# CELL SUBSTRATE CHARACTERISATION
# GENETIC STABILITY AND IDENTITY TESTING AT SGS (GLASGOW)

## INTRODUCTION

Genetic stability and identity testing are key requirements for regulatory approval of therapeutic biological products produced using a cell bank system.

Manufacturers of cell line recombinant DNA-derived biotherapeutics must demonstrate that the culture is pure, the cells are stable, and the recombinant DNA gene expression system is intact.[1] Characterisation of the cell bank should be performed as described in the International Conference on Harmonisation (ICH) Q5B prior to initiation of clinical trials.[2]

Genetic stability and identity testing of cell banks are key in the analysis required to maintain quality assurance of a production cell line. ICH guidelines state that testing of the cell substrate during production should be carried out from at least two time points, typically the master cell banks (MCB) or the working cell bank (WCB) and the end of production cell bank (EPCB).[3] Genetic stability testing allows comparison of the EPCB with the MCB to ensure the expression construct has not been mutated or lost during production and verifies consistency between the MCB and ECPB.

The European Pharmacopeia section 5.2.3 requires that DNA fingerprinting methods and nucleic acid amplification techniques (NAT) are used to characterise and identify cell substrates used for manufacturing gene therapy and vaccine products and to detect cross contamination of cell lines.[4] This is particularly critical when more than one cell line is present in an area used for cell banking procedures (including expansion, pooling or aliquoting of a cell line); therefore, these tests must discriminate between closely related species.

Characterisation of the cell bank should confirm the identity of the cell line and evaluate the copy number and stability of introduced nucleic acids, and the

quantity and quality of expressed proteins. Identity testing involves utilising real time polymerase chain reaction (PCR) to identify species-specific DNA target sequences within test samples. Genetic stability testing involves analysis of the transgene using techniques such as Southern blotting, Northern blotting, determination of transgene copy number, and sequencing of mRNA, plasmid, or genomic DNA; all of which provide the information necessary to demonstrate the stability of the recombinant DNA-derived products.[1,6]

SGS (Glasgow) offers a package of Genetic Stability and Identity assays, typically performed on a manufacturer's MCB, WCB and EPCB (see Table 1).

The standard SGS (Glasgow) Genetic Stability and Identity package includes:

- DNA and RNA sequencing
- Transgene copy number determination
- Northern blot and Southern blot analysis
- Cell line identification by gene-specific nucleic acid amplification techniques (NAT) and identification and DNA fingerprinting by Random Amplification of Polymorphic DNA (RAPD)

## SEQUENCING

### Nucleotide Sequencing Analysis of integrated expression cassette genomic DNA, mRNA and upstream and downstream Regulatory Control Genomic DNA

Where an expression construct is used to produce recombinant DNA-derived proteins, the manufacturer must demonstrate that the correct coding sequence of the insert encoding the recombinant DNA-derived product has been incorporated into the cell and that this is maintained to the end of production. Genetic stability testing of cell banks producing recombinant proteins or monoclonal antibodies from integrated expression cassettes requires direct nucleotide sequencing of the therapeutic gene, vector or vaccine, and the gene control regions should be verified by direct sequencing of the upstream and downstream locations.[2-5] The nucleic acid sequence must be verified and shown to be identical to that of the expression construct and should correspond to that expected for the protein sequence.[1-4] Sequencing of total genomic DNA is suitable for assessing cells with a single genomic copy of the transgene; however, for cells with multiple insertions of the transgene,

the integrity of the product transcript should be verified by sequencing of the mRNA.[4,5]

SGS (Glasgow) provides four-fold coverage nucleic acid sequence analysis of plasmids, genomic DNA, recombinant mRNA (cDNA), and upstream and downstream regulatory control regions of genomic DNA performed using the Applied Biosystems 3500 and 3500xL analysers. Using primers designed in-house or supplied by the Sponsor, complete sequence verification of both DNA strands is provided by contig alignment and comparison of the consensus sequence with the Sponsor's reference sequence.

### Nucleotide Sequencing Analysis of Sponsor's Viral Vector

Regulatory approval of biopharmaceuticals and vaccines requires extensive analysis to confirm the nucleotide sequence of the therapeutic gene, vector or vaccine.[1-4] SGS (Glasgow) provides DNA sequencing of viral vectors using DNA sequencing primers designed in-house from the expression vector map and electronic reference sequence provided by the Sponsor, or with Sponsor-supplied primer sequences. Viral DNA prepared from the Sponsor's sample is used as the template for PCR amplification of the viral vector, the resulting PCR products of which are purified and used as template(s) for DNA sequencing using the ABI 3500 or ABI 3500 xL DNA Analyser.

