BIOTECHNOLOGICAL PRODUCTS AND PROCESS ENGINEERING

Evidence that the insertion events of IS*2* transposition are biased towards abrupt compositional shifts in target DNA and modulated by a diverse set of culture parameters

**Geisa A. L. Gonçalves · Pedro H. Oliveira ·
Ana G. Gomes · Kristala L. J. Prather · Leslie A. Lewis ·
Duarte M. F. Prazeres · Gabriel A. Monteiro**

Abstract Insertion specificity of mobile genetic elements is a rather complex aspect of DNA transposition, which, despite much progress towards its elucidation, still remains incompletely understood. We report here the results of a meta-analysis of IS*2* target sites from genomic, phage, and plasmid DNA and find that newly acquired IS*2* elements are consistently inserted around abrupt DNA compositional shifts, particularly in the form of switch sites of GC skew. The results presented in this study not only corroborate our previous observations that both the insertion sequence (IS) minicircle junction and target region adopt intrinsically bent conformations in IS*2*, but most interestingly, extend this requirement to other families of IS elements. Using this information, we were able to pinpoint regions with high propensity for transposition and to predict and detect, de novo, a novel IS*2* insertion event in the 3′ region of the *gfp* gene of a reporter plasmid. We also found that during amplification of this plasmid, process parameters such as scale, culture growth phase, and medium composition exacerbate IS*2* transposition, leading to contamination levels with potentially detrimental clinical effects. Overall, our findings provide new insights into the role of target DNA structure in the mechanism of transposition of IS elements and extend our understanding of how culture conditions are a relevant factor in the induction of genetic instability.

Keywords Insertion sequence · GC skew · Genetic instability · Bacteria · Culture growth · Biopharmaceuticals

Geisa A. L. Gonçalves and Pedro H. Oliveira contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s00253-014-5695-6) contains supplementary material, which is available to authorized users.

G. A. L. Gonçalves · D. M. F. Prazeres · G. A. Monteiro (✉)
Institute for Biotechnology and Bioengineering (IBB), Centre for Biological and Chemical Engineering, Department of Bioengineering, Instituto Superior Técnico, Lisbon, Portugal
e-mail: gabmonteiro@ist.utl.pt

P. H. Oliveira
Institut Pasteur, Microbial Evolutionary Genomics, Département Génomes et Génétique, Paris, France

P. H. Oliveira
CNRS, UMR3525, Paris, France

A. G. Gomes
Instituto Politécnico de Setúbal, ESTBarreiro, 2839-001 Barreiro, Portugal

K. L. J. Prather
Department of Chemical Engineering, Massachusetts Institute of Technology, Room E17-504G, Cambridge, MA 02139, USA

L. A. Lewis
Department of Biology, York College of the City University of New York, Jamaica, NY 11451, USA

G. A. L. Gonçalves · K. L. J. Prather · D. M. F. Prazeres ·
G. A. Monteiro
MIT Portugal Program, Lisbon, Portugal

Introduction

Insertion sequences (ISs) are small and genetically compact autonomous mobile genetic elements (MGEs) that play an important role in prokaryotic genome evolution. IS2 is a 1.3 kb member of the large IS3 family of MGEs (reviewed in (Chandler and Mahillon 2002), see also ISfinder: https://www-is.biotoul.fr/) and was among the first bacterial transposable elements to be identified (Hirsch et al. 1972). It contains two partially overlapping open reading frames orfA and orfB from which the transposase (TPase) OrfAB is generated by a programmed −1 translational frameshift (Polard et al. 1991; Sekine et al. 1994; Vogele et al. 1991). The termini of IS2 consist of two 42-bp and 41-bp left and right imperfect inverted repeats (IRL and IRR), respectively (Fig. 1). IS2 transposition occurs by a circle forming variation of the cut-and-paste pathway that involves in a first step the production of a figure-8 molecule, which serves as the precursor of a minicircle intermediate, that in a second step is inserted into the target (Lewis et al. 2004; Lewis et al. 2001; Lewis and Grindley 1997). Despite reports in the late twentieth century of a regional preference for IS2 insertion in bacteriophage P1 genome (Sengstag and Arber 1983; Sengstag and Arber 1987; Sengstag et al. 1983), the published literature in the subsequent years did not bring any conclusive evidence regarding its insertional specificity (Bernard et al. 1991; Cohen et al. 1993; Lewis et al. 1994a; Lewis et al. 1994b; Rijavec et al. 2007; Szeverenyi et al. 1996a; Szeverenyi et al. 1996b; Whiteway et al. 1998). It was only recently that a GFP-tagged approach was successfully used to obtain, under native conditions, preparations of full length, soluble, and active IS2 TPase that was usually insoluble when prepared under native conditions, and refractory to whole structure-function or biophysical studies when solubilized (Lewis et al. 2011). This paved the way for generating hydroxyl radical footprinting data, which together with curvature propensity data obtained from several target regions, provided for the first time support for a structure-dependent mechanism for IS2 transposition (Lewis et al. 2012). Both minicircle junction and IS2 target sites were shown to adopt similar intrinsically curved structures, supporting the idea that a particular configuration is needed to regulate transposition by promoting the assembly of the transpososome (Lewis et al. 2012). In fact, a growing body of evidence has yielded similar observations in other MGEs such as IS10 (Kobori et al. 2009), IS231 (Hallet et al. 1994), Tn10 (Pribil and Haniford 2003), bacteriophage Mu, and Tn5 (reviewed in (Nagy and Chandler 2004)).

Asymmetries in DNA sequence composition have been previously linked to conformational changes capable of favorably influencing the integration of MGEs. For example, retroviral integration was shown to occur more efficiently in vitro at the outer face of curved DNA segments composed of alternate AT- and GC-rich tracts (Muller and Varmus 1994). In two other reports, integration of T-DNA from Agrobacterium tumefaciens was described to take place in regions of the Arabidopsis genome coincident with inversely correlated and sharply asymmetric GC and AT skews (Schneeberger et al. 2005; Zhang et al. 2007). The latter studies observed an over-abundance of guanine and adenine residues upstream of the insertion sites and an opposite trend downstream. Complete symmetry regarding both skews was concomitantly observed at the target site (Schneeberger et al. 2005; Zhang et al. 2007). More recently, a large-scale comparative genomic analysis performed in both bacteria and archaea revealed that switch sites of GC skew (ssGCs) are hotspots for genomic island integration (Du et al. 2011).

Taking all the above into consideration, we decided to perform an extensive literature search for IS2 target sites, in order to investigate whether these sites significantly correlate with the presence of nearby compositional asymmetries. Our results show a consistent and significant transposition bias towards abrupt DNA compositional shifts, particularly in the form of ssGCs. To further strengthen and validate our meta-analysis, we used the information provided by the compositional profiles to direct a de novo search of IS2 insertions. Because spontaneous contamination with bacterial ISs represents a serious concern during the amplification of plasmids for therapeutic purposes or protein expression studies, we decided to focus our search on a plasmid model (pVAX1GFP) typically used in the biopharmaceutical industry as backbone for DNA vaccine development. We found a region near the 3′ end of the gfp gene showing a skew profile similar to those found in our meta-analysis and into which IS2 inserted in a spontaneous fashion. Such finding represents, to the best of our knowledge, one of the very few reports of an IS-mediated instability event taking place in what is probably the most widely used reporter gene.

