4

Functionalizing Bacterial Artificial Chromosomes with Transposons to Explore Gene Regulation

Hope M. Wolf[1,2], Oladoyin Iranloye[1,3],
Derek C. Norford[1] and Pradeep K. Chatterjee[1,3]
[1]*Julius L. Chambers Biomedical/ Biotechnology Research Institute*
[2]*Department of Chemistry, University of North Carolina at Chapel-Hill, Chapel-Hill, NC*
[3]*Department of Chemistry, North Carolina Central University, Durham*
USA

1. Introduction

Strategies for altering sequences in large DNA inserts in BACs are fundamentally different from those traditionally used for small plasmids. Two factors are primarily responsible for this: the existence of a high multiplicity of sites in BACs recognized by DNA modifying enzymes, such as restriction endo-nucleases, as well as the brittle nature of large duplex DNA that is not packaged with DNA-binding proteins in the test-tube. The large number of DNA fragments generated by the common restriction enzymes, and the unavailability of robust separation techniques to isolate and keep track of the relative order of the pieces; precludes using the "cut-and-paste" mechanism to alter DNA sequences in BACs. Instead DNA recombination *in vivo* has become the method of choice for modifying BACs. Although the net result again is cutting and re-joining of DNA, the entire process is concerted in the bacterial host with no free ends of DNA to go astray; and occurs primarily in a nucleoprotein complex that protects it from shear forces which otherwise would break the large DNA during manipulations *in vitro*.

Two separate approaches based upon using sequence homology for recombining DNA were developed independently to alter sequences in BACs. The first method introduces the major recombination function of *E.coli*, RecA, back into the severely recombination deficient host DH10B, originally engineered to propagate vertebrate DNA in BACs (1). The second approach introduces recombination functions of phage λ, namely red α, red β and red γ into the DH10B host, and utilizes recombination of homologous sequences of shorter length to insert exogenous DNA cassettes into BACs (2). Both methods have been widely used to engineer a variety of alterations in BAC DNA, such as introducing reporter gene cassettes into the genomic insert, mutating sequences at a target site, and introducing loxP sites (1-11).

A different strategy for modifying BAC DNA also uses recombination, but does not require targeting vectors to carry sequences homologous to those in the genomic insert to introduce exogenous DNA cassettes. Insertions of the bacterial Tn10 transposon can introduce

exogenous DNA, including lox sites, at random locations in the BAC (12, 13). Site-specific recombination using the Cre-lox system on the other hand, can deliver reporter genes and other exogenous DNA cassettes such as sequencing primer sites, mammalian cell-selectable antibiotic resistance genes, enhancer-traps and sequences specific to the vertebrate transposon Tol2 precisely at the ends of the genomic DNA insert in BACs (14-16). It is significant that the recombinases involved in either of these approaches, Tn10-transposase and Cre protein respectively, do not act upon sequence repeats and/or other recombinogenic sites in the genomic DNA insert to rearrange it. This particular characteristic should make the approach applicable to a wider variety of BACs in the public domain, including those containing repetitive sequences [see reference (16) for an example].

Insertions of the Tn10 transposon into DNA of BACs from a wide variety of vertebrate genome libraries appear to be random, demonstrating little sequence specificity for transposition (16). It is unclear whether this lack of sequence specificity arises from the absence of selective pressure evolutionarily for insertions of a prokaryotic transposon into vertebrate DNA, because insertions into prokaryotic DNA have long been known to prefer a somewhat degenerate nevertheless consensus insertion site (17). The minor sequence preferences for insertions of Tn10 observed in BACs probably have more to do with availability of sites for Tn10 insertions in HU protein-packaged vertebrate DNA in the bacterium than specificity for sequences. On occasion, incorporation of Tn10 into insert DNA of a rare BAC clone displays apparent sequence selectivity (18). However, this was clearly shown to be due to a clonal selection process that arose from picking single colonies of transposon plasmid transformed BAC clones which had induced the transposase gene prior to actual induction with Isopropyl β-D-1-thiogalactopyranoside (IPTG). Slight modification of the procedure, that has since recommended inducing a large pool of transposon plasmid transformed BAC colonies instead of a single one, has rectified this potential problem completely (18). A detailed description of the transposon retrofitting approach for BACs follows.

1.1 Methodology for Tn10 transposon retrofitting of BACs

The Tn10 transposon modification procedure for BACs is conceptually simple, as illustrated in Figure 1. A loxP sequence is placed within the 70 base inverted repeat ends of the bacterial Tn10 mini-transposon (shown in the lower panel of Figure 1).

The plasmid DNA carrying the transposon is introduced into the same cell that houses the BAC plasmid and the transposase gene, located outside the inverted repeat ends in the transposon plasmid, is induced. Upon induction the transposase protein excises the DNA cassette flanked by the inverted repeats (shown as green and pink boxes L & R in the transposon plasmid) and inserts it into other nearby DNA in the bacterial host, including the bacterial genome and the BAC DNA. Insertions of the Tn10 into BAC DNA occur in either orientation, and are irreversible because the transposase gene is left behind in the excised transposon plasmid and destroyed. Important considerations that dictate subsequent steps in the procedure include 1) damage to the host genome from Tn10 insertions, and 2) the fraction of BACs actually modified is relatively low ~ 1 in 10,000. Therefore efficient steps are required to recover the low percentage of BACs containing insertions and subsequently transfer these into a new host. Both these challenges are met by packaging the Tn10-modified BAC DNA in phage P1 heads. Therefore the cells containing the BAC DNA with

transposon insertions are infected with P1 phage after induction of the transposase gene with IPTG [see reference (19) for a detailed description of the procedure].

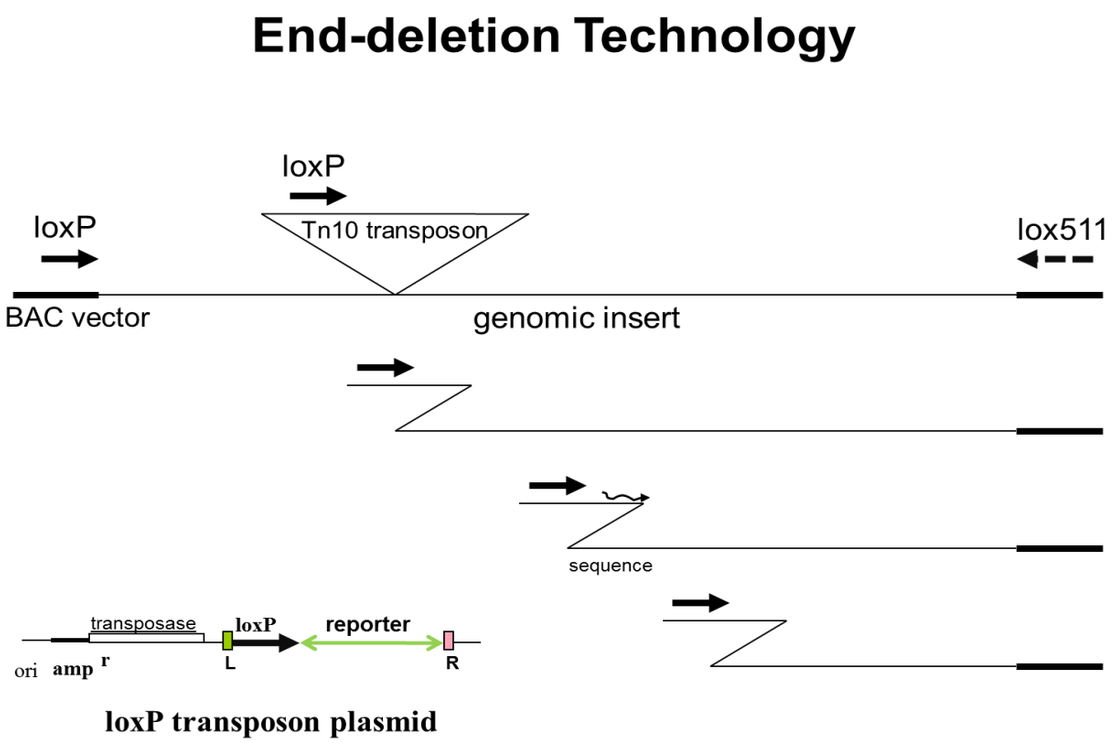


Fig. 1. Legend: Schematic representation of the BAC end-deletion technology using a loxP transposon. The inverted triangle represents the loxP transposon, which is shown in greater detail in the lower panel. The 70 bp inverted repeat ends of the transposon are indicated by the green and pink boxes marked L and R respectively. The thick black arrow represents the loxP site in both the transposon and BAC vector DNA. The broken arrow represents the lox511 site in BAC vector. Because transposon insertions are rare, single random transpositions into the BAC DNA occur when the transposase gene is induced. Upon Cre recombination, the transposon-inserted loxP of identical orientation to the loxP endogenous to the BAC generates a deletion of the DNA between them. Inversions are not shown. The pool of BAC DNA molecules therefore generates a library of end-deleted BACs from the random insertions of a loxP transposon into BAC DNA molecules.

