BioProcess International
SPECIAL REPORT

FUJIFILM
Diosynth biotechnologies

Application of Targeted Locus Amplification

for Enhanced Apollo X CHO Clone Screening

Application of Targeted Locus Amplification

for Enhanced Apollo X CHO Clone Screening

L.P. Pybus, J. Bergboer, M. van Min, and F.L. Saunders

Summary

- Fujifilm Diosynth Biotechnologies has integrated Cergentis's Targeted Locus Amplification (TLA) technology into Apollo X cell-line development as a powerful tool for completely sequencing transgenes and integration sites.
- TLA detects all single-nucleotide variants (SNVs) and structural variants in an integrated transgene sequence.
- TLA analyses can be used in development and quality control of cell lines, for clone selection, and for assessment of clonality and genetic stability.

Chinese hamster ovary (CHO) cells are used widely for production of recombinant therapeutic proteins, including monoclonal antibodies (MAbs). Efficient generation of high-quality recombinant CHO cell lines is a critical step in biopharmaceutical development.

Recombinant expression of therapeutic proteins requires stable integration of a gene of interest (GoI) into a CHO cell genome. An important determinant of the quality of the resulting recombinant CHO cell line is the genomic location and integrity of that transgene sequence. The most common approach is random GoI integration — as part of a linearized plasmid — into the host genome, followed by selection of stable integrants (**1**). That process can lead to undesirable integration events — e.g., multiple integration sites, unexpected integrations of backbone sequences, and undesired sequence or structural variants in the integrated transgene sequence (**2–6**).

Traditionally, the genetic characterization of cell lines has been carried out using Southern blot and/or fluorescence in situ hybridization (FISH) analysis. However, their results are incomplete and can be difficult to interpret. Next-generation sequencing (NGS) approaches such as Cergentis's proprietary Targeted Locus Amplification (TLA) technology is proving to be powerful for providing information-rich analysis of transgenes and their integration sites (**7**). TLA-based transgene and integration-site sequencing presents a cost-effective and high-quality alternative to conventional approaches.

TLA-based transgene analyses can

- identify the transgene integration site(s)
- detect structural changes in host DNA at the transgene integration site(s)
- sequence an entire transgene and detect all single-nucleotide variants (SNVs) as well as structural changes within the transgene
- provide an estimation of the transgene copy number.

Figure 1: Next-generation sequence (NGS) coverage across the transgene in Apollo X cell line 33-4, with primer sets 1 (TOP) and 2 (BOTTOM); black arrows indicate primer location; *y*-axis is limited to 500×.

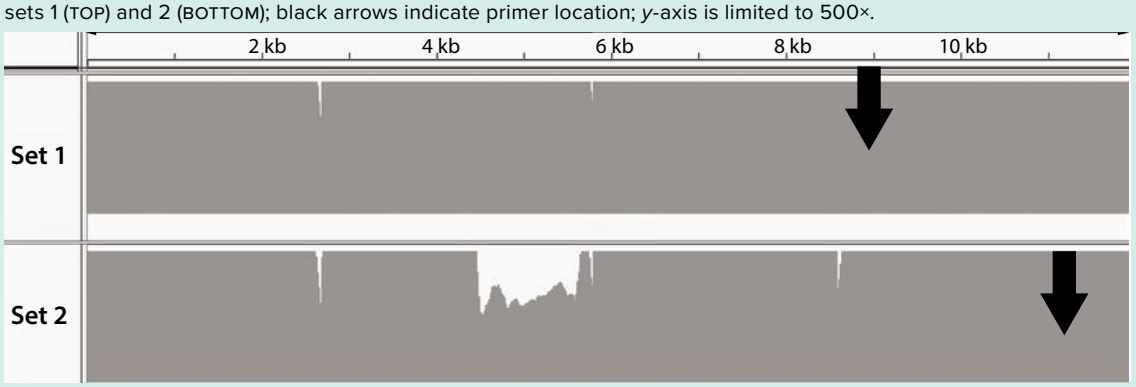


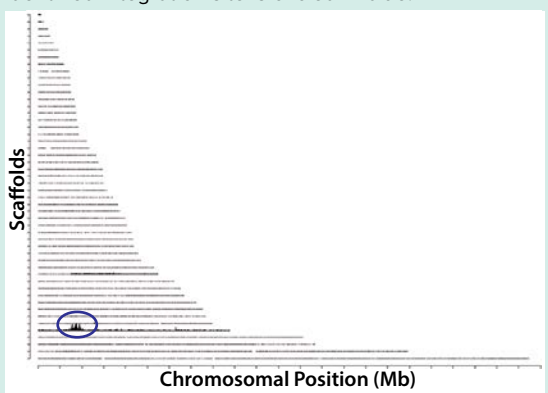
Table 1: Identified single-nucleotide variants (SNVs) in Apollo X cell line 33-158