Along with the need for a comprehensive knowledge of target preference and insertion specificity, a better understanding of the type of environmental conditions capable of exacerbating the transposition of MGEs is currently lacking. The latter is usually greatly suppressed under optimal growth conditions, but transiently or permanently activated in response to situations of cellular stress (Eichenbaum and Livneh 1998; Twiss et al. 2005). This so-called stress-induced transposition allows for fast genetic changes to take place, fuelling the emergence of novel and advantageous traits able to overcome stress. In this regard, we also evaluated the impact of parameters such as growth medium composition, production scale, and culture time on IS2 accumulation. Our results demonstrate that prolonged culturing or the use of a chemically defined medium and growth conducted under high-cell densities favor the accumulation of IS2-contaminant molecules. Altogether, these findings provide extensive

evidence to establish for the first time a clear basis for the nonrandom regional insertion specificity of IS2 and demonstrate the important roles played by a diverse set of culture parameters in promoting the transposition of IS elements.

Materials and methods

Bioinformatic analysis

We performed a literature search and gathered the nucleotide sequences of 41 different IS2 target sites from genomic, phage, and plasmid DNA (see Suppl. Material S1). The strands analyzed from each target site correspond to those in which the 5′ end of IRL and the 3′ end of IRR are arranged from left to right upon insertion. We have used the DNA base composition analysis tool (http://molbiol-tools.ca/Jie_Zheng/) to evaluate GC and AT skew profiles of ±700 bp regions flanking IS2 target sites. Sliding windows of 200 bp and a step of 100 bp were used. Nucleotide skew was calculated as $(F_A - F_B)/(F_A + F_B)$ where F_A and F_B are respectively the frequencies of nucleotides A and B. To statistically validate the significance of the compositional asymmetries found, the same dataset was randomly shuffled 30 times (k-tuples of

length 1) using the shuffleseq tool from the EMBOSS suite (Rice et al. 2000). Both GC and AT skews were then computed from the randomized sequences, and z scores were computed as:

$$z = \frac{S - \overline{S_{rand}}}{\sigma_{rand}}$$

where $\overline{S_{rand}}$ is the average of the randomized variable S, and σ_{rand} represents its standard deviation. The corresponding p values were obtained from $p = erfc(|z|/\sqrt{2})$, where $erfc$ is the complement error function. Curvature propensity plots were obtained by analyzing 200 bp sequences flanking the target sites with the bend.it server (Gabrielian et al. 1997) (http://hydra.icgeb.trieste.it/dna/bend_it.html) using the DNAse I-based parameters of (Brukner et al. 1995). Three-dimensional reconstruction was performed with the model.it server (http://hydra.icgeb.trieste.it/dna/model_it.html) (Vlahovicek et al. 2003), and the output was displayed and visualized with Polyview 3D (http://polyview.cchmc.org/polyview3d.html) (Porollo et al. 2004).
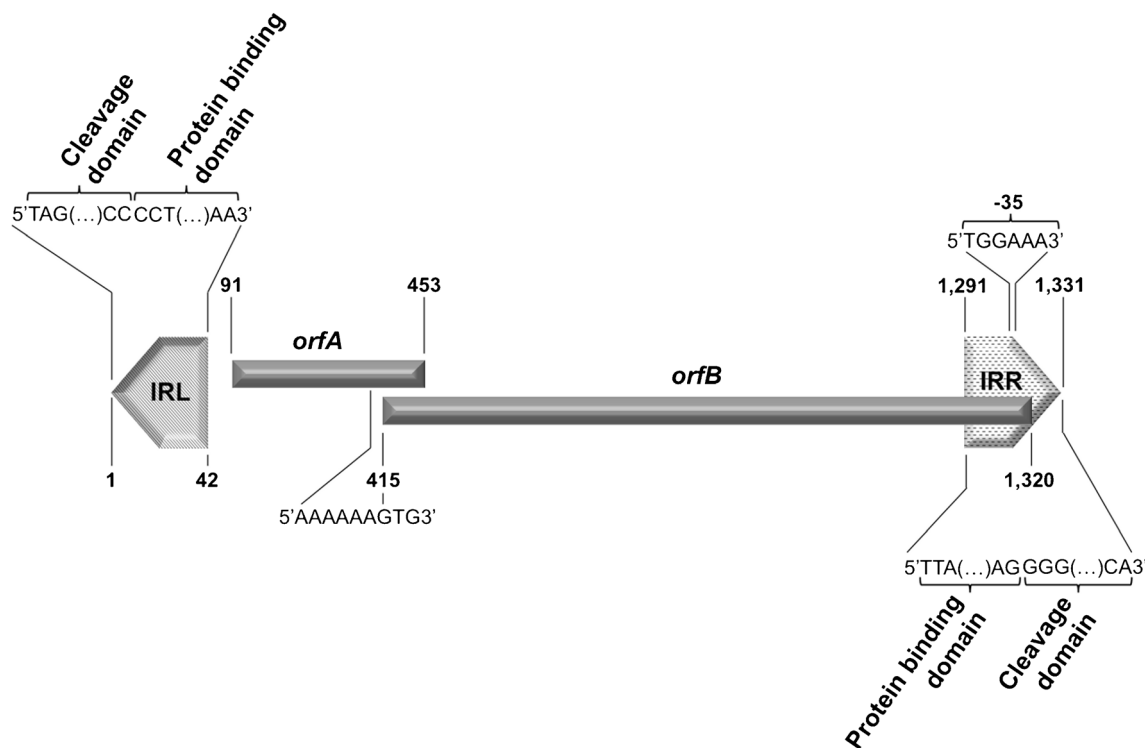

Fig. 1 Schematic illustrating the organization of the IS2 insertion sequence. The two partially overlapping open reading frames, *orfA* and *orfB* are flanked by imperfect left and right inverted repeats (IRL and IRR, respectively). The latter are bipartite structures, containing an internal protein binding domain and an outer cleavage domain (Lewis et al. 2001). *orfA* contains an AAAAAAG (A$_6$G) slippery signal located near its stop codon, which is changed to A$_7$G by a −1 translational frameshift to produce an OrfAB fusion protein (Lewis and Grindley 1997). Also shown is the outwardly directed −35 hexamer located in the IRR, which together with the terminal −10 hexamer of the catalytic domain of IRL, is involved in the formation of a junction promoter in the minicircle intermediate (Lewis et al. 2004)

Bacterial strains and plasmid DNA

The bacterial strains *Escherichia coli* MG1655 Δ*end*A Δ*rec*A (F^- λ^- *ilv*G *rfb*-50 *rph*1 Δ*end*A Δ*rec*A) and DH5α (F^- 80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F)U169 *rec*A1 *end*A1 *hsd*R17(r_k^-, m_k^+) *pho*A *sup*E44 *thi*-1 *gyr*A96 *rel*A1) were respectively obtained from Kristala Prather's Lab at MIT (Cambridge, MA) and from Invitrogen (Carlsbad, CA). The model plasmids used in this work were pCIneo, a 5,472 bp mammalian expression vector (Promega, Madison, WI), and pVAX1GFP, a 3,697 bp plasmid constructed from pVAX1LacZ (Invitrogen) as described previously (Azzoni et al. 2007).

Cell culture in shake flasks

E. coli cells were grown at 37ºC and 250 rpm in 1 L shake flasks containing 250 mL of Luria-Bertani (LB) medium (Sigma-Aldrich, St. Louis, MO). Ampicillin (Roche Diagnostics, Mannheim, Germany) or kanamycin (Calbiochem, La Jolla, CA) were routinely added for plasmid selection at final concentrations of 100 and 30 μg mL^{-1}, respectively. Cells were grown for a total cumulative time of 192 h (4 consecutive inoculations followed by growth for 48 h each).