Infection with P1 phage serves an additional purpose. The phage expresses Cre protein early in its life cycle to circularize the otherwise linear DNA within the phage head. Newly synthesized Cre protein acts *in trans* to also recombine the loxP site transposed into the BAC genomic insert with the loxP endogenous to the BAC and located at one end of the insert DNA [see step 1, Figure 2]. Thus the transposed loxP site of one orientation produces a deletion from one end of the genomic insert, while the loxP inserted in the opposite

orientation simply inverts the DNA between it and the one endogenous to the BAC. Because BACs in all modern libraries carry insert DNA of average size ~ 160 kb, and because the amount of DNA that can be packaged in a P1 head is ~110 kb (20), the P1 headful packaging step can serve as a selection strategy to isolate transpositions of loxP in the orientation identical to the endogenous one (12), as only that allows reducing the BAC DNA length to less than ~110 kb. Only those end-deletions that reduce the length of BAC DNA to less than ~110 kb are rescued, as shown in Figure 2.

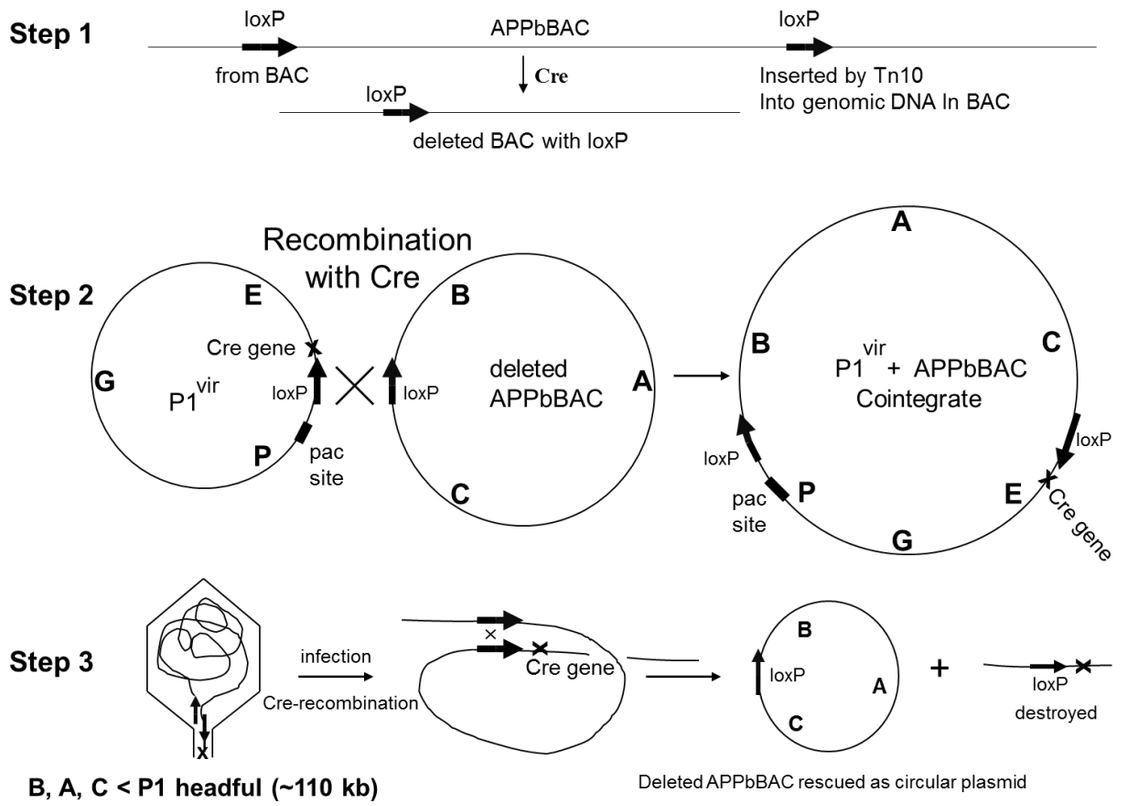


Fig. 2. Legend: Schematic representation of the DNA recombinations occurring in the transposon retrofitting of BACs. Step 1 shows creation of a deletion in the BAC DNA by Cre recombination of the transposed and endogenous loxP sites. Step 2 shows the Cre recombination of circularized phage P1 DNA (P, G, E) with the end-deleted BAC DNA (B, A, C) to generate the co-integrate. The phage packaging site "pac site" and the Cre gene are indicated in the phage DNA by the thick solid bar and the X, respectively. Step 3 shows the BAC DNA packaged in the phage P1 head, its recombination after entry into cells by newly synthesized Cre protein, and circularization into a BAC plasmid. If the length of DNA (B, A, C) exceeds the headful packaging capacity of ~110 kb, then the second loxP site (indicated by the thick arrows) would not fit in the phage head and the DNA cannot circularize by loxP-Cre recombination after entry into cells. This leads to the BAC DNA not being rescued. Note that the Cre gene is also lost upon circularization of the truncated BAC DNA.

Packaging of the end-deleted BAC DNA occurs from a large co-integrate plasmid in the cell which is described in detail elsewhere (21, 22). In co-integrate formation, the P1 phage DNA effectively contributes a packaging location named the "pac site" which is recognized by the phage packaging proteins [shown in step 2, Figure 2]. The co-integrate DNA is thought to replicate by a "rolling circle" mechanism. A cut is made in the newly synthesized co-integrate DNA at the "pac site", and the DNA end corresponding to piece loxP-B, A, C is stuffed into newly assembled empty heads of phage P1. Packaging of DNA from co-integrate is directional and the second cleavage is made after the P1 head is full regardless of sequence using a "headful-cleavage" mechanism (21, 22). Note that if Cre recombination of the transposed loxP with the loxP endogenous to the BAC did not reduce the length of BAC DNA segment B, A, C, to less than ~110 kb, this piece of BAC DNA inside the phage head would not be flanked by loxP on both ends [see step 3 of Figure 2]. In the absence of loxP sites flanking BAC DNA on both ends, the DNA would be unable to circularize by Cre recombination and would be destroyed when introduced into cells. Insertions of loxP in the other orientation, that cause inversion of the DNA, do not reduce the size of segment B, A, C to less than 110 kb. Consequently, the BAC DNA in P1 heads from inversions is destroyed upon entry into cells.

The bacterial lysate containing P1 heads packaged with end-deleted BAC DNA is treated with chloroform. This treatment not only facilitates lysis of P1 infected recombination deficient DH10B but also kills all cells harboring intact transposon and BAC plasmids. This killing is important for the selection of transposon inserted BACs in the next stage when antibiotic selection is employed, as otherwise cells without transposon insertions into BACs but merely carrying the two intact plasmids would get selected. The lysate after chloroform treatment is used to infect fresh bacteria. Upon entry of the phage packaged BAC DNA into cells; Cre protein is expressed from the DNA end adjacent to the second loxP site [indicated as X in phage head and co-integrate], but only transiently, because the Cre gene is lost after Cre-loxP recombination to circularize the linear BAC DNA [Step 3, Figure 2]. Cells are plated on LB plates containing chloramphenicol to select for clones of end-deleted BAC. Note that the phage DNA segment [P, G, E] from the co-integrate is incompletely packaged, if at all, and is destroyed after the loxP-Cre circularization in the new host. Even the rare phage particle containing a complete phage genome is unlikely to replicate or survive because of the chloramphenicol selection on the plates.