					Primer Set 1		Primer Set 2	
Variant	Region	Position	Reference	Mutation	Coverage	Percentage	Coverage	Percentage
1	Promoter	1728	G	+5ACTTA	357	73%	294	74%
2	Promoter	4172	G	+5AATTC	1986	9%	1194	11%
3	HC	4264	A	+1G	1561	94%	648	93%
4	HC	4316	T	+1C	1743	90%	610	85%
5	HC	5218	T	+1C	32	78%	14	71%
6	HC	5484	G	+1A	21	90%	9	89%
7	Backbone	7896	C	–1T	3824	95%	1665	94%

Table 2: TG–TG fusions in Apollo X cell line 33-158

Fusion	Transgene		Transgene		Fusion Orientation	Homology, Inserts
1	←	1	11966	←	head to tail	—, —
2	→	761	10104	→	tail to head	2, —
3	→	1310	9924	←	tail to tail	1, —
4	←	2619	9923	←	head to tail	1, —
5	←	2668	9924	←	head to tail	—, —
6	→	3198	10104	→	tail to head	2, —
7	→	3747	9924	←	tail to tail	1, —
8	←	5742	9923	←	head to tail	1, —
9	←	5791	9924	←	head to tail	—, —
10	←	7011	8137	←	head to tail	2, —
11	←	9554	9928	→	head to head	—, 3
12	→	9917	9980	→	tail to head	—, —
13	←	9921	10882	←	head to tail	—, —
14	→	9924	10616	→	tail to head	—, —

Figure 2: TLA sequence coverage across the top 50 scaffolds of the CHO genome using primer set 1 in Apollo X cell line 32-121; the scaffolds are indicated on the *y*-axis, the scaffold position on the *x*-axis. Similar results were obtained with primer set 2, and the identified integration site is circled in blue.



TLA in the Apollo X Cell-Line Development Process

Apollo X technology is the latest Fujifilm Diosynth Biotechnologies mammalian expression system that fast tracks the cell-line development process without compromising product quality. This iteration delivers MAb titers up to 10 g/L in a 10-week development timeline. As a contract development and manufacturing organization (CDMO), we deliver high-quality products to our partners as our main priority.

TLA technology can be applied at different stages of the Apollo X cell-line development process. Genetic characterization of transgenes and genomic integration sites gives additional assurance about the quality, efficiency, and reliability of the Apollo X work flow. TLA analyses can determine the number of integration sites with their exact positions and detect SNVs or structural changes.

Quality Control of Cell Lines During Clone Selection: TLA can increase the efficiency and quality of clone selection (**8**). It can be used to

• deselect siblings — determine which candidate cell lines share the same integration site(s) and the same genetic origin

• select cell lines without transgene sequence variation in the GoI — by enabling early detection of cell lines with potentially problematic structural variants or SNVs in the plasmid or GoI sequence)

• select cell lines with desired (numbers of) transgene integration sites.

The number and complexity of integration sites partly determine genetic stability of cell lines (**9**). Therefore, it can be reasonable to select those lines with the smallest number of clean integration sites.

Cell-Bank Genetic Characterization: TLA generates high-quality data regarding the transgene sequence. So it can be used to characterize the complete genetics of transgenes and integration sites in cell banks.

Clonality Assessment: TLA analyses detect transgene integration site(s) at base-pair resolution. Identified breakpoint-spanning reads are unique to each integration site. So breakpoint sequence-based polymerase chain reaction (PCR) analysis can be used to test subclones and assess the clonality of a

