 131–326) of an adjacent monomer.

The high fidelity of wild-type Cre for the 13-bp repeat restricts the range of available recombination targets. Efficient reactions are only realized with close matches to LoxP and even single base changes in the 13-bp repeats can nearly abolish efficient function.[31–33] This limitation has been circumvented by directed evolution of Cre variants with altered DNA specificity.[16,25,32,34] For example, a quintuple mutant CreALSHG ["C2(+/−) #4" from the work of Santoro and Schultz[25]], prefers to recombine the LoxP variant, LoxM7. LoxM7 contains three base-pair substitutions at a key protein–DNA interface in the 13-bp repeat (Fig. 1b), and is not recombined by wild-type Cre (CreWT).

The homotetrameric Cre-LoxP synapse exerts further substrate restrictions (Fig. 2a, i). The symmetric arrangement of protein subunits and their DNA-binding surfaces matches the arrangement of 13-bp repeats of the identical LoxP sites in the complex. As a result, CreWT cannot recombine chimeric Lox sites containing both LoxP and LoxM7 13-bp repeats[26] (Fig. 1b), since two of the 13-bp repeats do not match the subunit specificities.

Two strategies have been employed to bypass the symmetry restriction of YSSRs. The first strategy utilizes relaxed-specificity recombinases to simultaneously recognize substantially different 13-bp repeats.[16,25,26,35] While this is a convenient approach, such promiscuity could lead to off-target recombination that would be unacceptable for high-fidelity applications, such as gene therapy. The second strategy uses combinations of altered-specificity Cre mutants.[26] A four-variant mixture would permit recombination between two nonidentical asymmetric targets with the only requirement being identity in the central six base pairs that are swapped (Fig. 2a, ii). However, such a mixture can assemble into 70