Available online at www.sciencedirect.com
SCIENCE DIRECT
Biomolecular Engineering 22 (2005) 113–120
Biomolecular Engineering
www.elsevier.com/locate/geneanabioeng

Generation of variability by in vivo recombination of halves of a $(\beta/\alpha)_8$ barrel protein

Gloria Saab-Rincón, Eugenio Mancera [1], Gabriela Montero-Morán, Filiberto Sánchez, Xavier Soberón *

Instituto de Biotecnología, UNAM, Apartado Postal 510-3, Cuernavaca, Morelos 62271, México

Received 26 October 2004; received in revised form 25 December 2004; accepted 18 January 2005

Abstract

Similar to what has been achieved with nucleic acids, directed evolution of proteins would be greatly facilitated by the availability of large libraries and efficient selection methods. So far, host cell transformation efficiency has been a bottleneck, practically limiting libraries to sizes less than 10^9. One way to circumvent this problem has been implemented with antibody systems, where contribution to the binding site is provided by two different polypeptides (light and heavy chains). The central concept is the construction of binary systems in which the gene from the two chains are separated by a cre-lox recombinase recognition site, packaged in a phage, and subsequently introduced, by multiple infection, into a recombinase expressing cell [Sblattero D, Bradbury A. Nat Biotechnol 2000;18(1):75–80]. Here, we describe the development of a system which applies the same concept to a single-domain enzyme, the cytoplasmic $(\beta/\alpha)_8$ barrel protein phosphoribosyl anthranilate isomerase (PRAI) from *E. coli*. For that purpose, we identified the site at which a loop containing the recognition sequence for cre-lox recombinase could be inserted yielding a functional enzyme. We evaluated the effect of this insertion on the capability of the engineered gene to complement a *trp F − E. coli* strain and the efficiency of the system to recover the original sequence from an abundance of non-functional mutant genes.

Keywords: TIM barrels; Fragments complementation; Directed-evolution; PRAI; Protein engineering

1. Introduction

Directed evolution has proven to be a useful and promising approach to endow proteins with improved attributes [2–5]. Laboratory methods to evolve proteins operate on the same principles of natural evolution: variation and selection. Unfortunately, the timescale and the volume of sample where these experiments are performed are very small fractions of those occurring in nature. In those cases where true selection systems are available one could, in principle, aim for very large libraries, which can be easily generated in vitro, but are currently limited by the ability to introduce DNA variants into cells, i.e., by transformation efficiency [6]. One way to overcome this bottleneck has been implemented with antibodies, where the light and heavy chains are expressed from different vectors, at least one of which can be introduced into the cells by phage infection, thus overcoming the aforementioned limitation [7–9]. Variations of the same approach have been published subsequently [1], but always with the antibody system.

Most recent systems are based on Cre recombinase and its recognition sites, known as LoxP/Cre systems. In these, the Cre recombinase enzyme mediates recombination between two 34 bp LoxP sites with no additional factors required. LoxP sites are comprised of two 13 bp inverted repeats separated by a central, asymmetric 8 bp sequence. When the two LoxP sites are introduced in the same circular molecule, if they are in the same direction, the intervening sequence is excised by recombinase action; when they are in opposite directions the intervening sequence is reversed [10]. Mutations in the central sequence of LoxP sites allow the control of

* Corresponding author. Tel.: +52 777 329 1671; fax: +52 777 317 2388.
E-mail address: soberon@ibt.unam.mx (X. Soberón).
[1] Present address: Center for Molecular Biology of the University of Heidelberg (ZMBH).

1389-0344/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.bioeng.2005.01.002

recombination by preventing this phenomenon, while still allowing recombination between different sites present in separate replicons, for example with a G substitution by an A at the seventh base of the central sequence, the LoxP511 site generated retains 100% efficient recombination. [10]. This system has been successfully used to recombine light and heavy chains of an antibody library in vivo [11].

Following the same approach, our aim was to extend these capabilities of large library generation to other protein architectures, especially to enzymes. We chose the $(\beta/\alpha)_8$ barrel (also called TIM barrel) protein architecture, as it offers specially attractive opportunities for protein engineering. This protein fold is present in a significant fraction of the enzymes whose structure has been solved [12], thus providing a very large repertoire of catalytic activities implemented on the same basic scaffold. Notably, the active sites of TIM barrels are always located at the same face of the structure (the carboxyl termini of the β strands) [13] making their analysis and manipulation even more straightforward. These and other attributes have made TIM barrels favorite objects for protein engineering experiments over the past two decades [14–18].

The diversity of activities found on TIM barrels corresponds to 83% of the basic enzyme groups as classified by IUPABQ (EC numbers 1–6). It is tantalizing to observe that much of this diversity may have arisen from a divergent evolution process [13,19]. Further, this divergent process has been found to start at the level of half barrels for at least some members of this fold [20], which is not surprising, looking at the pseudo symmetric design of these proteins.

We have demonstrated that functional *E. coli* phosphoribosyl anthranilate isomerase (PRAI) can be expressed from different cistrons, and even different replicons, when fragmented in approximate halves but not when fragmented in other specific locations splitting it in 6 + 2 or 2 + 6 β/α subunits [21]. Here we report a genetic system that, through in vivo recombination of half protein genes, allows the generation of large libraries of protein variants ($>10^{10}$) and thus increases the utility of a TIM barrel scaffold in directed evolution studies.

2. Materials and methods

2.1. Enzymes and vectors

High fidelity polymerase Expand™, T4 DNA ligase, alkaline phosphatase and most of restriction enzymes, were purchased from Böheringer Mannheim. pDAN5 vector was generously provided by Dr. Andrew Bradbury from Los Alamos Laboratories. *Alw*NI and *Msc*I restriction enzymes were purchased from New England Biolabs.

2.2. Reagents

Buffers, reagents and agar noble were purchased from Sigma. Casamino acids were purchased from Difco Laboratories. Kits for plasmid and PCR purification were purchased from Qiagene.

2.3. PRAI gene

The engineered monofunctional *E. coli* PRAI gene ([ML256-452] PRAI) cloned into pUC18 through *Eco*RI/*Hin*dIII restriction sites, was generously provided by Dr. K. Kirshner's group of Basel University. It was engineered from the bifunctional gene expressing IGPS-PRAI, inserting a stop codon after residue 252, a Shine-Delgarno and codons for Met-Leu followed by the sequence coding for amino acids 256–452, that comprises the PRAI function [22].

2.4. Strains

E. coli strain Cre^+JL03, which constitutively expresses Cre recombinase, was a generous gift form Dr. Andrew Bradbury from Los Alamos Laboratories.

To evaluate functional genes we used *E. coli* strain JMB9 r^-m^+ *trpF*−, generously provided by Dr. K. Kirschner. This strain has a substitution of the IGPS-PRAI gene by the monofunctional version of the IGPS gene in its chromosome [16]. PRAI participates in the tryptophan biosynthetic pathway, and therefore, a strain *trpF*− is auxotrophic for tryptophan. Since a requirement of the system in development is the introduction of the genetic material by infection rather than by transformation we constructed an F′ selection strain by conjugation of strain JMB9 r^-m^+ *trpF*− with *E. coli* strain XL1Blue, which carries the F′ plasmid. To assess the conjugation, we screened for colonies that were *tetracycline*+, *trpF*−, F′.

2.5. Construction of plasmids for recombination

pDAN5 vector (a generous gift from Dr. A. Bradbury) contains the light and heavy variable chains of an antibody linked by the LoxP511 site (which codes for the sequence ITSYNVYYTKL), a bacterial leader sequence, a His_6 tag, and an amber stop codon followed by the g3 gene from fdTet (Fig. 1a). A wild-type Lox P site is also included downstream from the amber codon to shuffle the heavy chains of different constructs. The gene for the light variable chain is flanked by the *Hin*dIII and *Sal*I restriction sites, while the heavy chain gene is flanked by *Xho*I and *Nhe*I restriction sites at their 5′and 3′ ends, respectively.

2.6. Construction of the pDAN5B1A4PRAI vector[2]

The first half of the *E. coli* PRAI gene containing a ribosomal binding site was amplified from the pUC18-PRAI construction using oligonucleotides B1HindPRAI, 5′CAGT

[2] pDAN5B1A4PRAI vector comprises the DNA sequence coding from β1 to α4 of *E. coli* PRAI fused to the LoxP511 linker and the sequence for heavy chain of an antibody cloned into pDAN5 phagemide.