

ADVERTIMENT. L'accés als continguts d'aquesta tesi queda condicionat a l'acceptació de les condicions d'ús estableties per la següent llicència Creative Commons:  http://cat.creativecommons.org/?page_id=184

ADVERTENCIA. El acceso a los contenidos de esta tesis queda condicionado a la aceptación de las condiciones de uso establecidas por la siguiente licencia Creative Commons:  <http://es.creativecommons.org/blog/licencias/>

WARNING. The access to the contents of this doctoral thesis it is limited to the acceptance of the use conditions set by the following Creative Commons license:  <https://creativecommons.org/licenses/?lang=en>



**Universitat Autònoma
de Barcelona**

Escola d'Enginyeria

Departament d'Enginyeria Química, Biològica i Ambiental

Programa de Doctorat en Biotecnologia

Production of Virus-Like Particles in HEK 293 cells and functionalization with SARS-CoV-2

Arnaud Boix i Besora

PhD Thesis

June 2022

Advisors:

Dr. Francesc Gòdia Casablancas

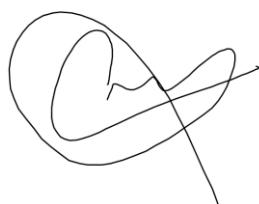
Dr. Laura Cervera Gracia

Prof. Dr. Francesc Gòdia Casablancas and Dr. Laura Cervera Gracia

Hereby certify:

That Arnau Boix i Besora has successfully completed his PhD entitled **“Production of Virus-Like Particles in HEK 293 cells and functionalization with SARS-CoV-2”** in the Chemical, Biological and Environmental Engineering Department at Universitat Autònoma de Barcelona under our supervision. This PhD thesis is hereby presented in this document in order to opt to the PhD title as Doctor in Biotechnology by the Universitat Autònoma de Barcelona.

Thereby and for the appropriate purposes, we sign this document in Bellaterra, June 2022.



Arnau Boix i Besora
(author)



Francesc Gòdia Casablancas
(co-director)



Laura Cervera Gracia
(co-director)

ADVERTIMENT. L'accés als continguts d'aquesta tesi queda condicionat a l'acceptació de les condicions d'ús establertes per la llicència Creative Commons Reconeixement-NoComercial-SenseObraDerivada 4.0 Internacional (CC BY-NC-ND 4.0).

Per a més informació: <https://creativecommons.org/licenses/by-nc-nd/4.0/deed.ca>.

ADVERTENCIA. El acceso a los contenidos de esta tesis queda condicionado a la aceptación de las condiciones de uso establecidas por la licencia Creative Commons Atribución-NoComercial-SinDerivadas 4.0 Internacional (CC BY-NC-ND 4.0).

Para más información: <https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>.

WARNING. The access to the contents of this doctoral thesis is limited to the acceptance of the use conditions set by the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International license (CC BY-NC-ND 4.0). To view a copy of this license, visit: <https://creativecommons.org/licenses/by-nc-nd/4.0/deed.en>.



Arnaud Boix i Besora, 2022

Table of contents

| | |
|--|----|
| Abbreviations..... | 9 |
| Summary..... | 15 |
| Resum..... | 17 |
| Resumen..... | 19 |
| Introduction..... | 21 |
| 1. Vaccination role in global health protection | 23 |
| 2. Vaccines..... | 24 |
| 2.1. Conventional vaccines | 24 |
| 2.2. New generation vaccines | 25 |
| 3. Virus-like particles (VLPs) | 28 |
| 4. Platforms for VLP production | 29 |
| 5. HIV-1 and AIDS..... | 30 |
| 5.1 HIV-1 treatment and vaccine development | 32 |
| 5.2 HIV-1 VLPs | 33 |
| 6. Gag-VLPs production in HEK293 cells..... | 34 |
| 6.1 Transient gene expression (TGE) | 35 |
| 6.2 Stable gene expression (SGE) | 35 |
| 6.2.1 Random integration | 37 |
| 6.2.2 Site-directed integration..... | 38 |
| 7. SARS-CoV-2 and COVID-19..... | 39 |
| Objectives..... | 53 |
| Results | 55 |
| Chapter one..... | 57 |
| Optimization, Production, Purification and Characterization of HIV-1 Gag-Based Virus-Like | |

Particles Functionalized with SARS-CoV-2

| | |
|--|-----|
| Chapter two | 93 |
| Study of the recognition of different SARS-CoV-2 functionalized VLPs by COVID-19 convalescent human sera | |
| Chapter three | 121 |
| Stable HEK293 cell line generation for the production of Gag::eGFP VLPs by random integration | |
| Chapter four | 157 |
| Stable HEK293 cell line generation for the production of Gag::eGFP VLPs with a site-directed approach using CRISPR / Cas9 technology | |
| Chapter five | 203 |
| Stable HEK293 RMCE cell line generation for the production of Gag::eGFP VLPs by lentiviral infection | |
| Chapter six | 249 |
| Study and process optimization of stable HEK293 production of Gag::eGFP VLPs by different strategies | |
| Chapter seven | 269 |
| Production of SARS-CoV-2 pseudotyped VLPs in HEK293 cell lines with stable Gag::eGFP expression | |
| Discussion and future work | 287 |
| Conclusions..... | 305 |

Abbreviations

| | |
|-----------------|---|
| 3D | Three-dimensional |
| AAV | Adeno-Associated Virus |
| AAVS1 | Adeno-Associated Virus Site 1 |
| Abs | Absorbance |
| ACE-2 | Angiotensin-Converting Enzyme-2 |
| ADE | Antibody Dependent Enhancement |
| AIDS | Acquired Immunodeficiency Syndrome |
| ALIX | ALG-2 Interacting Protein X |
| ARTs | Antiretroviral Treatments |
| ARV | Antiretroviral |
| AU | Absorbance Units |
| BCA | Bicinchoninic acid |
| BSA | Bovine Serum Albumin |
| C | Contaminants and aggregates fraction (DSP) |
| CA | Capsid (from HIV-1 Gag polyprotein) |
| CAG | CMV enhancer, Chicken beta-Actin promoter and rabbit beta-Globin splice acceptor site |
| CAP | CEVEC's Amniocyte Production (cell line) |
| Cas9 | CRISPR-associated protein 9 |
| CCD | Charge-Coupled Device |
| CCR5 | Chemokine Receptor 5 |
| CDE | Cell Density Effect |
| CDS | Coding DNA Sequence |
| CFDA | China Food and Drug Administration |
| cGMP | current Good Manufacturing Practice |
| CHO | Chinese Hamster Ovary (cell line) |
| CiP | Cleaning in Place |
| CMV | Cytomegalovirus |
| CNV | Copy Number Variation |
| COVID-19 | Coronavirus Disease 2019 |
| CpG | Cytosine and Guanine separated by a phosphate |
| cPPT/CTS | central Polypurine Tract/Central Termination Sequence |
| Cq | quantification cycle |
| Cre | "Causes recombination" recombinase protein |
| CRISPR | Clustered Regulatory Interspaced Short Palindromic Repeats |
| crRNA | CRISPR RNA |
| CryoTEM | Cryogenic Transmission Electron Microscopy |

| | |
|--------------------------------|---|
| CS | Cleavage Site |
| CV | Conjugate Vaccines |
| CV (DSP context) | Column Volumes |
| D | aspartic |
| DCs | Dendritic Cells |
| ddPCR | digital droplet PCR |
| DF | Degree of Freedom |
| DMSO | Dimethyl Sulfoxide |
| DNA | Deoxyribonucleic Acid |
| DoE | Design of Experiments |
| dpi | days post infection |
| dpt | days post-transfetion |
| DSB | Double Strand Break |
| dsDNA | double-stranded DNA |
| DSP | Downstream Process |
| DsRed | <i>Discosoma</i> Red Fluorescent Protein |
| DTT | Dithiothreitol |
| E | Envelope protein (from SARS-CoV-2) |
| E(1-4) | Elution fractions (1-4) |
| EF-1α | Elongation Factor 1-alpha |
| EGE | Extended Gene Expression |
| ELISA | Enzyme-Linked Immunosorbent Assay |
| EMA | European Medicines Agency |
| EMCV | Encephalomyocarditis Virus |
| Env | Envelope |
| ERGIC | ER-Golgi Intermediate Compartment |
| ESCRT | Endosomal Sorting Complexes Required for Transport |
| EV | Extracellular Vesicles |
| FACS | Fluorescence-Activated Cell Sorting |
| FBS | Fetal Bovine Serum |
| FcRs | Fc Receptors |
| FDA | Food and Drug Administration (U.S.) |
| FITC | Fluorescein Isothiocyanate |
| Flp | Flippase |
| FP | Fusion Peptide |
| FreestyleMix | FreeStyle™ 293 media (ThermoFisher) supplemented with an in-house formula |
| FRT | Flipase Recognition Target |
| FSC | Forward Scatter |

| | |
|-------------------|--|
| FT | Flow Through |
| FU | Fluorescence Units |
| G | Glycine |
| G418 | Geneticin disulphate |
| Gag | Group-specific Antigen |
| Gag::eGFP | Gag polyprotein fused in-frame with the enhanced GFP protein |
| gDNA | genomic DNA |
| GFP / eGFP | Green Fluorescence Protein / enhanced Green Fluorescence Protein |
| GMP | Good Manufacturing Practice |
| GOI | Gene Of Interest |
| gp120 | glycoprotein 120 (from HIV-1 Env polyprotein) |
| gp160 | glycoprotein 160 (from HIV-1 Env polyprotein) |
| gp41 | glycoprotein 41 (from HIV-1 Env polyprotein) |
| GSH | Genomic Safe Harbor |
| G-VLP | HIV-1 Gag-based VLP |
| h | hours |
| HA-L | Left Homology Arm |
| HA-R | Right Homology Arm |
| HBS | HEPES Buffered Salt solution |
| HBV | Hepatitis B Virus |
| HDR | Homology-Directed Repair |
| HEK/HEK293 | Human Embryonic Kidney 293 (cell line) |
| HeLa | Henrietta Lacks (cell line) |
| HEPES | 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid |
| HETP | Height Equivalent to a Theoretical Plate |
| HIV | Human Immunodeficiency Virus |
| HIV-1 | Human Immunodeficiency Virus type 1 |
| HIV-1 ψ | HIV-1 packaging signal |
| HIV-2 | Human Immunodeficiency Virus type 2 |
| hpi | hours post infection |
| hpt | hours post-transfection |
| HPV | Human Papillomavirus |
| HR | Homologous Recombination |
| HR1 | Heptad Repeat 1 |
| HR2 | Heptad Repeat 2 |
| hRosa26 | human orthologue of the mouse Reverse Orientation Splice Acceptor 26 |
| HyCell | HyCell TransFx-H media (HyClone) |
| ICC | Immunocytochemistry |

| | |
|--------------|--|
| IEX | Ion-Exchange chromatography |
| IF | Immunofluorescence |
| IgG | Immunoglobulin G |
| IN | Integrase (from HIV-1 Pol polyprotein) |
| IRES | Internal Ribosome Entry Site |
| IV | Inactivated Vaccines |
| Kana | Kanamycin |
| LA | Live Attenuated |
| LB | Lysogeny Broth |
| LNP | Lipid Nanoparticle |
| LoxP | Locus of crossover in phage P1 |
| LTR | Long Terminal Repeat |
| LV | Lentiviral Vector |
| M | Membrane protein (from SARS-CoV-2) |
| MA | Matrix (from HIV-1 Gag polyprotein) |
| MERS | Middle East Respiratory Syndrome |
| min | minute |
| MOI | Multiplicity Of Infection |
| MOPS | 3-(N-morpholino)propanesulfonic acid |
| mRNA | messenger Ribonucleic Acid |
| MS | Mean Square |
| MTT | 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide |
| N | Nucleocapsid protein (from SARS-CoV-2) |
| NC | Nucleocapsid (from HIV-1 Gag polyprotein) |
| NCBI | National Center for Biotechnology Information |
| Neo | Neomycin |
| NHEJ | Non-Homologous End Joining |
| nsps | non-structural proteins |
| NTA | Nanoparticle Tracking Analysis |
| NTD | N-Terminal Domain |
| NTU | Nephelometric Turbidity Unit |
| p17 | protein 17 (or matrix protein from HIV-1 Gag polyprotein) |
| p24 | protein 24 (or capsid protein from HIV-1 Gag polyprotein) |
| p6 | protein 6 (from HIV-1 Gag polyprotein) |
| PAGE | Polyacrylamide Gel Electrophoresis |
| PAM | Protospacer Adjacent Motif |
| PAMPs | Pathogen-Associated Molecular Pattern motifs |
| PBS | Phosphate-Buffered Saline |

| | |
|---------------------------|--|
| PCR | Polymerase Chain Reaction |
| PEI | Polyethylenimine |
| Pol | Polymerase |
| PPP1R12C | Phosphatase 1 Regulatory subunit 12C |
| PR | Protease (from HIV-1 Pol polyprotein) |
| PrEP | Pre-Exposure Prophylactic |
| PTMs | Post-Translational Modifications |
| q_p | Specific productivity |
| qPCR | quantitative Polymerase Chain Reaction |
| RBD | Receptor Binding Domain |
| RBM | Receptor-Binding Motif |
| RdRP | RNA-dependent RNA polymerase |
| RFU | Relative Fluorescence Units |
| RH | Relative Humidity |
| RMCE | Recombinase-Mediated Cassette Exchange |
| RNA | Ribonucleic Acid |
| rpm | revolutions per minute |
| RPP30 | Ribonuclease P Protein Subunit P30 |
| RRE | Rev Response Element |
| RSV | Respiratory Syncytial Virus |
| RSV | Rous Sarcoma Virus |
| RT | Room Temperature |
| RT (HIV-1 context) | Retro-Transcriptase (from HIV-1 Pol polyprotein) |
| RTs | Recombinase Target sites |
| S | Spike protein (from SARS-CoV-2) |
| s | second |
| S1 | Spike protein subunit 1 |
| S2 | Spike protein subunit 2 |
| SARS-CoV | Severe Acute Respiratory Syndrome Coronavirus |
| SARS-CoV-2 | Severe Acute Respiratory Syndrome Coronavirus 2 |
| SDS | Sodium Dodecyl Sulfate |
| SEC | Size-Exclusion Chromatography |
| Sf9 | <i>Spodoptera frugiperda</i> 9 (cell line) |
| SGE | Stable Gene Expression |
| sgRNA | single guide RNA |
| SIN | Self-Inactivating |
| SIV | Simian Immunodeficiency Virus |
| SS | Sum of Squares |

| | |
|-------------------------|--|
| SSC | Side Scatter |
| SSR | Site-Specific Recombinase |
| ssRNA | single-stranded RNA |
| STR | Stirred Tank Reactor |
| SV | Subunit Vaccines |
| SV40 | Simian Virus 40 |
| S-VLP | HIV-1 Gag-based SARS-CoV-2 spike functionalized VLP |
| TAG | Tagging |
| TALENs | Transcription Activator-Like Effector Nucleases |
| TAR | Targeting |
| TEM | Transmission Electron Microscopy |
| TGE | Transient Gene Expression |
| Tm | Temperature of melting |
| tracrRNA | trans-activating CRISPR RNA |
| TV | Toxoid Vaccines |
| v/v | volume for volume |
| vc | viable cells |
| VCD | Viable Cell Density |
| Vero | Verda reno (cell line) |
| VLPs | Virus-like particles |
| VOC | Variant Of Concern |
| v_p | volumetric productivity |
| V-SSC | Violet Side Scatter |
| VSV-G | Vesicular Stomatitis Virus Glycoprotein |
| VV | Vector-based Vaccines |
| VV (DSP context) | Void Volume fraction |
| W | wash |
| w/v | weight for volume |
| w/w | weight for weight |
| WB | Western Blot |
| WHO | World Health Organization |
| WPRE | Woodchuck hepatitis virus Posttranscriptional Regulatory Element |
| ZFN | Zinc Finger Nucleases |
| Δneo | neomycin phosphotransferase gene lacking the initial ATG codon |
| λDNA | DNA from Bacteriophage lambda |
| λ_{em} | emission wavelength |
| λ_{ex} | excitation wavelength |

Summary

Vaccines are the most cost-effective life-saving invention of the human history and their relevance in society had become evident during the recent COVID-19 pandemic. Virus-like particles (VLPs) are a promising approach for the development of new vaccine candidates. This work is focused on the production of VLPs by the recombinant expression of the HIV-1 Gag polyprotein, and their modification to present epitopes from pathogens of interest. The work can be divided in three main sections. The first one, comprising chapters one and two, studies the functionalization of Gag VLPs with the SARS-CoV-2 Spike (S) protein. The second one, comprising chapters three, four, five and six, focuses on the generation, characterization and production enhancement of different stable gene expression (SGE) cell lines for the production of Gag-VLPs. Finally, the third one, comprising chapter seven, merges the knowledge acquired in the previous sections to produce S-functionalized VLPs at bioreactor scale.

In chapter one, the methodology for the generation of Gag-VLPs incorporating the S protein (S-VLPs) by the transient co-expression of S and Gag proteins is presented. The functionalization of the VLPs is achieved, their production is enhanced using design of experiments (DoE), and a bioprocess consisting of a bioreactor followed by a scalable downstream purification process is performed.

In chapter two, several S-VLP variants harboring different S protein mutations are proposed. The influence of the mutations is studied in terms of VLP expression and quality. After that, COVID-19 convalescent human sera recognition assays are performed to select the best proposed variant based on its immunogenic potential.

In chapter three, a stable cell line for the production of Gag-VLPs is generated by means of illegitimate random integration. The selection of the 10H9 clone is based on growth kinetics, VLP production and expression stability.

In chapter four, a SGE cell line constitutively expressing Gag-VLPs is generated by locus-specific integration at the AAVS1 genomic safe harbor using CRISPR/Cas9. The selected clone, named 13++, greatly improves the production levels of the 10H9 cell line. Moreover the success of the used

integration approach postulates it as an interesting methodology for the production of new cell lines.

In chapter five, a Recombinase-Mediated Cassette Exchange (RMCE)-competent cell line stably expressing Gag-VLPs is generated by lentiviral transduction. The clones are screened in terms of growth kinetics, VLP production, stability, and ability to perform RMCE targeting. The selected clone, named SH5, can be used to easily generate high producer stable cell lines for the expression of virtually any gene of interest.

In chapter six, the VLP production and quality of the 10H9, 13++ and SH5 cell lines are compared with transient gene expression (TGE), while the culture media used and the length of the bioprocess is studied. 13++ cell line cultured in HyCell media and harvested at day 10 improves by 2.7-fold the VLP production levels of TGE.

Finally, in chapter seven, the production of S-VLPs by 13++ cell line transfection at 1 liter bioreactor is studied. The results are evaluated and discussed to define the steps to be followed in future processes for the generation of pseudotyped VLPs.

Resum

Els vaccins són l'invent més rendible de la història de la humanitat quant a vides salvades i la seva rellevància en la societat s'ha fet evident durant la recent pandèmia de COVID-19. En aquest sentit, les Virus-Like Particles (VLPs) constitueixen una prometedora aproximació per al desenvolupament de nous candidats vacunals. Aquest treball es centra en la producció de VLPs mitjançant l'expressió recombinant de la poliproteïna Gag del VIH-1 i la seva modificació amb l'objectiu de presentar epítops de patògens d'interès. El treball es pot dividir en tres apartats principals. El primer, format pels capítols u i dos, estudia la funcionalització de les VLPs de Gag amb la proteïna Spike (S) del SARS-CoV-2. El segon, format pels capítols tres, quatre, cinc i sis, es focalitza en la generació, caracterització i millora de la producció de diferents línies cel·lulars d'expressió gènica estable (SGE) per a la producció de VLPs de Gag. El tercer, format pel capítol set, fusiona els coneixements adquirits en els apartats anteriors per produir VLPs funcionalitzades amb la proteïna S a escala de bioreactor.

En el primer capítol, es presenta una metodologia que permet la generació de VLPs de Gag funcionalitzades amb la proteïna S (S-VLPs) mitjançant la co-expressió de les proteïnes S i Gag. S'assoleix la funcionalització de les VLPs, la seva producció és optimitzada mitjançant disseny d'experiments i es realitza un cultiu en bioreactor seguit d'un procés de purificació amb potencial per a ésser traslladat a escala industrial.

Al capítol dos, es proposen diverses variants de S-VLPs que presenten diferents mutacions de proteïna S. La influència de les mutacions en l'expressió i la qualitat de les VLPs és estudiada. Posteriorment, es realitzen assajos per a avaluar el reconeixement de les VLPs per part de sèrum de persones recuperant-se de la COVID-19 per a seleccionar la millor variant en funció del seu potencial immunogènic.

Al capítol tres, es genera una línia cel·lular estable per a la producció de VLPs de Gag, mitjançant integració aleatòria. La selecció del clon 10H9 és basada en

els seus paràmetres de creixement, la producció de VLPs i l'estabilitat de l'expressió.

Al capítol quatre, es genera una línia cel·lular estable que expressa VLPs de Gag de manera constitutiva, mitjançant la integració dirigida al *genomic safe harbor* AAVS1 mitjançant CRISPR/Cas9. El clon seleccionat, anomenat 13++, millora significativament els nivells de producció de la línia cel·lular 10H9. A més, l'èxit obtingut emprant aquesta tècnica, la postula com una metodologia interessant per a la producció de noves línies cel·lulars.

Al capítol cinc, es genera mitjançant transducció lentiviral una línia cel·lular competent per intercanvi de cassets mediat per recombinasa (RMCE) que expressa de forma estable VLPs de Gag. Els clons s'examinen en termes de cinètica de creixement, producció de VLPs, estabilitat i capacitat de dur a terme intercanvi de cassets per RMCE. El clon seleccionat, anomenat SH5, pot ser emprat per a generar fàcilment línies cel·lulars estables d'alta producció per a l'expressió de pràcticament qualsevol gen d'interès.

En el capítol sis, es comparen els nivells de producció i la qualitat de les VLPs de les línies cel·lulars 10H9, 13++ i SH5, tot comparant-les amb una producció per expressió gènica transitòria (TGE), mentre que s'estudien els medis de cultiu utilitzats i la durada del bioprocés. El cultiu de la línia cel·lular 13++ en medi HyCell durant 10 dies millora en 2,7 vegades els nivells de producció de VLPs de la TGE.

Finalment, al capítol set, s'estudia la producció de S-VLPs mitjançant la transfecció transitòria de la línia cel·lular 13++ en bioreactor d'un litre. Els resultats s'avaluen i es discuteixen per definir els passos a seguir en futurs processos per a la generació de VLPs pseudotipades.

Resumen

Las vacunas son el invento más rentable de la historia de la humanidad en cuanto a vidas salvadas y su relevancia en la sociedad se ha hecho evidente durante la reciente pandemia de COVID-19. En este sentido, las Virus-Like Particles (VLPs) constituyen una prometedora aproximación para el desarrollo de nuevos candidatos vacunales. Este trabajo se centra en la producción de VLPs mediante la expresión de la poliproteína Gag del VIH-1 y su modificación para presentar epítopos de patógenos de interés. El trabajo puede dividirse en tres apartados principales. El primero, formado por los capítulos uno y dos, estudia la funcionalización de las VLPs de Gag con la proteína Spike (S) del SARS-CoV-2. El segundo, formado por los capítulos tres, cuatro, cinco y seis, se focaliza en la generación, caracterización y mejora de la producción de diferentes líneas celulares de expresión génica estable (SGE) para la producción de VLPs de Gag. El tercero, formado por el capítulo siete, fusiona los conocimientos adquiridos en los apartados anteriores para producir VLPs funcionalizadas con la proteína S a nivel de biorreactor.

En el primer capítulo, se presenta una metodología que permite la generación de VLPs de Gag funcionalizadas con la proteína S (S-VLPs) mediante la co-expresión de las proteínas S y Gag. Se alcanza la funcionalización de las VLPs, su producción es optimizada mediante diseño de experimentos y se realiza un cultivo en biorreactor seguido de un proceso de purificación con potencial para su traslado a escala industrial.

En el capítulo dos, se proponen diversas variantes de S-VLPs que presentan diferentes mutaciones de proteína S. La influencia de las mutaciones en la expresión y calidad de las VLPs es estudiada. Posteriormente, se realizan ensayos para evaluar el reconocimiento de las VLPs por parte de suero de personas recuperándose de la COVID-19 para seleccionar la mejor variante en función de su potencial inmunogénico.

En el capítulo tres, se genera una línea celular estable para la producción de VLPs de Gag, mediante integración aleatoria. La selección del clon 10H9 está

basada en sus parámetros de crecimiento, la producción de VLPs y la estabilidad de la expresión.

En el capítulo cuatro, se genera una línea celular estable que expresa VLPs de Gag de forma constitutiva, mediante la integración dirigida al *genomic safe harbor* AAVS1 mediante CRISPR/Cas9. El clon seleccionado, 13++, mejora significativamente los niveles de producción de la línea celular 10H9. Además, el éxito obtenido utilizando esta técnica, la postula como una metodología interesante para la producción de nuevas líneas celulares.

En el capítulo cinco, se genera mediante transducción lentiviral una línea celular competente por intercambio de casetes mediado por recombinasa (RMCE) que expresa de forma estable VLPs de Gag. Los clones se examinan en términos de cinética de crecimiento, producción de VLPs, estabilidad y capacidad de llevar a cabo RMCE. El clon seleccionado, SH5, puede utilizarse para generar fácilmente líneas celulares estables de alta producción para la expresión de prácticamente cualquier gen de interés.

En el capítulo seis, se comparan los niveles de producción y la calidad de las VLPs de las líneas celulares 10H9, 13++ y SH5, comparándolas con una producción por expresión génica transitoria (TGE), mientras que se estudian los medios de cultivo utilizados y la duración del bioproceso. El cultivo de la línea celular 13++ en medio HyCell durante 10 días mejora en 2,7 veces los niveles de producción de VLPs de la TGE.

Por último, en el capítulo siete, se estudia la producción de S-VLPs mediante TGE de la línea celular 13++ en biorreactor de un litro. Los resultados se evalúan y se discuten para definir los pasos a seguir en futuros procesos para la generación de VLPs pseudotipadas.

Introduction

1. Vaccination role in global health protection

Since the introduction of clean water sanitization and the discovery of antibiotics, few measures can be compared with the vaccination impact in terms of global health protection [1,2]. It is the most cost-effective life-saving invention in the human story [3] and according to the World Health Organization (WHO), vaccines against 20 life-threatening diseases had been approved, preventing 2-3 million deaths every year [4]. In the US alone, analyses show that the vaccination against 13 diseases prevents approximately 20 million cases including 40.000 deaths for each annual vaccinated cohort, resulting in 68.8 billion dollars in social cost savings [5]. Vaccination does not only protect people as individuals, but it also provides protection to the entire population by reducing person-to-person infections. When the proportion of immune individuals in a population reaches a certain threshold, the spread of the disease is interrupted, protecting vaccinated and unvaccinated individuals from large outbreaks [1,6]. This phenomenon is known as herd immunity and benefits vulnerable groups such as pregnant women, immune-compromised individuals and babies too young to be vaccinated [6].

The ultimate goal when fighting a disease is its elimination or eradication [7]. Eradication not only saves people's lives, but also represents a more economic strategy in the long term, since once eradicated no more resources need to be invested to control the disease [7]. In the human history, two diseases have been eradicated thanks to vaccination: smallpox and rinderpest [8,9], while poliomyelitis and rubella are considered potentially eradicable as their incidences had been drastically reduced worldwide as a consequence of vaccination campaigns (Figure 1) [10].

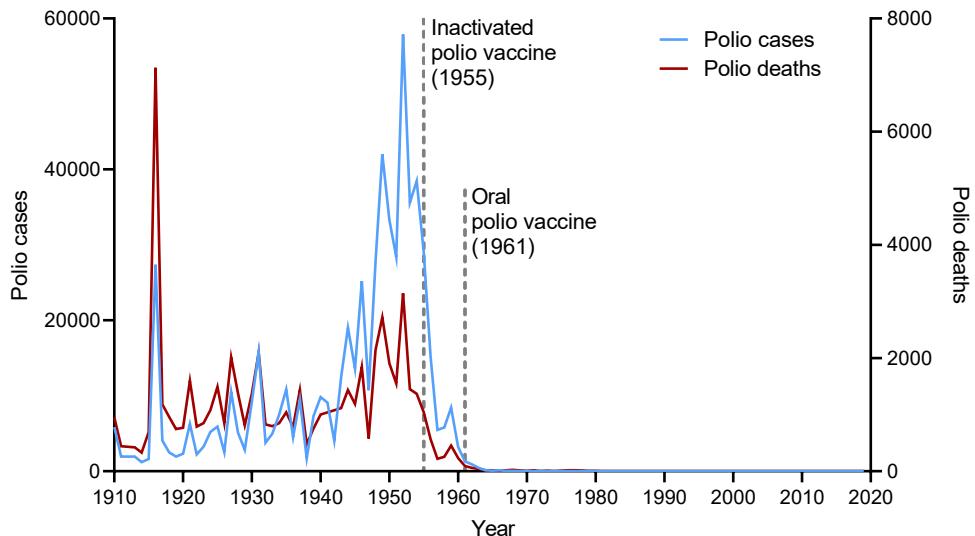


Figure 1: Polio cases and associated deaths in the United States of America from 1912 to 2019. Drastic drops in cases and mortality can be observed after the introduction of polio vaccines. Data was retrieved from Our World in Data (OWID) based on US Public Health Service and US Center for Disease Control.

2. Vaccines

Vaccines are biological entities that safely provide active induced immunity to a particular disease or infection, conferring protection on subsequent exposures to the pathogen that causes it [11]. In order to accomplish this, vaccines need to present the pathogenic antigens to the vaccinated individual immune system. These antigens can be an entire agent that resembles the disease-causing microorganism or parts of it [11]. This stimulates the host's immune system and if the inoculated individual is later exposed to the disease-causing pathogen, it will recognize it as exogenous and activate a much stronger and fast protective response [11,12]. From a general point of view, vaccines can be classified into two main families: conventional and new-generation vaccines [13].