Selecting with a single antibiotic for the BAC DNA is sufficient for the first round of end-deletions, using either a loxP or a lox511 transposon. Additional selection for transposition of lox sites is not necessary in this first round because the P1 headful packaging itself serves as the selection for the low frequency of lox-site transpositions in the previous step [see references (23, 24) for detailed discussions].

1.2 Analyzing retrofitted/ end-deleted BACs

BACs deleted from an end, or inserted with exogenous DNA cassettes at the newly created ends of genomic inserts, are analyzed by isolating their DNA and separating them by Field Inversion Gel Electrophoresis (FIGE). The DNA in BAC clones is easily isolated by simple mini-prep procedures, digested with Not I enzyme and analyzed on FIGE (13, 25), as shown in Figure 3. High throughput formats for preparing DNA from BAC deletion clones in 96 well chambers, suitable for subsequent FIGE analyses and end sequencing, have also been described earlier (13, 25).

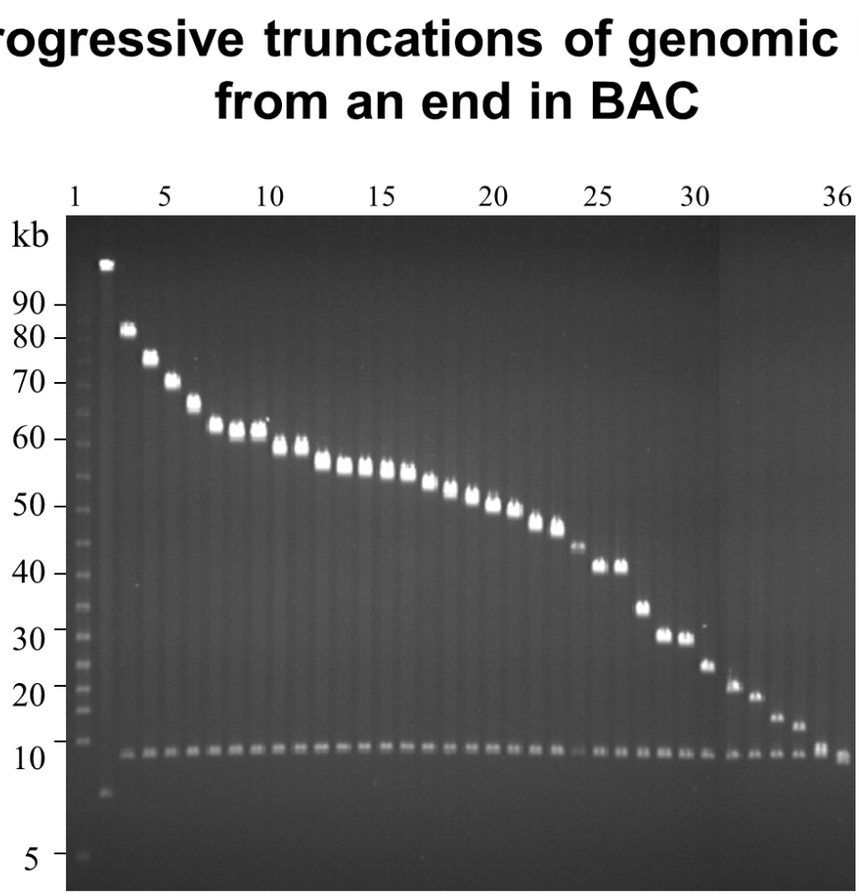


Fig. 3. Legend: FIGE display of an array of BAC deletion clone DNAs. DNA was isolated by the miniprep procedure from clones of a library of end-deleted BACs generated using a loxP transposon. The DNA was digested with Not I enzyme before electrophoresis. Lane 1 shows a 5 kb ladder, and lane 2 shows the DNA from the starting BAC clone.

1.3 Scoring authentic lox-Cre recombinations by unique size of BAC vector DNA upon Not I digestion

Whether the BAC DNA of reduced size arose from an authentic lox-Cre recombination, as opposed to an internal deletion between sequence repeats or other recombinogenic sites in the genomic insert, is one of the early determinations that need to be made. Note that lox-Cre independent deletions can also be packaged by the P1 headful mechanism when there is no selection for lox site transposition *per se;* such deletions are also efficiently isolated (23). Therefore all our loxP and lox511 transposons have been designed to alter the size of the Not I-Not I BAC-vector DNA band that arises from a Not I enzymatic digest of the deletion clone DNA [see references (23, 24, 27) for discussion]. An end-deletion from Cre-loxP recombination alters the size of BAC vector DNA, whereas deletions arising from within the genomic insert do not. End-deletions made with our lox511 transposons remove the Not I site at the lox511 end in the original BAC, resulting in a linear BAC with no separate vector DNA band (24).

Once authenticity of the retrofitted/ end-deleted BAC clone is established, the exact location of insertion of the transposon on genomic DNA is determined by sequencing the BAC DNA directly with unique primer sites located on either the loxP or lox511 transposon ends that remain after deletion formation (13, 24).

1.4 Truncations from the other end of insert DNA in BACs

The genomic DNA insert in BACs is flanked by a loxP site at one end and a mutant lox511 site at the other (24, 26). Modern genome libraries of DNA from numerous organisms such as the mouse, rat, human and zebrafish have used the vector pBACe3.6 or its derivatives [reference (26), link for pTARBAC2.1 vector *http://bacpac.chori.org/ptarbac21.htm*], and all share this characteristic (27-31). End-deletions have been made specifically from the lox511 side of genomic insert by a similar approach using a lox511 Tn10 mini-transposon (24).

It is important to note that second round deletions from the opposite end of insert DNA requires selecting for the transposition event in addition to the BAC DNA: because first round deletions are selected by P1 headfull packaging, all starting BACs for second round deletions would be less than ~110 kb, and hence require selecting for the low frequency of transposition [see reference (23, 24) for detailed discussions].