SGS (Glasgow) uses the BigDye™ Terminator 3.1 Ready Reaction Mix (Applied Biosystems) and where necessary, reagents such as DMSO, betaine, glycerol, dGTP, etc., may be used to relax secondary structures and to sequence through difficult DNA sequences. Sequence data will be analysed and assembled using appropriate software and analysis is performed making a comparison with the electronic reference sequence provided by the Sponsor.

## GENE COPY NUMBER

Cell lines produced by transfection of recombinant DNA often contain an array of multiple copy transgene inserts at a single genetic locus. This may give rise to problems with genetic stability when homologous recombination occurs at the locus between different sequences within the array, resulting in a reduction in the number of transgene copies at the site and a decrease in the quantity of mRNA and protein produced. A key indicator of stability in cell banks used in recombinant protein production is consistency of transgene copy number between the MCB, or WCB and EPCB.

SGS (Glasgow) offers a 2-step approach to determine gene copy number:

### 1. Validation of Quantitative Real Time PCR (Q-PCR) Assay to Determine the Transgene Copy Number

This validation is product-specific and is targeted to the Sponsor's transgene sequence. The results will determine the parameters that will be applied during the Gene Copy Number test performed simultaneously on both the MCB or WCB and EPCB. The assay may be completed prior to the EPCB and requires a sample of the plasmid and the sequence of the transgene.

### 2. Determination of Gene Copy Number of the Integrated Expression Cassette by Q-PCR

This assay uses the specific parameters determined during the validation process described above. The genetic stability of a cell bank can be assessed by determining the average number of copies per cell of a gene integrated into the host genome. SGS (Glasgow) uses real time Q-PCR to detect and quantify the presence of host cell DNA and integrated recombinant DNA gene copies within cell banks. This assay may be completed on both the MCB or WCB and EPCB simultaneously.

## STRUCTURAL ANALYSIS OF THE EXPRESSION CASSETTE

Regulatory authorities request information on the number of insertion sites, and structural integrity of the expression cassette, in addition to mRNA size:

### Structural Analysis of Integrated Recombinant Transgene by Southern Blot

Southern blots are used to confirm the presence of the transgene by restriction enzyme digest using enzymes selected specifically for the expression construct. For direct comparison, this assay can be performed on samples from both the MCB, or WCB and EPCB simultaneously and may allow the determination of insertion site number (the number of sites where the expression construct is integrated into the host cell genome) and Structure Map Confirmation, which confirms the absence of large insertions or deletions in the region of the integrated recombinant transgene. The Southern Blot method is validated and sensitive enough to detect a single copy integrated insertion site.

### Size Determination of the mRNA Expressed from the Recombinant Transgene

Northern blot analysis can reveal information about mRNA identity, and size. However, analysis of the mRNA coding sequence by RT-PCR and nucleic acid sequencing is considered more suitable. This is because it provides full information on the open reading frame and the predicated amino acid sequence of the therapeutic protein.

## CELL LINE IDENTIFICATION BY GENE-SPECIFIC NUCLEIC ACID AMPLIFICATION TECHNIQUES (NAT) AND RANDOM AMPLIFICATION OF POLYMORPHIC DNA (RAPD)

### Cell line identification by gene-specific nucleic acid amplification techniques (NAT)

SGS (Glasgow) offers a sensitive method of cell line DNA identification through the application of gene-specific nucleic acid amplification techniques (NAT). These procedures provide cell line identification by comparing genomic DNA from the Sponsor's test item and a positive control reference cell line of the same species using real time polymerase chain reaction (PCR). This technique identifies species-specific DNA target sequences within test samples.

### Identification and DNA Fingerprinting of Cell Lines by Random Amplification of Polymorphic DNA (RAPD)

RAPD is a PCR method in which genomic DNA from the Sponsor's test item and a reference cell line of the same species are analysed. Segments of cell line genomic DNA are amplified at random using arbitrary short primers, which are subsequently resolved by gel electrophoresis and the resulting banding patterns produce a semi-unique profile for the cell line being tested. No knowledge of the DNA sequence for the targeted gene is required, as the primers bind randomly in the cell genome sequence.

The reference cell lines comprise the appropriate species-specific DNA, which is included in the RAPD assay for comparison with the Sponsor's test item to confirm the identification of the species. Reference cell line DNA typically used include Sf21, 311, Sf9, TRA-171, CHO, HEK 293, Vero, A549, MDCK, Balb-C, MRC-5, Hi-5, Drosophila S2, BHK21, PG13, HeLa and HeLa S3.