2.1. Conventional vaccines

Conventional or first-generation vaccines generally consist on the entire pathogen that is being targeted, thus, the virus either in inactivated or attenuated form [14]. Live attenuated (LA) vaccines are those containing the wild or disease-causing virus weakened. This can be achieved by repeated culture in the laboratory under conditions that disable the pathogen virulent properties (typically passaging the pathogen through a series of chick embryo

cultures) without losing its immunogenicity or by using a viable related virus with equivalent immune response and reduced virulence properties [15]. Killed and inactivated vaccines (IV) are those containing a pathogen, inactivated by heat, radioactivity or treated with chemicals, among others [16]. These first generation vaccines have been proven very efficient to combat a series of infectious diseases. Mumps, measles and rubella vaccines are examples of live-attenuated; while rabies, injected polio and hepatitis A are inactivated vaccines (Table 1). However, virus-based conventional vaccines can potentially cause disease, either due to infection of immunocompromised individuals or to unlikely but possible genetic reversion of the attenuated or inactivated pathogen. Their manufacture is also hazardous as requires the manipulation of live viruses.

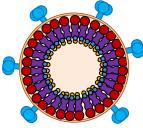
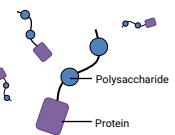
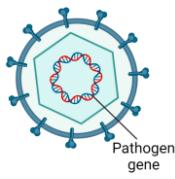
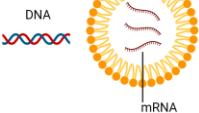
2.2. New generation vaccines

New generation vaccines encompass proteins or polysaccharides and generally use recombinant DNA technology to deliver specific pathogen components or genetic material instead of the whole microorganism. Toxoid vaccines (TV) are those containing an inactivated pathogen toxin, creating immunity against the toxin instead of the organism itself. If the presented components are peptides or polysaccharides, they are referred as subunit vaccines (SV) [17]. Peptide vaccines contain a synthetically-generated peptide predicted to provoke an effective immune response [18]. Polysaccharide and polysaccharide-conjugate vaccines are constituted by polymeric carbohydrate chains that bacteria present on their surface in large amounts. Conjugate vaccines (CV) contain immunogenic carrier proteins that have been covalently linked to the polysaccharide antigens, thereby inducing a stronger immune response. Examples of the TV, SV and CV vaccines are diphtheria and tetanus (TV), pneumococcal and typhoid (SV); and pneumococcal and meningococcal conjugates (CV) (Table 1) [14,19]. Those vaccines have difficulties mimicking the native conformation of antigen polymers resulting in a lower immunogenicity profile compared with LA or LI vaccines [16]. For that reason, SV are often co-administrated with adjuvants in order to boost the immune response [20].

Virus-like particles (VLPs) are a subclass of recombinant vaccines which include the viral structural proteins arranged without carrying any genetic material inside the particle itself, rendering it incapable of replicating but presenting the external appearance of a native virus. This approach triggers potent humoral and cellular immune responses making the use of adjuvants optional [16,21]. Hepatitis B and Human Papillomavirus (HPV) vaccines are examples of licensed VLP-based vaccines (Table 1) [22]. However different VLP vaccines such as Influenza Virus, Respiratory Syncytial Virus (RSV), HIV, Ebola or SARS-CoV-2, are currently under clinical trials [23,24].

Finally there is a group of new generation vaccines based on the delivery of genetic material into the vaccinated individual cells. This approach results in host cells producing the genetically encoded immunogenic protein from the pathogen and presenting it on their own membrane, where it can be then recognized by immune cells [25]. They are DNA, mRNA and Vector-based vaccines (VV). DNA vaccines contain genetically modified DNA encoded in bacterial plasmids. When injected, this DNA will be transcribed and traduced by the host cell. Additionally, a potent response against the bacterial DNA methylated motifs will further stimulate the immune system [16]. mRNA vaccines use a lipid nanoparticle as a carrier to deliver mRNA molecules which will be then traduced and presented on the dendritic cell membranes. DNA and mRNA vaccines are relatively easy to design and they avoid the production and purification of difficult proteins [26]. However, they require the use of adjuvants in order to boost the immune response. Vector vaccines (VV) deliver DNA by means of a non-replicating recombinant viral vector [25,27]. The Pfizer-BioNTech COVID-19 vaccine is an example of mRNA vaccine while the Vaxzevria COVID-19 vaccine (previously known as Oxford-AstraZeneca) is based on Adenoviral VVs (Table 1). There are no currently approved DNA vaccines for use in humans [27].

Table 1. Vaccines classified by type and first approval. Adapted from [11].

| Type of vaccine | Approved vaccines | First approved |
|-------------------------------|---|--|
| Live attenuated vaccines (LA) |  | Measles, mumps, rubella, yellow fever, influenza, oral polio, typhoid, Japanese encephalitis, rotavirus, <i>Mycobacterium bovis</i> bacillus Calmette–Guérin, varicella zoster |
| Inactivated vaccines (IV) |  | Whole-cell pertussis, polio, influenza, Japanese encephalitis, hepatitis A, rabies |
| Toxoid vaccines (TV) |  | Diphtheria, tetanus |
| Subunit vaccines (SV) |  | Pertussis, influenza, hepatitis B, meningococcal, pneumococcal, typhoid, hepatitis A |
| Virus-like particles (VLPs) |  | Hepatitis B, human papillomavirus |
| Conjugated vaccines (CV) |  | <i>H. influenzae</i> type B, pneumococcal, meningococcal, typhoid |
| Vector-based vaccines (VV) |  | Ebola, SARS-CoV-2 |
| Nucleic acid (DNA / mRNA) |  | SARS-CoV-2 (mRNA) |

3. Virus-like particles (VLPs)

Virus-like particles are well-defined molecular structures based on the intrinsic ability of viral proteins to self-assemble when expressed using recombinant systems [21,28]. The generated particles are composed by multiple proteins, they mostly range from 20-200 nm and can be either enveloped or non-enveloped [23,28,29], varying in complexity from non-enveloped single-protein simpler configurations to multilayered compositions [30]. VLPs preserve the native conformation and size of viral particles making them almost indistinguishable from viruses from an antigenic perspective [31,32]. They do not carry any viral genetic information so they are unable to infect or replicate, facilitating its potential use for clinical purposes [28]. This lack on genetic information represents a safer alternative to LA or IV, especially for immunocompromised or elder vaccinated individuals [23]. VLPs are also usually less costly vaccine candidates in economic terms and its industrial manufacture is easier and less hazardous, as no pathogenic viruses have to be manipulated unlike LA, IV and VV [33,34].

VLPs can be used to study conformational assembly of viral particles, drug delivery, delivery of bio- and nanomaterials or gene therapy among others [23,35]. However, the field where VLPs generate greater interest is in the VLP-based vaccine area, because the repetitive arrays on their surface constitute pathogen-associated molecular pattern motifs (PAMPs), which are recognized by the immune system generating strong cellular and humoral responses [21,32]. This makes the use of adjuvants optional [16,21], representing an advantage over SV, mRNA and DNA vaccines. VLPs elicit immune response pathways different from those activated by conventional inactivated viral vaccines, and its authentic conformation makes them more immunogenic than SV [24].

The first FDA approved VLP recombinant vaccine for human use was Recombivax HB®, against hepatitis B virus (HBV), in 1986 [22]. Nowadays HBV vaccine is produced by different companies in different host cell platforms including the mammalian Chinese Hamster Ovary (CHO). VLP-based vaccines against human papillomavirus (HPV) Gardasil® and Cervarix® (produced in

yeast and insect cells, respectively) have also been FDA-approved and licensed commercially [21,36,37]. Hepatitis E (Hecolin®) and Malaria (Mosquirix™) VLP-based vaccines have been approved recently by the China Food and Drug Administration (CFDA) [38,39] and the European Medicines Agency (EMA) [40], respectively.

4. Platforms for VLP production

VLPs can be produced in prokaryotic and eukaryotic heterologous expression systems, such as *E. coli* (bacteria), *S. Cerevisiae* (yeast), Sf9 (insect), *A. Tumefaciens* (plant) or HEK293 (mammalian) cells [21,41]. When producing a specific VLP, the advantages and disadvantages of each production platform must be considered.

Bacteria and yeast are simple, rapid, scalable, cost-effective and high-yield platforms suitable for non-enveloped VLP production, but can carry endotoxins, do not present or have inappropriate glycosylation patterns and can produce incorrectly-assembled or folded proteins.

Plants constitute a simple platform which cannot present human-derived viral contamination, but present low expression levels, cannot undergo correct post-translational modifications (PTMs) or VLP assembly and can derive to antigen degradation. They need to be properly modified and adjuvanted in order to be used as vaccines. The *Medicago Covifenz* plant-based VLP vaccine was proved to be a safe and effective against COVID-19 [42].

Insect cells can produce correctly assembled VLPs in high amounts and high cell density conditions. In addition, there is a reduced risk of culturing human opportunistic pathogens and baculovirus expression systems can ease the scale up of the bioprocesses. However, baculovirus-derived components are difficult to remove and can compete with the epitope of interest masking the immune response. Insect platforms also present glycosylation limitations [21].

Mammalian cells constitute an attractive production platform for enveloped or multimeric VLPs especially for their ability to perform complex PTMs and correct VLP assemblies [27]. Nevertheless, their culture productivities are relatively low,

they have a high production cost and they can potentially be contaminated by mammalian pathogens [24].

5. HIV-1 and AIDS

The Human Immunodeficiency Virus (HIV) is a lentivirus that targets the immune system and causes Acquired Immunodeficiency Syndrome (AIDS) [43], a chronic disease which targets and damages the immune system allowing the occurrence of various malignancies and opportunistic infections [44].

HIV virus is derived from the simian immunodeficiency virus (SIV) and phylogenetic studies have dated the first transmission between 1910 and 1930 in West Africa [45]. After its emergence it spread until the 1980 decade, when it was first recognized in the US [45]. Since then, HIV has claimed 36.3 million lives and it has been estimated that there are currently 37.7 million infected people worldwide [46].

HIV is transmitted by body fluids exchange, such as blood, breast milk, semen and vaginal secretions, or perinatal transition, however sexual transmission drives the HIV epidemic in most countries [47]. HIV most common strains are HIV type 1 and type 2 (HIV-1 and HIV-2, respectively), being HIV-1 the most widespread worldwide.

HIV virions are complex spherical particles with an average size of 120-145 nm in diameter formed by an internal cone-shaped protein core [48] surrounded by a lipid bilayer derived from the host cell membrane embedded with glycoproteins [49]. The conic capsid contains the HIV genome, composed by positive single-stranded RNA (ssRNA) chains of approximately 9.8kb in length [50] which are retrotranscribed into double-stranded DNA (dsDNA) and integrated into the infected cell's genome [51]. The viral ssRNA genomes are flanked by two long terminal repeats (LTR) and composed by nine open-reading frames (*gag*, *pol*, *env*, *vif*, *vpr*, *nef*, *vpu*, *tat* and *rev*) which encode for a total of 15 proteins after maturation (Figure 2) [49].

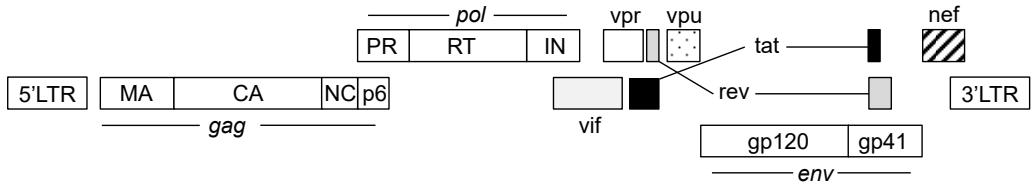


Figure 2: Scheme of the HIV-1 genome.

A promoter found in the LTR 5' of the integrated proviral genome initiates the transcription using the host cell machinery, resulting in a primary transcript which after alternative splicing become the mRNAs coding for env, vif, vpr, nef, vpu, tat and rev proteins [49,51]. Part of the pre-mRNA is not spliced, giving rise to Gag and Gag-Pol polyproteins [51]. As explained in Figure 3, the roles of the expressed proteins in the HIV-1 lifecycle are:

- Env (envelope protein): the env protein (gp160) is cleaved by host proteases resulting in gp120 and gp41 which are incorporated at the external part of the HIV-1 virion membrane. The first one mediates target cell recognition and binding in subsequent infections; and gp41 mediates the host cell membrane fusion [52,53].
- Pol (polymerase): part of the Gag-Pol polyprotein, it contains the protease (PR), retro-transcriptase (RT) and integrase (IN) domains. PR processes the polyproteins during virion maturation (Figure 3), while RT and IN drive the transcription of the genomic RNA into DNA and its integration into the host's cell genome, respectively [53].
- Gag (group-specific antigen): can be part of Gag or Gag-Pol polyproteins. It presents different domains which will result in p17 or matrix protein (MA), p24 or capsid protein (CA), nucleocapsid protein (NC) and p6 [49,52]. The MA domain targets Gag to the plasma membrane while recruiting the Env glycoproteins. CA domain promotes Gag multimerization, NC interacts and incorporates the viral genome into the virions while facilitating the viral assembly; and p6 recruits the endosomal sorting complexes required for transport (ESCRT), responsible for the release of the virion in a process known as budding. As represented in Figure 3, once the immature virion is released, viral protease PR cleaves and splits the Gag precursors into MA, CA, NC and

p6 structural proteins, which are rearranged resulting in the virion maturation [53].

- Rev and tat: Regulatory proteins. Rev promotes nuclear export of unspliced RNA while tat promotes transcription [52].
- vif, vpr, nef, vpu: Referred as accessory proteins, their presence is not mandatory for viral generation, but are required for *in-vivo* pathogenesis and infectivity [51,52].

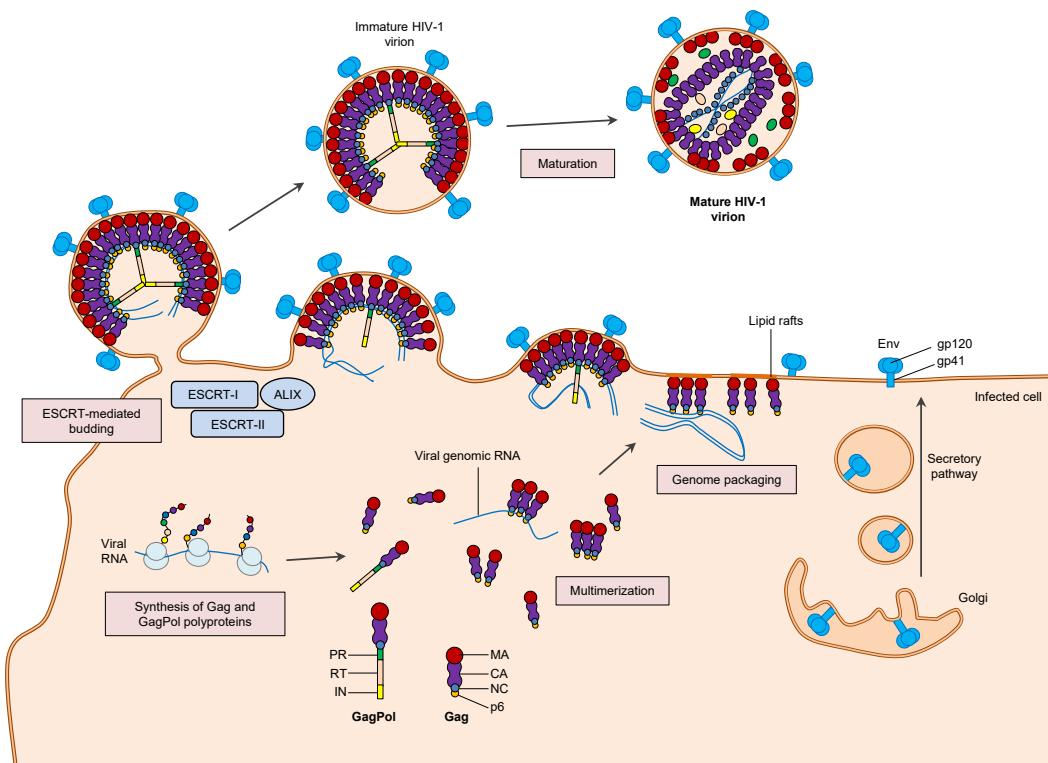


Figure 3. Generation of HIV-1 virions. Briefly, host cell mechanisms translate the viral RNA into Gag and GagPol polyproteins, which multimerize, bind to the lipid rafts of the cellular membrane and recruit the viral genome. Env is incorporated and then viral budding occurs facilitated by ESCRT and ALIX. Immature virion is released to the media and it undergoes maturation, promoted by the proteolytic cleavage of Gag and GagPol polyproteins by the PR protease, generating a mature HIV-1 virion. Based on the results of [53].

5.1 HIV-1 treatment and vaccine development

Apart from prevention, the current treatments against AIDS are based on a combination of antiretroviral (ARV) drugs [54]. Antiretroviral treatments (ARTs) suppress viral replication in the infected individuals, allowing their immune system to recover and regain capacity to function normally fighting cancer and opportunistic infections [46]. ARTs had a huge positive impact in life quality of

AIDS patients by prolonging their live span to almost this of non-infected individuals, converting a mortal disease into a chronic manageable condition [54]. Additionally, antiretroviral drugs can also be used in high risk population groups as a pre-exposure prophylactic (PrEP) measure [55]. It is estimated that nowadays combined ART and PrEP measures prevent more infections than any other strategy alone, but care must be taken since its long term general use is predicted to contribute to future HIV-1 drug resistances [56]. For this reason, the development of an HIV prophylactic vaccine would be of great convenience. However, the high mutation rates which allows HIV to escape adaptive immune responses, the genetic variability in circulating variants worldwide, and biological properties that difficult immune attack, makes this purpose a major challenge [57]. Traditional LA viruses might regain pathogenicity and infect the inoculated individual, while LI and SV vaccines cannot generate neutralizing antibodies against the genetically diverse variants currently circulating nor the necessary cell-mediated responses to counteract the HIV infection [58]. Although the vaccine candidates tested to date failed to prevent or counteract the HIV infection, huge efforts in vaccine research are being made with the goal of ultimately control and eradicate AIDS [58,59]. In this regard, HIV VLPs based on Gag polyprotein constitute an interesting approach for the generation of new vaccines against HIV or other diseases, since they do not represent a risk for patients and can generate strong immune responses [29,60].

5.2 HIV-1 VLPs

In 1989, Delchambre and colleagues described the ability of the SIV Gag polyprotein to be excreted in form of nanoparticles when expressed on its own [61]. During the next decade, Götlinger reported the same behavior for the HIV-1 Gag polyprotein: when recombinantly expressed in the absence of other viral proteins or RNA, it generated HIV-1 Gag VLPs [62]. These particles were spontaneously assembled using the intrinsic ability of the Gag polyprotein to travel to the plasmatic membrane, multimerize, recruit the ESCRT machinery and bud out of the cells [53].

Gag VLPs are enveloped nanoparticles of ~145 nm in diameter, composed by a core of about 3600 structural Gag monomers interacting with the inner part of

the membrane surrounding them (Figure 4) [63,64]. The VLP lipid bilayer is incorporated during the budding process from the producer cell membrane. In consequence, any protein present in the cellular membrane during the production process, will end up incorporated to the outer part of the generated Gag-VLP [27]. This can be exploited as an approach to functionalize VLPs by the co-expression of the Gag polyprotein with membrane proteins of interest, forming VLPs capable of conferring immunity to different viral variants or diseases.

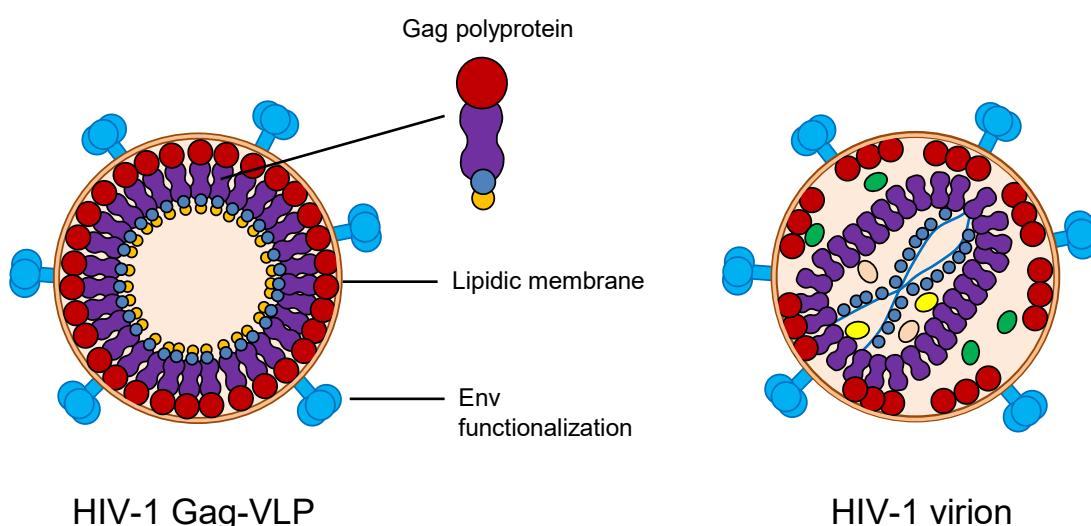


Figure 4. HIV-1 Gag-VLP and HIV-1 virion. The HIV-1 VLP is constituted by the immature Gag polyprotein attached to the inner part of the membrane. It can be further functionalized with proteins from pathogens of interest, such as Env from HIV-1. HIV-1 virion has undergone a process of maturation in which the Gag polyprotein is cleaved and its subunits rearranged to form an inner protein core containing the RNA genome together with other viral proteins. Unlike Gag-VLPs, HIV-1 virion can infect host cells, replicate and integrate.

6. Gag-VLPs production in HEK293 cells

As previously mentioned, mammalian platforms constitute a promising approach for the expression of multimeric enveloped VLPs due to their ability to perform complex PTMs and assemble VLPs in the desired native viral antigenic configurations. There are several mammalian cell platforms in which enveloped VLPs can be produced: HeLa, Vero, CAP, CHO or HEK293 [21,65,66]. HEK293 cell line, derived from human embryonic kidney (hence, its acronym HEK), is globally used for recombinant expression and viral-vector production due to its capacity to grow in suspension in bioreactors, good characterization, high transfection efficiency and its capability to grow in chemically-defined serum-

free media [27,67]. There are two main approaches for the production of VLPs in HEK293 cells: transient gene expression (TGE) and stable gene expression (SGE).

6.1 Transient gene expression (TGE)

TGE constitutes a rapid and easy approach in which multiple copies of the DNA coding for the gene of interest (GOI) are delivered into the cell facilitated by the use of physical or chemical methods, deriving in high protein productions [68]. The physical methods are mainly based on approaches like electroporation or micro injection, while chemical methods are based on positively charged reagents like polyethylenimine (PEI). PEI interacts with the DNA generating positively charged complexes that are then internalized by the cell [69,70]. In TGE, the foreign genetic material is not integrated into the transfected cell's genome. For that reason, the expression of the transgene is lost over time due to cell division and environmental factors [71]. TGE is mainly used for genetic studies and for small scale protein production pre-clinical stages [70]. Exceptionally, some pharmaceutical products like adeno-associated virus (AAV) vectors for gene therapy are produced by TGE [69]. TGE allows the expression and study of multiple proteins without the need of generating a stable cell line, which is a longer and more tedious process. TGE products are typically harvested between 48 and 72 hours after transfection [69]. TGE advantages and drawbacks are described in Table 2.

6.2 Stable gene expression (SGE)

Stable gene expression is a valuable strategy based on the direct integration of a gene or cassette of interest into the producer cell's genome, achieving a constitutive recombinant expression prolonged over a long period of time [70]. It is widely used at industrial processes once the biopharmaceutical or product of interest is selected, and the clonal homogeneous and well-defined nature of SGE cell lines is traduced in a high batch-to-batch consistency [72]. It avoids the need to generate high amounts of DNA when scaling up bioprocesses, avoids the economic cost of the transfections reagents at large scale and reduces the possibility of contaminations during transfection procedures [72]. Furthermore, it does not present limitations regarding the use of media, as in

TGE the formation and stabilization of transfection DNA complexes are greatly affected by media ionic composition [73]. This makes SGE an easier and advantageous approach for large scale production compared with TGE. However the use of SGE strategies need the generation of a stable cell line, a process which is laborious, time-consuming and expensive. Its advantages and drawbacks are described in Table 2.

Table 2. Advantages and drawbacks of transient and stable gene expression approaches.

| | Transient gene expression (TGE) | Stable gene expression (SGE) |
|----------------------------|---|---|
| Generation time | Weeks | Up to 12 months (the first time) |
| Difficulty | Easy | Laborious |
| Economic cost | Medium (repeated) | High (the first time), then low |
| Expression | Episomal: lost over time | Chromosome integrated: constitutive |
| Homogeneity | Heterologous: different GOI copies and presence of non-producer cells | Clonal: well-defined and characterized homogeneous cell line and product |
| Cell density | Limited to transfection procedure: cell density effect | No limitations |
| | Small to medium scale | No limitations |
| Scalability | DNA and transfection reagents cost drastically increases with the operation size | No transfection reagents or DNA needed |
| Common applications | Research and small-scale or pre-clinical recombinant protein or viral vector production | Industrial and large scale production |
| Media | Limited to transfection-compatible media | No limitations |
| | Useful for the expression of toxic products | Expression of toxic products must be performed using inducible strategies |
| Others | Facilitates the study of multiple candidates Cellular growth negatively affected by transfection reagents toxicity | New cell line for each product or variant |

SGE cell lines can be generated using different strategies and approaches, but all of them share a common pathway. The first step is the delivery of the DNA to be integrated into the cell by TGE. Viral transduction can also be used if the

integration of a single copy per clone is desired, by using low multiplicity of infection (MOI). Then, cells are clonally isolated and expanded in order to generate monoclonal cultures which will be subsequently screened for optimal attributes such as growth kinetics, specific productivity, stability, and product quality [74]. Clonal isolation can be performed using traditional methodology like limiting dilution, or by product-specific approaches such as fluorescence activated cell sorting (FACS) [75]. The selection of the clones harboring the construct of interest integration can be further facilitated by the presence of selection genes (antibiotic resistances) or by coupling the GOI with a reporter protein, such as green fluorescence protein (GFP) for its clonal sorting using FACS and the following product production characterization [76].

Regarding the integration of the DNA of interest, two main approaches can be used: random gene integration and rational site-directed gene integration.

6.2.1 Random integration

In random integration, the introduced recombinant DNA (linearized or circular) is spontaneously integrated into the host cell line chromosomal DNA [77]. This results in a cellular pool consisting of cells with no integration and clones presenting integration at different loci with varying copy numbers. Random integration into HEK293 cells is a low efficient and rare event as after its nuclear entry, the majority of the DNA is degraded or diluted in each cell division [77]. Studies suggest that the integration event occurs in one cell per thousands, although this may vary between cell lines and strategies [77,78]. Moreover, only a small part of the cellular population will accommodate the delivered construct in a highly-transcriptional and active locus, capable of producing the GOI in high titers. Even then, the expression of the transgene will be frequently silenced over time by epigenetic mechanisms such as changes in chromatin structure or DNA methylation [79], a phenomenon known as the “position effect” [80]. For that reason, selective pressure using markers and/or antibiotics followed by a screening process analyzing a large number of potential clones is highly recommended [79]. This facilitates finding clones with random integration into transcriptionally active locus (also known as “hot-spots”) [81].

6.2.2 Site-directed integration

Site-directed integration is based on the targeted gene integration into specific pre-defined loci, known to be transcriptionally active [74]. This rationally-designed locus-specific approach constitutes an attractive strategy that overcomes the dependence on a fortuitous optimal random integration and potential position effect drawbacks [82].

Genomic safe harbor (GSH) regions in the human genome constitute interesting targets for the generation of SGE. GSH can accommodate the integration of transgenes and support their solid, long-term and predictable gene expression without silencing or undesired interactions with adjacent chromosomal regions [83]. Currently, three main GSHs have been identified and reported: the chemokine receptor 5 (CCR5) gene locus, the human orthologue of the mouse ROSA26 locus (hRosa26) and the adeno-associated virus site 1 (AAVS1), [84,85].

Site-directed integration can be achieved by homologous recombination (HR) or homology-directed repair (HDR). The first strategy occurs when a construct is flanked by homology arms complementary with the targeted genomic region [86]. It is a spontaneous and inefficient event which takes place in one per 10^5 targeted cells [87]. However, the presence of double strand breaks (DSB) in the integration loci notably increases its occurrence [87]. DSBs trigger the eukaryotic mechanisms of cellular repair. Cells can repair the cleaved fragment by non-homologous end joining (NHEJ) resealing the cut joining the two broken DNA ends disrupting the original sequence [88–90]. However, if a homologous DNA template is present in the cell, it can be used by the homology-directed repair (HDR) mechanism as a guide to make the repair [90,91]. Therefore, SGE cell lines can be generated by inducing DSBs in a selected locus while delivering the construct of interest to be integrated surrounded by homology arms.

DSB can be generated using different targeted-location restriction enzyme technologies, like transcription activator-like effector nucleases (TALENs), Zinc Finger Nucleases (ZFN) and clustered regulatory interspaced short palindromic repeats associated with Cas9 (CRISPR/Cas9). In short, they rely on the guide

of customizable proteins (ZFN, TALENs) or RNA (CRISPR/Cas9) to find, bind and generate DSBs [91].