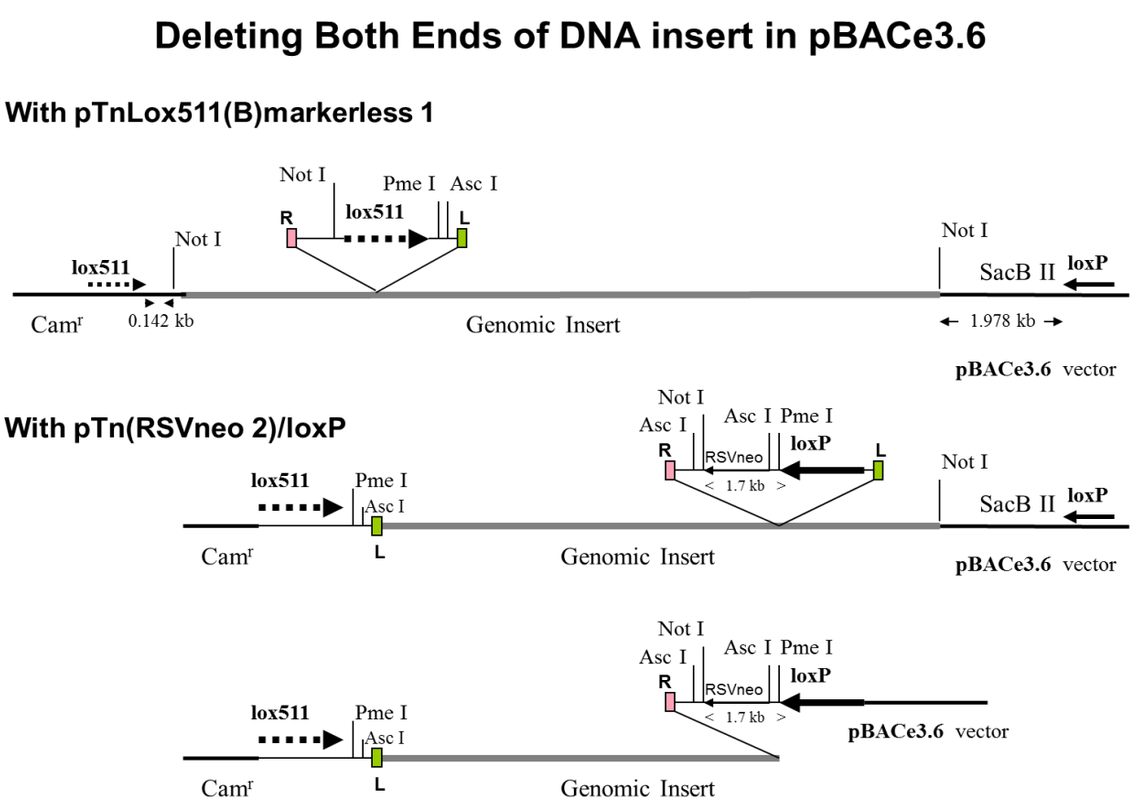


Fig. 4. Legend: Schematic representation of end-deletions generated from both ends of DNA insert in a BAC clone. The upper panel shows truncations from the lox511 end using a lox511 transposon. The largest clone from that library was then truncated from the loxP end with a loxP transposon. Figure is adapted from reference (24).

1.5 Potential Cre-mediated cross-recombination of loxP & lox511 sites does not occur in our method

The 34 bp sequence in the mutant lox site, lox511, differs by one nucleotide in the spacer region from that of loxP [see references (24, 32), for actual sequence of lox511]. Varying degrees of promiscuity in recombining different mutant *lox* sites, including the lox511 mutant, with wild type loxP have been reported in previous studies; using both partially purified Cre-extracts *in vitro* (32-36), or Cre over-expressed in cells (36-38). For example, cross recombination between loxP and lox511 has been reported to occur at efficiencies ranging from 5 to 100 % under those experimental conditions that express Cre constitutively (33, 35-38). We have not observed such cross-recombination, and high levels of stringency *in vivo* in recombining identical *lox* sites (21, 24), or *lox* sites with at least identical spacers (39), have been achieved with Cre protein expressed from its native source namely, a phage P1 infection (21, 24, 39). Depending on whether a loxP or a lox511 transposon is used, truncations from the corresponding lox- end can be made readily and exclusively with high stringency [shown schematically in Figure 4]. Thus truncations of genomic DNA from either end are not only efficient, but are highly specific for that end.

1.6 Exogenous DNA cassettes can be introduced precisely at one or both newly created BAC ends

An important feature of the end-deletion technology is the ease of determining exactly where in the BAC the loxP-transposon had inserted to create the truncation: primers for sequencing the new BAC end have been designed into the loxP and lox511 transposon-ends remaining after the lox-Cre recombinations [see Figure 1 and references (13, 24)], and these have been used to generate ~600 base reads by sequencing the retrofitted BAC DNA directly. An additional feature is the ability to introduce reporter genes and other DNA cassettes precisely at the new end created in the BAC clone. The sequence in front of the loxP or lox511 arrows, as shown in Figures 1 and 4, is retained after the recombination event that creates the deletion [note that orientation of arrow refers to directionality of loxP and lox 511 sequences]. This particular feature has been utilized to place numerous DNA cassettes, such as: i) mammalian cell-selectable antibiotic resistance genes in the BAC (24), ii) a basal promoter-containing EGFP gene in the BAC to serve as an enhancer-trap to functionally localize potential gene-regulatory elements further upstream (14), and iii) introduce iTol2 ends to generate functionalized BAC DNA ready for integration into zebrafish or mouse chromosomes (15).

The ability to truncate either end of a BAC insert progressively while keeping the other end intact is useful to a variety of mapping experiments. Thus genetic markers and polymorphisms have been mapped on a physical map of the chromosome using this technology (13, 16). Long-range gene regulatory elements have also been mapped functionally using the end-deletion approach by either generating transgenic mice with EGFP functionalized BAC DNA (40), or electroporating patient-derived cells in culture (41). The Nkx2-5 gene containing BAC from a mouse library was functionalized with EGFP reporter gene using the phage λ red-recombination system, and then truncated progressively from the far upstream 5′ end of the Nkx2-5 gene using a loxP transposon to identify transcription-enhancing factor binding sites (40). Monitoring mRNA levels of the gene of interest allows circumventing the use of a reporter gene such as EGFP in the BAC or PAC DNA, and a series of upstream deletions in a PAC clone was used to functionally identify potential far-upstream binding sites for transcription factors in the human Hox11 gene (41).

1.7 Simultaneous insertion of reporter gene & truncation of DNA from BAC ends

The two separate steps of introducing the EGFP reporter gene and truncating far upstream sequences of the gene of interest in the BAC can be done simultaneously using lox-Tn10 transposons (14). A basal promoter-containing EGFP gene was designed to serve as an enhancer-trap, and placed in front of the loxP arrowhead such that the cassette was retained after the end-deletion of the BAC DNA [see upper panel of Figure 5 for enhancer-trap transposon]. Progressive truncations from one end of the genomic insert with this enhancer-trap transposon generated a library of BAC DNA molecules that differ in the position of the enhancer-trap cassette with respect to the start site of transcription of the gene. The collection of well characterized BAC DNA can be introduced individually into zebrafish embryos for expression of the gene in specific tissues (14).

Such enhancer-trap containing BACs can be re-fitted again at the opposite end with DNA cassettes carrying repeat ends of the vertebrate transposon Tol2 as described below.

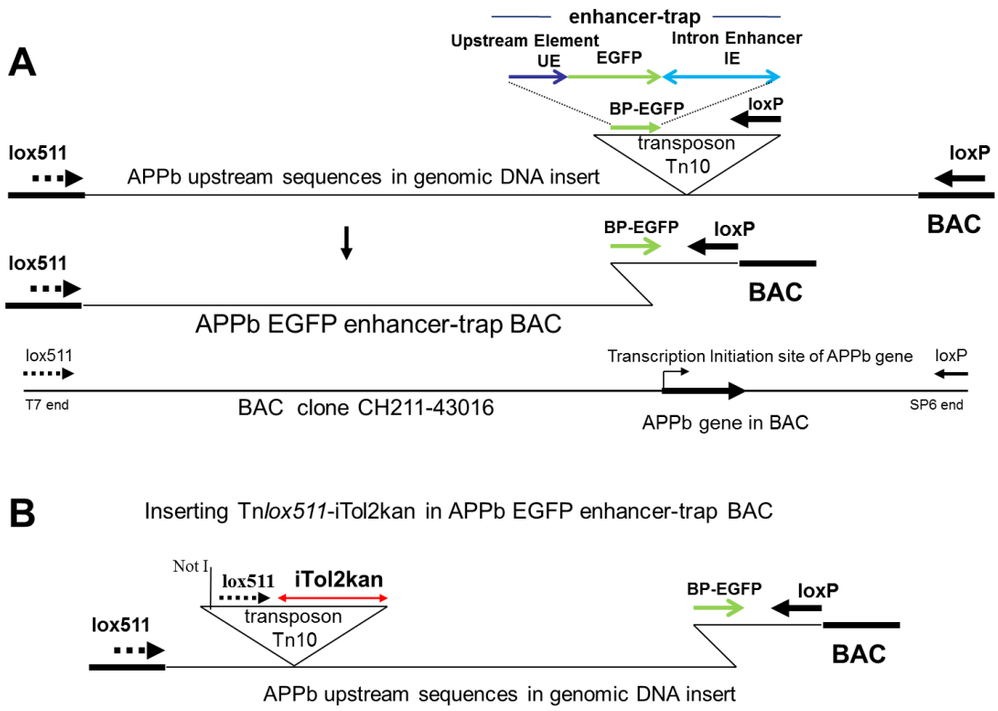


Fig. 5. Legend: Schematic representation of deletion formation in starting BAC APPb EGFP enhancer-trap by pTn*lox511*-iTol2kan: Panel A shows generation of starting BAC APPb EGFP enhancer-trap by insertion of the enhancer-trap transposon (inverted triangle), and the location of APPb gene within the BAC clone. Panel B: Shown schematically is the insertion of the Tn*lox511*-iTol2kan transposon into the enhancer-trap APPb BAC DNA. The thick black arrows represent loxP and lox511 sites (discontinuous arrow). DNA cassettes in front of the loxP or lox511 arrows are retained after the loxP-loxP or lox511-lox511 recombinations by Cre. The enhancer-trap DNA cassette is positioned in front of the loxP arrow (panel A), while the iTol2 cassette is located in front of the lox511 arrow. The colored arrowheads pointing outward in the iTol2 cassette correspond to the 200 bp inverted repeat end R and 150 bp inverted repeat end L of the Tol2 transposon [see Suster et al 2009 ref (42)]. The kanamycin resistance gene is between the Tol2 inverted repeat ends.