Table 1 MCB or WCB and EPCB Genetic Stability Testing

| SGS SP # | GENETIC STABILITY ASSAY | LAB TIME* | SAMPLE REQUIREMENTS** |
|---|---|---|---|
| SP-M.8401 | Identification and DNA Fingerprinting of Cell Lines by Random Amplified Polymorphic DNA (RAPD) Assay | 3 weeks | 2 vials of 1 x $10^7$ cell pellets per cell bank. |
| SP-M.8450 | Cell Line Identification by Gene-Specific Nucleic Acid Amplification Techniques (NAT) | 3 weeks | 2 vials of 1 x $10^7$ cell pellets per cell bank. |
| SP-M.8500V | Validation of Q-PCR Assay to Determine the Gene Copy Number of the Integrated Expression Cassette | 12 weeks (minimum) | Minimum 100 µl (~500 ng/µl) plasmid.<br>Full vector map with relevant information including gene insert flanking regions, resistance genes, promoters, etc. Full length nucleotide sequence in form of text file, Vector NTI, FASTA. |
| SP-M.8500T | Determination of the Gene Copy Number of the Integrated Expression Cassette by Q-PCR | 4 weeks | 2 vials of 2 x $10^7$ cell pellets per cell bank. |
| SP-M.8502 | Structural Analysis of Integrated Recombinant Transgene by Southern Blot | 12 weeks (minimum) with set up | 2 vials of 5 x$10^7$ cells. 200 µl of reference plasmid DNA (≥100 ng/µl).<br>Electronic DNA sequence of the entire vector and restriction map. Coding sequence coordinates. |
| SP-M.8517 | Nucleotide Sequencing of the Expressed Transgene mRNA (cDNA sequencing) | 10-12 weeks | 2 x $10^7$ – 5 x $10^7$ cells (project dependent; plus equivalent back-up sample).<br>Full vector map with relevant information including gene insert flanking regions, resistance genes, promoters, etc. Full length nucleotide sequence in form of text file, Vector NTI, FASTA. |
| SP-M.8518 *** | Nucleotide sequencing of the Upstream and Downstream Regulatory Control Genomic DNA (500 base pairs of 5' and 3' flanking regions) | 10-12 weeks | 2 x $10^7$ – 5 x $10^7$ cells (project dependent; plus equivalent back-up sample).<br>Full vector map with relevant information including gene insert(s) flanking regions, resistance genes, promoters, etc. Full length nucleotide sequence in form of text file, Vector NTI, FASTA. |
| SP-M.8516 | Nucleotide Sequence Analysis of Sponsor's Viral Vector | 8-10 weeks | 2 - 5 ml (project dependent); titre >$10^{12}$ (plus equivalent back-up sample).<br>Full vector map with relevant information including gene insert flanking regions, resistance genes, promoters, etc. Full length nucleotide sequence in form of text file, Vector NTI, FASTA. |
| SP-M.8515 | Nucleotide Sequence Analysis of Sponsor's Plasmid DNA Vector (purified) | 6-8 weeks | 300 ng – 500 ng/µl; minimum 250 µl.<br>(For *E. coli* - minimum of 500 µl of appropriate glycerol stock)<br>Full vector map with relevant information including gene insert flanking regions, resistance genes, promoters, etc. Full length nucleotide sequence in form of text file, Vector NTI, FASTA. |

*Lab times are estimated and may vary depending on specific Sponsor requirements.

** SP-V.0004. Cell culture expansion of client cells is available for this project when volumes stated cannot be supplied.

*** If the construct is present as a single genomic copy, then sequencing of the insert on the genome would also be determined after discussion with Sponsor.

*References*

1. *Good Manufacturing Practice Regulations European Council Derivatives 2003/94/EC and 91/412/EEC.*
2. *ICH Q5B (1995). Analysis of the Expression Construct in Cells used for the Production of rDNA Derived Protein Products.*
3. *ICH Q5D (1997). Derivation and Characterisation of Cell Substrates used for Production of Biotechnological/Biological products.*
4. *European Pharmacopoeia. Volume 1 and 2. Section 5.2.3. Cell Substrates for the Production of Vaccines for Human Use.*
5. *CBER FDA 1993 PTC in the Characterisation of Cell Lines used to Produce Biologicals.*
6. *CBER, FDA 1992 Supplement to PTC in the Production and Testing of New Drugs and Biologicals Produced by r-DNA Technology.*



**CONTACT INFORMATION**

SGS, GLASGOW
(SGS VITROLOGY)
5 South Avenue
Clydebank Business Park
Clydebank
G81 2LG

+44 (0)141 952 0022
biosafety@sgs.com
www.sgs.com/lifescience