7. SARS-CoV-2 and COVID-19

The Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a betacoronavirus that causes an acute respiratory disease, known as coronavirus disease 2019 (COVID-19) [92]. COVID-19 affects the upper gastrointestinal and respiratory tracks and it's characterized by fever, shortness of breath, cough, fatigue and effects on gastrointestinal, respiratory, musculoskeletal and nervous systems [93,94]. However, the symptomatology can significantly vary depending on the age, health and previous immunity of the patients, from asymptomatic to severe medical conditions [93].

SARS-CoV-2 emerged in Wuhan (Hubei, China) at the end of 2019 and since then has spread worldwide causing a global pandemic with devastating health and economic consequences [95]. Molecular evolutionary studies indicate that SARS-CoV-2 has a zoonotic origin, from a non-human mammalian reservoir [96]. However, the mechanism and host species implicated in the zoonotic event have not been elucidated. Since its emergence, SARS-CoV-2 has caused 6.3 million deaths, with 514 million reported cumulative cases [97].

SARS-CoV-2 is transmitted by inhalation of droplets and aerosols generated by individuals with an active infection (airborne transmission), droplets spray into eyes nose or mouth (close contact) or by deposition of ballistic droplets onto surfaces (fomite transmission) [98]. SARS-CoV-2 has a ~30 kbp genome, the second largest among all RNA viruses [99]. Despite this, it presents a slow mutation rate thanks to the proofreading function of its viral polymerase [100]. The massive worldwide spread resulted in the emergence of variants (some of them catalogued as variants of concern [VOC]) differing in its transmissibility, immune evasion and pathogenesis. However, SARS-CoV-2 cannot be considered a highly variable pathogen, as the actual variants can differ in less than 0.2% from the initial strains [101]. To put this numbers in perspective, HIV-1 subtypes can present differences of 15% of its genome [101].

SARS-CoV-2 virions are spherical enveloped particles presenting an average size of 80 nm in diameter characterized by the prominence of spike proteins on their surface [99]. They present a positive-sense single-stranded RNA genome of approximately 29,900 nucleotides composed by 14 open reading frames which encode for 31 proteins: 16 non-structural proteins (nsps), 4 structural proteins and 11 accessory proteins with different functions [102]. Nsps are coded in the 5' two-thirds of the genome, and regulate viral replication, containing the RNA-dependent RNA polymerase (RdRP) [99]. Membrane (M), envelope (E), nucleocapsid (N) and Spike (S) constitute the structural proteins (Figure 4). M protein is the most abundant structural particle in the virion and determines its structure and shape. E protein promotes viral particle release and together with M protein, orchestrates the viral assembly and the formation of mature envelopes [103], while N protein interacts and encapsulates the viral genome, forming an inner nucleocapsid [99]. S protein is a highly glycosylated transmembrane protein present in the surface of the viral particle forming homotrimeric spikes. It plays a crucial role in the viral infection as interacts with the host cell receptor ACE-2 and primes the fusion of the viral and cellular membranes, mediating viral entry (Figure 4) [99,104]. Due to its accessibility and role in viral infection, S protein is the main target for the development of neutralizing antibodies [105], being the main immunogen in all currently

approved

vaccines for
emergency
use by the
WHO [106].

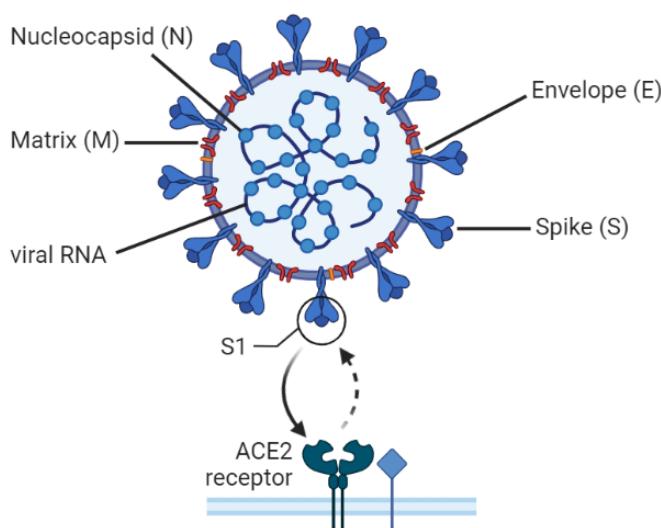


Figure 5. SARS-CoV-2 virion scheme. M, N, E and S proteins are represented. S1 subunit of the S protein interacts to the host cell receptor membrane protein ACE2 to bind and promote viral internalization

In this thesis the generation of HIV-1 Gag::eGFP VLPs functionalized with SARS-CoV-2 Spike protein was studied. First, the produced VLPs were characterized, its production was optimized, the bioprocess was scaled up and the obtained product was purified. Second, genetic modifications for the S protein were proposed and its immunogenicity was tested against sera extracted from COVID-19 patients. Third, different SGE platforms for the production of Gag::eGFP VLPs were generated using various approaches. Then, the cell lines were compared and their production process was enhanced in order to select the best SGE producers. Finally, a bioreactor of the best SGE cell line transfected with the S protein was performed.

References

- [1] Orensteina, W. A. & Ahmedb, R. Simply put: Vaccination saves lives. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 4031–4033 (2017)
- [2] Greenwood, B. The contribution of vaccination to global health: Past, present and future. *Philos. Trans. R. Soc. B Biol. Sci.* **369**, (2014)
- [3] Pulendran, B. & Ahmed, R. Immunological mechanisms of vaccination. *Nat. Immunol.* **12**, 509–517 (2011)
- [4] Organization, W. H. Vaccines and immunization. *World Health Organization* https://www.who.int/health-topics/vaccines-and-immunization#tab=tab_1 (2022)
- [5] Zhou, F. et al. Economic Evaluation of the Routine Childhood Immunization Program in the United States, 2009. *Pediatrics* **133**, 577–585 (2014)
- [6] Fontanet, A. & Cauchemez, S. COVID-19 herd immunity: where are we? *Nat. Rev. Immunol.* **20**, 583–584 (2020)
- [7] Dowdle, W. R. The principles of disease elimination and eradication. *Bull. World Health Organ.* **76 Suppl 2**, 22–5 (1998)
- [8] Henderson, D. A. The eradication of smallpox – An overview of the past, present, and future. *Vaccine* **29**, D7–D9 (2011)
- [9] Mariner, J. C. et al. Rinderpest Eradication: Appropriate Technology and Social Innovations. *Science/AAAS* **337**, 1309–1312 (2012)
- [10] Aylward, R. B. et al. Disease eradication as a public health strategy: a case study of poliomyelitis eradication. *Bull. World Health Organ.* **78**, 285–97 (2000)
- [11] Pollard, A. J. & Bijker, E. M. A guide to vaccinology: from basic principles to new developments. *Nat. Rev. Immunol.* **21**, 129–129 (2021)
- [12] Iwasaki, A. & Omer, S. B. Why and How Vaccines Work. *Cell* **183**, 290–295 (2020)

- [13] Chavda, V. P., Pandya, R. & Apostolopoulos, V. DNA vaccines for SARS-CoV-2: toward third-generation vaccination era. *Expert Rev. Vaccines* **20**, 1549–1560 (2021)
- [14] Plotkin, S. History of vaccination. *Proc. Natl. Acad. Sci.* **111**, 12283–12287 (2014)
- [15] Badgett, M. R., Auer, A., Carmichael, L. E., Parrish, C. R. & Bull, J. J. Evolutionary Dynamics of Viral Attenuation. *J. Virol.* **76**, 10524–10529 (2002)
- [16] Ada, G. Vaccines and Vaccination. *N. Engl. J. Med.* **345**, 1042–1053 (2001)
- [17] Ulmer, J. B., Valley, U. & Rappuoli, R. Vaccine manufacturing: challenges and solutions. *Nat. Biotechnol.* **24**, 1377–1383 (2006)
- [18] Slingluff, C. L. The Present and Future of Peptide Vaccines for Cancer. *Cancer J.* **17**, 343–350 (2011)
- [19] Josefsberg, J. O. & Buckland, B. Vaccine process technology. *Biotechnol. Bioeng.* **109**, 1443–1460 (2012)
- [20] Podda, A. The adjuvanted influenza vaccines with novel adjuvants: experience with the MF59-adjuvanted vaccine. *Vaccine* **19**, 2673–2680 (2001)
- [21] Fuenmayor, J., Gòdia, F. & Cervera, L. Production of virus-like particles for vaccines. *N. Biotechnol.* **39**, 174–180 (2017)
- [22] Emini, E. A. *et al.* Production and immunological analysis of recombinant hepatitis B vaccine. *J. Infect.* **13**, 3–9 (1986)
- [23] Nooraei, S. *et al.* Virus-like particles: preparation, immunogenicity and their roles as nanovaccines and drug nanocarriers. *J. Nanobiotechnology* **19**, 1–27 (2021)
- [24] Tariq, H., Batool, S., Asif, S., Ali, M. & Abbasi, B. H. Virus-Like Particles: Revolutionary Platforms for Developing Vaccines Against Emerging

- [25] Teijaro, J. R. & Farber, D. L. COVID-19 vaccines: modes of immune activation and future challenges. *Nat. Rev. Immunol.* **21**, 195–197 (2021)
- [26] Liu, M. A. DNA vaccines: a review. *J. Intern. Med.* **253**, 402–410 (2003)
- [27] Boix-Besora, A., Lorenzo, E., Lavado-García, J., Gòdia, F. & Cervera, L. Optimization, Production, Purification and Characterization of HIV-1 GAG-Based Virus-like Particles Functionalized with SARS-CoV-2. *Vaccines* **10**, 1–18 (2022)
- [28] Kushnir, N., Streatfield, S. J. & Yusibov, V. Virus-like particles as a highly efficient vaccine platform: Diversity of targets and production systems and advances in clinical development. *Vaccine* **31**, 58–83 (2012)
- [29] Roldão, A., Mellado, M. C. M., Castilho, L. R., Carrondo, M. J. & Alves, P. M. Virus-like particles in vaccine development. *Expert Rev. Vaccines* **9**, 1149–1176 (2010)
- [30] Lua, L. H. L. et al. Bioengineering virus-like particles as vaccines. *Biotechnol. Bioeng.* **111**, 425–440 (2014)
- [31] Zepeda-Cervantes, J., Ramírez-Jarquín, J. O. & Vaca, L. Interaction Between Virus-Like Particles (VLPs) and Pattern Recognition Receptors (PRRs) From Dendritic Cells (DCs): Toward Better Engineering of VLPs. *Front. Immunol.* **11**, (2020)
- [32] Crisci, E., Bárcena, J. & Montoya, M. Virus-like particle-based vaccines for animal viral infections. *Inmunología* **32**, 102–116 (2013)
- [33] Sandbrink, J. B. & Koblentz, G. D. Biosecurity risks associated with vaccine platform technologies. *Vaccine* **40**, 2514–2523 (2022)
- [34] Fontana, D., Kratje, R., Etcheverrigaray, M. & Prieto, C. Rabies virus-like particles expressed in HEK293 cells. *Vaccine* **32**, 2799–2804 (2014)
- [35] Chung, Y. H., Cai, H. & Steinmetz, N. F. Viral nanoparticles for drug delivery, imaging, immunotherapy, and theranostic applications. *Adv.*

Drug Deliv. Rev. **156**, 214–235 (2020)

- [36] Harper, D. M. *et al.* Efficacy of a bivalent L1 virus-like particle vaccine in prevention of infection with human papillomavirus types 16 and 18 in young women: a randomised controlled trial. *Lancet* **364**, 1757–1765 (2004)
- [37] Schiller, J. T. & Nardelli-Haefliger, D. Chapter 17: Second generation HPV vaccines to prevent cervical cancer. *Vaccine* **24**, S147–S153 (2006)
- [38] Cao, Y., Bing, Z., Guan, S., Zhang, Z. & Wang, X. Development of new hepatitis E vaccines. *Hum. Vaccin. Immunother.* **14**, 2254–2262 (2018)
- [39] Wu, T. *et al.* Hepatitis E vaccine development. *Hum. Vaccin. Immunother.* **8**, 823–827 (2012)
- [40] Laurens, M. B. RTS,S/AS01 vaccine (MosquirixTM): an overview. *Hum. Vaccin. Immunother.* **16**, 480–489 (2020)
- [41] Rodríguez-Limas, W. A., Sekar, K. & Tyo, K. E. Virus-like particles: the future of microbial factories and cell-free systems as platforms for vaccine development. *Curr. Opin. Biotechnol.* **24**, 1089–1093 (2013)
- [42] Hager, K. J. *et al.* Efficacy and Safety of a Recombinant Plant-Based Adjuvanted Covid-19 Vaccine. *N. Engl. J. Med.* **386**, 2084–2096 (2022)
- [43] Weiss, R. A. How Does HIV Cause AIDS? *Science/AAAS* **260**, 1273–1279 (1993)
- [44] Greene, W. C. A history of AIDS: Looking back to see ahead. *Eur. J. Immunol.* **37**, S94–S102 (2007)
- [45] Sharp, P. M. & Hahn, B. H. Origins of HIV and the AIDS Pandemic. *Cold Spring Harb. Perspect. Med.* **1**, a006841 (2011)
- [46] Organization, W. H. HIV/AIDS Global data. *World Health Organization* <https://www.who.int/news-room/fact-sheets/detail/hiv-aids> (2021)
- [47] Patel, P. *et al.* Estimating per-act HIV transmission risk: a systematic review. *AIDS* **28**, 1509–19 (2014)

-
- [48] Briggs, J. A. G. Structural organization of authentic, mature HIV-1 virions and cores. *EMBO J.* **22**, 1707–1715 (2003)
 - [49] Young, K. R., McBurney, S. P., Karkhanis, L. U. & Ross, T. M. Virus-like particles: Designing an effective AIDS vaccine. *Methods* **40**, 98–117 (2006)
 - [50] Xiao, Q., Guo, D. & Chen, S. Application of CRISPR/Cas9-Based Gene Editing in HIV-1/AIDS Therapy. *Front. Cell. Infect. Microbiol.* **9**, (2019)
 - [51] Caputi, M. The Regulation of HIV-1 mRNA Biogenesis. in *RNA Processing* (InTech, 2011). doi:10.5772/20899
 - [52] Votteler, J. & Schubert, U. Human Immunodeficiency Viruses: Molecular Biology. in *Encyclopedia of Virology* 517–525 (Elsevier, 2008). doi:10.1016/B978-012374410-4.00428-3
 - [53] Freed, E. O. HIV-1 assembly, release and maturation. *Nat. Rev. Microbiol.* **13**, 484–496 (2015)
 - [54] Broder, S. The development of antiretroviral therapy and its impact on the HIV-1/AIDS pandemic. *Antiviral Res.* **85**, 1–18 (2010)
 - [55] Koss, C. A. *et al.* HIV incidence after pre-exposure prophylaxis initiation among women and men at elevated HIV risk: A population-based study in rural Kenya and Uganda. *PLOS Med.* **18**, e1003492 (2021)
 - [56] Abbas, U. L., Glaubius, R., Mubayi, A., Hood, G. & Mellors, J. W. Antiretroviral Therapy and Pre-exposure Prophylaxis: Combined Impact on HIV Transmission and Drug Resistance in South Africa. *J. Infect. Dis.* **208**, 224–234 (2013)
 - [57] Walker, B. D. & Burton, D. R. Toward an AIDS Vaccine. *Science (80-.).* **320**, 760–764 (2008)
 - [58] Letvin, N. L. Progress and obstacles in the development of an AIDS vaccine. *Nat. Rev. Immunol.* **6**, 930–939 (2006)
 - [59] Picker, L. J., Hansen, S. G. & Lifson, J. D. New Paradigms for HIV/AIDS

- Vaccine Development. *Annu. Rev. Med.* **63**, 95–111 (2012)
- [60] Doan, L. X., Li, M., Chen, C. & Yao, Q. Virus-like particles as HIV-1 vaccines. *Rev. Med. Virol.* **15**, 75–88 (2005)
 - [61] Delchambre, M. *et al.* The GAG precursor of simian immunodeficiency virus assembles into virus-like particles. *EMBO J.* **8**, 2653–2660 (1989)
 - [62] Göttlinger, H. G. HIV-1 Gag: a Molecular Machine Driving Viral Particle Assembly and Release. *HIV Seq. Compend.* 2–28 (2001)
 - [63] Lavado-García, J. *et al.* Characterization of HIV-1 virus-like particles and determination of Gag stoichiometry for different production platforms. *Biotechnol. Bioeng.* **118**, 2660–2675 (2021)
 - [64] Cervera, L. *et al.* Generation of HIV-1 Gag VLPs by transient transfection of HEK 293 suspension cell cultures using an optimized animal-derived component free medium. *J. Biotechnol.* **166**, 152–165 (2013)
 - [65] Venereo-Sánchez, A. *et al.* Hemagglutinin and neuraminidase containing virus-like particles produced in HEK-293 suspension culture: An effective influenza vaccine candidate. *Vaccine* **34**, 3371–3380 (2016)
 - [66] Wurm, F. M. Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat. Biotechnol.* **22**, 1393–1398 (2004)
 - [67] Huh, S.-H. *et al.* Optimization of 25kDa linear polyethylenimine for efficient gene delivery. *Biologicals* **35**, 165–171 (2007)
 - [68] Salimzadeh, L., Jaberipour, M., Hosseini, A. & Ghaderi, A. Non-viral transfection methods optimized for gene delivery to a lung cancer cell. *Avicenna J. Med. Biotechnol.* **5**, 68–77 (2013)
 - [69] Cervera, L., Gutiérrez-Granados, S., Berrow, N. S., Segura, M. M. & Gòdia, F. Extended gene expression by medium exchange and repeated transient transfection for recombinant protein production enhancement. *Biotechnol. Bioeng.* **112**, 934–946 (2015)
 - [70] Kim, T. K. & Eberwine, J. H. Mammalian cell transfection: The present

and the future. *Anal. Bioanal. Chem.* **397**, 3173–3178 (2010)

- [71] Geisse, S. & Fux, C. Chapter 15 Recombinant Protein Production by Transient Gene Transfer into Mammalian Cells. in *Methods in Enzymology* vol. 463 223–238 (Elsevier Inc., 2009).
- [72] Hacker, D. L. & Balasubramanian, S. Recombinant protein production from stable mammalian cell lines and pools. *Curr. Opin. Struct. Biol.* **38**, 129–136 (2016)
- [73] González-Domínguez, I., Grimaldi, N., Cervera, L., Ventosa, N. & Gòdia, F. Impact of physicochemical properties of DNA/PEI complexes on transient transfection of mammalian cells. *N. Biotechnol.* **49**, 88–97 (2019)
- [74] Wurm, F. M. Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat. Biotechnol.* **22**, 1393–1398 (2004)
- [75] Ye, M., Wilhelm, M., Gentschev, I. & Szalay, A. A Modified Limiting Dilution Method for Monoclonal Stable Cell Line Selection Using a Real-Time Fluorescence Imaging System: A Practical Workflow and Advanced Applications. *Methods Protoc.* **4**, 16 (2021)
- [76] Priola, J. J. *et al.* High-throughput screening and selection of mammalian cells for enhanced protein production. *Biotechnol. J.* **11**, 853–865 (2016)
- [77] Würtele, H., Little, K. C. E. & Chartrand, P. Illegitimate DNA integration in mammalian cells. *Gene Ther.* **10**, 1791–1799 (2003)
- [78] Salman, H., Zbaida, D., Rabin, Y., Chatenay, D. & Elbaum, M. Kinetics and mechanism of DNA uptake into the cell nucleus. *Proc. Natl. Acad. Sci.* **98**, 7247–7252 (2001)
- [79] Büssow, K. Stable mammalian producer cell lines for structural biology. *Curr. Opin. Struct. Biol.* **32**, 81–90 (2015)
- [80] Mutskov, V. & Felsenfeld, G. Silencing of transgene transcription precedes methylation of promoter DNA and histone H3 lysine 9. *EMBO J.* **23**, 138–149 (2004)

- [81] Zhou, S. *et al.* Discovery of a stable expression hot spot in the genome of Chinese hamster ovary cells using lentivirus-based random integration. *Biotechnol. Biotechnol. Equip.* **33**, 605–612 (2019)
- [82] Hansen, J. *et al.* A large-scale, gene-driven mutagenesis approach for the functional analysis of the mouse genome. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 9918–9922 (2003)
- [83] Sadelain, M., Papapetrou, E. P. & Bushman, F. D. Safe harbours for the integration of new DNA in the human genome. *Nat. Rev. Cancer* **12**, 51–58 (2012)
- [84] Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823–6 (2013)
- [85] Aznauryan, E. *et al.* Discovery and validation of novel human genomic safe harbor sites for gene and cell therapies. *bioRxiv* 2021.03.04.433856 (2021)
- [86] Kandavelou, K. *et al.* Targeted manipulation of mammalian genomes using designed zinc finger nucleases. *Biochem. Biophys. Res. Commun.* **388**, 56–61 (2009)
- [87] Krämer, O., Klausing, S. & Noll, T. Methods in mammalian cell line engineering: From random mutagenesis to sequence-specific approaches. *Appl. Microbiol. Biotechnol.* **88**, 425–436 (2010)
- [88] Lieber, M. R., Ma, Y., Pannicke, U. & Schwarz, K. Mechanism and regulation of human non-homologous DNA end-joining. *Nat. Rev. Mol. Cell Biol.* **4**, 712–720 (2003)
- [89] Van Gent, D. C. & Van Der Burg, M. Non-homologous end-joining, a sticky affair. *Oncogene* **26**, 7731–7740 (2007)
- [90] Schmidt, F. & Grimm, D. CRISPR genome engineering and viral gene delivery: A case of mutual attraction. *Biotechnol. J.* **10**, 258–272 (2015)
- [91] LaFountaine, J. S., Fathe, K. & Smyth, H. D. C. Delivery and therapeutic applications of gene editing technologies ZFNs, TALENs, and

- [92] Hu, B., Guo, H., Zhou, P. & Shi, Z.-L. Characteristics of SARS-CoV-2 and COVID-19. *Nat. Rev. Microbiol.* **20**, 315–315 (2022)
- [93] Ali, I. & Alharbi, O. M. L. COVID-19: Disease, management, treatment, and social impact. *Sci. Total Environ.* **728**, 138861 (2020)
- [94] Çalıca Utku, A. *et al.* Main symptoms in patients presenting in the COVID-19 period. *Scott. Med. J.* **65**, 127–132 (2020)
- [95] Guan, D. *et al.* Global supply-chain effects of COVID-19 control measures. *Nat. Hum. Behav.* **4**, 577–587 (2020)
- [96] Singh, D. & Yi, S. V. On the origin and evolution of SARS-CoV-2. *Exp. Mol. Med.* **53**, 537–547 (2021)
- [97] Organization, W. H. WHO Coronavirus (COVID-19) Dashboard. *WHO Health Emergency Dashboard* <https://covid19.who.int/> (2022)
- [98] McNeill, V. F. Airborne Transmission of SARS-CoV-2: Evidence and Implications for Engineering Controls. *Annu. Rev. Chem. Biomol. Eng.* **13**, (2022)
- [99] Ashour, H. M., Elkhatib, W. F., Rahman, M. M. & Elshabrawy, H. A. Insights into the Recent 2019 Novel Coronavirus (SARS-CoV-2) in Light of Past Human Coronavirus Outbreaks. *Pathogens* **9**, 186 (2020)
- [100] Gorbatenko, A. E., Enjuanes, L., Ziebuhr, J. & Snijder, E. J. Nidovirales: Evolving the largest RNA virus genome. *Virus Res.* **117**, 17–37 (2006)
- [101] Jung, C. *et al.* Omicron: What Makes the Latest SARS-CoV-2 Variant of Concern So Concerning? *J. Virol.* **96**, (2022)
- [102] Kim, D. *et al.* The Architecture of SARS-CoV-2 Transcriptome. *Cell* **181**, 914–921 (2020)
- [103] Nal, B. *et al.* Differential maturation and subcellular localization of severe acute respiratory syndrome coronavirus surface proteins S, M and E. *J. Gen. Virol.* **86**, 1423–1434 (2005)

- [104] Yu, R. & Li, P. Screening of potential spike glycoprotein / ACE2 dual antagonists against COVID-19 in silico molecular docking. *J. Virol. Methods* **301**, 114424 (2022)
- [105] Weisblum, Y. *et al.* Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. *Elife* **9**, (2020)
- [106] Dolgin, E. Could computer models be the key to better COVID vaccines? *Nature* **604**, 22–25 (2022)

Objectives

The aim of this work is to increase the knowledge on the functionalization of Gag-based VLPs and the generation of HEK293 stable cell lines for their recombinant expression.

The specific objectives of this thesis are:

1. Functionalize HIV-1 Gag::eGFP VLPs with SARS-CoV-2 Spike protein.
 - a. Optimize the transient gene expression for the functionalization of Gag::eGFP VLPs with the spike protein from SARS-CoV-2 virus. Purify and characterize the produced vaccine candidate.
 - b. Study the immunogenicity of genetically engineered Spike proteins by their recognition by sera extracted from COVID-19 patients.
2. Generate and characterize stable cell lines for the production of HIV-1 Gag::eGFP VLPs using different techniques.
 - a. Generation of a stable cell line by illegitimate random integration.
 - b. Generation of a stable cell line by rational site-directed integration into a genomic safe harbor using CRISPR/Cas9 technology.
 - c. Generation of a recombinase-mediated cassette exchange (RMCE)-competent stable cell line by lentiviral transduction.
 - d. Selection of the best stable cell line in terms of growth kinetics, specific productivity and VLP quality. Optimize the culture media in order to enhance its production.
3. Define a bioprocess for the stable expression of Gag::eGFP VLPs and its functionalization with SARS-CoV-2 Spike protein.

Results

Chapter one

**Optimization, Production, Purification and Characterization of
HIV-1 Gag-Based Virus-Like Particles Functionalized with
SARS-CoV-2**

Published in Vaccines: DOI 10.3390/vaccines10020250

Abstract: Virus-like particles (VLPs) constitute a promising approach to recombinant vaccine development. They are robust, safe, versatile and highly immunogenic supra-molecular structures that closely mimic the native conformation of viruses without carrying their genetic material. HIV-1 Gag VLPs share similar characteristics with wild-type severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus, making them a suitable platform for the expression of its spike membrane protein to generate a potential vaccine candidate for COVID-19. This work proposes a methodology for the generation of SARS-CoV-2 VLPs by their co-expression with HIV-1 Gag protein. We achieved VLP functionalization with coronavirus spike protein, optimized its expression using a design of experiments (DoE). We also performed the bioprocess at a bioreactor scale followed by a scalable downstream purification process consisting of two clarifications, an ion exchange and size-exclusion chromatography. The whole production process is conceived to enhance its transferability at current good manufacturing practice (cGMP) industrial scale manufacturing. Moreover, the approach proposed could be expanded to produce additional Gag-based VLPs against different diseases or COVID-19 variants.

Keywords: VLP vaccines; HIV-1; SARS-CoV-2; COVID-19; transient transfection; HEK293; design of experiments; bioprocess; downstream process

1. Introduction

COVID-19 is a disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). It emerged in December 2019 in Wuhan and since then has spread around the globe causing a pandemic that had devastating health and economic consequences worldwide [1,2]. An enormous effort made by the scientific community resulted in more than 300 new vaccine candidates in less than a year since the outbreak, some of them being approved for emergency use [3], as well as the development of diagnosis methods for its detection [4] and treatment [5]. As of November 2021, more than 7 billion doses had been administrated [6], with an associated significant reduction of transmission and mortality among vaccinated populations [7]. Vaccination rollout offers a promising avenue for the pandemic and sanitary restrictions to come to an end. However, there are still some questions left to answer, like how long the immune memory lasts, the protective effect that current approved vaccines generate against emerging SARS-CoV-2 variants, or if it is possible to generate fully prophylactic vaccines against this new coronavirus [8].

FDA- and EMA-approved COVID-19 vaccines can be classified into mRNA, adenovirus-based or recombinant. The first group delivers mRNA into dendritic cells (DCs) using a lipid nanoparticle (LNP) as a carrier. The second one delivers DNA into DCs using a non-replicating recombinant adenovirus vector. Both strategies result in DCs producing the genetically encoded SARS-CoV-2 spike (S) surface glycoprotein and presenting it on their own membrane, where it is then recognized by the immune system cells [9]. The third strategy uses a saponin-based nanoparticle to present a recombinantly produced and purified spike glycoprotein, as well [10].

Virus-like particles (VLPs) are a promising, robust, safe, versatile and highly immunogenic approach that can be used to produce novel vaccines for emerging pandemics or diseases, like COVID-19. VLPs are supramolecular structures that closely mimic the native conformation of viruses without carrying genetic material (DNA or RNA), so they are unable to infect, replicate or integrate. They are generated by taking advantage of the intrinsic ability of some viral proteins to self-assemble when expressed in heterologous production platforms. Two

VLP-based vaccines have been proven successful and licensed for HPB and HPV [11–13]. VLPs can be functionalized to present pathogenic epitopes to generate immunity against diseases like dengue, influenza, etc. HIV-1 Gag VLPs have shown great potential for this purpose, since they are composed by a core of Gag molecules surrounded by a lipid bilayer, a membrane that can be further functionalized [14]. When administrated, due to their tridimensional configuration, VLPs drain and traffic within the immune system, interacting with cells such as DCs, B cells, T cells and macrophages [15,16]. Several proteins present at the VLP membrane interact with DC pattern-recognition receptors leading to a strong adaptative immune response, while multimeric epitopes promote the cross-linking of B cell receptors to induce antibody production [15,16]. Consequently, VLPs had been shown to induce potent humoral and cellular immune responses, and, although the use of adjuvants improves VLP-based vaccines' immunogenicity, their nature make adjuvant co-administration optional [15]. Therefore, VLP vaccination constitutes a promising approach compared to currently available vaccine technologies: VLPs are safer to manufacture and administrate than inactivated or attenuated vaccines due to their lack of viral genetical material, while they also provide a more potent and effective immune response compared to proteic or subunit vaccines since VLPs present conformationally authentic viral epitopes [17].