Figure 3: Sequencing reads

5′-integration site

Scaffold_5:59,971,578 (tail) fused to TG:9,654 (head) with 2 bp homology

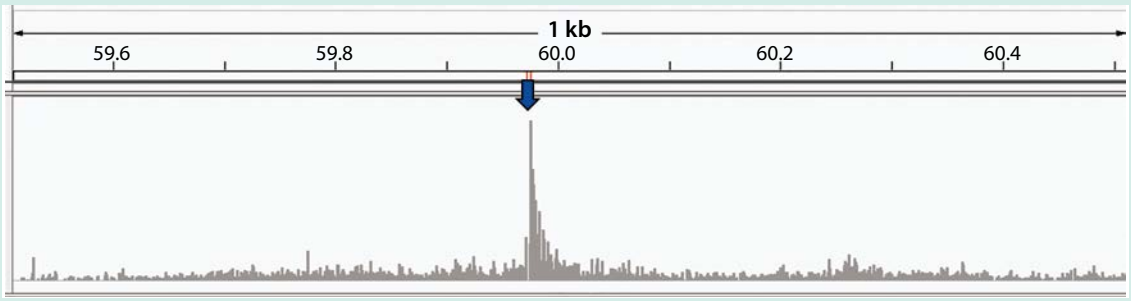
ATTGCTTGAATGGAGACACAGGGACCTCCATTGCTTTGGGGCTGATGGATGGGGTGAAAATGAATTCAGAA
TCATACGGACCACCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGT
CAGGGG

3′-integration site

TG:9,863 (tail) fused to scaffold_5:59,974,906 (head) with 31 bp inserted

GTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCAGCGC
GGTATCTGCGGGCGCTGCGCACGGTA TCAGGTATCTTCCTCCATCACTCTCCACTTTGTTGTTAGATTGGATCT
CTCACT GAACCCAGAGCTCACCGATTCTATTACACAGGCTGACCCATGAGCC

Figure 4: TLA sequence coverage across the transgene integration locus; CHO scaffold_5:59,509,768-60,517,018; sequence coverage (in gray) generated with primer set 1 in Apollo X cell line 31-187; blue arrow indicates the location of the breakpoint sequence identified with both primer sets. Similar results were obtained with primer set 2. The *y*-axis is limited to 50×.



master cell bank (MCB). If a MCB is clonal, then all subclones are expected to share the same integration site(s) (**10**).

Assessment of Genetic Stability: TLA-based analysis of multiple generations will provide information about the stability of a certain clone over time. Stable cell lines are expected to show the same integration site(s) and (if present) transgene sequence variants consistently (**9**).

Exemplary TLA-Based Sequencing of Apollo X Cell Lines

Transgene Sequence: TLA analyses result in high sequencing coverage across the integrated transgene site. Figure 1 depicts NGS coverage across the transgene sequence. *Coverage* is defined as the number of NGS reads that cover a given locus (in this case, the transgene sequence). Good coverage is observed across the complete transgene sequence TG:1–11966.

SNVs, Insertions, and Deletions: Table 1 specifies identified SNVs and small insertions or deletions (InDels). Note that SNVs 1 and 2 are located in the homologous regions and represent the same sequencing read — TCGTGACTTCCGAATTCGCCGCCACC — aligned to both positions at the border of the MAb light-chain (LC) and heavy-chain (HC) promoter. SNVs 1–7 are present in all 12 samples tested and probably represent variants present in the provided reference sequence.

Structural Variants: TLA enables detection of structural variants of the transgene sequence by detecting fusions of different transgene sequences. Identified transgene–transgene (TG–TG) fusions either represent a junction site of two transgene copies that have integrated as a concatemer or are the result of a structural rearrangement (inversion, deletion, or duplication) within one copy of a transgene.

Table 2 lists identified TG–TG fusion sites. Fusion read 1 represents the circularity of the plasmid. Fusion reads 2 and 6, 3 and 7, 4 and 8, and 5 and 9 each represent the same fusion reported twice because of transgene homology. At least one side of fusion reads 3–5, 7–9, and 11–14 is in close proximity to the linearization site at position TG:9920.

Table 3: Primers used in TLA analysis

Primer Set	Direction	Binding Position	Sequence
1	RV	8582	CGATCAAGGCGAGTTACAT
	FW	8952	TGCCTCACTGATTAAGCATT
2	RV	10730	CGACGATGCAGTTCAATG
	FW	11027	AACCGGAATTGGCAAGTAAA

Integration Sites: Integration sites were determined, and as Figure 2 shows, the TG has integrated into scaffold 4.

Locus-Wide Coverage and Breakpoint Sequences: The reads in Figure 3 were identified in Apollo X cell line 31-187 by marking the integration site on scaffold 5. The coverage profile in Figure 4 shows that a 3.3-kb genomic deletion with the coordinates in scaffold 5 of 659,971,581-59,974,905 has occurred in the region of the integration site.

Copy Number: An exact copy number cannot be determined using TLA. However, an estimation can be made based on the number of integration sites, the number of fusion reads, and the ratio of coverage on the TG and genome-integration site. The average copy number from all six Apollo X cell lines tested was 15.

Discussion

Early analytical assessment of gene-sequence integrity is important during CHO cell-line development of biopharmaceuticals. It enables selection of clonally derived cell lines with the most favorable stability and product quality attributes. The quality of recombinant protein produced is an important metric because deviations and variations in the primary protein sequence or higher-order structure can affect product efficacy and/or safety (**11, 12**).