Fed-batch fermentations

Fed-batch fermentations were performed in a Fermac 360 Bioreactor (Electrolab, Tewksbury, UK), equipped with equally spaced six-blade Rushton turbines and four baffles, with a working volume of 1.6 L. An initial volume of 1.1 L of medium containing 20 g L^{-1} of glucose (Merck KgaA, Darmstadt, Germany) was used during batch phase, and 0.5 L of feed solution containing 200 g L^{-1} of glucose was subsequently added. The exponential feeding rate was calculated according to the equation described in (Carnes et al. 2006), with μ=0.12 h^{-1}. Pre-inoculum was prepared by transferring cells from seed bank (1 % *v/v*) to 5 mL of medium supplemented with 30 μg mL^{-1} of kanamycin. Cells were grown overnight at 37ºC and 250 rpm. Next, an inoculum was prepared in 100 mL of medium with 1 % pre-inoculum culture and grown to early exponential phase (OD$_{600}$~1.5) at 37ºC and 250 rpm. One liter of medium was in situ autoclaved, and trace elements, kanamycin, and glucose (20 g L^{-1}) were added on the inoculation day. The reactor was inoculated at an initial OD$_{600}$~0.1 using the prepared inoculum. The temperature set point was 37ºC. A dissolved oxygen level of 30 % was cascade-controlled by agitation (250 to 800 rpm), and air was provided at a flow rate of 1 vvm. The pH was controlled at 7.10 using 2 M NaOH and 2 M H$_2$SO$_4$. Antifoam was manually added as required. Two different media were used: (i) a chemically defined medium

((NH$_4$)$_2$SO$_4$ (Panreac Quimica, Barcelona, Spain) 8 g L^{-1}, KH$_2$PO$_4$ (Merck) 13.3 g L^{-1}, citric acid (Sigma-Aldrich) 1.7 g L^{-1}, thiamine hydrochloride (Sigma-Aldrich) 120 mg L^{-1}, MgSO$_4$ (Riedel de Haën, Seelze, Germany) 1.2 g L^{-1}, trace elements solution 1 mL L^{-1}) (Listner et al. 2006) and (ii) a complex medium (yeast extract (Merck) 10 g L^{-1}, bacto tryptone (Becton, Dickinson and Company, Sparks, MD) 10 g L^{-1}, NaCl (Panreac) 10 g L^{-1}).

Plasmid DNA purification and analysis

Plasmid DNA was routinely purified using the High Pure Plasmid Isolation Kit (Roche Diagnostics). DNA extraction from agarose gels was performed using the QIAquick gel extraction kit (Qiagen, Crawley, UK). DNA sequencing was performed by STABVIDA (Portugal).

Detection and quantification of IS*2* insertions

Real-time PCR was used to detect and quantify the number of plasmid molecules with IS*2* insertions using outward IS*2*- and plasmid-specific primer pairs. For example, for the region of pVAX1GFP having an abrupt compositional asymmetry, we amplified a 255 bp fragment comprising the 3′ extremity of the *gfp* gene and the 5′ extremity of IS*2* using primers GFPFOR (5′ CCACTACCTGAGCACCCAG 3′) and IS2REV (5′ TACACCATGTTGCCGGGC 3′). For the pCIneo region having an abrupt compositional asymmetry (see text and also Figs. S1 and S2), this was performed with the amplification of a 244 bp fragment that comprised the entire inverted right repeat (IRR) and part of the 3′ extremity of the *neo*^{R} gene using primers IS2FOR (5′ GTGACTACAT CAGTATCATGC 3′) and IS2REV2 (5′ CCTGCGTGCAAT CCATCT 3′). Reactions were carried out using miniprep-purified pDNA according to the following program: 10 min at 95ºC followed by 45 cycles of 15 s at 95ºC, 15 s at 55ºC, and 15 s at 72ºC. Positive standards were prepared with increasing amounts of pure synthetic fragments containing the amplified region, ranging from 2.0×10^1 to 2.0×10^5. Appropriate dilutions were performed with milliQ water, and 4 μL of each of the former were mixed with the remaining PCR reagents as described below. Negative controls with no pDNA were also prepared. All PCR reactions were performed in a 20 μL final volume containing 1.6 μL of MgCl$_2$ solution (2.0 mM final concentration), 0.4 μL of each primer (0.2 μM final concentrations), 11.6 μL PCR grade water, 2.0 μL of 10× SYBR Green I mixture, and 4.0 μL of sample. Reactions were carried out in a Roche Light Cycler™ detection system, and threshold cycle (C_T) values were calculated by the LightCycler software version 3.4 (Roche Diagnostics) using the fit points method. Mean C_T values and standard deviations were calculated from at least three independent assays.

Results

Transposition of IS2 is biased towards regions harboring abrupt shifts in GC skew

We have shown in a recent study that IS2 preferentially transposes into intrinsically curved regions (Lewis et al. 2012). Because DNA composition asymmetries have been previously linked with transitions between rigid and flexible regions (Pedersen et al. 1998), we sought to understand in more detail if a similar relation could exist and explain the intrinsic curvature observed at IS2 target sites. For this purpose, we performed a literature search and gathered 41 IS2 insertion sites, each having 1.4 kb in length (±0.7 kb flanking the target site) (see Suppl. Material S1). During this process, we considered only recent transposition events (typically associated with major phenotypic changes and flanked by 5 bp duplicaous at the target site), thus avoiding including regions in which genomic erosion might have taken place. GC and AT skew profiles were subsequently computed using 200 bp windows and a 100 bp step. We consistently observed ssGCs and/or switch sites of AT skew (ssATs) occurring in the close vicinity of IS2 insertions. The former often consisted in abrupt and symmetric inversions with a crossing point close to the insertion site. To facilitate the analysis and the comparison with published data from other MGEs, we averaged the profiles over our entire data set (Fig. 2a). We observed a negative to positive transition in GC skew and a rather symmetric and less-prominent inversion in AT skew, both taking place at the position corresponding to the insertion site. We next evaluated if the average maximum amplitude observed at ssGCs and ssATs statistically departed from the average compositional fluctuation of the replicons from which they belong (Fig. 2b) (see "Materials and methods" section for details). Our results show significant differences between the observed and background maximum amplitudes for GC skew but not for AT skew, supporting a role of the former in the process of transposition of IS2.

De novo identification of IS2 transposition events occurring at ssGCs and ssATs

In order to validate the observations gathered in our meta-analysis, we performed a de novo identification of IS2 transposition events taking place in DNA regions exhibiting skew profiles resembling the one depicted in Fig. 2a. Since we are particularly interested in improving the stability of backbone sequences of biotechnologically and clinically relevant plasmids, we thought that pVAX1 would be a good candidate, as it is widely used as backbone for DNA vaccine development. Analysis of both GC and AT skew profiles did not reveal any conspicuous switches in the core sequence of pVAX1, but most interestingly, prominent ssGCs and ssATs were observed

within and in the close vicinity of the widely used reporter gene gfp (Fig. 3a). Although this gene is not present in therapeutic formulations of the vector, it is commonly used, for example, to assess transfection efficiencies in eukaryotic cells. Therefore, to investigate the possibility that IS2 transpositions may preferentially take place at this site (or in its close vicinity), we performed several PCR amplifications using plasmid- and IS2-specific primer pairs (see "Materials and methods" for details). The PCR amplification corresponding to this particular region showing a high asymmetry in GC and AT skew, led to the identification of a 255 bp amplicon, which after sequencing analysis, revealed the presence of an IS2 element disrupting the gfp gene (Fig. 3a, red triangle). No additional IS2 insertions were detected in the plasmid (data not shown). This region in the gfp gene is also shown to adopt an intrinsically bent conformation (Fig. 3b) in line with previous findings from our group (Lewis et al. 2012). The insertion produced a duplication of a 5 bp sequence (CCCCA) and resulted in the formation of a hybrid promoter with an outwardly directed −35 hexamer from IS2 and a resident −10 hexamer from gfp (Fig. 3c). The formation of such transient promoters is neither new nor exclusive to IS2 (Oliveira et al. 2009b; Zhang and Saier 2009) and represents one of the strategies employed by some ISs for gene activation. Our observations of IS2 insertions at ssGCs and ssATs are not exclusive to pVAX1. The only IS2 insertion site detected in the mammalian expression vector pCIneo (Fig. S1) (Oliveira et al. 2009b) is also located at a region showing a high asymmetry in GC and AT skew, and this further corroborates our observations.