1.8 BAC transgenesis in zebrafish and mice using Tol2

BACs have been integrated into the germlines of zebrafish and mice using the vertebrate transposon Tol2 system (42). Tol2 ends, in the inverted orientation and flanking a 1 kb spacer DNA (iTol2), were introduced into the BAC DNA within the bacterial host using recombination of homologous sequences between the targeting vector and the genomic insert. This approach used to introduce the iTol2 cassette can have unintended consequences: other sequence repeats existing in BACs from vertebrate DNA libraries were likely to rearrange as well. This would complicate the introduction of iTol2 cassettes at best, and were likely to present major challenges in BACs spanning chromosomal loci that contained highly repetitive DNA, such as the Npr3 gene locus analyzed earlier (16). Therefore a simpler and more flexible system was developed using our Tn10 transposon approach to deliver iTol2 ends into BACs. These iTol2-end containing BACs are suitable for transgenesis into zebrafish or mouse embryos, and have recently been reported (15).

The iTol2 DNA cassette was placed in front of a loxP or lox511-Tn10 transposon (15). Progressive truncations from an end of the genomic insert with this iTol2-Tn10 transposon generated a library of BAC DNA molecules that differed in the position of the iTol2 cassette with respect to upstream regulatory elements of the Amyloid Precursor Protein (APPb) gene in the BAC. The collection of well characterized BAC DNA was introduced into zebrafish embryos individually for expression of EGFP in neurons (Figure 6).

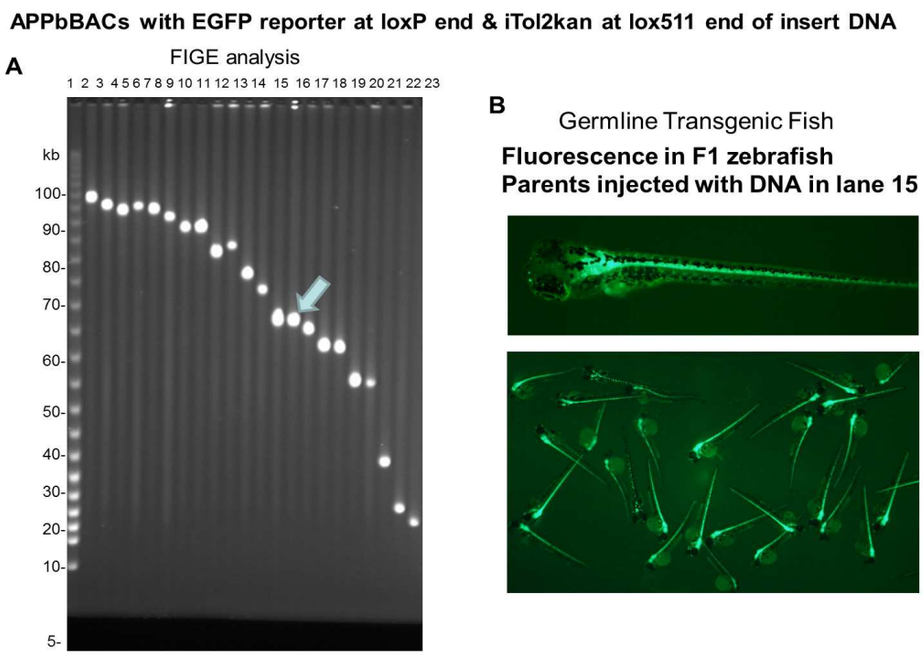


Fig. 6. Legend: Panel A: FIGE analysis of an array of APPb BACs with EGFP containing Enhancer-trap at loxP end and iTol2kan cassette at lox511 end of DNA insert. The BAC DNA was digested with Not I before FIGE. Note that there is no vector DNA band because the lox511-lox511 deletion with Cre removes both the Not I site in front of the BAC lox511 and the Not I site behind the lox511 in the transposon shown in panel B of Figure 5. Panel B: EGFP fluorescence in Germline transgenic F1 zebrafish obtained from injecting eggs (F0 zebrafish) with the BAC DNA shown by the blue arrowhead in lane 15.

2. Exploring gene regulation by distal *cis*-acting sequences using BACs

2.1 Rationale for using BACs to understand gene regulation

BACs offer a large span of DNA that can house most, if not all, the sequences recognized by DNA-binding regulatory proteins that act in concert to regulate expression of the gene. But more importantly, the binding sites of regulatory proteins in the BAC DNA exist in their chromosomal contexts, with sequences flanking *cis*-acting sites that have evolved for millions of years, preserved. This cannot be said of small plasmid constructs used for expressing genes where multiple, easily-recognizable distal *cis*-acting sequences are excised from their surrounding DNA in the chromosome and cobbled together to create artificial junctions.

2.2 Why is this important in regulation of gene expression?

We know that both the rate of a reaction, as well as the equilibrium that a reversible reaction establishes, are related inversely and logarithmically to the activation energy or the free energy difference between the reactants and products, respectively. It implies that small changes in the binding affinities of DNA-binding regulatory proteins, important for regulation of gene expression, can profoundly affect the rates and equilibrium established in the cell. Also, gene regulatory proteins in the cell do not exist as monomeric units, but as components of multiple different protein complexes with different physical dimensions destined for a variety of purposes in the cell. Thus it can be argued that the same protein in a multimeric complex might not be able to play identical gene-regulatory roles if the sequences flanking its binding site are altered compared to when they exist in their chromosomal context. It is conceivable that some of this disparity in DNA-binding affinities resulting from binding to sites with altered flanking sequences could eventually manifest itself in altered tissue specificity of expression of the gene.

There is also the issue of how one goes about selecting potential gene-regulatory elements to stitch together in a small plasmid designed for expressing a gene of interest. While one can make informed guesses one cannot be unbiased, as much of the players involved in the fine-tuning of gene expression remain largely unknown. Using BACs functionalized with suitable reporter genes, and integrated into the chromosome, to explore regulation of genes by *cis*-acting elements appears to circumvent many of these difficulties.

We have developed one such approach using BACs retrofitted with enhancer-traps (14). The overall effect of using this enhancer-trap technology in BACs is somewhat akin to scanning the BAC DNA from one end to the other with a "mine-sweeper" that is capable of unearthing transcription-enhancing factor binding sites. Functional comparisons between individual enhancer-trap BACs are thus more meaningful as the modifications in each remain constant. It is important to note that this approach to identifying gene-regulatory elements is likely to be less biased than a targeted approach, because there is no prior selection of sequences for testing their gene-regulatory potential (14, 15). In situations where regulatory function is conserved without sequence similarity (43), choosing the correct sequence to test might be a hit-or-miss phenomenon with a targeted approach. Thus using enhancer-traps to scan BACs offers an effective and relatively unbiased alternative to other targeted approaches to functionally map *cis*-acting gene regulatory elements. Important transcription enhancing elements have been discovered using this approach in non-coding DNA from the intron, and 30 kb upstream of the APPb gene transcription start site

[reference (14), and Shakes, Du, Sen, Abe, Kawakami, Wolf, Hatcher, Norford and Chatterjee manuscript in preparation].