Traditional methods of genetic characterization have limitations in data quantity, throughput, and turnaround time. Here, we have demonstrated successfully the applicability of TLA technology to CHO clone selection for biopharmaceutical production and the technology's ability more data than traditional genetic characterization techniques. Furthermore, previous studies have aligned the findings obtained from TLA data with those of other standard analytical methods (**8**).

Orthogonal product characterization methods such as liquid chromatography and mass spectrometry (LC-MS) still are preferred for detection and quantification of translational errors such as amino-acid substitutions. Those can be attributed to

cellular stress induced by suboptimal media and bioreactor conditions (**13, 14**). Although cell culture conditions can be optimized to improve translational errors later in process development, it is critical to screen and differentiate early — during cell-line development — for intrinsic mutations within the recombinant gene. That will help minimize negative impacts on development and costs.

Apollo X expression technology is based on a host cell line derived from CHO-DG44 and a dihydrofolate reductase (DHFR) selection system. Traditionally, DG44 cell-line development has used methotrexate (MTX)-based amplification to increase gene copy number and productivity (**15, 16**). The process can yield up to ~500 copies of recombinant-gene tandem repeats and directly affects development timelines and decreased cell-line stability (**17–19**). TLA analysis reveals that the approximate gene copy number for Apollo X–derived recombinant cell lines is in the tens rather than hundreds of gene copies.

That is likely to be a consequence of the holistic approach used during Apollo X development. Proprietary expression-vector technology was combined with directed evolution (to gain an improved host cell phenotype) and careful design of the cell-line development process (e.g., a single, low-concentration MTX selection step). That combination allowed us to accelerate cell-line development timelines without affecting the quality of the cell lines we produce.

TLA technology allows Fujifilm Diosynth Biotechnologies to support rapid genetic characterization of clonal cell lines during Apollo X cell-line development and clone selection. This includes early confirmation of recombinant-gene integrity and identification and elimination of clones that produce aberrant protein products.

Methods

TLA and Sequencing: Viable frozen CHO cells were used and processed according to the Cergentis TLA protocol (**7**). Two-primer sets were designed on the transgene (Table 3) and used in individual TLA amplifications. PCR products were purified and the library prepped using the Illumina Nextera flex protocol and sequences on an Illumina sequencer.

Alignment of Sequencing Reads: Reads were mapped using BWA-SW (version 0.7.15-r1140, settings bwasw–b7) (**20**). The NGS reads were aligned to the TG sequence and host genome. The CHO-K1GS assembly CHOK1GS_HZDv1 (GCA_900186095.1) was used as a reference host-genome sequence.

SNV Detection: The presence of SNVs was determined using SAMtools mpileup software (SAMtools version 1.3.1) (**21, 22**). SNVs are reported that meet the following criteria:

- allele frequency of at least 5% (relative amount of reads with the variant compared with total coverage on the variant position)
- variants present in the data of both primer sets
- coverage ≥ 30× for at least one primer set
- variants identified in both forward- and reverse-aligning sequencing reads.

TG–TG fusion detection: Fusion sequences consisting of two parts of the TG are identified using a proprietary Cergentis script. Fusions resulting from the TLA procedure itself are recognized by the restriction-enzyme–specific sequence at the junction site and removed. TG–TG fusions are reported that meet the following criteria:

- The TG–TG fusion is present in >1% of reads at the position of the fusion.
- The TG–TG fusion is observed in data of both primer sets unless those data provide a clear explanation why the fusion is not found in one dataset.
- Visual inspection is performed on the TG–TG fusions in an NGS data browser to remove fusions that are sequencing artifacts (e.g., fusions found at hairpin structures or low-complexity regions).

Integration Site Detection: Integration sites are detected based on coverage peak(s) in the genome and identification of fusion-reads between the TG sequence and host genome.

References

1 Wurm FM. Production of Recombinant Protein Therapeutics in Cultivated Mammalian Cells. *Nat. Biotechnol.* 22(11) 2004: 1393–1398.

2 Boyd D, et al. Isolation and Characterization of a Monoclonal Antibody Containing an Extra Heavy-Light Chain Fab Arm. *MAbs* 10(3) 2018: 346–353.

3 Cain-Hom C, et al. Efficient Mapping of Transgene Integration Sited and Local Structural Changes in Cre Transgenic Mice Using Targeted Locus Amplification. *Nucleic Acids Res.* 45(8) 2017: e62.