Percentage of IS2 insertions in E. coli plasmids increases with time under standard growth conditions

Understanding how external environmental stresses impact the behavior of ISs is central to predict and explain the evolvability of both natural and laboratory cell populations. In this and the following sections, we examine how IS2 accumulation in pVAX1GFP is shaped under growth conditions that are either routinely used in a laboratory setting (typically smaller working volumes, standard growth conditions, unmonitored pH, and aeration) or during a larger-scale production process (larger working volumes, higher cell densities, higher shear-stresses, monitored pH, and aeration). For this purpose, plasmid pVAX1GFP was gel-purified (to avoid starting with a heterogeneous population of parental and contaminant forms) and used to transform E. coli cells with two different genetic backgrounds: DH5α and MG1655 (see genotype details in the "Materials and methods" section). DH5α and MG1655 were chosen in this study as two examples of common laboratory strains, which have approximately the same number of full-length IS2 copies per genome
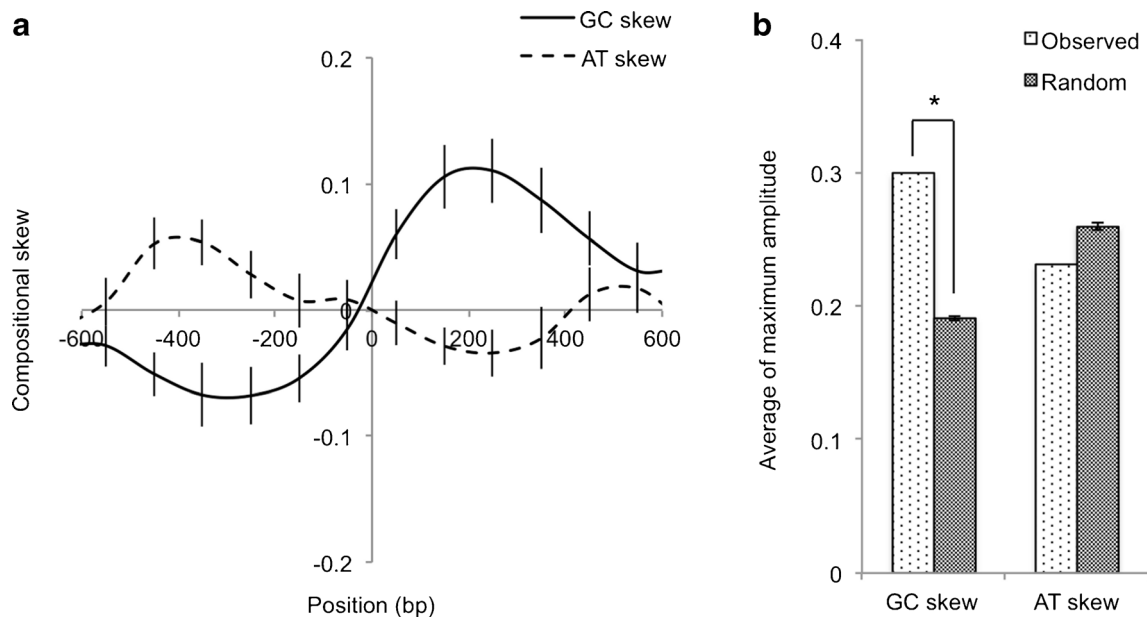
Fig. 2 Relation between DNA compositional asymmetries and IS2 transposition. **a** Average profiles of GC and AT skew 550 bp upstream and downstream of the insertion site of IS2. **b** Averages of maximum amplitude in GC and AT skews in the same window length used for skew analysis, both at observed and random IS2 target sites. *Error bars* represent the standard error of the mean. A more permissive threshold ($p < 0.1$) was allowed in this particular analysis to allow for increased sensitivity and in order not to miss positive results. *$p < 0.1$

(respectively 7.3±4.3 (PCR determined in this work) and 6 (Studier et al. 2009)).

To evaluate the kinetics of IS2 accumulation on a smaller scale, typical of many laboratory settings, cells were grown under standard conditions (shake flask culture, LB medium, 37°C) for an initial period of 48 h after which they were used to reinoculate a new shake flask for another 48 h. Such extended growth until late-stationary phase is frequently used during the overproduction of certain proteins or chemical compounds of biological interest, and it was used here as a worst-case scenario in terms of IS accumulation. The procedure was repeated four times until reaching a maximum accumulated growth time of 192 h. Samples taken at the end of each growth period were analyzed by real-time PCR for quantification of IS2 insertions (Fig. 4). Our results show that detectable levels of transposition (approximately 1 in 2×10^7 plasmid molecules) can be attained at very early stages of cell culture (soon after transformation and during inoculum growth) (time 0). Assuming a purely stochastic distribution of IS2-contaminated plasmids in the cell population, we can roughly estimate that at this stage, approximately 1 out of 6.7×10^4 cells will contain one plasmid with IS2 insertions (assuming an average number of 300 copies per cell). Our results also point out that growing *E. coli* for one single round of 48 h is enough to quickly increase the number of IS2 insertions as much as 1,000-fold (Fig. 4). This necessarily implies that our chance of finding one IS2-contaminated plasmid increased substantially, being now confined to a population of only 67 cells. By performing consecutive growth-reinoculation cycles, we observed a gradual accumulation of

IS2 insertions until a plateau stage was eventually reached (from 48 to 192 h after the first inoculation). We must emphasize that the profile of IS2 accumulation seen on Fig. 4 does not seem to be exclusive to pVA1GFP. We have performed a similar analysis using a different plasmid harboring a different target site, and the overall profile was found to be similar (Figs. S1 and S2) (Oliveira et al. 2009b). The observed dynamics of IS accumulation seems to reflect a balancing conflict between the need to fuel diversity under stress (transposition) and the requirement to assure genetic stability and to keep metabolic burden to bearable levels. Hence, we demonstrate that a single batch growth performed under standard growth conditions is enough to rapidly increase the population of plasmids with IS2 insertions. The fact that we have considered only one particular type of IS, raises concerns about the real proportion of contaminated plasmids.

Growth conducted under high-cell densities and in a chemically defined medium exacerbates IS2 transposition

We next investigated whether growth performed under conditions typical of a larger-scale production process could influence IS2 accumulation. For this purpose, fed-batch fermentations were conducted using both complex and chemically defined media (Fig. 5) (see Materials and Methods for details). Although the time-dependent accumulation of contaminants is similar to that previously observed under batch mode, we observe that at high-cell densities, contamination titers are substantially increased. In particular, when cells were grown
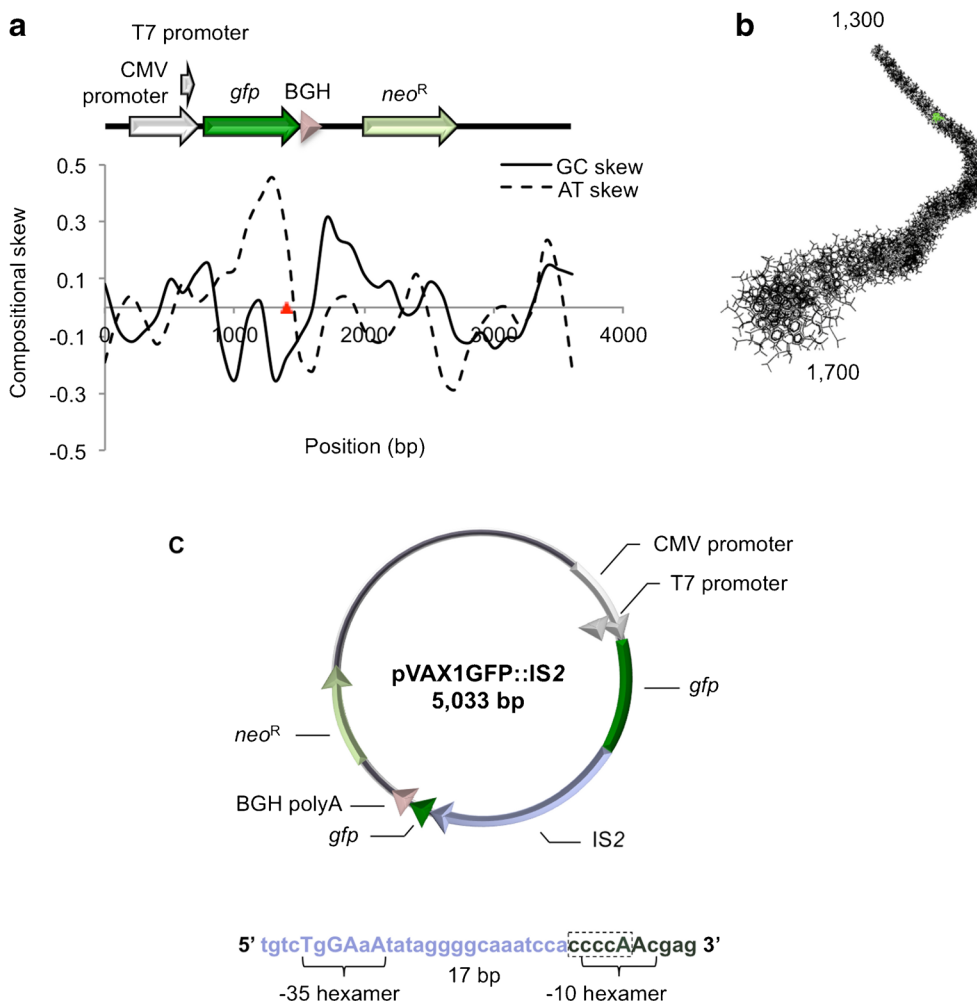
Fig. 3 Skew analysis and IS*2* insertion site in pVAX1GFP. **a** Profiles of GC and AT skew for plasmid pVAX1GFP. A large shift in DNA composition is observed close to the 3′ region of the *gfp* gene. PCR amplification of this region led to the identification of a spontaneous IS*2* insertion (*red arrow*) 83 bp upstream the stop codon. Less prominent and abrupt shifts can also be observed in the 5′ region of *gfp* and in the 5′ region of *neo*R (both only for GC skew). Although we have not detected transposition events in these latter regions, recent studies have done so (see main text). **b** Three-dimensional representation of a 0.4 kb region flanking the insertion site (shown in *green*). The region shows intrinsic curvature close to the insertion site, and concomitantly, in the vicinity of sequences showing abrupt compositional shifts. Start and end positions of the DNA fragment are indicated and refer to pVAX1GFP numbering. **c** Schematic representation of the pVAX1GFP::IS*2* plasmid, highlighting the hybrid promoter generated between the 3′ end of IS*2* (*purple*) and the *gfp* gene (*green*). *Capital letters* match the *E. coli* promoter consensus sequence TTGACAN$_{16-18}$TATAAT. The 5 bp duplication generated upon insertion is shown as a *stippled box*

in a chemically defined medium, the number of IS*2* contaminants surpassed 100,000 per 2×10^8 plasmid molecules after 36 h of growth (Fig. 5). If again, we assume an average plasmid copy number of 300 per cell, this would correspond to a situation in which we have approximately one out of seven cells contaminated with at least one mutant plasmid.

During a high-density fed-batch culture, cells are typically subjected to a scenario of substrate limitation. Under such stressful conditions, bacteria are known to undergo transcriptional, metabolic, morphological, and translational changes, as well as to increase their mutation rate (including the transposition of MGEs) in an SOS and RpoS-dependent manner (more on this below). Our data is in line with these observations of stress-induced transposition and also highlights the fact that invasion of a DNA molecule by a foreign MGE can take place at high frequencies and in short periods of time, even at target regions for which there is no clear advantage for the cell.

Discussion

In this study, we have performed a meta-analysis of published IS*2* target sites from genomic, phage, and plasmid DNA and found compelling evidence for the presence of this element within, or in the close vicinity of regions having abrupt shifts and/or inversions in GC skew. In earlier studies, DNA sequence asymmetries have been shown to be associated with
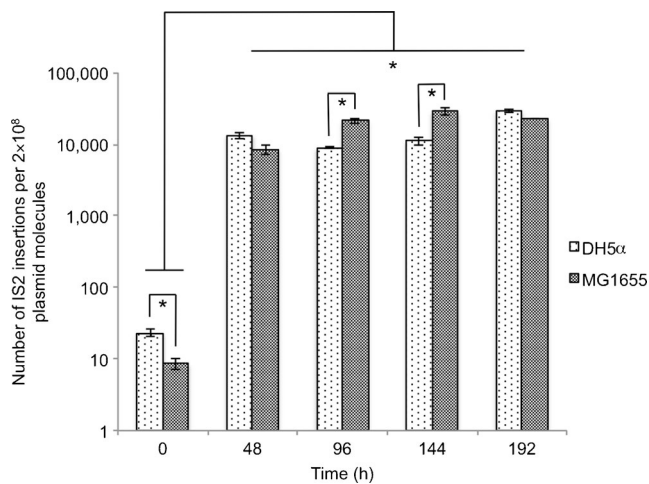
Fig. 4 Effect of cultivation time under standard growth conditions (shake flask, LB medium, 37ºC) on the accumulation of IS2 insertions in pVAX1GFP. Four sequential growths of 48 h each were performed with two different *E. coli* host strains: DH5α and MG1655 ΔendA ΔrecA. The number of insertions was measured by real-time PCR and expressed per 2×10^8 total plasmid molecules. *Error bars* represent the standard error of the mean and were calculated from at least three independent experiments. *$p < 0.05$

alternating rigid and flexible regions (Pedersen et al. 1998) as well as curved sequences (Brukner et al. 1995). The presence of sequence-directed bends has been implicated in functionally relevant cellular processes including transcription (Cress and Nevins 1996), replication (Gimenes et al. 2008), and recombination (Kusakabe et al. 2001). In this study, we present evidence that composition asymmetry directly affects the bendability and configuration of the target region, which in turn, likely modulates IS2 transposition in a way which we do not yet fully understand. Since both the IS minicircle junction
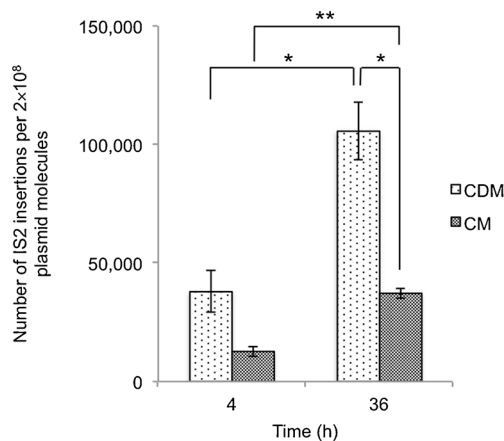