The BAC enhancer-trap technology described above has three additional features that should prove beneficial: i) allows sampling much larger sequences of DNA, and consequently multiple discontinuous regulatory domains simultaneously compared to small plasmids, ii) the context of regulatory DNA with respect to the gene and chromosome is preserved and, iii) can be used with BACs in established libraries from a wide variety of organisms, and tested in several species. Although the methodology does not allow generation of internal deletions, truncations from the end opposite to the enhancer-trap can be made with a lox511 transposon [as seen when iTol2-end DNA cassettes were introduced at the other end of BAC DNA, (Figures 5 and 6)] to explore functions of candidate regulatory regions in a limited way. Sequences bending DNA (45-47), or phasing nucleosomes and other transcription factors (48-50) are left unaltered using BACs compared to characteristics of the gene region found endogenously. Bringing exogenous pieces of DNA together to create artificial junctions in small plasmids to trans-activate reporter genes does not adequately address the endogenous role of the regulatory sequence; and this is avoided using BACs. It is probably worth noting that DNA structure surrounding regulatory factor binding sites have evolved over long periods, and these are also left unaltered here. We noted earlier that there were 52 sites with six or more consecutive A-residues, known to cause bends in unpackaged DNA (46), in the ~28 kb upstream regulatory sequence identified as important in regulating APPb (14).

2.3 Substitution of Lox sites flanking BAC inserts with Lox66

In a different application Tn10 transposons carrying 'arm' mutants of loxP, such as lox66, have been used to successfully substitute the loxP site at one end of the genomic insert DNA in a BAC (39). It should be possible also to substitute the lox511 site at the other end of insert DNA in a similar manner using a lox site with an identical 'spacer' as lox511 but carrying a different 'arm' sequence. The usefulness of such substitutions of loxP or lox511 sites in BACs with lox66 is realized from the fact that the integration of loxP plasmids to a loxP site on the chromosome is reversible, with the equilibrium favored for excision. Targeting lox66 substituted BACs specifically to pre-positioned lox71 sites in chromosomes via lox-Cre recombination should be irreversible. Such strategies are likely to be particularly useful in systems such as the zebrafish, where classical "knock-in" technology cannot be used because of genome duplication in an ancestral teleost [see reference (39) for discussion].

2.4 Pros and cons of the transposon retrofitting approach

The insertions of exogenous DNA cassettes into BAC DNA are performed using Tn10 transposons carrying loxP or lox511 sites, and do not rely upon sequences existing in the genomic inserts of BACs. Consequently, the Tn10 transposons developed for a particular objective are applicable to all BACs in the public domain. This feature is unlike the targeting vectors used in homologous recombination dependent strategies, where the targeting vector plasmids need to be constructed anew for each BAC.

Because the insertions of these Tn10 transposons into genomic DNA in BACs appear to be largely random [see references (16) for discussion], large collections of BACs with ends

progressively trimmed are generated in a single experiment. Such libraries, when made with Tn10 transposons carrying enhancer-traps or iTol2 cassettes, are uniquely suited for functionally mapping long-range gene regulatory sequences using transgenic animals. Thus multiple BACs from a contig spanning a genetic locus should enable such functional analyses to be extended over large sections of the genome. The approach does not require selecting sequences for mutational analysis to test their gene regulatory potential, thus it is an unbiased approach. It enables enhancer-trap containing BACs, or BACs retrofitted with EGFP cassettes by sequence homology based recombination, to be converted into deletion libraries with integrated iTol2 ready for chromosome integration. Thus compared to traditional approaches for enhancer-trapping used with whole genomes in animals (51-59), our approach using individual BACs, has the potential to allow a more uniform coverage of the genome because the baseline efficiency of trap insertion is reset for individual BACs in the bacterial host. Although there appear to be vast regions of the genome refractory to enhancer-trapping by traditional means, to date we have not encountered BACs refractory to Tn10 transposon insertions.

The absence of undesirable rearrangement of insert DNA due to the high content of sequence repeats in certain BACs is an additional advantage with our approach. We note that vertebrate DNA in general and mammalian DNA in particular, has large numbers of DNA sequences that are repeated (16). Use of radioactive isotopes is also avoided because no Southern blotting is required in retrofitting BACs with transposons.

A notable drawback of the transposon based approach appears to stem from the very feature that makes it readily doable: the P1 headful packaging strategy used to isolate the functionalized BAC so easily also limits the size of BAC DNA that can be analyzed to ~ 110 kb. Although the genomic DNA insert is truncated in the process, the resulting 103 kb insert DNA size, the remainder being modified BAC vector, is unlikely to be a disadvantage in most applications; because a majority of vertebrate genes can be housed in their entirety within this size limit. Almost half of evolutionarily conserved non-coding gene-regulatory sequences in vertebrate genomes (60), and probably a similar fraction of those that are conserved in function and shape but not in sequence (43, 44), are located within this span of DNA adjoining start sites of transcription of genes.

Lastly, both the lox sites flanking insert DNA in BACs can be readily substituted with lox sites with identical spacer but different arms using Tn10 transposons (39). Such replacements would make the ~110 kb retrofitted BACs amenable to “knock-in” strategies of a slightly different type: BACs with lox66 substituted for loxP should be targetable to lox71 sites in chromosomes of organisms such as zebrafish, a system where conventional homologous recombination-mediated “knock-in” technology is unavailable due to genome duplication in an ancestral teleost during evolution (61).

Clearly the transposon retrofitting strategy and those based on homologous recombination have strengths that appear somewhat complimentary in nature. Thus a judicious approach might be to use a combination of the two methodologies, as demonstrated in earlier studies (40, 15). Introducing the Tol2 inverted repeat ends to the ends of genomic DNA insert in a BAC that was previously functionalized with a reporter gene appears straightforward and easier to do with the Tn10 transposon approach. The libraries of iTol2 inserted BACs generated should also facilitate integration of trimmed single genes into the germline and

help to functionally map *cis*-acting gene regulatory sequences in animals. The approach should be applicable to a wider variety of BACs, including those with sequence repeats.

3. List of abbreviations

BAC- bacterial artificial chromosome/ **FIGE**- Field inversion gel electrophoresis**/ APPb**-amyloid precursor protein gene b/ **EGFP**-enhanced green fluorescent protein

4. Acknowledgements

The project described was supported by Award Number P20MD000175 from the National Center on Minority Health and Health Disparities (NCMHD) and funds from the North Carolina Biotechnology Center. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NCMHD or the National Institutes of Health. We thank Shanta Mackinnon and Charles Hatcher for zebrafish eggs, and Ms. Rosalind Grays, Connie Keys, Crystal McMichael and Darlene Laws for support and encouragement. PKC would like to thank Drs. Ken Harewood and Sean Kimbro for encouragement and support.

5. References

[1] Yang XW, Model P, Heintz N. (1997). Homologous recombination based modification in Escherichia coli and germline transmission in transgenic mice of a bacterial artificial chromosome. *Nat Biotechnol* 9: 859-865

[2] Zhang Y, Buchholz F, Muyrers JP, Stewart AF. (1998). A new logic for DNA engineering using recombination in Escherichia coli. *Nat Genet*. 20: 123-128.

[3] Jessen JR, Meng A, McFarlane RJ, Paw BH, Zon LI, Smith GR, Lin S. (1998). Modification of bacterial artificial chromosomes through chi-stimulated homologous recombination and its application in zebrafish transgenesis. *Proc Natl Acad Sci U S A*. 95: 5121-5126.

[4] Muyrers JP, Zhang Y, Testa G, and Stewart AF (1999). Rapid modification of Bacterial Artificial Chromosomes by ET recombination. *Nucleic Acids Res 27:* 1555-1557.

[5] Gong S, Yang XW, Li C, Heintz N. (2002) Highly efficient modification of bacterial artificial chromosomes (BACs) using novel shuttle vectors containing the R6Kgamma origin of replication. *Genome Res.* 12: 1992-1998.

[6] Warming S, Costantino N, Court DL, Jenkins NA, Copeland NG (2005) Simple and highly efficient BAC recombineering using galK selection. *Nucleic Acids Res. 33: e36.*

[7] Yang Z, Jiang H, Chachainasakul T, Gong S, Yang XW, Heintz N, Lin S. (2006). Modified bacterial artificial chromosomes for zebrafish transgenesis. *Methods.* 39:183-188.