4 Cartwright JF, et al. Highly Sensitive Detection of Mutations in CHO Cell Recombinant DNA Using Multi-Parallel Single Molecule Real-Time DNA Sequencing. *Biotechnol. Bioeng.* 115(6) 2018: 1485–1498.

5 Goodwin LO, et al. Large-Scale Discovery of Mouse Transgenic Integration Sites Reveals Frequent Structural Variation and Insertional Mutagenesis. *Genome Res.* 29(3) 2019: 494–505.

6 Kaas CS, et al. Deep Sequencing Reveals Different Compositions of mRNA Transcribed from the F8 Gene in a Panel of FVIII-Producing CHO Cell Lines. *Biotechnol. J.* 10(7) 2015: 1081–1089.

7 De Vree P, et al. Targeted Sequencing By Proximity Ligation for Comprehensive Variant Detection and Local Haplotyping. *Nat. Biotechnol.* 32(10) 2014: 1019–1025.

8 Aeschlimann SH, et al. Enhanced CHO Clone Screening: Application of Targeted Locus Amplification and Next Generation Sequencing Technologies for Cell Line Development. *Biotechnol. J.* 14(7) 2019: e1800371.

9 Patel NA, et al. Antibody Expression Stability in CHO Clonally Derived Cell Lines and Their Subclones: Role of Methylation in Phenotypic and Epigenetic Heterogeneity. *Biotechnol Prog.* 34(3) 2018: 635–649.

10 Aebischer-Gumy C, et al. Analytical Assessment of Clonal Derivation of Eukaryotic/CHO Cell Populations. *J. Biotechnol.* 286(5) 2018: 17–26.

11 Zhu J. Mammalian Cell Protein Expression for Biopharmaceutical Production. *Biotechnol. Adv.* 30(5) 2012: 1158–1170.

12 Ren D, et al. Detection and Identification of a Serine to Arginine Sequence Variant in a Therapeutic Monoclonal Antibody. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* 879(47) 2011: 2877–2884.

13 Raina M, et al. Reduced Amino Acid Specificity of Mammalian Tyrosyl-tRNA Synthetase Is Associated with Elevated Mistranslation of Tyr Codons. *J. Biol. Chem.* 289(25) 2014: 17780–17790.

14 Bullwinkle TJ, et al. Oxidation of Cellular Amino Acid Pools Leads to Cytotoxic Mistranslation of the Genetic Code. *eLife* 3, 2014: e02501.

15 Alt F, et al. Selective Multiplication of Dihydrofolate Reductase Genes in Methotrexate-Resistant Variants of Cultured Murine Cells. *J. Biol. Chem.* 253(3) 1978: 1357–1370.

16 Kellems R. Gene Amplification in Mammalian Cells: Strategies for Protein Production. *Curr. Opin. Biotechnol.* 2(5) 1991: 723–729.

17 Kim NS, Kim SJ, Lee GM. Clonal Variability Within Dihydrofolate Reductase-Mediated Gene Amplified Chinese Hamster Ovary Cells: Stability in the Absence of Selective Pressure. *Biotechnol. Bioeng.* 60(6) 1998: 679–688.

18 Yoshikawa T, et al. Evaluation of Stable and Highly Productive Gene-Amplified CHO Cell Line Based on the Location of Amplified Genes. *Cytotechnol.* 33(1–3) 2000: 37–46.

19 Yoshikawa T, et al. Amplified Gene Location in Chromosomal DNA Affected Recombinant Protein Production and Stability of Amplified Genes. *Biotechnol. Prog.* 16(5) 2000: 710–715.

20 Li H, Durbin R. Fast and Accurate Long-Read Alignment with Burrows-Wheeler Transform. *Bioinformatics* 26(5) 2010: 589–595.

21 Li H, et al. The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* 25(16) 2009: 2078–2079.

22 Li H. A Statistical Framework for SNP Calling, Mutation Discovery, Association Mapping and Population Genetical Parameter Estimation from Sequencing Data. *Bioinformatics* 27(21) 2011: 2987–2993.

L.P. Pybus *and* ***F.L. Saunders*** *are in mammalian cell culture process development at Fujifilm Diosynth Biotechnologies, Billingham, UK.* ***J. Bergboer*** *and* ***M. van Min*** *are with Cergentis B.V, Utrecht, The Netherlands. Apollo is a registered trademark of Fujifilm Diosynth Biotechnologies.*

BioProcess International
SPECIAL REPORT

FUJIFILM
Diosynth biotechnologies

Application of Targeted Locus Amplification for Enhanced Apollo X CHO Clone Screening