Fig. 5 Effect of medium composition and culture under high-cell densities on the number of IS2 insertions in pVAX1GFP. Fed-batch fermentations were performed with *E. coli* DH5α containing pVAX1GFP in two different media: chemically defined medium (CDM) and complex medium (CM). The number of insertions was measured by real-time PCR and expressed per 2×10^8 total plasmid molecules at 4 and 36 h of culture. *Error bars* represent the standard error of the mean and were calculated from at least three independent experiments. *$p < 0.05$, **$p < 0.01$

and target sites adopt bent structures that share identical curvature profiles (Lewis et al. 2012), it seems plausible that a similar conformation could facilitate the transposition process within the synaptic complex. Bending results in a compaction of the DNA structure that might contribute to an increased level of sequence discrimination and a proper alignment of the target DNA within the catalytic domains of the transpososome.

To further validate our in silico observations, we performed a de novo search for IS2 transpositions in a plasmid pertaining to the pVAX1 family. The rationale for this choice was twofold: first, multicopy systems such as plasmids render themselves more appropriate to an easier detection and to a more accurate quantification of spontaneous mutations that are not associated with any phenotypic change. Second, transposition of bacterial ISs into pDNA represents a major regulatory concern if clinical applications are envisaged, as these elements can seriously compromise the overall production yield obtained and alter the therapeutic potential and safety of the molecule (Oliveira et al. 2009a; Oliveira et al. 2012). Published examples of these are the detection of IS1 and IS5 in an HIV DNA vaccine (Prather et al. 2006), the contamination with IS1 of a plasmid used for cystic fibrosis therapy (Boyd et al. 1999), and the detection of IS2 transposition upstream of the *neo*[R] gene of the mammalian expression vector and DNA vaccine backbone pCIneo (Oliveira et al. 2009b). Interestingly, van der Heijden and coworkers recently observed an IS2 transposition event in a plasmid from the pVAX1 family (van der Heijden et al. 2013). In this case, the insert could be detected between the BGH polyadenylation signal and the *neo*[R] gene during a fed-batch fermentation step, but not in the inoculum from the master cell bank. The fact that the authors were not able to detect transposition at early stages of the production process seems to be in agreement with the existence of a less abrupt shift in GC skew in this region (see Fig. 3a).

To the best of our knowledge, there is only one other published study describing the loss of function of *gfp* by invading ISs (Yang et al. 2013). In that study, the authors observed the spontaneous transposition of IS2 into a region closer to the 3′ end of the gene, but did not provide further details about the exact insertion positions or transposition frequency.

In addition, we now have compelling evidence that transposition events of IS elements from families other than the IS3 family also correlate with the presence of abrupt compositional shifts, and thus with the intrinsic curvature of target DNA. One example is that of IS1, a 768 bp element ubiquitous within the Enterobacteriaceae family, for which we also found a preference for integration into ssGCs (see Fig. S3). Although the mechanisms behind the target specificity of IS1 are still poorly known, both its terminal inverted repeats as well as one of its insertion hotspots are known to contain binding sites for

the integration host factor (IHF) (Gamas et al. 1987), which in turn, has been shown to have a higher affinity for intrinsically bent/kinked DNA (Vivas et al. 2012). These findings are in line with our observations of abrupt shifts in GC skew, and both support a scenario similar to the one of IS2, in which intrinsically bent/kinked conformations are necessary both at the level of minicircle junction and cognate target DNA for binding site recognition (Lewis et al. 2012). These data then suggest a structural basis for the nonrandom regional specificity of the IS2 insertional pattern (Lewis et al. 1994a).

Bearing in mind the selfish nature of MGEs, it comes as no surprise that their mobility also appears to be associated with stress responses. In these situations, enhanced levels of transposition may prove advantageous in overcoming stress and allowing new traits to evolve. In this study, we demonstrate that variables of the production process such as culture time, process scale, and medium composition directly affect the transposition of IS2. The importance of these findings is of particular relevance in order to minimize unwanted insertion events in a high-quality pDNA production process for clinical or commercial applications. By performing sequential batch cultures of 48 h each (Fig. 4), we have allowed the cells to regularly reach an extended stationary phase, which is typically characterized by an extreme growth arrest, depletion of the carbon source, amino acid and oxygen limitations, pH shifts, secondary metabolite production, and profound gene expression changes (Kolter et al. 1993). Under such conditions, many Proteobacteria strongly upregulate the rpoS-encoded general stress response sigma factor σ^S, which is known to control the transcription of up to 10 % of the genome (Weber et al. 2005) and to play a key role in the adaptation to new environments. A further observation is that previous reports have implicated σ^S in the regulation of transcription from the Tn4652 transposase promoter of Pseudomonas putida (Ilves et al. 2001) and in stress-induced transposition-mediated deletions in E. coli (Gomez-Gomez et al. 1997). Also, nutrient starvation has been reported to enhance translational errors, and consequently, to facilitate the programmed translational frameshifting needed for transposition (Kharat et al. 2006). It is likely that in our case, both factors (σ^S dependency and attenuation of ribosome fidelity) may account for the enhanced levels of IS2 transposition observed under prolonged and consecutive batch cultures.

In this study, we also observed higher levels of IS2 transposition during high-cell density fed-batch fermentations in comparison with shake flask cultures; these were particularly pronounced when a chemically defined medium was used instead of complex medium (Fig. 5). In contrast with several evolution experiments performed in bacteria (typically in batch mode and smaller scale), the genomic stability with respect to the dynamics of MGEs in bioreactors has been poorly studied (Gaffe et al. 2011). E. coli grown at slow specific growth rates (μ=0.1 h^{-1}) in chemostats have shown

high levels of σ^S accumulation (Gaffe et al. 2011). The latter could at least partially explain the high transposition values observed in our fed-batch experiments, in which a very similar specific growth rate was used (μ=0.12 h^{-1}). On the other hand, the richer complex medium contains biosynthetic precursors that can be channeled directly into anabolic pathways, thus saving metabolic energy to the cell. The higher levels of transposition observed in chemically defined media could be explained on the basis of a nutritional down shift, followed by the initiation of the stringent response by rapid (p)ppGpp synthesis and consequent σ^S accumulation. Prather and collaborators showed a similar accumulation of IS1 transposition during a fed-batch fermentation using chemically defined medium (Prather et al. 2006).

Ultimately, this study extends the current knowledge on the mechanisms driving transposition of MGEs and opens up exciting perspectives for further research. A deeper understanding of the detailed mechanism of IS transposition may allow us to conceive of redirecting insertion specificity in a predictable way (Guynet et al. 2009), or instead, pave the way for the design of safer and more evolutionary robust genetic circuits. Our results with bench-scale bioreactors also raise important issues and concerns on the safety of DNA-based therapeutics upon amplification at a larger scale.