[8] Mortlock DP, Guenther C, Kingsley DM. (2003) A general approach for identifying distant regulatory elements applied to the Gdf6 gene. *Genome Res.* 9:2069-2081.

[9] Jessen, J. R., Willett, C. E., and Lin, S. (1999) Artificial chromosome transgenesis reveals long-distance negative regulation of rag1 in zebrafish. *Nat Genet*. 23: 15-16.

[10] Carvajal, J. J., Cox, D., Summerbell, D., and Rigby, P. W. (2001) A BAC transgenic analysis of the Mrf4/Myf5 locus reveals interdigitated elements that control activation and maintenance of gene expression during muscle development. *Development*. 128: 1857-1868.

[11] Orford M, Nefedov M, Vadolas J, Zaibak F, Williamson R, Ioannou PA. (2000) Engineering EGFP reporter constructs into a 200 kb human beta-globin BAC clone using GET Recombination. *Nucleic Acids Res* 18: E84

[12] Chatterjee, P.K., Coren, J.C. (1997). Isolating large nested deletions in bacterial and P1 artificial chromosomes by *in vivo* P1 packaging of products of Cre-catalyzed recombination between the endogenous and a transposed *lox*P site. *Nuc. Acids Res.* 25: 2205-2212.

[13] Chatterjee, P.K., Yarnall, D.P., Haneline, S.A., Godlevski, M.M., Thornber , S.J., Robinson, P.S., Davies, H.E., White, N.J., Riley, J.H. and Shepherd, N.S. (1999). Direct Sequencing of Bacterial and P1 Artificial Chromosome Nested-deletions for Identifying Position-Specific Single Nucleotide Polymorphisms. *Proc. Natl. Acad. Sci. (USA)* 96: 13276-13281.

[14] Shakes, L.A., Malcolm, T.L., Allen, K.L., De, S., Harewood, K.R., & Chatterjee, P.K. (2008). Context dependent function of APPb Enhancer identified using Enhancer Trap-containing BACs as Transgenes in Zebrafish. *Nucleic Acids Research* 36: 6237-6248.

[15] Shakes LA, Abe G, Eltayeb MA, Wolf HM, Kawakami K and Chatterjee PK (2011). Generating libraries of iTol2-end insertions at BAC ends using loxP and lox511 Tn10 transposons. *BMC Genomics* 12: 351

[16] Gilmore, R.C., Baker Jr., J., Dempsey, S., Marchan, R., Corprew Jr., R.N.L., Maeda, N., Smithies, O., Byrd, G., Bukoski, R.D., Harewood, K.R. & Chatterjee, P.K. (2001). Using PAC nested-deletions to order contigs and microsatellite markers at the high repetitive sequence containing Npr3 gene locus. *Gene* 275: 65-72.

[17] Craig, NL. (1997). Target site selection in transposition. *Annu Rev Biochem* 66: 437-474.

[18] Chatterjee, PK, and Briley, LP (2000) Analysis of a Clonal Selection Event during Transposon-Mediated Nested-Deletion Formation in Rare BAC and PAC Clones. *Analytical Biochemistry*, 285: 121-126.

[19] Chatterjee, P.K. (2004). Retrofitting BACs and PACs with LoxP Transposons to Generate Nested Deletions. "Bacterial Artificial Chromosomes" vol 1, pp 231-241. *Methods in Mol. Biology* series vol 255, The Humana Press Inc. Editors: Shaying Zhao & Marvin Stodolsky.

[20] Yarmolinsky, M.B. & Sternberg N. (1988). Bacteriophage P1 in *The Bacteriophages,* Vol 1 edited by Richard Calendar, *Plenum Publishing Corporation.*

[21] Chatterjee, P. K., Shakes, L. A., Srivastava, D.K., Garland, D.M., Harewood, K.R., Moore, K.J., and Coren, J.S. (2004). Mutually Exclusive Recombination of Wild Type and Mutant *lox*P Sites *in vivo* Facilitates Transposon-Mediated Deletions from Both Ends of Genomic DNA in PACs. *Nucleic Acids Research* 32: 5668-5676.

[22] Chatterjee, P.K. and Sternberg, N.L. (1996). Retrofitting high molecular weight DNA cloned in P1: introduction of reporter genes, markers selectable in mammalian cells and generation of nested deletions. *Genet Anal Biomol. Eng.* 13: 33-42.

[23] Chatterjee, P.K., Mukherjee, S., Shakes, L.A., Wilson, III, W., Harewood, K.R. & Byrd, G. (2004). Selecting Transpositions of a Markerless Transposon Using Phage P1 Headful Packaging: New Transposons for Functionally Mapping Long Range Regulatory Sequences in BACs. *Analytical Biochemistry* 335: 305-315.

[24] Shakes, L. A., Garland, D.M., Srivastava, D.K., Harewood, K.R. and Chatterjee, P. K. (2005). Minimal Cross-recombination between wild type and loxP511 sites *in vivo*

facilitates Truncating Both Ends of Large DNA Inserts in pBACe3.6 and Related Vectors. *Nucleic Acids Research.* 33: *e118*.

[25] Chatterjee, P.K. and Baker Jr. J. C. (2004). Preparing Nested Deletions Template DNA for Field Inversion Gel Electrophoresis Analyses and Position-Specific End Sequencing With Transposon Primers. "Bacterial Artificial Chromosomes" vol 1, pp 243-254. *Methods in Mol. Biology* series vol 255, The Humana Press Inc. Editors: Shaying Zhao & Marvin Stodolsky.

[26] Frengen E, Weichenhan D, Zhao B, Osoegawa K, van Geel M, de Jong PJ (1999). A modular, positive selection bacterial artificial chromosome vector with multiple cloning sites. *Genomics* 58: 250-253.

[27] Osoegawa K., Woon P.Y., Zhao B., Frengen E., Tateno M., Catanese J.J. de Jong P.J. (1998). An improved approach for construction of bacterial artificial chromosome libraries *Genomics* 52:1-8.

[28] Osoegawa K, Tateno M, Woon PY, Frengen E, Mammoser AG, Catanese JJ, Hayashizaki Y, de Jong PJ. (2000). Bacterial artificial chromosome libraries for mouse sequencing and functional analysis. *Genome Res.* 10: 116-128.

[29] Osoegawa K, Mammoser AG, Wu C, Frengen E, Zeng C, Catanese JJ, de Jong PJ. (2001). A bacterial artificial chromosome library for sequencing the complete human genome. *Genome Res.* 11: 483-496.

[30] Osoegawa K, Zhu B, Shu CL, Ren T, Cao Q, Vessere GM, Lutz MM, Jensen-Seaman MI, Zhao S, de Jong PJ (2004). BAC resources for the rat genome project. *Genome Res.* 4:780-785.

[31] Krzywinski M, Bosdet I, Smailus D, Chiu R, Mathewson C, Wye N, Barber S, Brown-John M, Chan S, Chand S, Cloutier A, Girn N, Lee D, Masson A, Mayo M, Olson T, Pandoh P, Prabhu AL, Schoenmakers E, Tsai M, Albertson D, Lam W, Choy CO, Osoegawa K, Zhao S, de Jong PJ, Schein J, Jones S, Marra MA. (2004). A set of BAC clones spanning the human genome. *Nucleic Acids Res.* 32: 3651-3660.

[32] Hoess RH, Wierzbicki A, Abremski K (1986). The role of the *loxP* spacer region in P1 site-specific recombination. *Nucleic Acids Res.* 14: 2287–2300.

[33] Lee G. and Saito,I., (1998) Role of nucleotide sequences of *loxP* spacer region in Cre-mediated recombination. *Gene* 216: 55–65

[34] Langer SJ, Ghafoori AP, Byrd M, Leinwand L.(2002) A genetic screen identifies novel non-compatible loxP sites. *Nucleic Acids Res. 30: 3067-3077.*

[35] Missirlis PI, Smailus DE, Holt RA (2006). A high-throughput screen identifying sequence and promiscuity characteristics of the *lox*P spacer region in Cre-mediated recombination. *BMC Genomics.* 7: 73–85.

