Available online at www.sciencedirect.com
ScienceDirect
Molecular & Biochemical Parasitology 154 (2007) 103–109
MOLECULAR & BIOCHEMICAL PARASITOLOGY

Short communication

Functional genomics in *Trypanosoma brucei*: A collection of vectors for the expression of tagged proteins from endogenous and ectopic gene loci

Steven Kelly [a], Jenny Reed [b], Susanne Kramer [b], Louise Ellis [b], Helena Webb [b], Jack Sunter [b], Jeanne Salje [b], Nina Marinsek [b], Keith Gull [a], Bill Wickstead [a], Mark Carrington [b,*]

[a] *Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK*
[b] *Department of Biochemistry, University of Cambridge, 80 Tennis Court Road, Cambridge CB2 1GA, UK*

Received 12 February 2007; received in revised form 20 March 2007; accepted 21 March 2007
Available online 27 March 2007

Keywords: *Trypanosoma brucei*; Functional genomics; Vectors; Transgene expression

The expression of transgenes encoding proteins modified to contain residues that impart a particular property, 'tagged proteins', is central to the post-genomic analysis of any organism. *Trypanosoma brucei* is a model kinetoplastid protozoan pathogen and has the most advanced repertoire of tools for reverse genetic analysis available for any protozoan. The vast majority of these tools take advantage of the predominance of homologous over non-homologous recombination to target constructs to specific genomic loci. Initially the targeting was used to direct unregulated transgenes to transcribed regions of the genome [1] and to perform gene deletions [2,3]. A leap forward in the sophistication of reverse genetic experiments occurred with the development of trypanosome cell lines expressing the tetracycline repressor (TetR) protein which facilitated tetracycline-regulated expression of transgenes [4,5]. Further development of cell lines expressing both TetR and bacteriophage T7 RNA polymerase (T7RNAP) allowed transgenes to be transcribed and expressed at very high levels [6]. The TetR- and T7RNAP-expressing cell lines are also central to most RNA interference-based analyses of gene function currently performed in trypanosomes [7–10]. In nearly all cases, an antibiotic resistance gene is used as the selectable marker. The DNA used for targetted integration is usually a linearised plasmid; targetting using PCR products directly is possible [11–13] but integration is less efficient and does not usually offer inducible expression.

The expression of tagged proteins has become central to the technologies that have developed to analyse the function of individual genes. The tag can have a range of functions falling into two main categories: the first is to provide evidence for the subcellular localisation of a protein in living cells using a fluorescent protein tag (see, for example [14]) or in fixed cells using a fluorescent protein or epitope tag (see, for example [15]). The second is to facilitate the purification of complexes which, when allied with mass spectrometric analysis and knowledge of the genome sequence, allows the identification of components of multimeric proteins. Two types of tag have been developed successfully in yeast: (i) the tandem affinity tag where two successive rounds of purification are used [16], and (ii) a tandem epitope tag which is used in a single step purification [17]. To date, the former has been exploited more in trypanosomes [18,19]. In the future, the investigation of transient protein-protein interactions *in vivo* could be analysed by techniques such as fluorescence resonant energy transfer (FRET) which is dependent on tagging the two target proteins with different fluorescent proteins [20]. Here, we describe five sets of plasmids that represent a substantial collection of publicly available vectors for adding tags to the N- and C-termini of proteins in *T. brucei*.

1. New vectors for inducible expression of transgenes from ectopic loci

Vectors were based on three different backbones: pLEW100 [6] (kind gift of George Cross, Rockerfeller University), and two new vectors, pDex377 and pDex577. Transgene expression in pLEW100 is regulated by a tetracycline-inducible EP1 procyclin promoter [6]. The plasmid was designed to integrate into the non-transcribed spacer between ribosomal RNA (rRNA) genes and requires cell lines expressing T7RNAP, as the bleomycin

* Corresponding author. Tel.: +44 1223 333683; fax: +44 1223 766002.
E-mail address: mc115@cam.ac.uk (M. Carrington).

0166-6851/$ – see front matter
doi:10.1016/j.molbiopara.2007.03.012

Table 1
Tags and vectors

Tag	Reference	Vectors for tetracycline-inducible transgenes				Vectors for tagging an endogenous locus	
		pLEW100 C-terminal	pDEX377 C-terminal	pLEW100 N-terminal	pDEX377 N-terminal	C-terminal	N-terminal
Enhanced yellow FP	Clontech	p2216	p2663	p2625	p2628	p2710	p2675
Cerulean FP	[24]	p2619	p2622	p2627	p2630	p2709	p2677
Cherry FP	[25]	p2686	p2664	p2665	p2666	p2705	p2679
TAP	[13]	p2289		p2626	p2629		p2676
TAP PTP	[16]	p2662	p2687			p2706	p2678
3 tandem MYC	[27]	p2280	p2674				
2 tandem TEV:6 tandem HA	[28,29]	p2477	p2623				
2 tandem TEV:12 tandem HA	[28,29]	p2620				p2708	
TAP Strep-tag:2 tandem TEV:12 tandem HA	IBA [28,29]	p2621	p2624			p2707	
Enhanced green FP: TY	Clontech [19,20]	pDex577-G				pEnT5-G pEnT6B-G pEnT6P-G	
Enhanced yellow FP: TY	Clontech [19,20]	pDex577-Y				pEnT5-Y pEnT6B-Y pEnT6P-Y	
Cerulean FP: TY	[19,20,24]	pDex577-C				pEnT5-C pEnT6B-C pEnT6P-C	