References

Azzoni AR, Ribeiro SC, Monteiro GA, Prazeres DMF (2007) The impact of polyadenylation signals on plasmid nuclease-resistance and transgene expression. J Gene Med 9:392–402. doi:10.1002/jgm.1031

Bernard N, Ferain T, Garmyn D, Hols P, Delcour J (1991) Cloning of the D-lactate dehydrogenase gene from Lactobacillus delbrueckii subsp. bulgaricus by complementation in Escherichia coli. FEBS Lett 290(1–2):61–64. doi:10.1016/0014-5793(91)81226-X

Boyd AC, Popp F, Michaelis U, Davidson H, Davidson-Smith H, Doherty A, McLachlan G, Porteous DJ, Seeber S (1999) Insertion of natural intron 6a-6b into a human cDNA-derived gene therapy vector for cystic fibrosis improves plasmid stability and permits facile RNA/DNA discrimination. J Gene Med 1(5): 312–321. doi:10.1002/(SICI)1521-2254(199909/10) 1:5<312::AID-JGM55>3.0.CO;2-#

Brukner I, Sànchez R, Suck D, Pongor S (1995) Sequence-dependent bending propensity of DNA as revealed by DNase I: parameters for trinucleotides. EMBO J 14(1812–1818)

Carnes AE, Hodgson CPA, Williams J (2006) Inducible Escherichia coli fermentation for increased plasmid DNA production. Biotechnol Appl Biochem 45:155–166. doi:10.1042/BA20050223

Chandler M, Mahillon J (2002) Insertion sequences revisited. In: Craig NL, Craigie R, Gellert M, Lambowitz AM (eds) Mobile DNA II. ASM Press, Washington DC, pp 305–366

Cohen SP, Hachler H, Levy SB (1993) Genetic and functional analysis of the multiple antibiotic resistance (*mar*) locus in *Escherichia coli*. J Bacteriol 175(5):1484–1492. doi:10.1053/rvsc.2001.0537

Cress WD, Nevins JR (1996) A role for a bent DNA structure in E2F-mediated transcription activation. Mol Cell Biol 16(5):2119–2127

Du P, Yang Y, Wang H, Liu D, Gao GF, Chen C (2011) A large scale comparative genomic analysis reveals insertion sites for newly acquired genomic islands in bacterial genomes. BMC Microbiol 11:135. doi:10.1186/1471-2180-11-135

Eichenbaum Z, Livneh Z (1998) UV light induces IS*10* transposition in *Escherichia coli*. Genetics 149(3):1173–1181

Gabrielian A, Vlahovicek K, Pongor S (1997) Distribution of sequence-dependent curvature in genomic DNA sequences. FEBS Lett 406(1–2):69–74. doi:10.1016/S0014-5793(97)00236-6

Gaffe J, McKenzie C, Maharjan RP, Coursange E, Ferenci T, Schneider D (2011) Insertion sequence-driven evolution of *Escherichia coli* in chemostats. J Mol Evol 72(4):398–412. doi:10.1007/s00239-011-9439-2

Gamas P, Chandler MG, Prentki P, Galas DJ (1987) *Escherichia coli* integration host factor binds specifically to the ends of the insertion sequence IS*1* and to its major insertion hot-spot in pBR322. J Mol Biol 195(2):261–272. doi:10.1016/0022-2836(87)90648-6

Gimenes F, Takeda KI, Fiorini A, Gouveia FS, Fernandez MA (2008) Intrinsically bent DNA in replication origins and gene promoters. Genet Mol Res 7(2):549–558

Gomez-Gomez JM, Blazquez J, Baquero F, Martinez JL (1997) H-NS and RpoS regulate emergence of Lac Ara+mutants of *Escherichia coli* MCS2. J Bacteriol 179(14):4620–4622

Guynet C, Achard A, Hoang BT, Barabas O, Hickman AB, Dyda F, Chandler M (2009) Resetting the site: redirecting integration of an insertion sequence in a predictable way. Mol Cell 34(5):612–619. doi:10.1016/j.molcel.2009.05.017

Hallet B, Rezsohazy R, Mahillon J, Delcour J (1994) IS*231A* insertion specificity—consensus sequence and DNA bending at the target site. Mol Microbiol 14(1):131–139. doi:10.1111/j.1365-2958.1994.tb01273.x

Hirsch HJ, Starlinger P, Brachet P (1972) Two kinds of insertions in bacterial genes. Mol General Genet 119(3):191–206. doi:10.1007/BF00333858

Ilves H, Horak R, Kivisaar M (2001) Involvement of sigma(S) in starvation-induced transposition of *Pseudomonas putida* transposon Tn*4652*. J Bacteriol 183(18):5445–5448. doi:10.1128/JB.183.18.5445-5448.2001

Kharat AS, Coursange E, Noirclerc-Savoye M, Lacoste J, Blot M (2006) IS*1* transposition is enhanced by translation errors and by bacterial growth at extreme glucose levels. Acta Biochimica Polonica 53(4):729–738

Kobori S, Ko Y, Kato M (2009) A target site for spontaneous insertion of IS10 element in pUC19 DNA located within intrinsically bent DNA. Open Microbiol J 3:146–150. doi:10.2174/1874285800903010146

Kolter R, Siegele DA, Tormo A (1993) The stationary phase of the bacterial life cycle. Annual Review of Microbiology 47:855–874. doi:10.1146/annurev.mi.47.100193.004231

Kusakabe T, Sugimoto Y, Maeda T, Nakajima Y, Miyano M, Nishikawa J, Tone S, Kawaguchi Y, Koga K, Ohyama T (2001) Linearization and integration of DNA into cells preferentially occurs at intrinsically curved regions from human LINE-1 repetitive element. Gene 274(1–2):271–281. doi:10.1016/S0378-1119(01)00631-X

Lewis LA, Grindley ND (1997) Two abundant intramolecular transposition products, resulting from reactions initiated at a single end, suggest that IS2 transposes by an unconventional pathway. Mol Microbiol 25(3):517–529. doi:10.1046/j.1365-2958.1997.4871848.x

Lewis LA, Gopaul S, Marsh C (1994a) The non-random pattern of insertion of IS2 into the *hemB* gene of *Escherichia coli*. Microbiol Immunol 38(6):461–465. doi:10.1111/j.1348-0421.1994.tb01808.x

Lewis LA, Lewis D, Persaud V, Gopaul S, Turner B (1994b) Transposition of IS2 into the *hemB* gene of *Escherichia coli* K-12. J Bacteriol 176(7):2114–2120

Lewis LA, Gadura N, Greene M, Saby R, Grindley ND (2001) The basis of asymmetry in IS2 transposition. Mol Microbiol 42(4):887–901. doi:10.1046/j.1365-2958.2001.02662.x

Lewis LA, Cylin E, Lee HK, Saby R, Wong W, Grindley ND (2004) The left end of IS2: a compromise between transpositional activity and an essential promoter function that regulates the transposition pathway. J Bacteriol 186(3):858–865. doi:10.1128/JB.186.3.858-865.2004

Lewis LA, Astatke M, Umekubo PT, Alvi S, Saby R, Afrose J (2011) Soluble expression, purification and characterization of the full length IS2 transposase. Mob DNA 2:14. doi:10.1186/1759-8753-2-14

Lewis LA, Astatke M, Umekubo PT, Alvi S, Saby R, Afrose J, Oliveira PH, Monteiro GA, Prazeres DM (2012) Protein-DNA interactions define the mechanistic aspects of circle formation and insertion reactions in IS2 transposition. Mob DNA 3(1):1. doi:10.1186/1759-8753-3-1

Listner K, Bentley LK, Chartrain M (2006) A simple method for the production of plasmid DNA in bioreactors. Meth Mol Med 127:295–309. doi:10.1385/1-59745-168-1:295