Gag VLPs share a similar particle diameter (~145 nm) with wild-type SARS-CoV-2 virus (~80 nm), making them a suitable platform for the expression of its epitopes in order to generate a new vaccine candidate for COVID-19 [14,18]. For this purpose, the SARS-CoV-2 spike (S) protein is a promising lead, since is the major structural protein anchored at the exterior of the membrane of native viruses, carries B cell and T cell epitopes and is the main target for neutralizing antibodies generated from natural infection that protect against viral infection and currently approved COVID-19 vaccines [19,20]. S protein monomers are 180 kDa and assemble to form trimeric units at the surface of native virions, giving them their characteristic crown-shape [21]. They contain a variable receptor-binding domain (RBD), responsible for binding to angiotensin-converting enzyme 2 receptor (ACE-2), facilitating viral entry into target cells [21]. As the disruption of the RBD-ACE2 interaction can block

SARS-CoV-2 cell entry [22], most of the reported neutralizing antibodies against SARS-CoV-2 bind to the RBD [23].

VLPs can be produced in prokaryotic and eukaryotic heterologous expression systems depending on their nature and use. Mammalian cells constitute an attractive production platform for enveloped or multimeric VLPs due to their capacity to perform complex post-translational modifications. There are several mammalian cell platforms in which VLPs can be produced, such as HeLa, Vero, CAP, CHO or HEK293 [13]. HEK293 can be cultured in suspension in bioreactors using chemically defined media free of animal components and can also be easily transfected. This makes them a good choice to produce Gag-based COVID-19 VLPs in large-scale bioreactors, in order to satisfy the needs for pre-clinical trials, clinical trials, and eventually large-scale production for its manufacture. For this purpose, a robust and scalable downstream process (DSP) needs to be implemented in order to obtain a high-purity vaccine product in its final buffer formulation. Finally, for the initial steps of proof of concept, HEK 293 cells transient transfection methodologies have been well established to provide a robust and fast approach to the generation of VLPs for testing, of special relevance when developing and comparing several candidates.

In the last decades and especially after the COVID-19 outbreak, different published works have focused on the generation of SARS-CoV and SARS-CoV-2 VLPs by the co-expression of the coronavirus S, M and E proteins [24]. This work focuses on the production, purification and characterization of a potential COVID-19 vaccine candidate, based on HIV-1 Gag-based SARS-CoV-2 spike VLPs (from now on S-VLPs), a never-before reported approach to our knowledge. S-VLP production scale-up and its DSP have been achieved by HEK293 transient transfection in a 1 L bioreactor and a purification process consisting of two clarification steps, an ion-exchange affinity step and a size-exclusion polishing and buffer exchange step. The production process and the obtained S-VLPs have been studied and characterized in this work.

2. Materials and Methods

2.1. Cell Line, Media and Culture Conditions

The serum-free suspension-adapted HEK293 cell line (HEK293SF-3F6) was used, kindly provided by Dr. Amine Kamen from the Biotechnology Research Institute at the National Research Council of Canada (Montreal, Canada) and McGill University. This cell line was derived from a current good manufacturing practice (cGMP) master cell bank available for manufacturing of clinical material.

The medium used for HEK293 cellular growth was the chemically defined and free from animal components HyCell TransFx-H from HyClone (GE Healthcare, Chicago, IL, USA) supplemented with 4 mM GlutaMAX (Gibco, Thermo Fisher Scientific, Waltham, MA, USA) and 0.1% Pluronic F-68 Non-ionic Surfactant (Gibco, Thermo Fisher Scientific, Waltham, MA, USA).

Suspension cell cultures were maintained routinely in exponential growth phase in 125mL or 1L disposable polycarbonate Erlenmeyer flasks with a vent cap (Corning, Tewksbury, MA, USA) in a LT-X Kuhner shaker (LT-X Kuhner, Birsfelden, Switzerland) shaking at 130 rpm, at 37 °C, 5% CO₂ and 85% RH. Cell counts and viability determinations were performed using the NucleoCounter NC-3000 automatic cell counter (Chemometec, Lillerød, Denmark) following the manufacturer's instructions.

2.2. Plasmids and Transfection

2.2.1. Plasmid Expression Vectors

The pGag::eGFP plasmid codes for a codon-optimized Rev-independent HIV-1 Gag protein fused in frame to the enhanced GFP driven by the CMV enhancer and promoter. The plasmid from the NIH AIDS Reagent Program (Cat 11468) (Hermida-Matsumoto and Resh, 2000) was constructed by cloning the Gag sequence from pCMV55M1-10 (Schwartz et al., 1992) into the pEGFP-N1 plasmid (Clontech, Palo Alto, CA, USA).

The pSpike plasmid codes for a mammalian cell codon optimized nucleotide sequence coding for the spike protein of SARS-CoV-2 driven by the CAG enhancer and β-actin promoter. It was produced under HHSN272201400008C

and obtained through BEI Resources, NIAID, NIH: Vector pCAGGS Containing the SARS-Related Coronavirus 2, Wuhan-Hu-1 spike Glycoprotein Gene, NR-52310.

pMock plasmid does not have any mammalian promoter or coding DNA sequence (CDS). It was constructed by the ligation of the pGag::eGFP backbone.

2.2.2. Plasmid Amplification and Purification

Plasmids were amplified in *Escherichia coli* DH5 α strain grown in LB medium (Conda, Madrid, Spain) supplemented with kanamycin (10 μ g/mL, Sigma, St. Louis, MO, USA) or ampicillin (100 μ g/mL, Sigma, St. Louis, MO, USA) depending on the *E. coli* antibiotic resistance present on each plasmid. Plasmid purification was carried out using the Endofree Plasmid Mega kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions.

2.2.3. PEI-Mediated Transient Transfection

Exponentially growing HEK293 cells were passaged in order to have a cell density of $2 \cdot 10^6$ cells/mL at transfection time. PEIpro (Polyplus-transfection SA, Illkirch-Graffenstaden, France) was used as a transfection reagent. PEI-DNA complexes were formed under sterile conditions, by adding PEI to a plasmid DNA mixture diluted for a total DNA concentration of 1 μ g/mL in fresh culture media (10% of the total volume of cell culture to be transfected). The mixture was incubated for 15 min at RT and then added to cell culture. The ratio between plasmids and transfection reagent was optimized using a Box–Behnken design of experiments and described in the next section.

2.3. Box–Behnken Design

A Box–Behnken design was used in order to define the optimal concentration for three independent variables in the cell transfection step: pGag::eGFP, pSpike and PEI. These variables were screened at three levels: a low level, coded as -1; a medium level, coded as 0; and a high level coded as +1, as indicated in Table 1.

Table 1. Box–Behnken experimental design: code levels, matrix design, response and regression coefficients.

| Independent variables | | | Coding levels | | |
|-----------------------|------------------------------|--------------------------|-----------------------|---|-----|
| | | | -1 | 0 | 1 |
| pGag::eGFP (µg/mL) | | | 0.2 | 0.8 | 1.4 |
| pSpike (µg/mL) | | | 0.2 | 0.8 | 1.4 |
| PEI (µg/mL) | | | 2 | 3.5 | 5 |
| Number | pGag::eGFP (X ₁) | pSpike (X ₂) | PEI (X ₃) | Double-Positive Population ^a (%) | |
| 1 | 0 | 0 | 0 | 46.9 | |
| 2 | 1 | 0 | 1 | 43.1 | |
| 3 | 0 | 0 | 0 | 47.2 | |
| 4 | -1 | -1 | 0 | 36.5 | |
| 5 | 0 | -1 | 1 | 44.9 | |
| 6 | 1 | -1 | 0 | 28.3 | |
| 7 | -1 | 0 | 1 | 42.7 | |
| 8 | 0 | 0 | 0 | 47.8 | |
| 9 | 0 | 1 | -1 | 51.3 | |
| 10 | -1 | 1 | 0 | 50.1 | |
| 11 | -1 | 0 | -1 | 56.4 | |
| 12 | 0 | -1 | -1 | 46.4 | |
| 13 | 1 | 1 | 0 | 37.2 | |
| 14 | 0 | 1 | 1 | 50.3 | |
| 15 | 1 | 0 | -1 | 39.0 | |

| Factor | Coefficient | f Value | p Value |
|----------------|-------------|---------|---------|
| Intercept | 47.3 | | |
| A-pGag::eGFP | -4.7625 | 33.078 | 0.0022 |
| B-pSpike | 4.1 | 24.516 | 0.0043 |
| C-PEI | -1.5125 | 3.336 | 0.1273 |
| AB | -1.175 | 1.007 | 0.3617 |
| AC | 4.45 | 14.440 | 0.0126 |
| BC | 0.125 | 0.011 | 0.9191 |
| A ² | -6.1 | 25.046 | 0.0041 |
| B ² | -3.175 | 6.785 | 0.0480 |
| C ² | 4.1 | 11.315 | 0.0200 |

| | DF | SS | MS | F value | p-value |
|-------|----|--------|-------|---------|---------|
| Model | 9 | 668.29 | 74.25 | 13.536 | 0.0052 |
| Error | 2 | 0.42 | 0.21 | | |

Abbreviations: DF degree of freedom, SS sum of squares, MS mean square. ^a Response is ICC-stained double-positive population at 72 h post transfection.

Box–Behnken experimental results were fitted to a second-order polynomial equation described below by non-linear regression analysis:

$$Y = \beta_0 + \sum \beta_i X_i + \sum \beta_{ii} X_i^2 + \sum \beta_{ij} X_i X_j \quad (1)$$

Where Y is the response (in this work, the percentage of cells expressing simultaneously Gag::eGFP and spike at 72 hpt); β_0 is the offset term; β_i is the linear coefficient; β_{ii} is the quadratic coefficient; β_{ij} the interaction coefficient, and X_i and X_j are the independent variables (pGag::eGFP, pSpike and PEI). The equation was used to predict the concentration of the independent variables in order to maximize the desired response. Three-dimensional response surface plots were generated using Design Expert version V8.0.6 software (Stat-Ease Inc., Minneapolis, MN, USA). Statistical analyses of the model were performed using Design Expert. The coefficient values corresponding to the generated response model are shown at Table 1.

Control groups to be transfected with just one plasmid coding for a protein, were co-transfected with pMock plasmid in order to deliver the same gene copies of the protein being expressed in the other conditions. Transfections associated with Box–Behnken optimization studies, validation and bioreactor production were carried out following the later described plasmid and PEI concentrations. Expression was analyzed at 0, 24, 48 and 72 hpt.

2.4. Stirred Tank Reactor (STR) Bioprocess

A BioStat B Plus bioreactor (Sartorius Stedim Biotech, Göttingen, Germany) equipped with a 3-blade segment dual impeller with UP-DP configuration [25] was used for HEK293 cell cultivation and production. The agitation was set at 200 rpm; the temperature was set at 37 °C, and the pH was set at 7.1, controlled with CO₂ and NaHCO₃ (7.5% w/v). Dissolved oxygen was controlled at 40% of air saturation by supplementing air by sparger at a constant flow of 0.1 L/min and additional pure oxygen when needed. HEK293 growing exponentially in disposable polycarbonate 1 L shake flasks (Corning, Tewksbury, MA, USA) were used to seed the bioreactor at 0.5·10⁶ cells/mL in 1 L of working volume.

2.5. Sucrose Cushion Small-Scale Purification

Culture harvests were performed at 72 hpt and centrifuged at $10,000 \times g$ for 10 min, and the supernatant was stored at -80°C for further analysis or stored at 4°C for its purification in less than 24 h. The supernatants containing VLPs were placed on a 30% (w/v) sucrose cushion for ultracentrifugation at 31,000 rpm for 2 h at 4°C . The supernatant was carefully discarded, and pellets were resuspended and placed on a new sucrose cushion for a second ultracentrifugation following the same protocol. The pellets were collected and resuspended in PBS.

2.6. Downstream Processing

2.6.1. Cell Harvest and Culture Supernatant Clarification

After 2 h of sedimentation, the culture medium from transfected cells was subjected to primary clarification using Supracap 50 V100 depth filter capsules (Pall Corporation, Port Washington, NY, USA) to remove cellular debris and other contaminants. For secondary clarification, the Supor EAV—Mini Kleenpak 20 filter capsules (Pall Corporation, Port Washington, NY, USA) was used. For both clarifications, a K2Ri pump (Repligen, Waltham, MA, USA) with MasterFlex 96410-13 silicon tubes (Cole-Parmer, Vernon Hills, IL, USA) connected to the filter inlet and outlet; and a pressure sensor (Cole-Parmer, Vernon Hills, IL, USA) connected to the filter inlet. The turbidity of the clarification samples was measured using a portable Eutech TN-100 turbidimeter (Thermo Fisher Scientific, Waltham, MA, USA).

2.6.2. Ion-exchange chromatography (IEX)

A prepacked 0.86 mL Mustang Q XT Acrodisc column (Pall Corporation, Port Washington, NY, USA) was used to capture the S-VLPs from the secondary clarification. Before loading, the column was pre-equilibrated with 5 column volumes (CV) of 5% buffer B (50 mM HEPES, 2M NaCl, pH = 7.2: Buffer B). The sample was directly loaded into the column via the sample pump. After sample application, the column was washed with 5 CV of buffer B at 5%. Elution was achieved by a salt step gradient consisting of 20 CV of 15%, 35%, 45% and 65% of buffer B (300 mM NaCl, 700 mM NaCl, 900 mM NaCl and 1300 mM NaCl). Solutions were filtered using 0.22 μm filters. Chromatographic runs were

performed with a flow rate of $1 \text{ mL}\cdot\text{min}^{-1}$, except for the sample application ($10 \text{ mL}\cdot\text{min}^{-1}$). Fractions of 1 mL were collected and pooled according to the chromatograms.

2.6.3. Size-Exclusion Chromatography

The collected peak containing the desired product from the IEX was loaded into a sepharose 4 Fast Flow (GE Healthcare, Chicago, IL, USA) in-house packed XK 16/40 desalting column of 48 mL . A column performance test with 1% acetone confirmed the correct values of asymmetry 10% and height equivalent to a theoretical plate (HETP). The column was pre-equilibrated with 5 CV of the formulation buffer ($20 \text{ mM NaH}_2\text{PO}_4$, 50 mM NaCl , 2 mM MgCl_2 , 2% sucrose, pH 7.5). Subsequently, the sample was injected onto the column via its sample pump. Elution was achieved with an isocratic elution (0–100%) of 2 CV of the formulation buffer. The column was sanitized with 5 CV of 0.5M NaOH. The chromatographic run was performed at a $2 \text{ mL}\cdot\text{min}^{-1}$ flow rate. Fractions of 1 mL were collected and pooled according to the chromatograms.

2.7. Immunocytochemistry Staining for Flow Citometry and Confocal Microscopy

For IF-ICC staining, cells were centrifuged 5 min at $300\times g$ and rinsed with staining solution (1.5% (v/v) fetal bovine serum (FBS) 1X phosphate-buffered saline (PBS)) before primary antibody incubation for 20 min at $4 \text{ }^\circ\text{C}$ in the dark. After rinsing twice, cells were incubated with the corresponding secondary antibody for 20 min at $4 \text{ }^\circ\text{C}$. After IF-ICC staining, fixation was performed using 2% (v/v) formaldehyde 1X PBS for 10 min at RT. Cells were resuspended in staining solution and stored at $4 \text{ }^\circ\text{C}$ prior to analysis.

Primary human anti-SARS-CoV-2 spike glycoprotein RBD domain antibody (ab272854, AbCam, Cambridge, UK) was diluted 1:1000. The secondary antibody used for flow citometry analysis was an anti-human IgG (H+L) coupled with CyTM5, produced in donkey (709-175-149, Jackson ImmunoResearch, West Grove, PA, USA), diluted 1:400. The secondary antibody used for confocal microscopy imaging was an anti-human IgG (H+L) coupled with Alexa Fluor 568,

produced in goat (#A-21090, Thermo Fisher Scientific, Waltham, MA, USA), diluted 1:400. All IF-ICC antibodies were diluted using staining solution.

2.7.1. Flow Cytometry

The transfected cellular populations of previously IF-ICC stained cells were assessed by flow cytometry using a BD FACS Canto flow cytometer (BD BioSciences, San Jose, CA, USA), at Servei de Cultius Cel·lulars, Producció d'Anticossos i Citometria (Universitat Autònoma de Barcelona, Bellaterra, Catalonia, Spain).

2.7.2. Confocal Microscopy

The imaging of previously IF-ICC-stained cells was performed using Leica TCS SP5 confocal fluorescence microscope (Leica Microsystems, Wetzlar, Germany) at Servei de Microscòpia (Universitat Autònoma de Barcelona, Bellaterra, Catalonia, Spain). Prior to visualization, cells were treated with 0.1% (v/v) of Hoechst 33342 (Thermo Fisher Scientific, Waltham, MA, USA) and 0.1% (v/v) of CellMask Deep Red (Thermo Fisher Scientific, Waltham, MA, USA) in order to stain cell nuclei and lipid membranes, respectively. Samples were placed in 35 mm glass bottom Petri dishes with 14 mm microwells (MatTek Corporation, Ashland, MA, USA) prior to their visualization under the microscope. 3D images were generated and analyzed using Imaris software (Bitplane, Oxford Instruments, Zurich, Switzerland).

2.8. Transmission Electron Microscopy

TEM analyses were performed at Servei de Microscòpia (Universitat Autònoma de Barcelona, Bellaterra, Catalonia, Spain). Samples were visualized in a JEOL 2011 transmission electron microscope (Jeol, Tokio, Japan) operating at an accelerating voltage of 200 kV. Electron micrographs were recorded with the Digital Micrograph software package (Gatan, Pleasanton, CA, USA). Images were recorded by a Gatan US4000 (Gatan, Pleasanton, CA, USA) cooled charge-coupled device (CCD) camera.

2.8.1. Transmission Electron Microscopy: Negative Staining

For negative staining, samples were prepared by means of the air-dried method. Briefly, an aliquot of purified VLPs was absorbed by flotation onto freshly glow discharged 400 mesh carbon film copper grids (22-1MC040-100, MicrotоНano, Haarlem, Netherlands). After standing for 1 min at RT, excess sample was drained carefully off the grid using Whatman filter paper, Grade 1 (WHA1001325, Merck, Kenilworth, NJ, USA). Samples were then stained with 5 μ L of uranyl acetate (2%) by incubation for 1 min at RT. The excess uranyl acetate was drained off as previously described.

2.8.2. Transmission Electron Microscopy: Immunogold Labeling

For immunogold labeling, 8 μ L of purified VLPs were loaded onto copper grids as previously described. After absorption, two wash cycles were performed. Each wash cycle consisted of adding by flotation 2% (w/v) BSA in PBS and removing the excess sample, followed by the addition of 1X PBS at RT. Then, primary human anti-SARS-CoV-2 spike glycoprotein antibody (ab272854, AbCam, Cambridge, UK) diluted 1:50 was added, and the grids were incubated for 1 h at RT. Following three wash cycles, grids were incubated with 6 nm gold-conjugated anti-human IgG (109-195-088, Jackson ImmunoResearch, West Grove, PA, USA) diluted 1:20 for 1 h at RT. After three wash cycles, grids were stained with uranyl acetate as mentioned before.

2.9. Nanoparticle Tracking Analysis

NTA-based Gag::eGFP VLP quantification and characterization was performed using a NanoSight®NS300 (Nanosight Ltd., Amesbury, UK) equipped with a blue filter module (488 nm) and a neutral filter at the Soft Material Service of the Institut de Ciència de Materials de Barcelona (ICMAB-CSIC, Bellaterra, Catalonia, Spain). Samples were previously diluted to a concentration of approximately 10^8 particles/mL. Sample injection was performed using a pump in order to improve the robustness of the measurement by continuous addition, and to minimize the photobleaching effect due to fluorescence depletion over time. 60 second videos were recorded at RT and analyzed with the NTA 3.4 software (Malvern Panalytical, Malvern, UK). Tracked particles size was determined from its Brownian motion. Three independent experimental replicas were carried out

for each sample. Camera level and detection threshold were manually adjusted for each replica.

2.10. Total Protein and dsDNA Quantification

A BCA Protein Assay (#23225, Thermo Fisher Scientific, Waltham, MA, USA) was performed following manufacturer's instructions using the provided BSA as standard. Colorimetric absorbance at 562 nm was read on a Multilabel Plate Reader Victor3 (Perkin Elmer, Waltham, MA, USA).

A Quant-iT PicoGreen dsDNA Assay Kit (#P11496, Thermo Fisher Scientific, Waltham, MA, USA) was performed following the manufacturer's instructions using the provided λ DNA as standard. Fluorescence ($\lambda_{\text{ex}} = 488 \text{ nm}$, $\lambda_{\text{em}} = 520 \text{ nm}$) was read on a Multilabel Plate Reader Victor3 (Perkin Elmer, Waltham, MA, USA). The fluorescence value of the reagent blank was subtracted for each sample before calculating the dsDNA concentration using the generated standard curve.

2.11. Western Blot and SDS-PAGE

Samples were mixed with 4x Laemili Buffer (#1610747, Bio-Rad, Hercules, CA, USA) and 1.4 M DTT (10708984001, Merck, Kenilworth, NJ, USA) to a final concentration of 1% v/v. Each sample was incubated at 96 °C for 20 min and stored at 4 °C until gelled. Precision Plus Protein WesternC (#1610376, Bio-Rad, Hercules, CA, USA) was used as molecular weight standard. A total of 20 μL of sample per lane was loaded and ran on precast SDS-polyacrylamide (4–20%) gel electrophoresis (#4561093, Bio-Rad, Hercules, CA, USA) at 200V, 400mA for 45 min. Running buffer used was Tris/Glycine/SDS (25 mM Tris, 192 mM glycine, 0.1% SDS, pH 8.3) (#1610772, Bio-Rad, Hercules, CA, USA).

For SDS-PAGE, proteins were stained with Coomassie Brilliant Blue EZBlueTM Gel Staining Reagent (G1041, Sigma Aldrich, St. Louis, MO, USA).

For Western blot, electrophoresis gel was transferred onto a polyvinylidene difluoride membrane for 7 min using the Trans-Blot Turbo Transfer System (#17001918, Bio-Rad, Hercules, CA, USA) following the manufacturer's instructions. Transferred membranes were then blocked with 5% (w/v) nonfat dry

milk in wash buffer (1× PBS 0.1% Tween-20). All the incubations and wash steps between incubations were performed at 40 rpm in a Polymax 1040 rocker shaker (Polymax 1040, Heidolph Instruments, Schwabach, Germany). For anti-HIV-1 Gag WB, blocking was performed overnight at 4 °C and incubated 2 h at RT with primary antibody. For SARS-CoV-2 spike WB, blocking was performed 40 min at RT, and it was incubated overnight at 4 °C with primary antibody. Primary antibodies used were rabbit polyclonal Anti-SARS-CoV-2 spike glycoprotein antibody (ab272504, AbCam, Cambridge, UK) and mouse monoclonal antibody to HIV-1 p24 (A2-851-500, Icosagen, Tartu, Estonia), both diluted 1:1000 in wash buffer. After primary incubation, membranes were incubated using anti-mouse IgG coupled with alkaline phosphatase antibody produced in goat (A3562, Merck, Kenilworth, NJ, USA) or anti-rabbit IgG coupled with alkaline phosphatase antibody produced in goat (A9919, Merck, Kenilworth, NJ, USA), as required, in wash buffer for 1 h at RT. Protein bands were visualized using NBT-BCIP solution (#1706432, Bio-Rad, Hercules, CA, USA) after 2–3 min incubation. Membranes were let to dry and then scanned and analyzed using the software ImageJ2 Fiji (National Institutes of Health, Bethesda, MD, USA).

2.12. Dot Blot

Samples were charged into Bio-Dot Apparatus (#1706545, Bio-Rad, Hercules, CA, USA) while a low vacuum was applied. Nitrocellulose membrane (#88018, Thermo Fisher Scientific, Waltham, MA, USA) was placed at the top of humidified filter paper. Once samples were transferred, membrane was incubated with anti-SARS-CoV-2 spike glycoprotein S2 monoclonal antibody (Ab281312, AbCam, Cambridge, UK) and an anti-rabbit secondary antibody (A9919, Merck, Kenilworth, NJ, USA) following the same procedure previously mentioned for Western blot. Once dried, membranes were scanned, and the pixel density for each loaded sample was analyzed using software ImageJ2 Fiji (National Institutes of Health, Bethesda, MD, USA). The standard used for quantification was a recombinant human coronavirus SARS-CoV-2 spike glycoprotein S2 subunit (Ab272106, AbCam, Cambridge, UK).

3. Results

3.1. SARS-CoV-2 Spike Protein Co-Expression and Localization

Chimeric VLPs were produced in HEK293 cells growing in suspension culture in a chemically defined and animal-component-free media, by transient transfection as a proof of concept. To produce SARS-CoV-2 spike Gag::eGFP VLPs (S-VLPs), cells were co-transfected with plasmids pGag::eGFP and pSpike, using PEI as the transfection reagent. A control cell group was co-transfected with pGag::eGFP and an empty plasmid to generate Gag::eGFP VLPs (from now on, G-VLPs). In order to easily track the HIV-1 Gag polyprotein expression and characterize S- and G-VLPs, Gag was fused in frame with eGFP, as previously reported [26]. S-VLP producer cells showed no significant difference in viable cell density compared with G-VLP producer cells control group (data not shown). Viabilities between 70–80% at 72 hpt were in agreement with values previously observed in PEI-mediated Gag-based VLP productions [26,27]. These results show that the expression of the CoV-2 spike protein does not have a toxic effect on the HEK293 platform used. Otherwise, low viabilities (<70%) could be indicative of toxicity caused by the spike protein's incomplete maturation through the secretory pathway [28].

Cells were analyzed by confocal microscopy in order to track Gag::eGFP and spike protein localizations at the time of harvest. As can be seen in Figure 1A, the green fluorescence channel shows Gag::eGFP along the cytoplasm to the vicinity of the plasmatic membrane. This corresponds to what was already known about Gag polyprotein maturation, which occurs at the cytoplasm until it reaches the plasmatic membrane surroundings, where budding occurs in order to generate the Gag-based VLPs [29]. To determine spike localization, cells were immunostained using an anti-S primary antibody and a fluorochrome-conjugated secondary antibody. By staining the lipid membrane with CellMask, the strong co-localization of the S protein (red) and cell membrane (grey) was observed (Figure 1B), as well as the co-localization of the S protein (red) and Gag::eGFP (green) in membrane (Figures 1A,C). These results suggest that the expressed S protein could be dragged and incorporated at the surface of the produced

S-VLPs, as they bud from the plasmatic membrane [29], where we observe that spike is present.

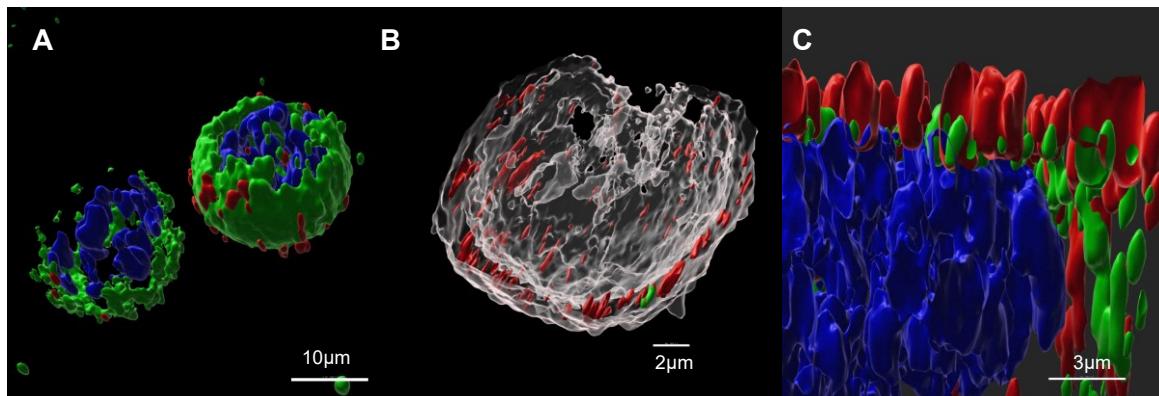


Figure 1. Spike protein cellular localization. Cells were treated with an anti-SARS-CoV-2 S1 spike subunit monoclonal antibody, followed by an Cy5 anti-human secondary incubation. **(A)**: Co-localization of spike (red) and Gag (green) can be observed at cells' external boundary. Cell nuclei were stained and shown as blue. **(B)**: By staining the lipid membrane with CellMask, the strong co-localization of S protein (red) and cell membrane (grey) was observed. **(C)**: Cross-sectional observation of the plasmatic membrane, where the co-localization of spike (red) and Gag (green) can be observed. Cell nuclei were stained and shown as blue.

3.2. Characterization of the Produced S-VLPs

Small-scale production, followed by sucrose cushion purification, was performed to study if the produced S-VLPs incorporate the S protein. The mode diameter of the purified chimeric S-VLPs was 134.9 ± 1.2 nm, as measured by nanoparticle-tracking analysis (NTA) (Figure 2A). Both purified S-VLPs and G-VLPs showed HIV-1 Gag bands by Western blot, while only S-VLPs showed intense SARS-CoV-2 spike protein bands, as shown in Figure 2B.