resistance selectable marker (*bleR*) is transcribed by a modified T7RNAP promoter. pDex377 is a new plasmid derived from pLEW100 and uses the same tetracycline-inducible EP1 procyclin promoter for control of transgene expression [6]. However, in pDex377 the selectable marker was changed to a hygromycin resistance gene (*hygR*) under the control of a rRNA promoter and the targeting sequence was changed to a repetitive DNA present on minichromosomes and intermediate-sized chromosomes (the 177 bp repeat). Regulated expression of the transgene from pDex377 requires cell lines expressing the tetracycline repressor only, but the vector can be used for unregulated expression in any cell line. pDex577 is a new plasmid that was designed to produce high level over-expression of proteins. It was derived in part from p2T7-177 [21] and pLEW100; transgene expression is directed by a tetracycline-inducible T7 promoter. pDex577 contains a *bleR* gene transcribed by a rRNA promoter and is targeted to 177 bp repeats [21]. The new plasmids pDex377 and pDex577 were sequenced to completion (4× coverage) and all additional tags/modifications were verified by sequencing. The plasmids are available from the authors and the sequences of all vectors are available from http://web.mac.com/mc115/iWeb/mclab/home.html and http://users.path.ox.ac.uk/~kgull/index.html.

The derivatives made from the three vectors for the expression of tagged proteins are shown in Fig. 1 and listed in Table 1. All the vectors derived from pLEW100 and pDex377 were designed to accept an open reading frame (ORF) as a *Hin*dIII *Bam*HI fragment and in all these vectors the open reading frames can be transferred from one vector to another without loss of reading frame. The fusion protein has a linker of 5–8 residues between ORF and the tag.

2. New vectors for constitutive expression of transgenes from the endogenous gene locus

We have developed two sets of vectors for tagging genes at the endogenous gene locus. The first set were derived from pN-PTPpuro and pC-PTPneo [19] (kind gifts from Arthur Günzl, University of Connecticut) and are listed in Table 1. In these vectors, part of the targeted ORF is cloned in frame with the tag and then the plasmid is linearised using a unique restriction enzyme site within the targeted ORF [19] (Fig. 2). The N-terminal *in situ* tagging vectors were designed to be digested with *Hin*dIII and *Eco*RV such that an N-terminal portion of the targeted ORF could be cloned after digestion with *Hin*dIII and any restriction enzyme that leaves a blunt end. Once the targeted ORF fragment has been inserted, there is a 10-residue linker (–GGGGSQASAT–) between the end of the tag and the initiation codon of the ORF. The C-terminal *in situ* tagging vectors were designed to be digested with *Swa*I and *Bam*HI and the C-terminal part of the targeted ORF cloned as a blunt end-*Bam*HI fragment. Once the targeted ORF fragment has inserted, there is a 5-residue linker (–GGGSG–) between the targeted protein and tag. The reading frames in these vectors are compatible with the *Hin*dIII and *Bam*HI sites present in the pLEW100 and pDex377 derived vectors above and the same amplified ORF can be used.

The second set of vectors for tagging genes at the endogenous locus were based on the new plasmids pEnT5 and pEnT6. These vectors were designed to be highly modular in nature to facilitate: (i) movement of DNA between plasmids; (ii) use of novel tags and (iii) use of endogenous intergenic sequence for tagged protein regulation. The same vector can be used for either N-terminal (using *Xba*I and *Bam*HI) or C-terminal (between *Hin*dIII and *Spe*I) tagging to generate chimeras tagged with both a fluorescent protein and the TY epitope for use in immunolocalisation [22,23]. The plasmids are designed to be used as replacement rather than insertional vectors, as outlined below, to avoid the generation of unwanted gene fragments. This strategy also removes the need for an endogenous linearization site in the targeting fragment (Fig. 2C). For example, to tag proteins at their N-terminus using pEnT5/6, two fragments are amplified from genomic DNA. The first encompasses 250–500 bases from the 5′ end of the target ORF beginning directly at the start or second codon of the ORF. The six bases necessary to form the consensus *Xba*I (or compatible site *Spe*I, *Avr*II, *Nhe*I) site are added to the 5′ end of the 5′ primer in frame with the