[36] Sheren J, Langer SJ, Leinwand LA (2007) A randomized library approach to identifying functional lox site domains for the Cre recombinase. *Nucleic Acids Res.* 35: 5464-5473.

[37] Siegel R.W., Jain,R. and Bradbury,A. (2001) Using an *in vivo* phagemid system to identify non-compatible *loxP* sequences. *FEBS lett.* 505: 467–473.

[38] Wang Z., Engler,P., Longacre,A. and Storb,U. (2001) An efficient method for high-fidelity BAC/PAC retrofitting with a selectable marker for mammalian cell transfection. *Genome Res.* 11: 137–142.

[39] Chatterjee, PK, Shakes, LA, Stennett, N, Richardson, VL, Malcolm, TL & Harewood, KR. (2010). Replacing the wild type *loxP* site in BACs from the public domain with *lox66* using a *lox66* transposon. *BMC Res Notes 3: 38*

[40] Chi, X., Chatterjee, P.K., Wilson III, W., Zhang, S-X., DeMayo, F., and Schwartz, R.J. (2005). Complex Cardiac Nkx2-5 Gene Expression Activated By Noggin Sensitive Enhancers Followed By Chamber Specific Modules. *Proc. Natl. Acad. Sci. (USA)* 102: 13490-13495.

[41] Brake, R.L., Chatterjee, P.K., Kees, U.R & Watt, P.M. (2004). The functional mapping of long-range transcription control elements of the *HOX11* proto-oncogene. *Biochem. Biophys. Res. Com.* 313: 327-335.

[42] Suster ML, Sumiyama K, Kawakami K (2009) Transposon-mediated BAC transgenesis in zebrafish and mice. *BMC Genomics* 10:477

[43] Fisher S, Grice EA, Vinton RM, Bessling SL, McCallion AS (2006) Conservation of RET regulatory function from human to zebrafish without sequence similarity *Science 14: 276-279.*

[44] Parker SC, Hansen L, Abaan HO, Tullius TD, Margulies EH. (2009) Local DNA topography correlates with functional noncoding regions of the human genome. Science. Apr 17; 324 (5925): 389-92. Epub 2009 Mar 12.

[45] Fried MG and Crothers DM (1983) CAP and RNA polymerase interactions with the lac promoter: binding stoichiometry and long range effects. *Nucleic Acids Res.* 11: 141–158

[46] Ulanovsky L, Bodner M, Trifonov E.N., Choder M.(1986) Curved DNA: design, synthesis, and circularization. *Proc Natl Acad Sci U S A*. 83: 862-866

[47] Gartenberg M.R., Ampe C, Steitz T.A., Crothers D.M. (1990) Molecular characterization of the GCN4-DNA complex *Proc Natl Acad Sci U S A*. 87: 6034-6038

[48] Hebbar P.B., Archer T.K. (2007) Chromatin-dependent cooperativity between site-specific transcription factors *in vivo*. *J Biol Chem* 282: 8284-8291.

[49] Ganapathi M, Singh G.P., Sandhu K.S., Brahmachari S.K, Brahmachari V. (2007) A whole genome analysis of 5' regulatory regions of human genes for putative cis-acting modulators of nucleosome positioning. *Gene.* 391: 242-251.

[50] S Hardy and T Shenk (1989) E2F from adenovirus-infected cells binds cooperatively to DNA containing two properly oriented and spaced recognition sites. *Mol Cell Biol* 9: 4495–4506.

[51] O'Kane CJ, Gehring WJ. (1987). Detection in situ of genomic regulatory elements in Drosophila. *Proc Natl Acad Sci U S A.* 84: 9123-9127.

[52] Wilson C, Pearson RK, Bellen HJ, O'Kane CJ, Grossniklaus U, Gehring WJ. (1989). P-element-mediated enhancer detection: an efficient method for isolating and characterizing developmentally regulated genes in Drosophila. *Genes Dev.* 3: 1301-1313.

[53] Korn R, Schoor M, Neuhaus H, Henseling U, Soininen R, Zachgo J, Gossler A. (1992). Enhancer trap integrations in mouse embryonic stem cells give rise to staining patterns in chimaeric embryos with a high frequency and detect endogenous genes. *Mech Dev*. 39: 95-109.

[54] Grabher C, Henrich T, Sasado T, Arenz A, Wittbrodt J, Furutani-Seiki M. (2003). Transposon-mediated enhancer trapping in medaka. *Gene.* 322: 57-66.

[55] Balciunas D, Davidson AE, Sivasubbu S, Hermanson SB, Welle Z, Ekker SC. (2004). Enhancer trapping in zebrafish using the Sleeping Beauty transposon. *BMC Genomics*. 5: 62.

[56] Kawakami, K., Takeda, H., Kawakami, N., Kobayashi, M., Matsuda, N., and Mishina, M. (2004) A transposon-mediated gene trap approach identifies developmentally regulated genes in zebrafish. *Developmental Cell* 7, 133-144

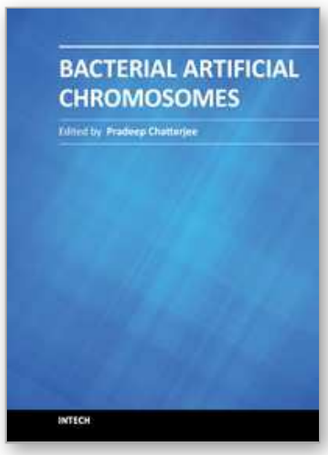
[57] Kawakami, K. (2005) Transposon tools and methods in zebrafish. *Developmental Dynamics* 234, 244-254

[58] Ellingsen S, Laplante MA, Konig M, Kikuta H, Furmanek T, Hoivik EA, Becker TS (2005) Large-scale enhancer detection in the zebrafish genome *Development.* 132: 3799-3811.

[59] Nagayoshi, S., Hayashi, E., Abe, G., Osato,N., Asakawa, K., Urasaki, A., Horikawa, K., Ikeo, K., Takeda, H., and Kawakami, K. (2008) Insertional mutagenesis by the Tol2 transposon-mediated enhancer trap approach generated mutations in two developmental genes: tcf7 and synembryn-like. *Development* 135(1), 159-169

[60] Woolfe A, Goodson M, Goode DK, Snell P, McEwen GK, Vavouri T, Smith SF, North P, Callaway H, Kelly K, Walter K, Abnizova I, Gilks W, Edwards YJ, Cooke JE, Elgar G. (2005) Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol.* (2005) Jan; 3: e7.

[61] Kleinjan DA, Bancewicz RM, Gautier P, Dahm R, Schonthaler HB, Damante G, Seawright (2008). A, Hever AM, Yeyati PL, van Heyningen V, Coutinho P. Subfunctionalization of duplicated zebrafish pax6 genes by cis-regulatory divergence. *PLoS Genet.* 4: e29



Bacterial Artificial Chromosomes

Edited by Dr Pradeep Chatterjee

ISBN 978-953-307-725-3
Hard cover, 148 pages
Publisher InTech
Published online 25, November, 2011
Published in print edition November, 2011

This book focuses on the numerous applications of Bacterial Artificial Chromosomes (BACs) in a variety of studies. The topics reviewed range from using BAC libraries as resources for marsupial and monotreme gene mapping and comparative genomic studies, to using BACs as vehicles for maintaining the large infectious DNA genomes of viruses. The large size of the insert DNA in BACs and the ease of engineering mutations in that DNA within the bacterial host, allowed manipulating the BAC-viral DNA of Varicella-Zoster Virus. Other reviews include the maintenance and suitable expression of foreign genes from a Baculovirus genome, including protein complexes, from the BAC-viral DNA and generating vaccines from BAC-viral DNA genomes of Marek's disease virus. Production of multi-purpose BAC clones in the novel Bacillus subtilis host is described, along with chapters that illustrate the use of BAC transgenic animals to address important issues of gene regulation in vertebrates, such as functionally identifying novel cis-acting distal gene regulatory sequences.

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen