Bioengineered 4:4, 258–261; July/August 2013; © 2013 Landes Bioscience

Recent advances in recombinant protein production

BAC-based expression vectors, the bigger the better

Renate Kunert[1] and Emilio Casanova[2,*]

[1]Department of Biotechnology; Vienna Institute of Biotechnology; University of Natural Resources and Life Sciences, Vienna; Vienna, Austria; [2]Ludwig Boltzmann Institute for Cancer Research (LBI-CR); Vienna, Austria

Designing appropriate expression vectors is one of the critical steps in the generation of stable cell lines for recombinant protein production. Conventional expression vectors are severely affected by the chromatin environment surrounding their integration site into the host genome, resulting in low expression levels and transgene silencing. In the past, a new generation of expression vectors and different strategies was developed to overcome the chromatin effects. Bacterial artificial chromosomes (BACs) are cloning vectors capable of accommodating up to 350 Kb. Thus, BACs can carry a whole eukaryotic locus with all the elements controlling the expression of a gene; therefore, BACs harbor their own chromatin environment. Expression vectors based on BACs containing open or permissive chromatin loci are not affected by the chromatin surrounding their integration site in the host cell genome. Consequently, BAC-based expression vectors containing the appropriate loci confer predictable and high levels of expression over time. These properties make BAC-based expression vectors a very attractive tool applied to the recombinant protein production field.

Keywords: bacterial artificial chromosome, chromatin, CHO cells, mammalian, expression vector, producer cell line, transgene, recombinant protein production

Submitted: 01/31/13

Revised: 02/19/13

Accepted: 02/20/13

http://dx.doi.org/10.4161/bioe.24060

*Correspondence to: Emilio Casanova; Email: emilio.casanova@lbicr.lbg.ac.at

Addendum to: Mader A, Prewein B, Zboray K, Casanova E, Kunert R. Exploration of BAC versus plasmid expression vectors in recombinant CHO cells. Appl Microbiol Biotechnol 2012; 97:4049-54; 23081777; http://dx.doi.org/ 10.1007/s00253-012-4498-x

Commentary

For some time, recombinant protein production in the mammalian system has been a central topic in biotechnology.[1] Indeed, more than a hundred therapeutic proteins are produced nowadays in mammalian systems, and it is expected that this number will increase rapidly due to the development of new therapeutic antibodies.[2] Therefore, an enormous effort to improve the generation of mammalian cell lines expressing recombinant proteins has been carried on during the past decades. As with many biological processes, the generation of producer cell lines is hampered by several bottle-necks (transcription, translation, protein folding, protein secretion, etc.). Indeed, one of the most important issues is the choice of the expression vector to generate the producer cell line.[1] In principle, an ideal expression vector should fulfill the following criteria: (1) expression should be independent of the integration site into the host genome, (2) expression should not be silenced over time, (3) expression should be directly proportional to the number of integrated copies into the host genome and (4) expression levels should be high. The most conservative method to generate stable cell lines comprises the transfection of a plasmid containing a promoter which drives the expression of a gene of interest and an antibiotic marker for selection (**Fig. 1A**). Generation of stable cell lines using these kinds of vectors results in unpredictable and generally low expression levels. Furthermore, expression from such vectors tends to be silenced over time. These disadvantages are due to the so called "positional chromatin effects."[3] Transgene expression using such plasmid-based vectors is highly affected by the surrounding chromatin of their integration site into the host genome. Thus, if the vector integrates in "silent chromatin" regions, its expression will be low or

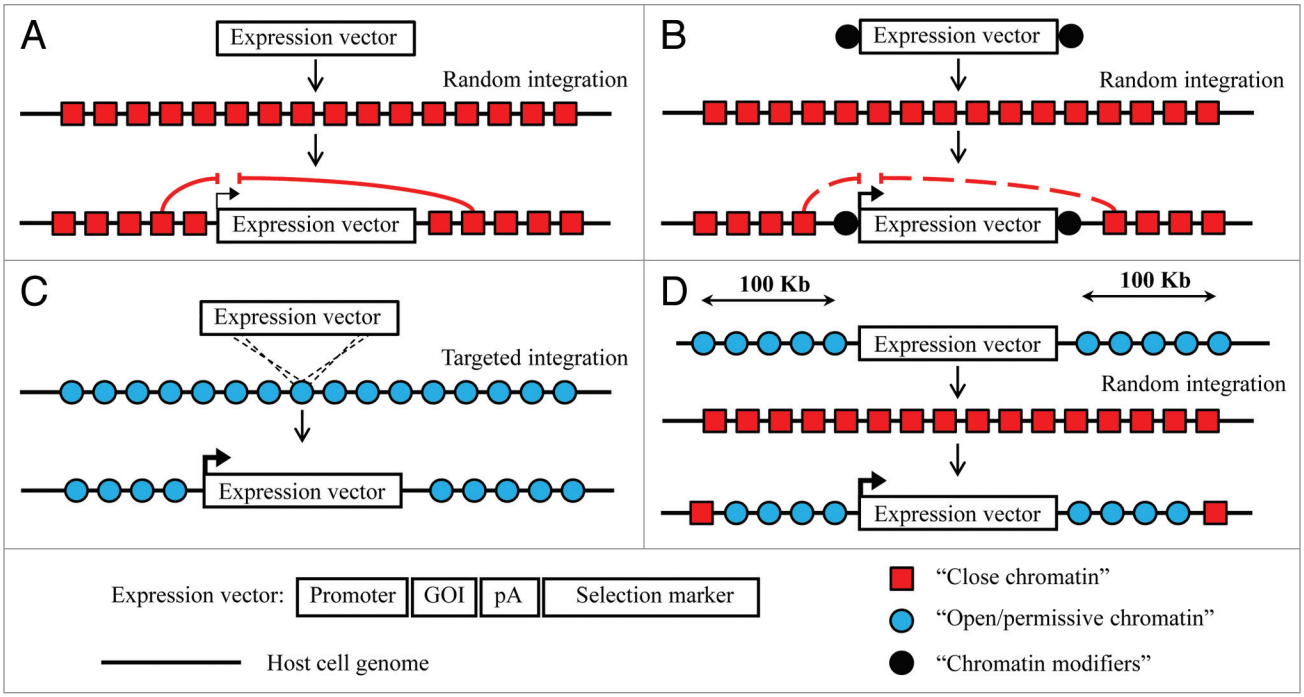

Figure 1. Strategies to generate stable cell lines. (**A**) Random integration of a plasmid-based vector: a conventional expression vector containing a promoter, a gene of interest (GOI), a polyadenylation signal (pA) and a selection marker (e.g., neomycin) transfected into cells integrates randomly into the host cell genome. These vectors are highly affected by the surrounding chromatin of their integration site, often resulting in low or no expression and silencing of the transgene over time. (**B**) Random integration of a plasmid-based expression vector flanked by "chromatin modifiers" (CM). The chromatin modifiers shield the expression vectors from the effects of the chromatin surrounding their integration site into the cell host genome. This results in better expression and stability of the transgene compared with (**A**). (**C**) Targeted integration of an expression vector into a chromatin permissive region (hot spot). By means of recombinase-mediated cassette exchange or somatic homologous recombination, an expression vector is targeted (integrated) to a hot spot known to be a permissive chromatin region. This results in predictable and stable expression of the transgene, but it is a more laborious method than described in (**A**) or (**B**). (**D**) Random integration of a BAC containing an expression vector. An expression vector is inserted into a 200 kb BAC containing an open chromatin locus (e.g., *Rosa26* locus), transfected into the cells and randomly integrated into the host cell genome. The BAC-based expression vector carries on its own chromatin environment and it is not affected by the surrounding chromatin of its integration site. Several copies of the BAC-based vector can be co-integrated, thus resulting in high and stable expression levels of the transgene.

suppressed at some time. This translates in the necessity of analyzing large numbers of cell clones to find an optimal producer. Furthermore, several rounds of genomic amplification of the transgene using selection markers such us dihydrofolate reductase (*Dhfr*) or glutamine synthetase (*Gs*) are needed to reach high levels of expression.[4] This makes the generation of stable cell clones a tedious and time-consuming process. During the past years, researchers have developed several strategies to avoid the positional chromatin effects and improve the efficacy of the expression vectors. One of the most popular approaches consists of flanking the expression vectors with cis-elements that modify the chromatin environment: "chromatin modifiers" (CM) (**Fig. 1B**). The rationale behind this approach is that expression vectors containing CM are less affected by the detrimental effects of the chromatin surrounding the integration site into the host genome. Thus, by placing CM in the expression vectors, less numbers of clones need to be analyzed. In this regard, several CM have been described and successfully used for the establishment of recombinant cell lines, e.g., insulators,[5] matrix attachment regions,[6] chromatin opening elements[7] or antirepressor elements.[8] A second strategy consists of the identification of a genomic region on the host cell genome that allows reproducible and high expression levels of the transgene, the so-called "hot spots" or "chromatin permissive regions." Once an optimal hot spot has been identified, the expression vector can be knocked-in into such a pre-defined genomic region (**Fig. 1C**). This is usually achieved using "site-specific recombinase mediated cassette exchange"[9,10] or by stimulating homologous recombination into pre-predefined genomic regions using zinc-finger nucleases (ZFNs)[11] or transcription activator-like effector nucleases (TALENs).[12] The advantage of this method is that the expression levels of the transgenes are predictable. However, since this method introduces one single copy of the expression vector into the host genome, the expression levels may be limited and genomic amplification may be needed to obtain higher/sufficient expression. A third strategy is to use expression vectors with large cloning capacity able to accommodate a whole mammalian genetic locus. Such vectors will carry a pre-defined open chromatin locus and its own chromatin environment. Therefore, these vectors will not be affected at all by the surrounding chromatin of their integration site (**Fig. 1D**).[3] The most popular vectors with large cloning capacity are yeast artificial chromosomes (YACs) and

bacterial artificial chromosomes (BACs). YACs can accommodate up to 2 Mb, thus making them interesting candidates for expression vectors.[3] However, manipulation of YACs is very labor-intensive and time consuming; therefore, we do not consider YACs qualified for routine expression vectors for recombinant protein production. BACs, circular plasmids maintained in *Escherichia coli,* are derived from the F-factor and are able to accommodate up to 350 kb. BACs are easy to manipulate, and yields of purified BAC DNA are reasonable and can be transfected into mammalian cells using conventional methods, although due to their large size, transfection efficiencies are lower compared with smaller plasmids.[13] Indeed, BACs can be handled almost like "conventional plasmids"; however, due to their large size, modification of the BACs cannot be done using standard cloning procedures (digestion with restriction enzymes, ligation, etc.) but using homologous recombination in *E. coli* (recombineering), a very simple and well-established method.[14-16] Furthermore, there are large BAC genomic DNA libraries annotated to the human and mouse genome available that greatly alleviate identification and obtainment or retrieval of a BAC containing a locus of interest.[17]

As mentioned above, the major advantage of using BACs as expression vectors for recombinant protein production is their large cloning capacity. BACs can contain an entire locus, including most (if not all) of the elements that regulate gene expression (promoters, enhancers, silencers, insulators, etc.).[18] Thus, BAC-based vectors carry on its own chromatin environment and they are not affected by the surrounding chromatin upon random integration into the host cell genome (**Fig. 1D**). Therefore, BAC-based vectors confer copy-number dependent and chromatin position independent expression, making BACs very attractive candidates for expression vectors. Indeed, due to their characteristics, BAC-based expression vectors have been vastly used during the past 15 years in the transgenic mouse field;[19] surprisingly, only a few examples have been published using BACs as expression vectors applied to recombinant protein production in mammalian cells.[20-23]

A critical issue of using BACs as expression vectors is the carried locus. Obviously, in order to exploit the full benefits of BACs as expression vectors they must carry an open chromatin locus. In this sense, BAC-based vectors can be considered as the genetic equivalent to a knock-in of an expression vector in a genomic hot spot (**Fig. 1C**). The advantage is that the BAC-based vectors carry the hot spot themselves, thus making BAC-based vectors more flexible than the knock-in strategy. Furthermore, the expression levels of BACs are not limited to one integrated copy as it is in the case of the knock-in strategy. We have observed cell lines with at least 50 BAC copies integrated, thus initial expression levels are already high, avoiding the requirement for transgene amplification.

In our initial studies, we used a 200 kb BAC containing the *Rosa26* (*Thumpd3*) locus.[24] The *Rosa26* locus is a region known to be open chromatin and widely used to express genes in transgenic mice.[25] By placing an expression cassette (promoter, gene of interest, polyadenylation signal and selection marker) into the *Rosa26* BAC we were able to obtain HEK293 stable pools with a 10× increase in productivity compared with the plasmid-based vectors. Furthermore, we observed that protein expression levels were directly proportional to the integrated BAC-copy number and that protein expression was stable for at least 30 passages.[21] Recently, using two anti-HIV1 single-chain fragment crystallizable (scFc) antibody derivates of human IgGs as model proteins, we have shown that the *Rosa26* BAC-based vector improves recombinant protein production in DUKX-B11 CHO cells and, importantly, the expression levels are maintained over time.[23] These results illustrate the robustness of the BAC vector system and its flexibility regarding the host cell line. Furthermore, our results have been elegantly confirmed by other researchers. A 175 kb BAC-based expression vector containing the dihydrofolate reductase (*Dhfr*) resulted in copy number-dependent and stable expression in NIH 3T3 cells.[20] Thus, other BAC-based expression vectors containing other loci than the *Rosa26* can also be successfully used to establish cell lines.

In conclusion, BAC-based vectors are not affected by the chromatin surrounding their genomic integration site, their expression levels positively correlate with the integrated BAC copy numbers and they confer high and stable expression. Thus, BAC-based vectors satisfy the conditions of an ideal expression vector for recombinant protein production in mammalian cells. Future developments may include the use of isogenic BACs to the host cell line. Indeed, the *Rosa26* locus used in our BAC vectors is from murine origin and was used to generate cell lines in HEK293 and CHO cells from human and hamster origin, respectively. Although we have obtained good results using a non-isogenic BAC, it seems more reasonable to use isogenic BACs to the host cells, e.g., BACs containing genes from CHO origin in CHO cells. The choice of the BAC-based expression vector containing the *Rosa26* locus was the consequence of an educated guess. The BAC-based expression vector containing the *Rosa26* locus is not necessarily the best option to generate CHO cell lines. Therefore, a very attractive approach will be to screen CHO cells to find genomic hot spots, retrieve the BACs encompassing such genomic locations and use such BACs as expression vectors. The establishment of BAC libraries from CHO origin and the availability of the CHO genome should facilitate this approach in the future.[26,27]

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

We thank Dr Dagmar Stoiber-Sakaguchi for correcting this manuscript. The work of E.C. is supported by the Austrian Research Promotion Agency (FFG, www.ffg.at), grant 829611.

References

1. Wurm FM. Production of recombinant protein therapeutics in cultivated mammalian cells. Nat Biotechnol 2004; 22:1393-8; PMID:15529164; http://dx.doi.org/10.1038/nbt1026.
2. Rita Costa A, Elisa Rodrigues M, Henriques M, Azeredo J, Oliveira R. Guidelines to cell engineering for monoclonal antibody production. Eur J Pharm Biopharm 2010; 74:127-38; PMID:19853660; http://dx.doi.org/10.1016/j.ejpb.2009.10.002.
3. Giraldo P, Montoliu L. Size matters: use of YACs, BACs and PACs in transgenic animals. Transgenic Res 2001; 10:83-103; PMID:11305364; http://dx.doi.org/10.1023/A:1008918913249.
4. Kim JY, Kim YG, Lee GM. CHO cells in biotechnology for production of recombinant proteins: current state and further potential. Appl Microbiol Biotechnol 2012; 93:917-30; PMID:22159888; http://dx.doi.org/10.1007/s00253-011-3758-5.
5. Pikaart MJ, Recillas-Targa F, Felsenfeld G. Loss of transcriptional activity of a transgene is accompanied by DNA methylation and histone deacetylation and is prevented by insulators. Genes Dev 1998; 12:2852-62; PMID:9744862; http://dx.doi.org/10.1101/gad.12.18.2852.
6. Girod PA, Nguyen DQ, Calabrese D, Puttini S, Grandjean M, Martinet D, et al. Genome-wide prediction of matrix attachment regions that increase gene expression in mammalian cells. Nat Methods 2007; 4:747-53; PMID:17676049; http://dx.doi.org/10.1038/nmeth1076.
7. Antoniou M, Harland L, Mustoe T, Williams S, Holdstock J, Yague E, et al. Transgenes encompassing dual-promoter CpG islands from the human TBP and HNRPA2B1 loci are resistant to heterochromatin-mediated silencing. Genomics 2003; 82:269-79; PMID:12906852; http://dx.doi.org/10.1016/S0888-7543(03)00107-1.
8. Kwaks TH, Barnett P, Hemrika W, Siersma T, Sewalt RG, Satijn DP, et al. Identification of anti-repressor elements that confer high and stable protein production in mammalian cells. Nat Biotechnol 2003; 21:553-8; PMID:12679786; http://dx.doi.org/10.1038/nbt814.
9. Huang Y, Li Y, Wang YG, Gu X, Wang Y, Shen BF. An efficient and targeted gene integration system for high-level antibody expression. J Immunol Methods 2007; 322:28-39; PMID:17350648; http://dx.doi.org/10.1016/j.jim.2007.01.022.
10. Nehlsen K, Schucht R, da Gama-Norton L, Krömer W, Baer A, Cayli A, et al. Recombinant protein expression by targeting pre-selected chromosomal loci. BMC Biotechnol 2009; 9:100; PMID:20003421; http://dx.doi.org/10.1186/1472-6750-9-100.
11. Carroll D. Genome engineering with zinc-finger nucleases. Genetics 2011; 188:773-82; PMID:21828278; http://dx.doi.org/10.1534/genetics.111.131433.
12. Boch J. TALEs of genome targeting. Nat Biotechnol 2011; 29:135-6; PMID:21301438; http://dx.doi.org/10.1038/nbt.1767.
13. Montigny WJ, Phelps SF, Illenye S, Heintz NH. Parameters influencing high-efficiency transfection of bacterial artificial chromosomes into cultured mammalian cells. Biotechniques 2003; 35:796-807; PMID:14579745.
14. Ciotta G, Hofemeister H, Maresca M, Fu J, Sarov M, Anastassiadis K, et al. Recombineering BAC transgenes for protein tagging. Methods 2011; 53:113-9; PMID:20868752; http://dx.doi.org/10.1016/j.ymeth.2010.09.003.
15. Copeland NG, Jenkins NA, Court DL. Recombineering: a powerful new tool for mouse functional genomics. Nat Rev Genet 2001; 2:769-79; PMID:11584293; http://dx.doi.org/10.1038/35093556.
16. Zhang Y, Buchholz F, Muyrers JP, Stewart AF. A new logic for DNA engineering using recombination in Escherichia coli. Nat Genet 1998; 20:123-8; PMID:9771703; http://dx.doi.org/10.1038/2417.
17. Johansson T, Broll I, Frenz T, Hemmers S, Becher B, Zeilhofer HU, et al. Building a zoo of mice for genetic analyses: a comprehensive protocol for the rapid generation of BAC transgenic mice. Genesis 2010; 48:264-80; PMID:20143345.
18. Maston GA, Evans SK, Green MR. Transcriptional regulatory elements in the human genome. Annu Rev Genomics Hum Genet 2006; 7:29-59; PMID:16719718; http://dx.doi.org/10.1146/annurev.genom.7.080505.115623.
19. Yang XW, Model P, Heintz N. Homologous recombination based modification in Escherichia coli and germline transmission in transgenic mice of a bacterial artificial chromosome. Nat Biotechnol 1997; 15:859-65; PMID:9306400; http://dx.doi.org/10.1038/nbt0997-859.
20. Bian Q, Belmont AS. BAC TG-EMBED: one-step method for high-level, copy-number-dependent, position-independent transgene expression. Nucleic Acids Res 2010; 38:e127; PMID:20385594; http://dx.doi.org/10.1093/nar/gkq178.
21. Blaas L, Musteanu M, Eferl R, Bauer A, Casanova E. Bacterial artificial chromosomes improve recombinant protein production in mammalian cells. BMC Biotechnol 2009; 9:3; PMID:19144171; http://dx.doi.org/10.1186/1472-6750-9-3.
22. Blaas L, Musteanu M, Grabner B, Eferl R, Bauer A, Casanova E. The use of bacterial artificial chromosomes for recombinant protein production in mammalian cell lines. Methods Mol Biol 2012; 824:581-93; PMID:22160921; http://dx.doi.org/10.1007/978-1-61779-433-9_31.
23. Mader A, Prewein B, Zboray K, Casanova E, Kunert R. Exploration of BAC versus plasmid expression vectors in recombinant CHO cells. Appl Microbiol Biotechnol 2012; PMID:23081777; http://dx.doi.org/10.1007/s00253-012-4498-x.
24. Zambrowicz BP, Imamoto A, Fiering S, Herzenberg LA, Kerr WG, Soriano P. Disruption of overlapping transcripts in the ROSA beta geo 26 gene trap strain leads to widespread expression of beta-galactosidase in mouse embryos and hematopoietic cells. Proc Natl Acad Sci U S A 1997; 94:3789-94; PMID:9108056; http://dx.doi.org/10.1073/pnas.94.8.3789.
25. Soriano P. Generalized lacZ expression with the ROSA26 Cre reporter strain. Nat Genet 1999; 21:70-1; PMID:9916792; http://dx.doi.org/10.1038/5007.
26. Omasa T, Cao Y, Park JY, Takagi Y, Kimura S, Yano H, et al. Bacterial artificial chromosome library for genome-wide analysis of Chinese hamster ovary cells. Biotechnol Bioeng 2009; 104:986-94; PMID:19575438; http://dx.doi.org/10.1002/bit.22463.
27. Xu X, Nagarajan H, Lewis NE, Pan S, Cai Z, Liu X, et al. The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. Nat Biotechnol 2011; 29:735-41; PMID:21804562; http://dx.doi.org/10.1038/nbt.1932.