Muller HP, Varmus HE (1994) DNA bending creates favored sites for retroviral integration: an explanation for preferred insertion sites in nucleosomes. EMBO J 13(19):4704–4714

Nagy Z, Chandler M (2004) Regulation of transposition in bacteria. Res Microbiol 155(5):387–398. doi:10.1016/j.resmic.2004.01.008

Oliveira PH, Prather KJ, Prazeres DM, Monteiro GA (2009a) Structural instability of plasmid biopharmaceuticals: challenges and implications. Trends Biotechnol 27(9):503–511. doi:10.1016/j.tibtech.2009.06.004

Oliveira PH, Prazeres DM, Monteiro GA (2009b) Deletion formation mutations in plasmid expression vectors are unfavored by runaway amplification conditions and differentially selected under kanamycin stress. J Biotechnol 143(4):231–238. doi:10.1016/j.jbiotec.2009.08.002

Oliveira PH, Prazeres DM, Monteiro GA (2012) Structural and segregational instability in plasmid biology. In: Gonzalez FER, Lopez MI (ed) Plasmids: genetics, applications and health. Nova Publishers, NY, pp. 79–100

Pedersen AG, Baldi P, Chauvin Y, Brunak S (1998) DNA structure in human RNA polymerase II promoters. J Mol Biol 281(4):663–673. doi:10.1006/jmbi.1998.1972

Polard P, Prere MF, Chandler M, Fayet O (1991) Programmed translational frameshifting and initiation at an AUU codon in gene expression of bacterial insertion sequence IS911. J Mol Biol 222(3):465–477. doi:10.1016/0022-2836(91)90490-W

Porollo AA, Adamczak R, Meller J (2004) POLYVIEW: a flexible visualization tool for structural and functional annotations of proteins. Bioinformatics 20(15):2460–2462. doi:10.1093/bioinformatics/bth248

Prather KL, Edmonds MC, Herod JW (2006) Identification and characterization of IS1 transposition in plasmid amplification mutants of *E. coli* clones producing DNA vaccines. Appl Microbiol Biotechnol 73(4):815–826. doi:10.1007/s00253-006-0532-1

Pribil PA, Haniford DB (2003) Target DNA bending is an important specificity determinant in target site selection in Tn*10* transposition. J Mol Biol 330(2):247–259. doi:10.1016/S0022-2836(03)00588-6

Rice P, Longden I, Bleasby A (2000) EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet 16(6):276–277. doi:10.1016/S0168-9525(00)02024-2

Rijavec M, Budic M, Mrak P, Muller-Premru M, Podlesek Z, Zgur-Bertok D (2007) Prevalence of ColE1-like plasmids and colicin K production among uropathogenic *Escherichia coli* strains and quantification of inhibitory activity of colicin K. Appl Environ Microbiol 73(3):1029–1032. doi:10.1128/AEM.01780-06

Schneeberger RG, Zhang K, Tatarinova T, Troukhan M, Kwok SF, Drais J, Klinger K, Orejudos F, Macy K, Bhakta A, Burns J, Subramanian G, Donson J, Flavell R, Feldmann KA (2005) *Agrobacterium* T-DNA integration in *Arabidopsis* is correlated with DNA sequence compositions that occur frequently in gene promoter regions. Funct Integr Genomics 5(4):240–253. doi:10.1007/s10142-005-0138-1

Sekine Y, Eisaki N, Ohtsubo E (1994) Translational control in production of transposase and in transposition of insertion sequence IS3. J Mol Biol 235(5):1406–1420. doi:10.1006/jmbi.1994.1097

Sengstag C, Arber W (1983) IS2 insertion is a major cause of spontaneous mutagenesis of the bacteriophage P1: non-random distribution of target sites. EMBO J 2(1):67–71

Sengstag C, Arber W (1987) A cloned DNA fragment from bacteriophage P1 enhances IS2 insertion. Mol General Genet 206(2):344–351. doi:10.1007/BF00333593

Sengstag C, Shepherd JC, Arber W (1983) The sequence of the bacteriophage P1 genome region serving as hot target for IS2 insertion. EMBO J 2(10):1777–1781

Studier FW, Daegelen P, Lenski RE, Maslov S, Kim JF (2009) Understanding the differences between genome sequences of Escherichia coli B strains REL606 and BL21(DE3) and comparison of the E. coli B and K-12 genomes. J Mol Biol 394(4):653–680. doi:10.1016/j.jmb.2009.09.021

Szeverenyi I, Bodoky T, Olasz F (1996a) Isolation, characterization and transposition of an (IS2)₂ intermediate. Mol General Genet 251(3):281–289. doi:10.1007/BF02172518

Szeverenyi I, Hodel A, Arber W, Olasz F (1996b) Vector for IS element entrapment and functional characterization based on turning on expression of distal promoterless genes. Gene 174(1):103–110. doi:10.1016/0378-1119(96)00407-6

Twiss E, Coros AM, Tavakoli NP, Derbyshire KM (2005) Transposition is modulated by a diverse set of host factors in Escherichia coli and is stimulated by nutritional stress. Mol Microbiol 57(6):1593–1607. doi:10.1111/j.1365-2958.2005.04794.x

van der Heijden I, Gomez-Eerland R, van den Berg JH, Oosterhuis K, Schumacher TN, Haanen JB, Beijnen JH, Nuijen B (2013) Transposon leads to contamination of clinical pDNA vaccine. Vaccine 31(32):3274–3280. doi:10.1016/j.vaccine.2013.05.022

Vivas P, Velmurugu Y, Kuznetsov SV, Rice PA, Ansari A (2012) Mapping the transition state for DNA bending by IHF. J Mol Biol 418(5):300–315. doi:10.1016/j.jmb.2012.02.028

Vlahovicek K, Kajan L, Pongor S (2003) DNA analysis servers: plot.it, bend.it, model.it and IS. Nucleic Acids Res 31(13):3686–3687

Vogele K, Schwartz E, Welz C, Schiltz E, Rak B (1991) High-level ribosomal frameshifting directs the synthesis of IS150 gene products. Nucleic Acids Res 19(16):4377–4385. doi:10.1093/nar/gkg559

Weber H, Polen T, Heuveling J, Wendisch VF, Hengge R (2005) Genome-wide analysis of the general stress response network in Escherichia coli: sigmaS-dependent genes, promoters, and sigma factor selectivity. J Bacteriol 187(5):1591–1603. doi:10.1128/JB.187.5.1591-1603.2005

Whiteway J, Koziarz P, Veall J, Sandhu N, Kumar P, Hoecher B, Lambert IB (1998) Oxygen-insensitive nitroreductases: analysis of the roles of nfsA and nfsB in development of resistance to 5-nitrofuran derivatives in Escherichia coli. J Bacteriol 180(21):5529–5539

Yang S, Sleight SC, Sauro HM (2013) Rationally designed bidirectional promoter improves the evolutionary stability of synthetic genetic circuits. Nucleic Acids Res 41(1):e33. doi:10.1093/nar/gks972

Zhang Z, Saier MH Jr (2009) A novel mechanism of transposon-mediated gene activation. PLoS Genet 5(10):e1000689. doi:10.1371/journal.pgen.1000689

Zhang J, Guo D, Chang Y, You C, Li X, Dai X, Weng Q, Zhang J, Chen G, Li X, Liu H, Han B, Zhang Q, Wu C (2007) Non-random distribution of T-DNA insertions at various levels of the genome hierarchy as revealed by analyzing 13 804 T-DNA flanking sequences from an enhancer-trap mutant library. Plant J 49(5):947–959. doi:10.1111/j.1365-313X.2006.03001.x