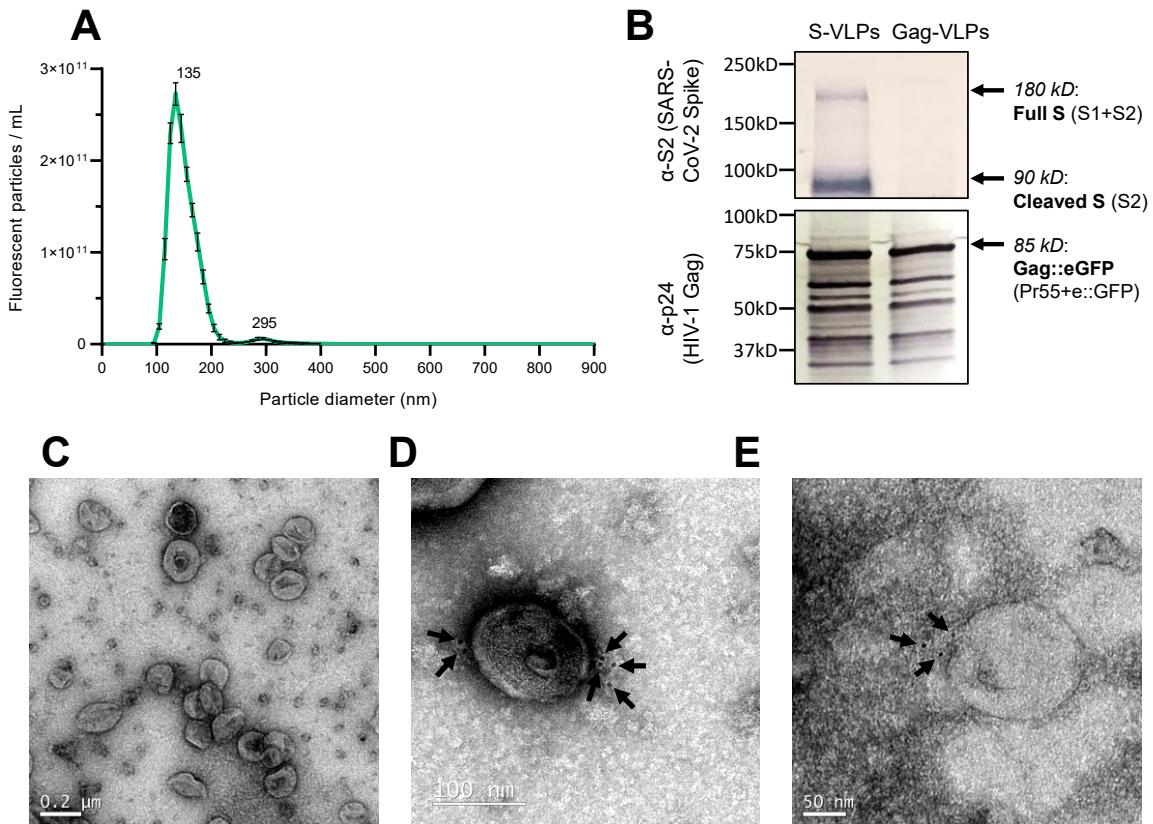


Figure 2. Characterization of the produced S-VLPs. **(A)**: Particle-size distribution of the purified S-VLPs, analyzed by NTA. **(B)**: Western blot of purified S-VLPs and G-VLPs. Top: Membrane was treated with an anti-SARS-CoV-2 S2 spike subunit polyclonal antibody, followed by a goat anti-rabbit secondary incubation. Bottom: Membrane was treated with an anti-HIV-1 p24 monoclonal antibody, followed by goat anti-mouse secondary incubation. Both S-VLPs and G-VLPs showed HIV-1 Gag bands by Western blot, while only S-VLPs showed intense SARS-CoV-2 S protein bands. **(C–E)**: Electron microscopy images of the purified VLPs. **(C)**: CryoTEM of untreated purified S-VLPs. **(D)**: Gold-immunolabeled TEM. **(E)**: Gold-immunolabeled CryoTEM. Grids were treated with an anti-S protein monoclonal antibody, followed by incubation with goat anti-human secondary coupled with 6 nm gold particles. Gold-immunolabeled S-VLPs showed S protein localization (arrows) on the surface of the chimeric VLPs.

The purified VLP concentrates contained spherical enveloped particles with no significant structural differences compared to non-functionalized G-VLPs, as seen under EM by negative staining (Figure 2C). Immunogold labeling using anti-SARS-CoV-2 spike protein primary antibody and 6 nm gold-labeled secondary antibody showed S protein localization on the surface of the chimeric VLPs (Figures 2D,E). These EM and Western blot results confirm the functionalization of the Gag-based produced S-VLPs with SARS-CoV-2 S antigens, which has not been previously described in any published work to the best of our knowledge. Further, we also achieved the generation of a

spike-presenting enveloped VLP without the need for co-expressing M and E coronavirus proteins, as described in the literature [24].

3.3. Transient Transfection Optimization by a Box–Behnken Design of Experiments

Further, the effect of different PEI and plasmid DNA concentrations and their effect on cell transfection were studied using a three-factor, three-level Box–Behnken design of experiments with the aim of finding an optimal condition that maximizes the percentage of cells expressing both Gag::eGFP and spike proteins. The three independent variables at the transfection mix included pGag::eGFP, pSpike and PEI. The experimental design matrix in coded values, response and statistical analysis is shown at Table 1. Experimental data were fitted to a second-order polynomial equation (Eq 1) using a non-linear regression analysis. The generated equation for the percentage of double-positive transfected cells analyzed by flow cytometry after IF-ICC staining at 72 hpt is shown below (Eq 2):

$$Y = 47.3 - 4.763 X_1 + 4.1 X_2 - 1.513 X_3 - 1.175 X_1 \cdot X_2 + 4.45 X_1 \cdot X_3 + \\ 0.125 X_2 \cdot X_3 - 6.1 X_1^2 - 3.175 X_2^2 + 4.1 X_3^2 \quad (2)$$

Where Y is the percentage of double positive transfected cells; X_1 is the coded value for pGag::eGFP; X_2 is the coded value for pSpike, and X_3 the coded value for PEI.

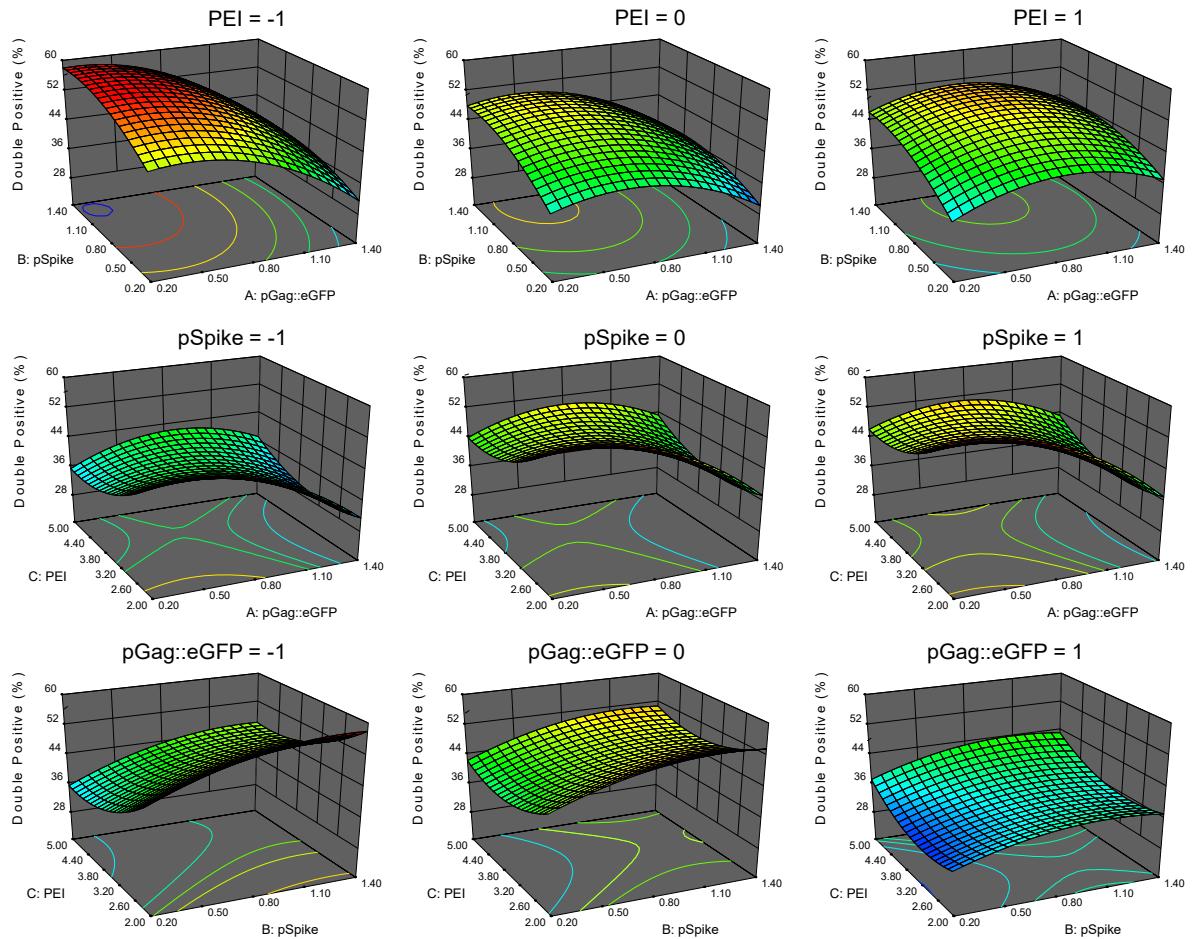


Figure 3. Analysis response surface graphs for the transfection optimization using the Box-Behnken design of experiments. The Y-axis accounts for the analyzed response: the percentage of cells expressing both Gag::eGFP and spike proteins (double-positive population) at 72 h post-transfection. Surface graphs describe the effect of the pGag::eGFP plasmid, pSpike plasmid and PEI concentrations (in $\mu\text{g/mL}$) in the double-positive population (% of total cells). Each combination of two independent variables is presented at the X- and Z-axis, while the response is shown at the Y-axis. (**Top**): pSpike vs. pGag::eGFP. (**Middle**): PEI vs. pGag::eGFP. (**Bottom**): PEI vs. pSpike. The third independent variable for each pair is modulated at each of the three coded levels (-1, 0, 1), as indicated above each individual surface response graph.

The model fitted the data with a R^2 of 0.9606, which corroborates its consistency with 96% of the variability in the data. The obtained F-value of 13.54 indicates that the model is also significant. There is only a 0.52% chance that a “Model F-Value” this large could occur due to signal noise. The Fisher’s F-test associated p-value of <0.0052 indicates the model was significant. Values of the terms A, B, AC, A^2 , B^2 and C^2 have a “Prob>F” less than 0.05, which also indicates that they are significant.

The model was used to plot response surface graphs (Figure 3) and to calculate the optimal factor levels that resulted in the highest double-positive transfected cell population. Response evaluation over the experimental region illustrates that the optimal concentration for pSpike is near the center of the range of concentrations tested, while optimal concentrations for pGag::eGFP and PEI are near their boundaries.

The optimum concentrations found for pGag::eGFP, pSpike and PEI were 0.308 µg/mL, 1.058 µg/mL and 2.045 µg/mL, respectively. With these concentrations, the model predicts $57.5 \pm 2.3\%$ of the total cell population would express both Gag::eGFP and spike protein at 72 hpt. To validate the generated model, a verification experiment was performed for the optimal conditions ($n=3$). A total of $58.9 \pm 0.4\%$ of the double-positive population was obtained at 72 hpt, corroborating the predictability of the model and setting the conditions to maximize double-positive cell population.

3.4. Production in Stirred-Tank Bioreactor

In order to evaluate production at a bioreactor scale, a culture in a 1 L stirred-tank bioreactor (STR) was performed. Cells were inoculated at a concentration of $0.5 \cdot 10^6$ viable cells / mL (vc/mL). The bioreactor was set to operate at 200 rpm, pH 7.1, 37 °C and a dissolved oxygen concentration over 40%. The STR culture was transfected with the previously established optimal DNA and PEI concentrations at a cell density of $2 \cdot 10^6$ vc/mL. After transfection, viable cells continued slowly growing until their harvest at 72 hpt, reaching a final density of $3.95 \cdot 10^6$ vc/mL, while their viability decreased to 76.1%. As Figure 4A shows, similar behavior was also observed in the parallel runs in 20 mL Erlenmeyer shake flasks ($n=3$), wherein cells reached slightly higher concentrations of $4.2 \cdot 10^6$ vc/mL at their peak but with a lower viability of 63.8% at 72 hpt.

The positive green fluorescent population by flow cytometry was evaluated at different time points in order to assess the transfection kinetics. As it can be observed in Figure 4B, fluorescent protein expression during the 48 h after transfection was slightly lower in the reactor in comparison with the Erlenmeyer flasks. At 72 hpt, both reached the same total Gag::eGFP producer population,

around 70%. At that time point, culture samples of STR and shake flask were IF-ICC stained and analyzed by flow cytometry in order to assess what percentage of cells was expressing HIV-1 Gag and/or the SARS-CoV-2 spike. Reactor and flasks showed similar percentages of double-transfected cells expressing Gag::eGFP and spike (55.1% and 55.8%, respectively), together with a ~27% population of single Gag::eGFP positive cells and a ~13.7% percentage of single S-expressing cells (Figure 4C). Therefore, no statistically significant differences were observed between the reactor and shake flasks cultures. Viability, growth and transfection analysis allow the conclusion that the production was successfully transferred to a 1 L bioreactor scale, which represents a very promising outcome for a potential scale-up of the process for the production of large amounts of the vaccine candidate for pre-clinical and clinical trials.

Supernatants from the bioreactor and shake flasks had very similar VLP concentrations ($\approx 3.5 \cdot 10^9$ VLPs/mL) at 72 hpt harvest, with an almost identical level of purity (~16.75% of VLPs over total particles) as evaluated by NTA fluorescent particle analysis (Figure 4D,E). The harvested 1 L work volume of the reactor contained $3.58 \cdot 10^{12}$ VLPs with no significant difference in mode particle diameter (data not shown), as evaluated by NTA. The spike concentration of the harvested supernatants was also determined by quantitative dot blot: the reactor showed a concentration of 1.78 µg spike/mL, while flask supernatants had a slightly lower concentration of 1.46 µg spike/mL (Figure 4F).

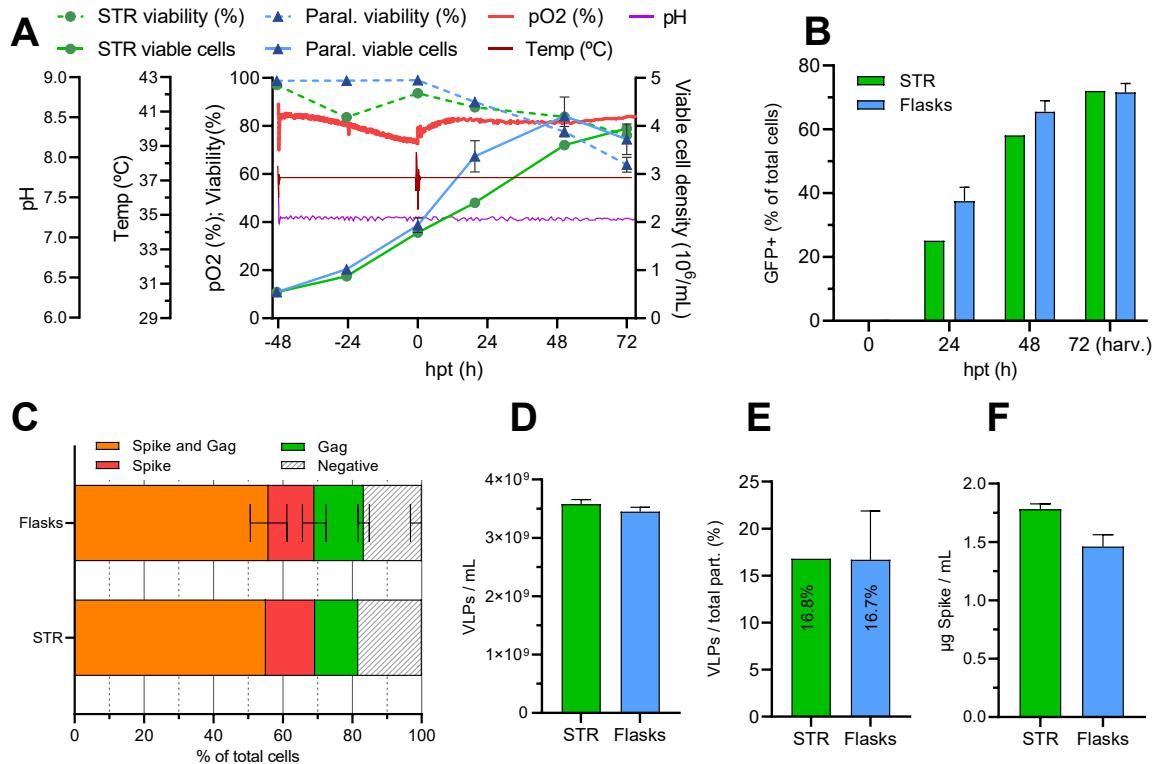


Figure 4. Production of S-VLPs in a stirred 1 L bioreactor and harvest analysis. (A): Online profile measurements of different process parameters: Temperature, pH and dissolved oxygen concentration (pO₂). Cell density and cell viability values are also shown. The arrow indicates the moment of transfection. (B): Transfection kinetics: percentage of Gag::eGFP-expressing cells (single- or double-positive) at different time points after transfection. (C): Immunocytochemistry analysis at 72 hpt. STR and shake flasks show a similar percentage of double-transfected populations (55.1% and 55.8%, respectively), together with a ~27% population of single Gag::eGFP-positive cells and a ~13.7% percentage of single S-expressing cells. (D): Harvest VLP concentrations, measured by NTA fluorescent particle analysis. (E): Harvest purity, which illustrates the percentage of VLPs from the total particles present at the harvest, by NTA fluorescent and non-fluorescent particle analysis. (F): Harvest SARS-CoV-2 spike concentrations, analyzed by dot blot.

3.5. Downstream Process of the Produced S-VLPs

A downstream purification process (DSP) that could be scalable in order to facilitate its potential use at industrial levels has been considered. For this purpose, a DSP consisting of two clarification steps, a capture ion-exchange step and a polishing size-exclusion final step was tested for the S-VLP purification of the harvested product from the 1L bioreactor.

The two initial clarification steps are necessary since the harvested supernatant presents turbidity caused by cellular debris and contaminating particles that can interfere when loaded into chromatographic columns. After two hours of

sedimentation, the harvested media was subjected to primary clarification using a depth filter for the removal of cell debris, intact cells, aggregates, impurities and other contaminating particulate materials from the harvested product. After the primary clarification, the clarified bulk was used as a secondary clarification feed. The filter used in this step is designed for bioburden and particle removal and can favor reductions in the levels of precipitates, as required for proper chromatographic performance. To evaluate the efficacy of the clarification steps, the turbidity was measured during the process. The harvested supernatant had a turbidity of 22.5 NTU after sedimentation. The first clarified bulk showed a turbidity of 4.21 NTU and second clarified 2.13 NTU, which is a desirable value for the good performance of the following purification steps [30].

The capture step consisted of an anion exchange (IEX) chromatography to separate molecules based on their net surface charge, concentrating the desired S-VLPs while decreasing the contaminants' contents. As the ion concentration was changed in different steps, the expected elution peaks were observed in the chromatograms at 488 nm absorbance (Figure 5A). Absorbance at 488 nm is caused by Gag::eGFP proteins, which allow monitoring S-VLP presence at the different stages of the process. The desired highly VLP-concentrated elution peak was collected and loaded into the next step.

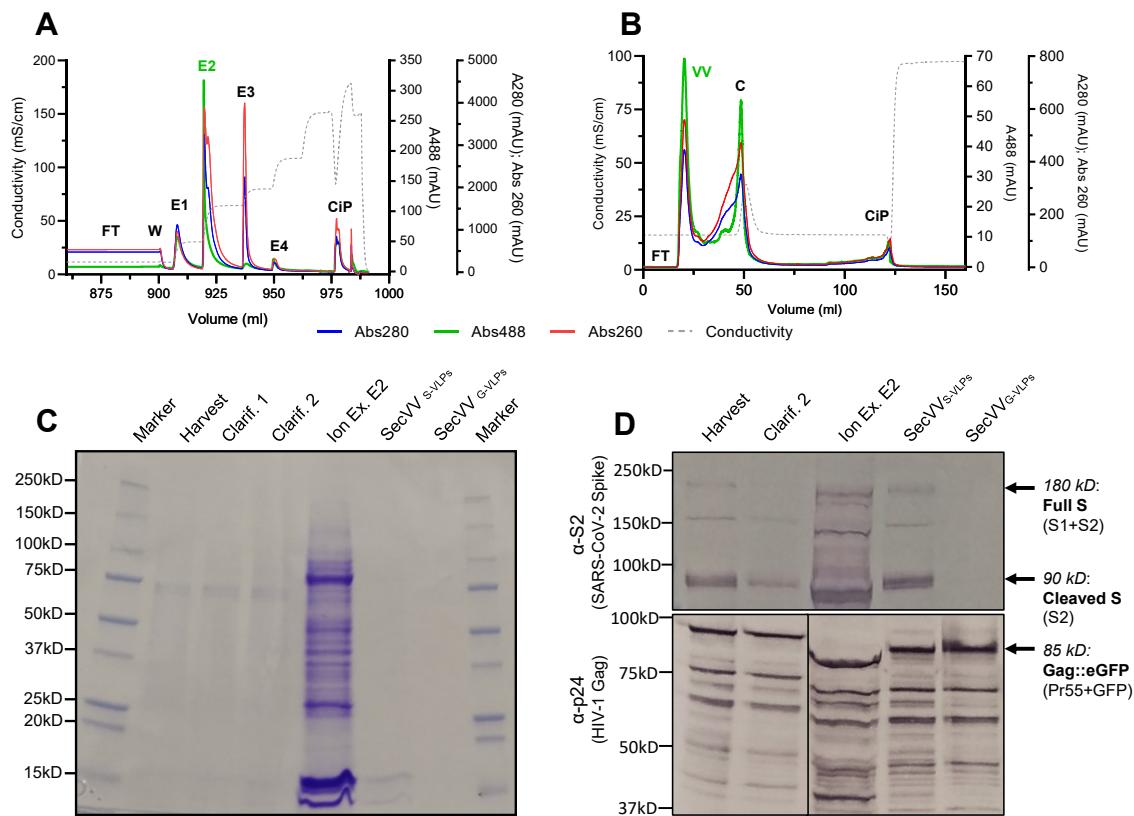


Figure 5. (A): Ion-exchange capture step chromatogram: FT (flow through), W (wash), E1-4 (eluted fractions resulting from different conductivity step increases, with E2 being the IEX purified product), CiP (cleaning in place). **(B)**: Size-exclusion polishing-step chromatogram: FT (flow through), VV (void volume, which is the final SEC-purified product), C (contaminants and aggregates), CiP (cleaning in place). **(A, B)**: Blue line: absorbance at 280 nm, indicating protein presence. Red line: absorbance at 260 nm, indicating DNA presence. Green line: absorbance at 488 nm, indicating Gag::eGFP presence. Dotted grey line: conductivity. **(C)**: SDS-Page of samples collected after each purification step: harvest, primary clarification, secondary clarification, ion exchange E2 fraction and size-exclusion VV fraction. A total protein concentration increase can be observed after the IEX step, caused mostly by undesired protein contaminants. As can be observed, those proteins are successfully eliminated after the following SEC polishing step. **(D)**: Western blot of samples after purification steps. Top: Membrane was treated with an anti-SARS-CoV-2 S2 spike subunit polyclonal antibody, followed by a goat anti-rabbit secondary incubation. Bottom: Membrane was treated with an anti-HIV-1 p24 monoclonal antibody, followed by a goat anti-mouse secondary incubation. Purified S-VLPs and G-VLPs show HIV-1 Gag bands, while only S-VLPs show SARS-CoV-2 S protein bands.

The size-exclusion (SEC) polishing step is intended for bulk impurities removal, the elimination of VLP aggregates, desalting and buffer exchange to achieve the final product with the desired level of purity [31]. The polishing step was performed successfully, as the chromatograms show (Figure 5B). Two main 488nm peaks can be observed: the first one is the void volume fraction, which was collected as the final product. The second peak corresponds to

contaminants “C” fraction and contains undesired VLP aggregates as analyzed by NTA (data not shown), which explains the notably high 488 nm absorbance levels.

As shown in Table 2, although a significant loss in terms of total number of VLPs, especially due to the low yield during the capture step, the overall DSP increased the VLP concentration from the initial bulk to the final product. The capture and polishing steps also had a significant positive impact in increasing the VLP purity over total particles, from 17.9% at the harvest to 31.1% at the final product.

The whole downstream process also succeeded in the reduction of undesired residual cellular contaminants, a crucial quality requirement for vaccines produced in cellular platforms [32]. The final purified sample contains 1.03% of the initial DNA and 0.22% of the initial protein (Table 2). This reduction of undesired protein concentration can also be observed by SDS-Page, especially after the polishing step (Figure 5C).

Along the different purification process steps, a drop in spike concentration higher than that expected due to VLP loss can be observed (Table 2), as analyzed by dot blot. This can be explained by the fact that, after VLP budding, unincorporated spike protein will remain present in the cellular membrane of the cells and be present in the cellular debris. This fraction is then removed during clarification steps. In addition, spike monomers solved at the clarified bulk are removed during the following capture and polishing steps. These decreasing values indicate that the S protein present at the final formulation could be due to properly VLP-incorporated and folded S proteins, which will present the immunogenic epitopes to the vaccinated patient’s immune system in a disposition that resembles the native S protein present at the SARS-CoV-2 virus. Western blot analyses for different DSP fractions using antibodies against HIV-1 Gag and SARS-CoV-2 spike confirmed the presence of those proteins with no significant migration pattern changes along the purification process (Figure 5D). The final purified product was also compared with identically purified G-VLPs, showing an almost identical Gag pattern together with no spike presence, as was expected for the negative control.

Table 2. Data of the different purification steps.

| Step | Volume (mL) | Total Protein (mg) | Protein Yield (%) ^a | Total DNA (µg) | DNA Yield (%) ^a | VLP Conc. (VLPs/mL) | VLP Purity (%) | Total VLPs | VLP Yield (%) ^b | Spike Conc. (µg/mL) |
|-------------|-------------|--------------------|--------------------------------|----------------|----------------------------|---------------------|----------------|----------------------|----------------------------|---------------------|
| Harvest | 1000 | 1074.9 | 100 | 2232.9 | 100 | $2.76 \cdot 10^9$ | 17.9 | $2.76 \cdot 10^{12}$ | - | 1.780 |
| Clarified 1 | 965 | 958.5 | 89.17 | 1558.0 | 69.77 | $2.94 \cdot 10^9$ | 21.2 | $2.84 \cdot 10^{12}$ | 102.79 | 1.000 |
| Clarified 2 | 960 | 930.5 | 86.57 | 1457.4 | 65.27 | $2.35 \cdot 10^9$ | 17.9 | $2.26 \cdot 10^{12}$ | 79.52 | 0.970 |
| Ion Ex. | 6 | 7.1 | 0.66 | 553.7 | 24.80 | $3.53 \cdot 10^9$ | 24.9 | $2.12 \cdot 10^{11}$ | 10.01 | 5.828 |
| Size Ex. | 7 | 0.2 | 0.02 | 23.1 | 1.03 | $8.61 \cdot 10^9$ | 31.3 | $6.03 \cdot 10^{10}$ | 37.94 | 2.198 |

^a From initial step. ^b From previous step.

Although the overall purification process shows very good purification results in terms of purity, a further optimization of the capture step would help to establish a methodology that will increase downstream process yield by reducing VLP loss.

4. Discussion

In this article, we first evaluated the cellular co-expression of SARS-CoV-2 spike glycoprotein with HIV-1 Gag, concluding that it has no significant negative effect in cell growth and viability. This suggests that it has no cytotoxic effect caused by protein secretory pathway failures. Confocal microscopy analysis showed that, after its expression, native envelope spike glycoprotein travels to the plasmatic membrane, wherein it co-localizes with Gag::eGFP. As Gag-based VLP generation occurs at the plasmatic membrane via budding, those results lay the groundwork to hypothesize that the S protein is incorporated to the VLPs. After that, we analyzed the produced and sucrose-cushion-purified VLPs by Western blot to find that S protein is present on the produced Gag-VLPs, confirming the incorporation of this SARS-CoV-2 antigen in our vaccine candidate. The produced VLPs had a mode diameter of 134.9 ± 1.2 nm, as measured by NTA. EM observations led us to conclude that they had no significant structural differences from Gag-based non-functionalized G-VLPs. Further, S protein presence was confirmed by immunogold labeling at the surface of S-VLPs, a key feature in order to present immunogenic SARS-CoV-2 epitopes to a patient's immune system when used as a vaccine. This was also relevant as, to the best of our knowledge, this study is the first report of Gag-based VLP functionalization with SARS-CoV-2 epitopes in order to generate a vaccine candidate against COVID-19.

Further, we optimized the production bioprocess using design of experiments in order to increase S-VLP productivity. We identified the transfection conditions

maximizing the cellular population co-expressing simultaneously Gag and S proteins. This is important in order to maximize the percentage of cells responsible for the production of the S-VLPs and to minimize the single-expressing population that generates non-functionalized VLPs. The model predicted a double-positive population of $57.5 \pm 2.3\%$ for the optimal transfection condition, which was validated and then implemented to transfet a 1 L stirred tank bioreactor.

The bioprocess was carried out satisfactorily, achieving good cellular growth and viabilities comparable to the parallel Erlenmeyer shake flasks. The double-transfected population accounted for 55.1% of the total cells, which is concordant with what was predicted by the previously generated model. The bioreactor showed similar behavior with the parallel Erlenmeyer flasks at 72 hpt in terms of spike concentration, VLP concentration and the purity of the produced S-VLPs. This confirmed that production in a 1 L reactor was achieved successfully.

Finally, the 1 L bioreactor harvested product was purified using a downstream process consisting of two clarification steps, an anion-exchange capture step and a size-exclusion final polishing step. Clarification steps succeeded at reducing the turbidity of the sample by removing undesired contaminants, aggregates, intact cells and debris. Capture and polishing steps reduced the presence of undesired proteins, dsDNA and VLP aggregates while increasing VLP purity. The final purified product presents a significant reduction of dsDNA (1.03%) and host cell protein presence (0.22%) relative from the initial sample. Western blot analysis helped to track spike and Gag presence along the purification process, while dot blot analyses were also performed in order to quantify spike concentrations, obtaining a concentration of 2.198 ng/ μ L at the final purified product. Overall, the DSP process had a low yield in terms of VLP recovery but highly succeeded at concentrating and purifying the desired S-VLPs while generating a final product with little undesired contaminant presence.

5. Conclusions

This work established a method for the production of SARS-CoV-2 VLPs by their co-expression with Gag::eGFP. We studied their expression, demonstrated the incorporation of the desired S proteins onto the produced VLPs and optimized the production process. Then, we successfully performed the bioprocess at a 1 L bioreactor scale and purified the produced harvest using a scalable DSP process. Furthermore, the reagents used in this work were animal-free, and all the materials and equipment used throughout the whole bioprocess are cGMP. This facilitates the potential transfer of the product manufacture to the industrial scale.

The bioprocess defined in this work will be subsequently used to produce new VLP candidates against recently emerged COVID-19 variants, since it harbors the potential to produce different Gag-based chimeric VLPs. The future work will also be focused on testing the produced S-VLPs by evaluating its immunogenic potential against convalescent COVID-19 patient sera and mice animal model.

6. Acknowledgments and funding

The authors thank Dr. Amine Kamen (McGill University, Montreal, Canada) for providing the HEK293 SF-3SF6 cells. The development of SARS-CoV-2 reagents was partially supported by the NIAID Centers of Excellence for Influenza Research and Surveillance (CEIRS) contract HHSN272201400008C. The pGag::eGFP (Cat#11468) reagent from Dr. Merilyn Resh was obtained through the National Institutes of Health AIDS Reagent Program, Division AIDS, National Institute of Allergy and Infectious Diseases, National Institutes of Health. We would like to thank Dr. Jose Amable Bernabé from the Institut de Ciència de Materials de Barcelona (ICMAB-CSIC, Bellaterra, Catalonia, Spain) for its support in nanoparticle tracking analysis techniques. The support of Martí de Cabo Jaume and Núria Barba Bosch from the Servei de Microscòpia (Universitat Autònoma de Barcelona, Bellaterra, Catalonia, Spain) in the development of electron and optical microscopy images and analysis is greatly appreciated. We would also like to thank Manuela Costa Garcia from the Servei de Cultius Cel·lulars, Producció d'Anticossos i Citometria (Universitat Autònoma de Barcelona, Bellaterra, Catalonia, Spain) for its precious help with flow cytometry.

We would also like to thank Jesus Lavado Garcia for its help setting up and inoculating the bioreactor, Erianet Lorenzo for performing the downstream process and Paula Grau García for its help with unpublished experiments.

This research has been partially funded by the Universitat Autònoma de Barcelona crowd funding campaign “Ajuda’ns a desenvolupar una vacuna contra el SARS CoV-2”.

References

- [1] Hu, B., Guo, H., Zhou, P. & Shi, Z. L. Characteristics of SARS-CoV-2 and COVID-19. *Nat. Rev. Microbiol.* **19**, 141–154 (2021)
- [2] Guan, D. *et al.* Global supply-chain effects of COVID-19 control measures. *Nat. Hum. Behav.* **4**, 577–587 (2020)
- [3] Forni, G. *et al.* COVID-19 vaccines: where we stand and challenges ahead. *Cell Death Differ.* **28**, 626–639 (2021)
- [4] Omidifar, N. *et al.* Different laboratory diagnosis methods of covid-19: A systematic review. *Arch. Clin. Infect. Dis.* **16**, 1–12 (2021)
- [5] Mousavi, S. M. *et al.* Recent biotechnological approaches for treatment of novel COVID-19: from bench to clinical trial. *Drug Metab. Rev.* **53**, 141–170 (2021)
- [6] Jhaveri, R. One Year of COVID-19 mRNA Vaccines: Incredible Progress and Unfinished Business. *Clin. Ther.* **43**, 2041–2043 (2021)
- [7] Rossman, H. *et al.* COVID-19 dynamics after a national immunization program in Israel. *Nat. Med.* **27**, 1055–1061 (2021)
- [8] Smit, M., Marinosci, A., Agoritsas, T. & Calmy, A. Prophylaxis for COVID-19: a systematic review. *Clin. Microbiol. Infect.* **27**, 532–537 (2021)
- [9] Teijaro, J. R. & Farber, D. L. COVID-19 vaccines: modes of immune activation and future challenges. *Nat. Rev. Immunol.* **21**, 195–197 (2021)
- [10] Heath, P. T. *et al.* Safety and Efficacy of NVX-CoV2373 Covid-19 Vaccine. *N. Engl. J. Med.* **385**, 1172–1183 (2021)
- [11] Garg, H., Mehmetoglu-Gurbuz, T. & Joshi, A. Virus Like Particles (VLP) as multivalent vaccine candidate against Chikungunya, Japanese Encephalitis, Yellow Fever and Zika Virus. *Sci. Rep.* **10**, 4017 (2020)
- [12] Wang, J. W. & Roden, R. B. S. Virus-like particles for the prevention of human papillomavirus-associated malignancies. *Expert Rev. Vaccines* **12**, 129–141 (2013)

- [13] Venereo-Sánchez, A. *et al.* Hemagglutinin and neuraminidase containing virus-like particles produced in HEK-293 suspension culture: An effective influenza vaccine candidate. *Vaccine* **34**, 3371–3380 (2016)
- [14] Lavado-García, J. *et al.* Characterization of HIV-1 virus-like particles and determination of Gag stoichiometry for different production platforms. *Biotechnol. Bioeng.* **118**, 2660–2675 (2021)
- [15] Zepeda-Cervantes, J., Ramírez-Jarquín, J. O. & Vaca, L. Interaction Between Virus-Like Particles (VLPs) and Pattern Recognition Receptors (PRRs) From Dendritic Cells (DCs): Toward Better Engineering of VLPs. *Front. Immunol.* **11**, 1100 (2020)
- [16] Du, L. *et al.* The spike protein of SARS-CoV - A target for vaccine and therapeutic development. *Nat. Rev. Microbiol.* **7**, 226–236 (2009)
- [17] Noad, R. & Roy, P. Virus-like particles as immunogens. *Trends Microbiol.* **11**, 438–444 (2003)
- [18] Yao, H. *et al.* Molecular Architecture of the SARS-CoV-2 Virus. *Cell* **183**, 730-738.e13 (2020)
- [19] Baden, L. R. *et al.* Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. *N. Engl. J. Med.* **384**, 403–416 (2021)
- [20] Farnós, O. *et al.* Rapid high-yield production of functional sars-cov-2 receptor binding domain by viral and non-viral transient expression for pre-clinical evaluation. *Vaccines* **8**, 654 (2020)
- [21] Rabi, F. A., Al Zoubi, M. S., Al-Nasser, A. D., Kasasbeh, G. A. & Salameh, D. M. Sars-cov-2 and coronavirus disease 2019: What we know so far. *Pathogens* **9**, 231 (2020)
- [22] Chen, Y. *et al.* Potent RBD-specific neutralizing rabbit monoclonal antibodies recognize emerging SARS-CoV-2 variants elicited by DNA prime-protein boost vaccination. *Emerg. Microbes Infect.* **10**, 1390–1403 (2021)

- [23] Berry, J. D. *et al.* Neutralizing epitopes of the SARS-CoV S-protein cluster independent of repertoire, antigen structure or mAb technology. *MAbs* **2**, 53–66 (2010)
- [24] Mortola, E. & Roy, P. Efficient assembly and release of SARS coronavirus-like particles by a heterologous expression system. *FEBS Lett.* **576**, 174–178 (2004)
- [25] Buffo, M. M. *et al.* Influence of dual-impeller type and configuration on oxygen transfer, power consumption, and shear rate in a stirred tank bioreactor. *Biochem. Eng. J.* **114**, 130–139 (2016)
- [26] Gutiérrez-Granados, S., Cervera, L., Gòdia, F., Carrillo, J. & Segura, M. M. Development and validation of a quantitation assay for fluorescently tagged HIV-1 virus-like particles. *J. Virol. Methods* **193**, 85–95 (2013)
- [27] Cervera, L., Gutiérrez-Granados, S., Berrow, N. S., Segura, M. M. & Gòdia, F. Extended gene expression by medium exchange and repeated transient transfection for recombinant protein production enhancement. *Biotechnol. Bioeng.* **112**, 934–946 (2015)
- [28] Esposito, D. *et al.* Optimizing high-yield production of SARS-CoV-2 soluble spike trimers for serology assays. *Protein Expr. Purif.* **174**, 105686 (2020)
- [29] Lavado-García, J. *et al.* Metabolic engineering of HEK293 cells to improve transient transfection and cell budding of HIV-1 virus-like particles. *Biotechnol. Bioeng.* **118**, 1649–1663 (2021)
- [30] Carvalho, S. B. *et al.* Efficient filtration strategies for the clarification of influenza virus-like particles derived from insect cells. *Sep. Purif. Technol.* **218**, 81–88 (2019)
- [31] Nweke, M. C., McCartney, R. G. & Bracewell, D. G. Mechanical characterisation of agarose-based chromatography resins for biopharmaceutical manufacture. *J. Chromatogr. A* **1530**, 129–137 (2017)

- [32] Knezevic, I., Stacey, G. & Petricciani, J. WHO Study Group on cell substrates for production of biologicals, Geneva, Switzerland, 11-12 June 2007. *Biologicals* **36**, 203–211 (2008)

Chapter two

**Study of the recognition of different SARS-CoV-2 functionalized
VLPs by COVID-19 convalescent human sera**

Abstract: The robustness, safety, versatility and high immunogenicity of Virus-like particles (VLPs) makes them a promising approach for the generation of vaccines against a broad range of pathogens. VLPs are recombinant macromolecular structures that closely mimic the native conformation of viruses without carrying viral genetic material. Particularly, HIV-1 Gag-based VLPs are a suitable platform for the presentation of the SARS-CoV-2 Spike (S) protein on their surface. In this context, this work studies the effect of different rationally engineered mutations of the S protein to improve some of its characteristics. The studied variants harbored mutations such as proline substitutions for S stabilization, D614G from the early dominant pandemic form, the elimination of the S1/S2 furin cleavage site to improve S homogeneity, the suppression of a retention motif to favor its membrane localization; and cysteine substitutions to increase its immunogenicity and avoid potential undesired Antibody-Dependent Enhancement (ADE) effects. The influence of the mutations on the VLP expression was studied, as well as their immunogenic potential, by testing the recognition of the generated VLP variants by COVID-19 convalescent patients' sera. The results of this work are conceived to give insights on the selection of S protein candidates for their use as immunogens.

Keywords: SARS-CoV-2, spike, Virus-like particles, COVID-19

1. Introduction

The coronavirus disease 2019 (COVID-19) pandemic caused by severe acute respiratory syndrome 2 (SARS-CoV-2) has had a brutal health and economic impact worldwide [1,2]. Vaccines constitute a major part of the solution to mitigate the expansion of this virus, its emerging variants or the zoonotic threat posed by new coronaviruses with pandemic potential [3,4]. The approval of multiple commercial vaccines, their intensive manufacturing and the deployment of the corresponding vaccination programs, resulted in the reduction of transmissions, associated hospitalizations and deaths, making possible the sanitary restrictions to come to an end [5]. However, none of them confer a full prophylactic protection for a long period of time, and new emerging variants are evolving with potential to escape the immune protective effect of the vaccinated population [6–8]. Virus-like particles (VLPs) constitute a high-immunogenic, versatile, robust and safe approach with great potential [9,10]. They induce potent cellular and humoral responses and can be pseudotyped to present epitopes of interest when used as vaccine [11]. HIV-1 Gag based VLPs are particles of ~145 nm of diameter which have been successfully functionalized to present SARS-CoV-2 proteins [12]. They are generated by the recombinant expression of the HIV-1 polyprotein which accumulates at the membrane of the producer cells and buds from them, taking part of the cell's plasmatic membrane as its lipidic envelope [13,14]. If the producer cell is simultaneously expressing other membrane proteins, those are incorporated into the surface of Gag VLPs [12]. This can be taken as an advantage to functionalize them with SARS-CoV-2 proteins.

SARS-CoV-2 virus is formed by four structural proteins: membrane (M), envelope (E), nucleocapside (N) and Spike (S) (Figure 1.A). The spike glycoprotein is present on the surface of the viral particles forming prominent homotrimers. It is a type I transmembrane fusion protein composed by 1273 aminoacids (aa) divided in two subunits: S1 (1-685 aa) and S2 (686-1273 aa). S1 is constituted by the N-terminal domain (NTD) and the receptor binding domain (RBD) (Figure 1.B). The RBD receptor-binding motif (RBM) interacts and binds to the angiotensin-converting enzyme-2 (ACE-2) receptor present at the surface of some host cell types [15]. S2 is a less exposed S subunit

responsible for the fusion between the host cell and viral membranes [16], whose domains are detailed in Figure 1.B. After RBD interaction with ACE-2 receptor, host-cell protease cleavage induces large conformational changes resulting in the exposure of the S2 fusion machinery allowing membrane fusion and viral entry [17]. Overall, this critical role of the spike in the life cycle of the viral infection makes it the primary target for the development of preventive therapies and vaccines [17,18]. Small mutations can affect its transmission, pathogenicity and immunogenicity. For that reason they need to be studied and considered for the advancement in SARS-CoV-2 vaccine development [15].

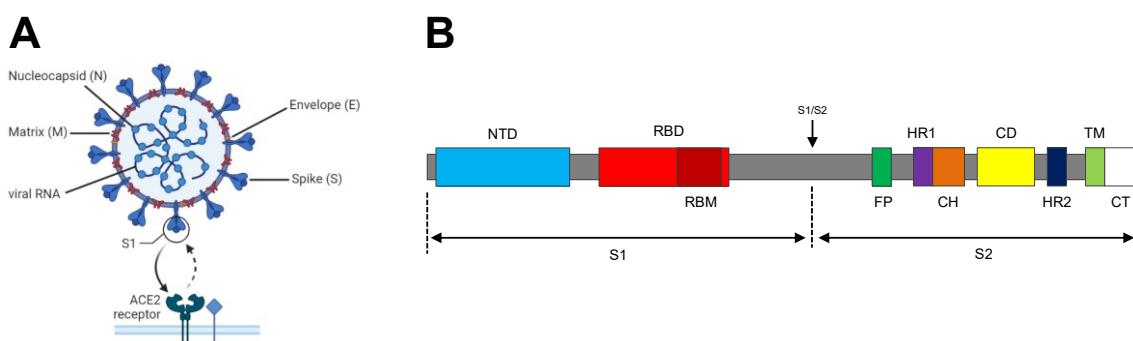


Figure 1. (A): SARS-CoV-2 virion scheme. M, N, E and S proteins are represented. S1 subunit of the S protein interacts to the host cell receptor membrane protein ACE2 to bind and promote viral internalization. (B): Schematic representation of the Spike protein. S1/S2 cleavage site is indicated with an arrow. Abbreviations: NTD, N-terminal Domain; RBD, receptor-binding domain; RBM, receptor-binding motif; FP, fusion peptide; HR1 and HR2, heptad repeat 1 and 2; CH, central helix; CD, connector domain; TM, transmembrane domain; CT, cytoplasmic tail.

The incorporation of the S-protein at the surface of the Gag-based VLPs generates SARS-CoV-2 functionalized VLPs (S-VLPs) [12]. In this work, S was rationally engineered by modifying its nucleotide sequence to propose and study different S variants. The introduced mutations consist in stabilizing proline substitutions [19,20]; D614G as the early dominant pandemic form [21]; the substitution of three arginine codons in order to eliminate the S1/S2 polybasic cleavage site [17,22]; lysine and histidine substitutions at the C-terminal in order to eliminate the ER-Golgi intermediate compartment (ERGIC) retention dibasic motif [23]; and two cysteine substitutions to create a disulfide bond in order to avoid antibody dependent enhancement (ADE) response [17], a phenomenon that can occur when the antibodies generated after immunization are not able to

stop the infection and instead act as a “Trojan horse” facilitating pathogen cellular entry [24,25]. In this study three S-VLP candidates harboring different S protein mutations were generated and their recognition by sera from COVID-19 patients was tested. This allowed determining its immunologic potential and selecting the more promising immunogen for its further study as a potential vaccine candidate.

2. Materials and methods

2.1. Cell Line, Media and Culture Conditions

The serum-free suspension-adapted HEK293 cell line (HEK293SF-3F6) was used, kindly provided by Dr. Amine Kamen from the Biotechnology Research Institute at the National Research Council of Canada and McGill University (Montreal, Canada). This cell line was derived from a current good manufacturing practice (cGMP) master cell bank available for manufacturing of clinical material.

The medium used for HEK293 cellular growth was the chemically defined and free from animal components HyCell TransFx-H from HyClone (GE Healthcare, Chicago, IL, USA) supplemented with 4 mM GlutaMAX (Gibco, Thermo Fisher Scientific, Waltham, MA, USA) and 0.1% Pluronic F-68 Non-ionic Surfactant (Gibco, Thermo Fisher Scientific, Waltham, MA, USA).

Suspension cell cultures were maintained routinely in exponential growth phase in 125mL or 1L disposable polycarbonate Erlenmeyer flasks with a vent cap (Corning, Tewksbury, MA, USA) in a LT-X Kuhner shaker (LT-X Kuhner, Birsfelden, Switzerland) shaking at 130 rpm, at 37 °C, 5% CO₂ and 85% RH. Cell counts and viability determinations were performed using the NucleoCounter NC-3000 automatic cell counter (Chemometec, Lillerød, Denmark) following the manufacturer's instructions.

2.2. Plasmids and Transfection

2.2.1. Plasmid Expression Vectors

The pGag::eGFP plasmid codes for a codon-optimized Rev-independent HIV-1 Gag protein fused in frame to the enhanced GFP driven by the CMV enhancer and promoter. The plasmid from the NIH AIDS Reagent Program (Cat 11468) (Hermida-Matsumoto and Resh, 2000) was constructed by cloning the Gag sequence from pCMV55M1-10 (Schwartz et al., 1992) into the pEGFP-N1 plasmid (Clontech, Palo Alto, CA, USA).

The pSpike plasmid codes for a mammalian cell codon optimized nucleotide sequence of the spike protein of SARS-CoV-2 driven by the CAG enhancer and β-actin promoter. It was produced under HHSN272201400008C and obtained through BEI Resources, NIAID, NIH: Vector pCAGGS Containing the SARS-CoV-2, Wuhan-Hu-1 spike Glycoprotein Gene, NR-52310.

The pSpikemut2 plasmid codes for a mammalian cell codon optimized nucleotide sequence of the Spike protein of SARS-CoV-2 (Wuhan-Hu-1 spike Glycoprotein Gene, NR-52310) harboring K968P, V969P, S383C, D985C, D614G, and R682_R685delinsGSAS mutations. The expression is driven by the CAG enhancer and β-actin promoter. It was designed by A. Boix-Besora, and produced by Gene Synthesis & DNA Synthesis Services of GenScript (GenScript, Leiden, Netherlands), derived from pSpike. Sequencing and restriction analysis were carried out to validate the construct.

The pSpikemut3 plasmid codes for a mammalian cell codon optimized nucleotide sequence of the Spike protein of SARS-CoV-2 (Wuhan-Hu-1 spike Glycoprotein Gene, NR-52310) harboring K968P, V969P, S383C, D985C, D614G, R682_R685delinsGSAS and K1269A H1271A mutations. The expression is driven by the CAG enhancer and β-actin promoter. It was designed by A. Boix-Besora, and produced by Gene Synthesis & DNA Synthesis Services of GenScript (GenScript, Leiden, Netherlands), derived from pSpike. Sequencing and restriction analysis were carried out to validate the construct.

pMock plasmid does not have any mammalian promoter or coding DNA sequence (CDS). It was constructed by the ligation of the pGag::eGFP backbone.

2.2.2. Plasmid Amplification and Purification

Plasmids were amplified in *Escherichia coli* DH5 α strain grown in LB medium (Conda, Madrid, Spain) supplemented with kanamycin (10 μ g/mL, Sigma, St. Louis, MO, USA) or ampicillin (100 μ g/mL, Sigma, St. Louis, MO, USA) depending on the *E. coli* antibiotic resistance present on each plasmid. Plasmid purification was carried out using the Endofree Plasmid Mega kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions.

2.2.3. PEI-Mediated Transient Transfection

Exponentially growing HEK293 cells were passaged in 1 L polycarbonate erlenmeyer flasks in order to have a cell density of $2 \cdot 10^6$ cells/mL at transfection time. A media exchange was performed prior to transfection by centrifugation of the cells at 300 $\times g$ for 5 minutes. 25 kDa linear polyethylenimine (PEI Max, PolySciences, Warrington, PA, USA) was used as transfection reagent. PEI-DNA complexes were formed under sterile conditions. Briefly, DNA was diluted in culture media (10% of the total volume of cell culture to be transfected) for a final total DNA concentration of 1 μ g/mL and vortexed for 10 s. Then, polyethylenimine (PEI), was added for a final concentration of 2 μ g/mL (a 2:1 PEI:DNA ratio (w/w)) and vortexed three times for 3 s. The mixture was incubated for 15 minutes at RT and then added to the culture.

2.7. Immunocytochemistry Staining for Flow Citometry

For IF-ICC staining, cells were centrifuged 5 min at 300 $\times g$ and rinsed with staining solution (1.5% (v/v) fetal bovine serum (FBS) 1X phosphate-buffered saline (PBS)) before primary antibody incubation for 20 min at 4 °C in the dark. After rinsing twice, cells were incubated with the corresponding secondary antibody for 20 min at 4 °C. After IF-ICC staining, fixation was performed using 2% (v/v) formaldehyde 1X PBS for 10 min at RT. Cells were resuspended in staining solution and stored at 4 °C prior to analysis.

Primary human anti-SARS-CoV-2 spike glycoprotein RBD domain antibody (ab272854, AbCam, Cambridge, UK) was diluted 1:1000. The secondary antibody used for flow citometry analysis was an anti-human IgG (H+L) coupled with CyTM5, produced in donkey (709-175-149, Jackson ImmunoResearch, West Grove, PA, USA), diluted 1:400. The secondary antibody used for confocal microscopy imaging was an anti-human IgG (H+L) coupled with Alexa Fluor 568, produced in goat (#A-21090, Thermo Fisher Scientific, Waltham, MA, USA), diluted 1:400. All IF-ICC antibodies were diluted using staining solution.

The transfected cellular populations of previously stained cells were analyzed by flow cytometry using a BD FACS Canto flow cytometer (BD BioSciences, San Jose, CA, USA), at Servei de Cultius Cel·lulars, Producció d'Anticossos i Citometria (Universitat Autònoma de Barcelona, Bellaterra, Catalonia, Spain).

2.9. Sucrose Cushion Purification

Culture harvests were performed at 72 hpt and centrifuged 10,000x g for 10 min. The supernatants containing VLPs were placed on a 30% (w/v) sucrose cushion for ultracentrifugation at 31,000 rpm for 2 h at 4 °C. The supernatant was carefully discarded, and pellets containing the VLPs were resuspended in PBS.

2.10 Nanoparticle Tracking Analysis

NTA-based Gag::eGFP VLP quantification and characterization was performed using a NanoSight®NS300 (Nanosight Ltd., Amesbury, UK) equipped with a blue filter module (488 nm) and a neutral filter at the Soft Material Service of the Institut de Ciència de Materials de Barcelona (ICMAB-CSIC, Bellaterra, Catalonia, Spain). Samples were previously diluted to a concentration of approximately 10⁸ particles/mL. Sample injection was performed using a pump to improve the robustness of the measurement by continuous addition, and to minimize the photobleaching effect due to fluorescence depletion over time. 60 second videos were recorded at RT and analyzed with the NTA 3.4 software (Malvern Panalytical, Malvern, UK). Tracked particles size was determined from its Brownian motion. Three independent experimental replicas were carried out

for each sample. Camera level and detection threshold were manually adjusted for each replica.

2.11. Total Protein and Spike quantifications

A BCA Protein Assay (#23225, Thermo Fisher Scientific, Waltham, MA, USA) was performed following manufacturer's instructions using the provided BSA as standard. Colorimetric absorbance at 562 nm was read on a Multilabel Plate Reader Victor3 (Perkin Elmer, Waltham, MA, USA).

For SARS-CoV-2 Spike quantification, samples were charged into Bio-Dot Apparatus (#1706545, Bio-Rad, Hercules, CA, USA) while a low vacuum was applied. Nitrocellulose membrane (#88018, Thermo Fisher Scientific, Waltham, MA, USA) was placed at the top of humidified filter paper. Once samples were transferred, membrane was incubated with anti-SARS-CoV-2 spike glycoprotein S2 monoclonal antibody (Ab281312, AbCam, Cambridge, UK) and an anti-rabbit secondary antibody (A9919, Merck, Kenilworth, NJ, USA) following the same procedure previously mentioned for Western blot. Once dried, membranes were scanned, and the pixel density for each loaded sample was analyzed using software ImageJ2 Fiji (National Institutes of Health, Bethesda, MD, USA). The standard used for quantification was a recombinant human coronavirus SARS-CoV-2 spike glycoprotein S2 subunit (Ab272106, AbCam, Cambridge, UK).

2.12. Human sera assay

The research procedure involving the use of sera of human origin was reviewed and approved by the Ethics Committee on Animal and Human Experimentation (CEEAH) of the Universitat Autònoma de Barcelona (UAB) to meet the ethical and legal requirements regarding research with biological samples from humans. The approved CEEAH reference number is 5293. The sera used in this study were provided by the Biobank of the Banc de Sang i Teixits (BST) and samples were de-identified from patient information. Eight anonymized sera samples from convalescent non-vaccinated COVID-19 patients (confirmed with RT-qPCR) were used in this work. Four sera collected from COVID-19 uninfected and unvaccinated individuals were used as negative control.

Briefly, VLP variants and controls were charged into Accutran-Cross Blot-System for Cross Blot (#448100, Schleicher&Schuell, Dassel, Germany) containing vertical lane-shaped wells above an immobilized nitrocellulose membrane (#88018, Thermo Fisher Scientific, Waltham, MA, USA) and incubated overnight at 4 °C with agitation. Once antigens were transferred, membranes were incubated for 1 h in agitation with blocking buffer (2% (w/v) nonfat dry milk in 1x PBS). After blocking, membranes were incubated with human sera (1:60 dilutions in blocking buffer) for 2 h, charged in horizontal lane-shaped wells. After sera incubation, membranes were incubated with an anti-Human IgG (Fab specific)-peroxidase antibody produced in goat (#A0293, Sigma Aldrich, St. Louis, MO, USA) (1:2000 in blocking buffer) for 1 h. Then, they were revealed with Pierce™ ECL Plus western blotting substrate (32132, Thermo Fisher Scientific, Waltham, MA, USA). Once dried, membranes were imaged using a ChemiDoc™ Touch Imaging System (#1708370, Bio-Rad, Hercules, CA, USA) and pixel densities were analyzed using the software ImageJ2 Fiji (National Institutes of Health, Bethesda, MD, USA). All the wash steps between incubations were performed with agitation in 0.05% Tween-20 in 1x PBS. The antigen used as positive control was a SARS-CoV-2 Spike S1-His Recombinant Protein (#40591-V08H, SinoBiological Europe GmbH, Eschborn, Germany).

3. Results

3.1 Engineered spike protein variants

Three S protein variants were generated and studied, named S_{WT} , S_{mut2} and S_{mut3} . S_{WT} codes for the original Wuhan sequence. S_{mut2} incorporates two proline substitutions, D614G as the early dominant pandemic form, the elimination of the S1/S2 cleavage site and the creation of a disulfide bond to avoid ADE, by substitution of two cysteines (Figure 2). S_{mut3} incorporates the same mutations as S_{mut2} with additional substitutions in its C-terminus in order to eliminate its ER-Golgi intermediate compartment (ERGIC) retention motif (Figure 2).

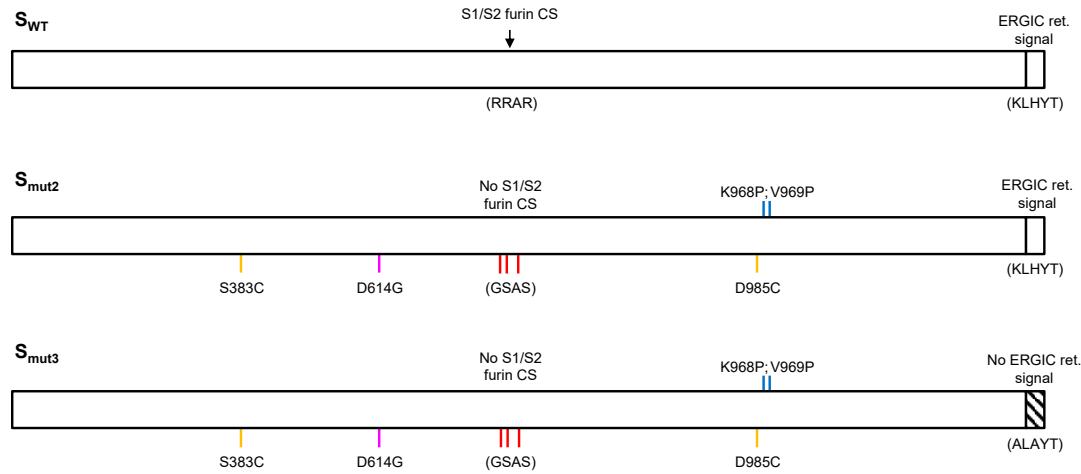


Figure 2. Schematic representation of the S_{WT} protein and the S_{mut2} and S_{mut3} variants with its mutations indicated. Furin cleavage site (CS) is indicated by an arrow. ER-Golgi intermediate compartment retention signal (ERGIC ret. signal) is represented as a box.

3.1.1 Proline substitutions

The spike protein can transition between an unstable prefusion state to a postfusion stable conformation, as a consequence of its role in membrane fusion [20]. The protein structural design generated with K968P and V969P proline substitutions is used and described as a strategy that can be applied in different Betacoronavirus S proteins, for stabilizing and retaining them in the antigenically optimal prefusion conformation by inactivating its membrane fusion activity [19,26]. This approach improves its expression yields and conformational homogeneity [19]. Prefusion-stabilized antibody epitopes are more likely to lead to neutralizing antibody responses and it was demonstrated that these mutations were able to elicit high neutralizing antibody titers against MERS-CoV [19,20,26]. Indeed, several SARS-CoV-2 vaccines incorporate stabilizing proline substitutions for the mentioned reasons [27,28].

3.1.2 Cysteine substitutions

Antibodies are generally beneficial and protective against viral infections [24]. However, in SARS-CoV-2 infection, sub-optimal antibody production and early seroconversion had been reported to correlate with disease severity by ADE phenomenon, which can occur mediated by the engagement of Fc receptors (expressed on monocytes, macrophages and B cells among others) [24]. This

phenomenon may promote ACE-2-independent viral entry to cells expressing Fc receptors (FcRs) [29]. As the quality and quantity of the antibody is crucial in order to elicit a good and effective immune protection against SARS-CoV-2, literature suggests that the disulfide bond created by cysteine substitutions S383C and D985C will hide some of the SARS-CoV-2 S immunogen non-neutralizing epitopes that might cause an ADE response [17,30]. Although low expression yields had been reported when expressing S protein harboring S383C and D985C mutations, stabilizing mutations like proline substitutions may increase its immunogenic potential [17,31].

3.1.3 Early dominant pandemic form

Despite not presenting alarmingly high mutation ratios, during the COVID-19 pandemic SARS-CoV-2 has experienced different sequence variations which resulted in changes on its transmissibility, severity and immune escape [15,32]. D614G variant, presenting a glycine (G) substitution of the aspartic (D) present at the 614 position of the original S sequence was rare on February 2020 [21]. However, it quickly replaced the ancestral virus becoming the dominant pandemic form worldwide by April 2020 [33]. The D614G mutation resulted in a fitness advantage without an increased severity, appearing to correlate with higher viral loads in patients, and increasing in-vitro infectivity [21,34]. It has been suggested that the higher infectivity of the G614 variant is mostly caused by its increased stability when forming S trimmers, preventing its premature loss and hence, effectively increasing the amount of S proteins that can facilitate the infection [33]. Studies had shown that G614 has a higher neutralization sensitivity to COVID-19 convalescent human sera compared to the ancestral variant, suggesting an epitope exposure increase [35]. Its conformational changes improving stability, preventing premature loss and favoring pre-fusion conformations make D614G an interesting mutation to incorporate for the generation of new vaccine candidates [33].

3.1.4 Furin cleavage site removal

As suggested in different studies, probably all coronavirus S proteins are cleaved at some point during infection and, in many cases this cleavage occurs at the S1/S2 position [22]. SARS-CoV-2 S protein sequence contains a S1/S2

polybasic cleavage site (CS) which can be recognized and cleaved by host cell furin protease [36]. Upon S1/S2 cleavage, a second CS present at S2 domain becomes exposed, which after further cleavage, activates S2 membrane fusion machinery [36]. It has been demonstrated that the S1/S2 CS increases SARS-CoV-2 pathogenesis [37,38] and promotes its entry into lung cells [36,39]. However, studies on pre-fusion stabilized S immunogens presenting proline substitutions (Section 3.1.1) concluded that no large conformational changes were imparted consequence of the furin S1/S2 cleavage absence [20]. This justifies its removal in order to generate a more homogeneous vaccine immunogens [20]. To do so, the 682-685 a.a. RRAR sequence of the S protein was substituted by GSAS [17].

3.1.5 Dibasic motif removal

SARS-CoV-2 protein presents a dibasic motif (KXHXX) in the last four amino acids of its cytoplasmic tail, which specifies intracellular localization [40]. This motif reduces the S rate of traffic through the Golgi complex and promotes its retention at ERGIC [41,42]. Studies showed that when lysine (K) and histidine (H) residues present in the dibasic motif were substituted by alanines, efficient transport and localization at the plasmatic membrane occurred [23,41,42]. S-VLPs are generated by the incorporation of the S protein present at the cell membrane into the budding Gag-based particles. For that reason, the removal of the ERGIC retention signal was hypothesized to favor the generation of S-VLPs.

3.2 S-VLPs variants production and quantification

To generate the S-VLP candidates, exponentially growing HEK293 cells in 1L suspension Erlenmeyer flasks were transfected with PEI as transfection reagent according to Table 1. Negative control VLPs produced by the co-transfection of Gag::eGFP and an empty plasmid (G-VLPs) were also generated to assess and subtract the unspecific binding of the polyclonal human sera to the non-functionalized VLP scaffold. The HIV-1 Gag polyprotein used in this work was fused in frame with the eGFP reporter, in order to facilitate tracking and quantification of the produced VLPs [13].

Table 1. Summary of the tested VLP candidates, its mutations and plasmids used to generate them.

| VLP | Spike mutations | Plasmids transfected |
|------------------------|---|-------------------------|
| S _{WT} -VLP | Original Wuhan-Hu-1 spike glycoprotein sequence | pSpike + pGag::eGFP |
| S _{mut2} -VLP | Stabilizing proline substitutions, disulfide bond (no ADE), D | pSpikemut2 + pGag::eGFP |
| S _{mut3} -VLP | K968P, V969P, S383C, D985C, D614G, R682_R685delinsGSAS, K1269A H1271A | pSpikemut3 + pGag::eGFP |
| G-VLP | - | pGag::eGFP + pMock |

Cells transfected with Gag and Spike proteins showed viabilities between 70-80% at harvest time (Figure 3.A). Interestingly, cells transfected with S_{mut2} and S_{mut3} variants showed a ~10% viability improvement compared to S_{WT} expression, reaching higher cellular densities as well. This may be favored by the stabilizing mutations which facilitate protein expression. No significant differences in terms of viability and cell growth were observed between S_{mut2} and S_{mut3} variants, indicating that the suppression of the ERGIC retention motif does not have any effect on this aspect.

Additionally, ICC analysis showed that the expression of the mutated S variants was translated in an increase of the total populations expressing S protein and the double positive population co-expressing Gag and S, which rose from 55.2% for the S_{WT} VLP variant to 60.6% and 62.2% for S_{mut2} and S_{mut3} VLP variants respectively (Figure 3.B). In contrast, confocal microscopy analysis did not show significant differences of S protein internal compartment or membrane localization between variants (data not shown).

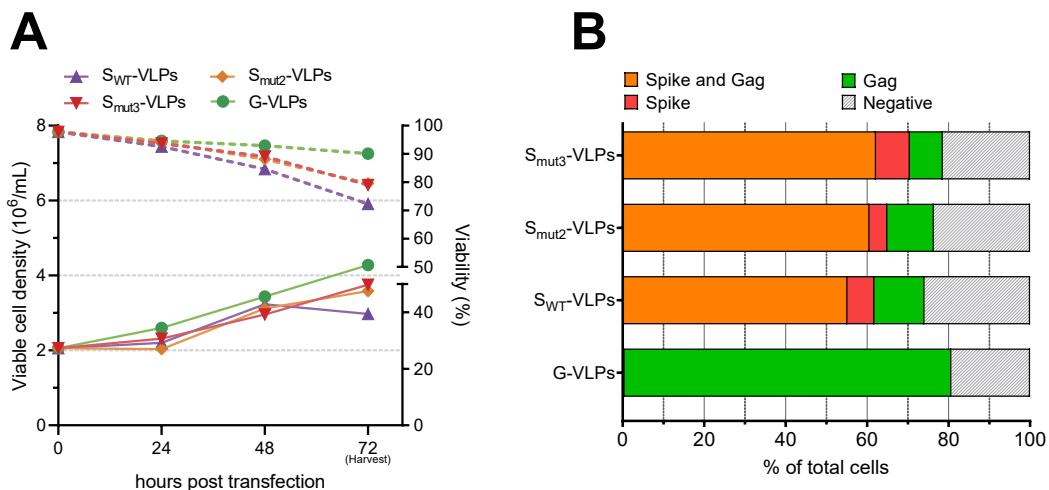


Figure 3. (A): Cell concentrations (solid lines) and viabilities (dotted lines) of HEK293 cells transfected for the production of different VLP variants, described at Table 1. (B): Transfected populations, analyzed by ICC at 72 hpt.

VLPs were purified from the harvested supernatants by sucrose cushion ultracentrifugation. After purification, the group transfected with the unmodified S_{WT} protein showed VLP concentrations of $2.9 \cdot 10^{10}$ particles/mL, compared with $6.38 \cdot 10^{10}$ and $4.69 \cdot 10^{10}$ VLPs/mL for S_{mut2} and S_{mut3} VLP variants, representing 2.2 and 1.6 fold increases, respectively. Additionally, the VLP ratio among total extracellular particles improved from 8.3% for the $S_{\text{WT}}\text{-VLPs}$ to 16.2% and 14% for the S_{mut2} and S_{mut3} VLPs (Table 2), suggesting a positive effect of the mutations harbored by the two mutants enhancing the VLP production and improving its ratio among the total extracellular particles.

Table 2. Characteristics of each produced S-VLP candidate and control (G-VLPs).

| | VLPs/mL | Total part | VLP/part ratio (%) | Total protein conc. ($\mu\text{g/mL}$) | Spike conc. ($\mu\text{g/mL}$) | Spike/VLP* (units/VLP) |
|-------------------------------|---------------------------------|---------------------------------|--------------------|--|----------------------------------|------------------------|
| $S_{\text{WT}}\text{-VLPs}$ | $(2.90 \pm 0.13) \cdot 10^{10}$ | $(3.51 \pm 0.13) \cdot 10^{11}$ | 8.3 | $566,14 \pm 19$ | 12.08 ± 0.99 | 1776.8 |
| $S_{\text{mut2}}\text{-VLPs}$ | $(6.38 \pm 0.05) \cdot 10^{10}$ | $(3.94 \pm 0.2) \cdot 10^{11}$ | 16.2 | $263,99 \pm 15$ | 13.90 ± 0.35 | 929.3 |
| $S_{\text{mut3}}\text{-VLPs}$ | $(4.69 \pm 0.05) \cdot 10^{10}$ | $(3.35 \pm 0.07) \cdot 10^{11}$ | 14.0 | $297,88 \pm 104$ | 27.05 ± 4.03 | 2460.2 |
| G-VLPs | $(1.85 \pm 0.04) \cdot 10^{11}$ | $(5.15 \pm 0.24) \cdot 10^{11}$ | 35.9 | $305,83 \pm 46$ | - | - |

*Assuming a Mw of 141.178 kDa for the Spike protein

The total protein concentration of the purified S_{mut2} and S_{mut3} VLPs was approximately half of that for S_{WT} group (Table 2), although presenting a higher VLP concentration. Considering that functionalized VLP variants presented higher VLP concentrations, this seems to indicate that the VLPs generated by the expression of the S_{mut2} S_{mut3} variants, present less undesired host cell proteins. A profound VLP characterization including proteomic analysis could help to better define the host cell protein composition of the produced particles to explain the observed phenomena [43].

Dot Blot analysis allowed to determine the spike concentration of each VLP purified candidate. The S protein concentration was similar for the purified S_{mut2} -VLPs and S_{WT} -VLPs. Interestingly; it was increased by 2.2 fold when expressing the S_{mut3} variant, suggesting a positive effect of the removal of the ERGIC retention motif favoring the S localization at the plasmatic membrane and subsequently facilitating its incorporation to the produced VLPs. This was translated in a greater functionalization of the S_{mut3} -VLPs, with ~2460 spike proteins per VLP (Table 2).

3.3 Assessment of the immunogenic potential of the S-VLP candidates

To assess the immunogenic potential of the generated S_{WT} , S_{mut2} and S_{mut3} VLPs, they were tested for its recognition by human convalescent COVID-19 patient sera. G-VLPs were used as negative control and a recombinant commercial S protein was used as a positive control. 900ng (total protein) of each VLP candidate were transferred into a nitrocellulose membrane and treated with different patient sera as detailed in section 2.12 (Figure 4).

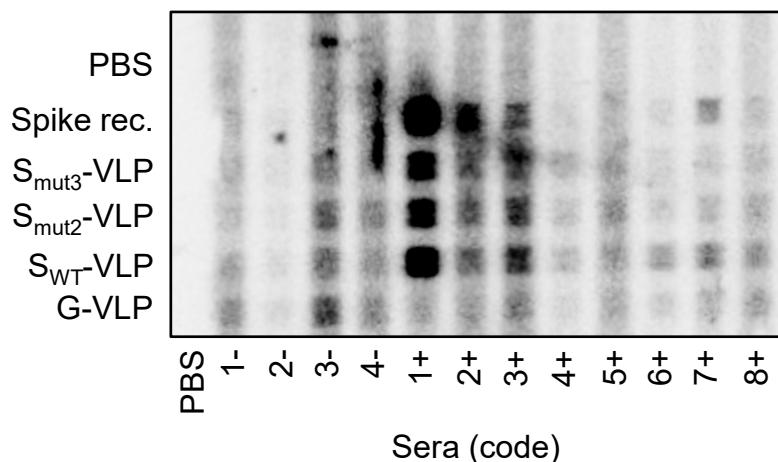


Figure 4. Membrane of the antigenic recognition assay. G-VLPs were used as negative control and recombinant soluble spike protein (S rec.) was used as positive control.

Previously, the level of antibodies against SARS-CoV-2 for each sera sample was typified in order to determine each patient's immune response level (Table 3).

The analysis of the pixel density from Figure 4 allowed the quantification of the antigen recognition by each individual serum. The unspecific signal against the control G-VLPs was used to determine the S-specific immune response threshold for each serum (dashed lines, Figure 5).

Table 3. Summary of positive and negative sera samples. Its antibody levels against SARS-CoV-2 are indicated in arbitrary units (AU) per mL.

| Negative sera | | | Positive sera | | |
|-----------------|------|---|-----------------|------|---|
| Reference (BST) | Code | Antibody positivity against SARS-CoV- 2 (AU/mL) | Reference (BST) | Code | Antibody positivity against SARS-CoV- 2 (AU/mL) |
| E002520158146 | 1- | 2.5 (negative) | E002520158136 | 1+ | 4046.4 |
| E002520158147 | 2- | 8.1 (negative) | E002520158137 | 2+ | 307.5 |
| E002520158149 | 3- | 1.1 (negative) | E002520158139 | 3+ | 214.21 |
| E002520158150 | 4- | 0.1 (negative) | E002520158141 | 4+ | 98.6 |
| | | | E002520158142 | 5+ | 105.7 |
| | | | E002520158143 | 6+ | 135.1 |
| | | | E002520158144 | 7+ | 427.3 |
| | | | E002520158145 | 8+ | 132.8 |

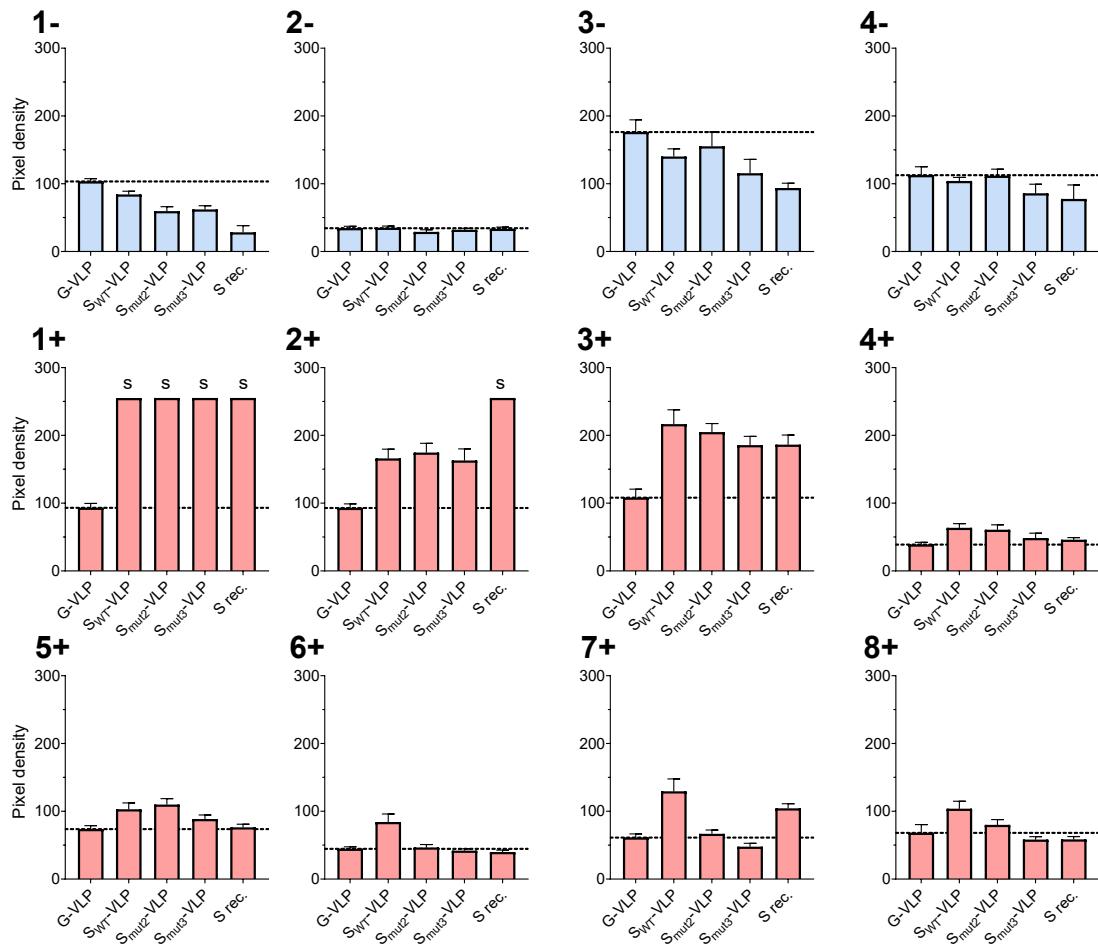


Figure 5. Relative pixel densities of the VLP recognition assays by negative (blue) and positive (red) tested sera samples. Horizontal dashed line indicates the unspecific recognition threshold, marked by the pixel density of the G-VLP negative control. Saturated pixel densities are indicated with an “s”. Soluble spike recombinant protein (S rec.) was used as positive control.

As expected, negative sera did not present S-specific antibody recognition of the S-VLP variants nor the S positive control (Figure 4). Generally, all the COVID-19 convalescent sera showed specific immune responses against S-VLP variants greater than its unspecific G-VLP threshold. Patients coded as 1+, 2+ and 3+, presented the highest antigen recognitions in agreement with their previously determined high anti-SARS-CoV-2 antibody positivity (Table 3). Particularly, 1+ reached the pixel saturation limit for all the studied S immunogens (not allowing their comparative study), whereas sera 2+ only showed pixel saturation for the S recombinant positive control, which was significantly more recognized than the S-VLPs (Figure 4). Interestingly, and except for serums 1+ and 2+, S-VLPs were generally better recognized by the

convalescent sera than the recombinant S protein control. This can be explained as the S-VLPs present the S protein in a more genuine conformation and proteic context than the individual soluble S protein. Overall, the sera assays had better recognized the S_{WT} -VLP immunogens than the S_{mut2} and S_{mut3} VLP variants (Figure 4). This was not the case for serums 2+ and 5+ although they did not show large significant differences between S_{WT} and S_{mut2} VLP recognition. Additionally, -and contrarily to what might be expected given the large number of S protein units presented on its surface- S_{mut3} -VLPs presented the worst recognition levels by all the tested convalescent sera (Figure 4). The obtained results made us conclude that the unmodified wild-type S protein was the best candidate for the functionalization of Gag-based VLPs for their use as immunogens.

4. Conclusions

In this work, three SARS-CoV-2 Gag-based VLP variants were rationally engineered and studied. The candidates, named S_{WT} -VLP, S_{mut2} -VLP and S_{mut3} -VLP were generated by the co-expression of the Gag::eGFP and a spike protein variant: S_{WT} , S_{mut2} and S_{mut3} , respectively (Figure 2). S_{WT} codes for the original SARS-CoV-2 spike Wuhan sequence. S_{mut2} incorporates stabilizing proline substitutions, the early dominant pandemic form mutation D614G, the elimination of the S1/S2 polybasic furin cleavage site to generate more homogeneous vaccine candidates; and two cysteine substitutions to increase its immunogenicity and avoid potential undesired ADE responses. S_{mut3} harbors the presented mutations with the additional suppression of its C-terminal ERGIC retention motif, with the aim to promote its membrane localization and thereby its incorporation to the generated VLPs.

VLPs were produced by transient co-transfection. Cells expressing the two mutated S variants behaved similarly, showing a ~10% viability improvement and an increased cell density at harvest time compared the S_{WT} group (Figure 3.A). This behavior could be consequence of the stabilizing mutations introduced to the mutated spike proteins facilitating its expression. Furthermore, ICC analysis showed an increase of the S presence in cell groups expressing the mutated S variants, resulting in the increase of the double positive

expresser populations from 55.2% for the S_{WT} VLP expresser cells to 60.6% and 62.2% for S_{mut2} and S_{mut3} groups (Figure 3.B). The increase of the double positive population is a desirable outcome as denotes that a higher proportion of the cultured cells are generating functionalized Gag-VLPs. VLP purification was carried on by sucrose cushion ultracentrifugation. The VLP concentration of the purified S_{mut2} and S_{mut3} VLPs represented 2.2- and 1.6-fold increases ($6.38 \cdot 10^{10}$ and $4.69 \cdot 10^{10}$ VLPs/mL respectively) compared with S_{WT} -VLPs (Table 2). In addition, the VLP ratio among total extracellular particles for the S_{mut2} and S_{mut3} VLPs improved to 16.2% and 14% compared with the 8.3% obtained for the S_{WT} -VLPs. Dot Blot analysis of the purified VLP candidates allowed to determine that S_{mut3} -VLPs presented 2.2-fold increase of S protein concentration compared to the S_{WT} -VLPs and S_{mut2} -VLPs. The S ratio per VLP showed a clear advantage of the ERGIC retention signal removal as S_{mut3} presented remarkably greater functionalization levels than its homologous variant S_{mut2} (Table 2). Overall, the transfection and VLP characterization studies showed evidence that the introduced mutations contributed positively to the generation of VLPs and its functionalization with S protein.

Further, the obtained VLP immunogens were tested for their recognition by COVID-19 convalescent human sera. Non-functionalized Gag-VLPs (G-VLPs) were used to determine the unspecific immune response of the sera and to establish a threshold beyond which the immune response was considered SARS-CoV-2 S-specific. All the different COVID-19 patient's convalescent sera recognized the VLP variants above the G-VLP threshold, varying on intensity depending on their antibody positivity against SARS-CoV-2 (Table 3), as exemplified by serums 1+ and 2+ (Figure 5). In general, S-VLPs were better recognized by the convalescent sera compared with the soluble version of the S protein, due to the advantageous nature of the VLPs when it comes to proper antigen presentation in terms of conformation and protein context. Surprisingly, the assay determined that most of the sera showed the best recognition levels for the unmodified S_{WT} -VLPs (Figure 5). In contrast, and contrarily to what might be expected given its S functionalization levels (Table 2), S_{mut3} -VLPs presented the worst convalescent sera recognition levels (Figure 5). This work concludes that although improving the expression of VLPs and their functionalization with

the S immunogen, the presented S_{mut2} and S_{mut3} VLP candidates did not show immune responses improving or equaling the conventional S_{WT} -VLPs. However, further studies need to be performed to fully understand the immunogenic potential of the presented VLP variants, as the low recognition by convalescent sera could be -in part- due to the concealment of some regions containing epitopes with potential to trigger undesired ADE responses.

5. Acknowledgements and funding

The authors thank Dr. Amine Kamen (McGill University, Montreal, Canada) for providing the HEK 293 SF-3SF6 cells. The pGag::eGFP (Cat#11468) reagent from Dr. Merilyn Resh was obtained through the National Institutes of Health AIDS Reagent Program, Division AIDS, National Institute of Allergy and Infectious Diseases, National Institutes of Health. We would like to thank Dr Jose Amable Bernabé from the Institut de Ciència de Materials de Barcelona (ICMAB-CSIC, Bellaterra, Catalonia, Spain) for his support in nanoparticle tracking analysis techniques. We also would like to thank Toni Iborra from Servei de Cultius Cel·lulars, Producció d'Anticossos i Citometria (Universitat Autònoma de Barcelona, Bellaterra, Catalonia, Spain) for his precious help performing the human sera antigen recognition assays. Figures and Graphs were generated using GraphPad Prism version 8.0.2 for Windows (San Diego, CA, USA).

This research has been partially funded by the Universitat Autònoma de Barcelona crowd funding campaign “Ajuda’ns a desenvolupar una vacuna contra el SARS CoV-2”.

References

- [1] Ali, I. & Alharbi, O. M. L. COVID-19: Disease, management, treatment, and social impact. *Sci. Total Environ.* **728**, 138861 (2020)
- [2] Hu, B., Guo, H., Zhou, P. & Shi, Z.-L. Characteristics of SARS-CoV-2 and COVID-19. *Nat. Rev. Microbiol.* **20**, 315–315 (2022)
- [3] Wang, C. *et al.* A conserved immunogenic and vulnerable site on the coronavirus spike protein delineated by cross-reactive monoclonal antibodies. *Nat. Commun.* **12**, 1715 (2021)
- [4] Forni, G. & Mantovani, A. COVID-19 vaccines: where we stand and challenges ahead. *Cell Death Differ.* **28**, 626–639 (2021)
- [5] Mohammed, I. *et al.* The efficacy and effectiveness of the COVID-19 vaccines in reducing infection, severity, hospitalization, and mortality: a systematic review. *Hum. Vaccin. Immunother.* **18**, (2022)
- [6] Bezbaruah, R. *et al.* Developmental Landscape of Potential Vaccine Candidates Based on Viral Vector for Prophylaxis of COVID-19. *Front. Mol. Biosci.* **8**, (2021)
- [7] Weisblum, Y. *et al.* Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. *Elife* **9**, (2020)
- [8] Geers, D. *et al.* SARS-CoV-2 variants of concern partially escape humoral but not T cell responses in COVID-19 convalescent donors and vaccine recipients. *Sci. Immunol.* **6**, (2021)
- [9] Zepeda-Cervantes, J., Ramírez-Jarquín, J. O. & Vaca, L. Interaction Between Virus-Like Particles (VLPs) and Pattern Recognition Receptors (PRRs) From Dendritic Cells (DCs): Toward Better Engineering of VLPs. *Front. Immunol.* **11**, (2020)
- [10] Virus-like Particles in Vaccine Development. *Virus-like Part. Vaccine Dev.* **9**, 1–136 (2014)
- [11] Nooraei, S. *et al.* Virus-like particles: preparation, immunogenicity and

- their roles as nanovaccines and drug nanocarriers. *J. Nanobiotechnology* **19**, 1–27 (2021)
- [12] Boix-Besora, A., Lorenzo, E., Lavado-García, J., Gòdia, F. & Cervera, L. Optimization, Production, Purification and Characterization of HIV-1 GAG-Based Virus-like Particles Functionalized with SARS-CoV-2. *Vaccines* **10**, 1–18 (2022)
 - [13] Cervera, L. *et al.* Generation of HIV-1 Gag VLPs by transient transfection of HEK 293 suspension cell cultures using an optimized animal-derived component free medium. *J. Biotechnol.* **166**, 152–165 (2013)
 - [14] Fuenmayor, J., Gòdia, F. & Cervera, L. Production of virus-like particles for vaccines. *N. Biotechnol.* **39**, 174–180 (2017)
 - [15] Thakur, V. *et al.* Waves and variants of SARS-CoV-2: understanding the causes and effect of the COVID-19 catastrophe. *Infection* **50**, 327–327 (2022)
 - [16] Shah, P., Canziani, G. A., Carter, E. P. & Chaiken, I. The Case for S2: The Potential Benefits of the S2 Subunit of the SARS-CoV-2 Spike Protein as an Immunogen in Fighting the COVID-19 Pandemic. *Front. Immunol.* **12**, (2021)
 - [17] Henderson, R. *et al.* Controlling the SARS-CoV-2 spike glycoprotein conformation. *Nat. Struct. Mol. Biol.* **27**, 925–933 (2020)
 - [18] Farnós, O. *et al.* Rapid High-Yield Production of Functional SARS-CoV-2 Receptor Binding Domain by Viral and Non-Viral Transient Expression for Pre-Clinical Evaluation. *Vaccines* **8**, 654 (2020)
 - [19] Pallesen, J. *et al.* Immunogenicity and structures of a rationally designed prefusion MERS-CoV spike antigen. *Proc. Natl. Acad. Sci.* **114**, E7348–E7357 (2017)
 - [20] Kirchdoerfer, R. N. *et al.* Stabilized coronavirus spikes are resistant to conformational changes induced by receptor recognition or proteolysis. *Sci. Rep.* **8**, 1–11 (2018)

- [21] Korber, B. *et al.* Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell* **182**, 812–827.e19 (2020)
- [22] Millet, J. K. & Whittaker, G. R. Host cell proteases: Critical determinants of coronavirus tropism and pathogenesis. *Virus Res.* **202**, 120–134 (2015)
- [23] Lontok, E., Corse, E. & Machamer, C. E. Intracellular Targeting Signals Contribute to Localization of Coronavirus Spike Proteins near the Virus Assembly Site. *J. Virol.* **78**, 5913–5922 (2004)
- [24] Iwasaki, A. & Yang, Y. The potential danger of suboptimal antibody responses in COVID-19. *Nat. Rev. Immunol.* **20**, 339–341 (2020)
- [25] Singh, M. *et al.* SARS-CoV-2 and Trojan horse phenomenon—caveat in vaccine quest? *Scand. J. Immunol.* **92**, (2020)
- [26] Wrapp, D. *et al.* Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science (80-.)* **367**, 1260–1263 (2020)
- [27] Yu, J. *et al.* DNA vaccine protection against SARS-CoV-2 in rhesus macaques. *Science (80-.)* **369**, 806–811 (2020)
- [28] Sanders, R. W. & Moore, J. P. Virus vaccines: proteins prefer prolines. *Cell Host Microbe* **29**, 327–333 (2021)
- [29] Jaume, M. *et al.* Anti-Severe Acute Respiratory Syndrome Coronavirus Spike Antibodies Trigger Infection of Human Immune Cells via a pH- and Cysteine Protease-Independent Fc γ R Pathway. *J. Virol.* **85**, 10582–10597 (2011)
- [30] Wang, L. *et al.* Importance of Neutralizing Monoclonal Antibodies Targeting Multiple Antigenic Sites on the Middle East Respiratory Syndrome Coronavirus Spike Glycoprotein To Avoid Neutralization Escape. *J. Virol.* **92**, (2018)
- [31] Hsieh, C.-L. *et al.* Structure-based design of prefusion-stabilized SARS-CoV-2 spikes. *Science (80-.)* **369**, 1501–1505 (2020)

- [32] Gorbatenko, A. E., Enjuanes, L., Ziebuhr, J. & Snijder, E. J. Nidovirales: Evolving the largest RNA virus genome. *Virus Res.* **117**, 17–37 (2006)
- [33] Zhang, J. *et al.* Structural impact on SARS-CoV-2 spike protein by D614G substitution. *Science (80-).* **372**, 525–530 (2021)
- [34] Hou, Y. J. *et al.* SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission in vivo. *Science (80-).* **370**, 1464–1468 (2020)
- [35] Weissman, D. *et al.* D614G Spike Mutation Increases SARS CoV-2 Susceptibility to Neutralization. *Cell Host Microbe* **29**, 23-31.e4 (2021)
- [36] Peacock, T. P. *et al.* The furin cleavage site in the SARS-CoV-2 spike protein is required for transmission in ferrets. *Nat. Microbiol.* **6**, 899–909 (2021)
- [37] Johnson, B. A. *et al.* Loss of furin cleavage site attenuates SARS-CoV-2 pathogenesis. *Nature* **591**, 293–299 (2021)
- [38] Lau, S.-Y. *et al.* Attenuated SARS-CoV-2 variants with deletions at the S1/S2 junction. *Emerg. Microbes Infect.* **9**, 837–842 (2020)
- [39] Hoffmann, M., Kleine-Weber, H. & Pöhlmann, S. A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells. *Mol. Cell* **78**, 779-784.e5 (2020)
- [40] Walls, A. C. *et al.* Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **181**, 281-292.e6 (2020)
- [41] McBride, C. E., Li, J. & Machamer, C. E. The Cytoplasmic Tail of the Severe Acute Respiratory Syndrome Coronavirus Spike Protein Contains a Novel Endoplasmic Reticulum Retrieval Signal That Binds COPI and Promotes Interaction with Membrane Protein. *J. Virol.* **81**, 2418–2428 (2007)
- [42] Wang, X. *et al.* Deletion of ER-retention motif on SARS-CoV-2 spike protein reduces cell hybrid during cell–cell fusion. *Cell Biosci.* **11**, 114 (2021)

- [43] Lavado-García, J. *et al.* Characterization of HIV-1 virus-like particles and determination of Gag stoichiometry for different production platforms. *Biotechnol. Bioeng.* **118**, 2660–2675 (2021)

Chapter three

**Stable HEK293 cell line generation for the production of
Gag::eGFP VLPs by random integration**

Abstract: Virus-like particles (VLPs) have shown a great potential as new-generation vaccines. They are capable of generating strong immune responses and represent a safer and cheaper alternative to conventional vaccines. Mammalian cells offer an attractive production platform for enveloped VLPs due to their ability to perform complex post-translational modifications (PTM). HIV-1 VLPs are enveloped nanoparticles constituted by recombinant-expressed structural Gag polyprotein. Its production can be achieved by transient gene expression (TGE) or stable gene expression (SGE). The former is based on the episomal plasmid DNA expression while the latter achieves a prolonged and constitutive gene expression via direct integration of the gene of interest into the cell genome. TGE is a rapid and easy approach mostly used for pre-clinical studies or small scale production of a product of interest. It represents a more economic and simpler approach for scale up and large productions, does not have the limitations found in TGE regarding the need to use transfection-compatible media and reduces batch-to-batch variability. In this work, a stable cell line producing Gag::eGFP has been generated through illegitimate random integration. Sorted individual clone candidates were isolated, screened and the more promising ones were successfully adapted to suspension and serum-free media. Growth kinetic study of the four best producers allowed the selection of the best clone which presents a duplication time of 29.2 hours, a maximum cell density peak of $6.16 \cdot 10^6$ viable cells/mL and a specific productivity of 35.7 VLPs/cell·day. The generated cell line presented 4 copies of the integrated construct as assessed by digital droplet PCR (ddPCR). Its stability was studied, concluding that VLP expression remained stable during the 60 studied days.

Keywords: VLPs, stable gene expression, random integration, HEK293, Gag::eGFP, HIV-1.

1. Introduction

Virus-like particles (VLPs) represent a safer alternative to conventional vaccines as they can present epitopes in a natural conformation, which can generate strong cellular and humoral immunoresponses [1]. They do not carry viral genetic material and are formed by the intrinsic ability of viral proteins to self-assemble [1,2]. VLPs can be produced in prokaryotic and eukaryotic expression systems. However, mammalian cells offer an attractive production platform for enveloped or multimeric VLPs due to their ability to perform complex post-translational modifications (PTM) which play a crucial role in the antigen recognition by the immune system [2]. There are several mammalian cell platforms in which VLPs can be produced: HeLa, Vero, CAP, CHO or HEK293 [2,3]. HEK293 cell line, derived from human embryonic kidney, is globally used due to its capacity to grow in suspension, high transfection efficiency and its capability to grow in serum-free media [4].

Human Immunodeficiency Virus-1 (HIV-1) VLPs are spherical enveloped particles of approximately ~145nm in diameter constituted by recombinantly expressed structural Gag polyprotein [5,6]. Gag is capable of self-assemble forming a protein core inside the VLP interacting with its lipid bilayer, taken from the producer cell in a process known as budding [7]. This lipid bilayer can be further functionalized with other HIV-1 antigens such envelope proteins or antigens from other viral species forming chimeric VLPs capable of conferring immunity to different diseases [6]. In order to produce VLPs or any recombinant proteins in mammalian platforms, there exist two main approaches: transient gene expression (TGE) and stable gene expression (SGE).

TGE is a rapid and easy approach for expression of the product of interest. DNA enters to cells facilitated by the use of a transfection method, which can be divided into biological, physical or chemical [8]. Chemical methods are widely used and are based on positively charged reagents like polyethylenimine (PEI) [9]. In TGE, the cells do not integrate the foreign genetic information into its genome and as a consequence the expression of the gene of interest decreases due to cell division and environmental factors, hence expression is lost over time [10]. TGE is mainly used for the study of genes, to inhibit or

enhance specific genes and for recombinant production, especially in small scale or pre-clinical stages [11]. It delivers multiple copies of the gene of interest (GOI) to each transfected cell, deriving in high protein production. The TGE product is typically harvested between 48 and 72 hours after transfection and gives enough product quantity and quality to perform pre-clinical assays. Even some pharmaceutical products that require small volumes in their productive stage (like adeno-associated virus vectors for gene therapy) can be produced using TGE [9].

SGE is a valuable strategy which achieves a prolonged and constitutive expression via direct integration of the gene or cassette of interest into the host cell genome [11]. The genomic integration is a rare event, so clones that express the desired protein have to be selected, isolated and grown from one single cell to achieve monoclonal cultures. Clones have to be subsequently screened for optimal attributes such as cell growth rate, specific productivity, stability and product quality [12]. As SGE cell population is defined and well characterized thanks to its monoclonal nature, cell culture media can be modified or supplemented in order to enhance cellular growth and optimize product production. In large scale bioprocesses, SGE is a more economic approach since it avoids the high amounts of transfection reagents and plasmid DNA needed for TGE, while simplifying the complexity of the bioprocess and decreasing batch-to-batch variability [13]. Additionally, it does not have the TGE limitation regarding the use of media, as the formation and stabilization of transfection DNA complexes are greatly affected by its ionic composition [14].

It represents a more economic and simpler approach for scale up and large productions, does not have the limitations found in TGE regarding the need to use transfection-compatible media and reduces batch-to-batch variability

There are two main approaches for the generation of a stable cell line: by random gene integration and directed gene integration [15]. The first method takes advantage of the spontaneous illegitimate integration into the host cell genome, while the second consists in a rationally-designed, homology-dependent and locus-specific genetic integration [16]. Several options are used for the generation of a random integrated stable cell line. The standard method

is to transfect the cells with a plasmid or linearized vector containing the GOI, together with a gene allowing its subsequent selection, for example an antibiotic resistance. Random integration is a rare event, as the efficiency of the genetic integration into HEK293 cells is very low. Moreover, only a small proportion of the cells will accommodate the GOI into a locus with a high transcriptional activity capable of producing sufficient amounts of the desired protein. Even then, most of the integrated transgenes will be silenced over time by epigenetic mechanisms [15]. This is known as the “position effect” phenomenon [17]. In order to find clones with an integration into a transcriptionally active locus (or “hot-spot”), a selection using a marker or/and an antibiotic followed by the isolation of and characterization of a large number of clones is mandatory [15].

The generation of a stable cell line expressing the HIV-1 Gag polyprotein for Gag-based VLP production is of great interest, since it can ease its industrial scale up and facilitate its functionalization with other pathogenic membrane proteins in order to generate vaccine candidates against different diseases [6].

In this work, we generated a HEK293 stable cell line for the stable expression of the Gag::eGFP polyprotein via random integration. The effect of using circular and linear DNA was also studied. Gag was fused in frame with the green fluorescence protein (GFP) reporter in order to facilitate its tracking and therefore clone screening. G418 was used for the selection of the polyclonal pool and the subsequent clonal screening, amplification and characterization process. The best cell lines were adapted to suspension and serum-free culture and characterized in terms of cell growth, specific productivity and stability to select the best cell line.

2. Materials and methods

2.1 Cell lines, media and culture conditions

A serum-free suspension-adapted HEK293 cell line (HEK293SF-3F6) was used, kindly provided by Dr. Amine Kamen from the Biotechnology Research Institute at the National Research Council of Canada (Montreal, Canada) and McGill University. This cell line was derived from a Current Good Manufacturing

Practice (cGMP) master cell bank available for manufacturing of clinical material.

The medium used for HEK293 growth was the chemically defined and free from animal components FreeStyle 293 (ThermoFisher, Waltham, MA, USA) supplemented with 0.1% Pluronic (ThermoFisher, Waltham, MA, USA), 1.6 mg/L of r-transferrin (Merck, Kankakee, IL, USA), 19.8 mg/L of r-insulin (Novo Nordisk Pharmatech, Køge, Denmark) and 0.9X of an in-house developed lipid mixture to maximize cell growth [5]. For cell selection 450 µg/mL of G418 (ThermoFisher, Waltham, MA, USA) was used. After cell sorting the medium was supplemented with 10% Fetal Bobine Serum (FBS) (26140079, ThermoFisher, Waltham, MA, USA), gentamicin (ThermoFisher, Waltham, MA, USA) and antibiotic-antimycotic (ThermoFisher, Waltham, MA, USA).

Suspension cell cultures were maintained routinely in exponential growth phase in 125ml disposable polycarbonate Erlenmeyer flasks with vent cap (Corning, Tewksbury, NY, USA) or 6-well plates (ThermoFisher, Waltham, MA, USA), placed in an orbital shaker (orbit diameter 16mm, Stuart, Stone, UK) at 130 rpm in a humidified Steri-Cult CO₂ incubator (3310, ThermoFisher, Waltham, MA, USA) at 37°C, 85% RH and 5% CO₂ in air.

Adherent cell cultures were maintained in exponential grown phase in 25 mm² T-flasks, 75 mm² T-flasks, 96-well, 24-well and 6-well plates (Sigma-Aldrich, Saint Louis, MO, USA) in a humidified incubator at 37°C, 85% RH and 5% CO₂ in air. Adherent cell monolayers were re-suspended without trypsinization, by gently hitting the T-flask or plate and pipetting.

Cell counts and viability determinations were performed using the NucleoCounter NC-3000 automatic cell counter (Chemometec, Lillerød, Denmark) following manufacturer's instructions. Adherent clones and cultures were observed under the Olympus CKX41 optical inverted microscope (Olympus, Tokyo, Japan) to qualitatively analyze cell growth and vitality. Gag::eGFP producer cells were observed under the Nikon Eclipse TE2000E optical fluorescence inverted microscope (Nikon, Tokyo, Japan) to qualitatively assess its Gag::eGFP expression.

Specific growth rate (μ) was calculated using the Equation 1. Specific productivity (q_p) and volumetric productivity (v_p) were calculated using Equation 2 and 3 respectively, where P is the product, X the viable cell density (VCD), t the time, and μ the specific growth rate of the cells during the selected time interval.

$$\mu = \frac{\ln(X_t - X_0)}{t_t - t_0} \quad (1)$$

$$q_p = \frac{\int_{t_0}^t P(t) \cdot dt}{\int_{t_0}^t X(t) \cdot dt} \cdot \mu \quad (2)$$

$$v_p = P/t \quad (3)$$

2.2 Plasmids and transfection

2.2.1 Plasmid expression vectors, amplification

The pGag::eGFP plasmid (Figure 2.A) codes for a codon optimized Rev-independent HIV-1 Gag protein fused in frame to the enhanced GFP driven by CMV enhancer and promoter. It also codes for the Aminoglycoside phosphotransferase from Tn5 (NeoR/KanaR) which confers resistance to kanamycin in bacteria, or neomycin and G418 (Geneticin) in eukaryotic cells, driven by SV40 promoter. The plasmid from the NIH AIDS Reagent Program (Cat 11468) [18] was constructed by cloning the Gag sequence from pCMV55M1-10 [19] into the pEGFP-N1 plasmid (Clontech, Palo Alto, CA, USA).

Plasmids were amplified in *Escherichia coli* Top10 strain grown in LB medium at 37 °C supplemented with 10 µg/ml of kanamycin (Sigma, St. Louis, MO, USA). Plasmid purification was carried out using the Endofree Plasmid Midi/Mini kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions.

2.2.2 PEI-mediated transient transfection

Exponentially growing HEK293 cells were passaged in 6-well plates in order to have a cell density of $2 \cdot 10^6$ cells/mL at transfection time. A media exchange

was performed prior to transfection by centrifugation of the cells at 300 x g for 5 minutes. 25 kDa linear polyethylenimine (PEI Max, PolySciences, Warrington, PA, USA) was used as transfection reagent. PEI-DNA complexes were formed under sterile conditions. Briefly, DNA was diluted in culture media (10% of the total volume of cell culture to be transfected) for a final total DNA concentration of 1 µg/mL and vortexed for 10 s. Then, polyethylenimine (PEI), was added for a final concentration of 2 µg/mL (a 2:1 PEI:DNA ratio (w/w)) and vortexed three times for 3 s. The mixture was incubated for 15 minutes at RT and then added to the culture.

Transfection was performed using circular or linearized DNA. The pGag::eGFP plasmid was used for circular DNA transfection, while its linearization (Figure 2.B) was achieved by FastDigest Vspl (ThermoFisher, Waltham, MA, USA) restriction, according to manufacturer recommendations.

2.3 Toxicity assays

100 µl of HEK293 cells were seeded in a 96-well plate at $0.5 \cdot 10^6$ cells/ml together with 10 µl of increasing antibiotic concentrations, all points triplicated. The plate was cultured under standard HEK293 culture conditions. After two days 20 µl of MTT reagent (Promega, WI, USA) was added to each well and the plate was incubated for one hour at 37°C with an agitation of 130 rpm. Just before the MTT addition, a standard curve of cell concentrations from 0 to $4 \cdot 10^6$ viable cells/ml was performed in duplicated columns. Absorbance at 490 nm was read on Multilabel Plater Reader VICTOR3® (Perkin Elmer, MA, USA).

2.4 Flow cytometry and cell sorting

For flow cytometry, cells were analyzed using a BD FACS Canto flow cytometer (BD BioSciences, San Jose, CA, USA), at Servei de Cultius Cel·lulars, Producció d'Anticossos i Citometria (Universitat Autònoma de Barcelona, Bellaterra, Catalonia, Spain). Cell sorting was carried out using a fluorescence activated BD FACSJazz Cell Sorter (BD Biosciences/ThermoFisher, Waltham, MA, USA). Data was processed using BD FACSDiva software (BD Biosciences, San Jose, CA, USA).

2.5 Spectrofluorometry-based product quantitation

Green fluorescence intensity in Gag-GFP samples was measured using a Cary Eclipse Fluorescence Spectrophotometer (Agilent Technologies, Santa Clara, CA, USA). The instrument parameters were set as follows: $\lambda_{\text{ex}} = 488$ nm (slit 5 nm), $\lambda_{\text{em}} = 510$ nm (slit 10 nm). Readings were carried out at room temperature. Relative fluorescence units (RFU) values were calculated by subtracting fluorescence units (FU) of negative control samples from that given by the sample.

2.6 Quantitation of Gag::eGFP and VLP concentrations by p24 ELISA

Gag::eGFP polyprotein concentrations were determined by p24 ELISA using the commercially available kit Innotest® HIV antigen mAb (Innogenetics NV, Gent, Belgium). The assay was performed according to manufacturer's instructions. Concentrations obtained by ELISA were corrected by an underestimation factor of 36, as previously reported [20]. This correction factor takes into account the underestimation of the kit when detecting the Gag::eGFP polyprotein and the difference in molecular weight between p24 and Gag::eGFP.

2.6 Nanoparticle Tracking Analysis

NTA-based Gag::eGFP VLP quantification and characterization was performed using a NanoSight®NS300 (Nanosight Ltd., Amesbury, UK) equipped with a blue filter module (488 nm) and a neutral filter at the Soft Material services of the Institut de Ciència de Materials de Barcelona (ICMAB-CSIC, Bellaterra, Catalonia, Spain). Samples were previously diluted to a concentration of approximately 10^8 particles/mL. Sample injection was performed using a pump in order to improve the robustness of the measurement by continuous addition, and to minimize the photobleaching effect due to fluorescence depletion over time. 60 second videos were recorded at RT and analyzed with the NTA 3.4 software (Malvern Panalytical, Malvern, UK). Tracked particles size was determined from its Brownian motion. Three independent experimental replicas were carried out for each sample. Camera level and detection threshold were manually adjusted for each replica.

2.7 Flow virometry

Flow virometry was used for the stability analysis of the clone 10H9. VLP quantification by Flow Virometry was performed in a CytoFLEX (Beckman Coulter Inc., Brea, CA, USA) equipped with a violet side scatter (V-SSC) 405 nm filter configuration. The threshold of area trigger signal in V-SSC was set to 300 and the laser gains were set as follows: 72 for FSC, 135 for SSC, 9 for V-SSC, and 500 for FITC. Samples to be analyzed were diluted in 0.22 µm filtered HyClone™ PBS (GE Healthcare, Chicago, IL, USA) in order to be in the range of 500 - 5000 events/µL with an abort rate below the 5%. 300.000 events were analyzed per sample at a flow rate of 10 µl/mL. V-SSC vs B525-FITC density plots were used to gate the VLPs. Results were analyzed with the CytExpert v.2.3 software (Beckman Coulter Inc., Brea, CA, USA) and VLP concentrations were calculated with Equation (4), where $C_{NTA/FV}$ is the standard for data harmonization between NTA and Flow Virometry instruments.

$$\text{VLP conc.} \left(\frac{\text{VLPs}}{\text{mL}} \right) = \frac{\text{events}}{\mu\text{L}} \cdot \frac{1000 \mu\text{L}}{\text{mL}} \cdot \text{Dilution} \cdot C_{NTA/FV} \quad (4)$$

2.8 Digital droplet PCR (ddPCR)

2.8.1 Genomic DNA extraction and digestion

Genomic DNA was purified using Wizard Genomic DNA Purification Kit (Promega, WI, Madison, USA) according to manufacturer's instructions. Purity and concentration of the extracted genomic DNA was assessed with a Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA).

A pre-digestion of the purified genomic DNA was performed in order to ensure that all potentially linked tandem gene copies would be randomly and independently distributed into the ddPCR droplets. For this purpose 2000 ng of each cell line gDNA were restricted with 14 units of HindIII (New England Biolabs, Ipswich, MA, USA) in a total digestion mixture of 50 µl for 1 hour at 37°C. HindIII was selected as it does not cut inside the RPP30 nor Gag::eGFP designed amplicons, is a high-fidelity enzyme with low star activity, has no

methylation-sensitivity, and presents 100% activity in NEBuffer r2.1 (New England Biolabs, Ipswich, MA, USA) which does not interfere with ddPCR. No inactivation or further purification was performed.

2.8.2 Primer design for ddPCR

Two pairs of primers were designed to target and amplify fragments of 101 and 141 bp from Gag::eGFP and Ribonuclease P Protein Subunit P30 (RPP30) genes respectively, using the online PrimerQuest Tool from IDT (Integrated DNA Technologies IDT-DNA, Leuven, Belgium) website (<https://eu.idtdna.com/pages/tools>). The salt-adjusted Tm and primer self-complementarity were studied using the OligoCalc v3.27 online calculator [21] (<http://biotools.nubic.northwestern.edu/OligoCalc.html>) and primer-dimer estimation was assessed by using Multiple Primer Analyzer tool from ThermoFisher (Waltham, MA, USA) website (<https://www.thermofisher.com/>). RPP30 sequence was retrieved from the NCBI GenBank database (Gene ID: 10556, National Centre for Biotechnology Information, Bethesda, MD, USA) website (www.ncbi.nlm.nih.gov/genbank/). RPP30 assay primer sequences were (forward primer) 5'-CCAACCTCATGCCACCCAGACCATC-3' and (reverse primer) 5'-CCCGCGCTAGGAATCAGACCAACAC-3'. Gag::eGFP assay primer sequences were (forward primer) 5'-GTCCAGGAGCGCACCATCTTCTTC-3' and (reverse primer) 5'-TTCAGCTCGATGCGGTTCACCAAG-3'.

2.8.3 qPCR for ddPCR annealing determination

In order to optimize the annealing temperatures of the primers to be used in the ddPCR, a qPCR with a temperature gradient of 55.4 to 60.7 °C was performed in a C1000 Touch thermal cycler (#1851197, Bio-Rad, Hercules, CA, USA). Each 10 µl PCR mixture contained 2x iTaq Universal SYBR Green (#1725151, Bio-Rad, Hercules, CA, USA), primers at a final concentration of 200 nM (IDT-DNA, Leuven, Belgium) and 80 ng of gDNA. The content was then transferred to a 96-well PCR plate, sealed and cycled using the following conditions: 95 °C for 5 min (1 cycle); 95 °C for 30 s and 55.4–60.7 °C for 1 min (40 cycles); 4 °C for 5 min, 90 °C for 5 min (1 cycle), 4 °C hold. On the basis of this assessment, the annealing temperature was determined to be 58 °C for both RPP30 and Gag::eGFP primers. Additionally; its melting curves showed a single peak,

indicating the presence of a single specific amplicon, as it is crucial to avoid unspecific amplifications such as primer dimers when performing ddPCRs. The tested gDNA samples had Cq values varying between 20.5 and 23.5 cycles at 58 °C. As a consequence, samples were diluted for the subsequent ddPCRs in order to avoid saturation by an excessive number of positive droplets.

2.8.4 ddPCR for Copy Number Variation (CNV) determination

Each ddPCR mixture contained 2x EvaGreen ddPCR Supermix (#1864034, Bio-Rad, Hercules, CA, USA), the corresponding pair of primers at a final concentration of 200 nM (IDT-DNA, Leuven, Belgium) and 16 ng of gDNA (as the previous qPCR assay determined). The 20 µl PCR mixtures were loaded into DG8 Cartridges together with 70 µl of Droplet Generation Oil (#1863009 and #1863005 respectively, Bio-Rad, Hercules, CA, USA). After gasket attachment, the cartridges were placed into the QX200 Droplet Generator (#1864002 Bio-Rad, Hercules, CA, USA), which generated ~20.000 droplets per sample. The content was then transferred to a 96-well PCR plate, sealed by means of a PX1 PCR Plate Sealer (#1814000, Bio-Rad, Hercules, CA, USA) and cycled using the following conditions: 95 °C for 5 min (1 cycle); 95 °C for 30 s and 58°C for 1 min (40 cycles); 4 °C for 5 min, 90 °C for 5 min (1 cycle), 4 °C hold. The QX200 Droplet Reader (#1864003, Bio-Rad, Hercules, CA, USA) was used to read the individual cycled droplets. The described process was carried on at Laboratori de Luminescència i Espectroscòpia de Biomolècules from the Universitat Autònoma de Barcelona.

The generated data was analyzed using QuantaSoft Software 1.7.4.0917, which determined the absolute quantity of target DNA molecules by fitting the number of positive and negative droplets after amplification (>31.000 accepted droplets per condition) into a Poisson distribution. The used statistical analysis took into account and compensated for the droplets presenting two or more target DNA copies due to the random nature of partitioning when droplets were generated.

2.9 Cryopreservation

The cryopreservation mixture consisted of 7.5% DMSO (Sigma-Aldrich, Saint Louis, MO, USA), 46.25% of conditioned medium and 46.25% of fresh medium.

Each cryotube™ (ThermoFisher, MA, USA) contained 1 mL of cryopreservation mix and $5 \cdot 10^6$ cells. Cryotubes were placed in a Mr. Frosty™ (ThermoFisher, MA, USA) to reach -80°C with a rate of 1°C/minute. After 24 hours, cryotubes were stored in liquid nitrogen.

3. Results and discussion

3.1 Toxicity assay

The pGag::eGFP plasmid used for the generation of the cell line, codes for the Gag::eGFP gene under the CMV promoter, whose expression derives in the Gag-based VLP production. It also contains the aminoglycoside phosphotransferase gene (NeoR/KanaR), which confers resistance to neomycin and G418 (geneticin) in eukaryotic cells. Therefore, in order to generate the HEK293-GagEGFP random stable cell line, a cellular pool under G418 selection was carried out. Since antibiotic effect and tolerance differs between cell lines, toxicity profiles were performed with the aim to characterize cellular response to geneticin and further determine its minimal selective concentration. The chosen concentration range was from 25 µg/ml to 1225 µg/ml, as determined from bibliography [22].

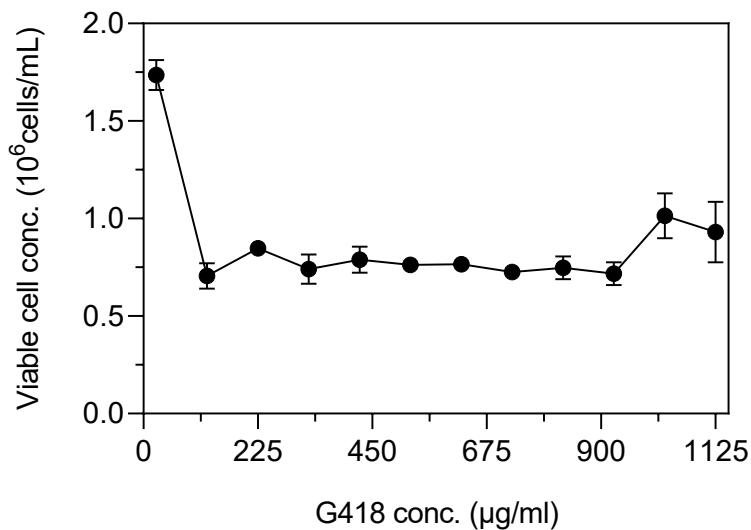


Figure 1. Geneticin (G418) toxicity assay. Viable cellular concentrations after being cultured for 48 hours under different puromycin concentrations are shown. Above 25 µg/ml of G418, no cellular growth was observed.

Cellular growth was only observed at the 25 µg/ml condition over the two days in which the assay was performed (Figure 1). 125 µg/ml was stipulated as the minimal geneticin concentration for selection experiments with this specific HEK293 cell line. Nevertheless, 225 µg/ml was the chosen concentration for pool generation and cellular selection in order to increase the selective pressure.

3.2 Pool generation

To generate the stable cell line producing Gag::eGFP-VLPs, pGag::eGFP plasmid was transfected both circular and linearized (Figure 2) into suspension HEK293 cells as previously described in 6-well plates. Delivered DNA remain as an episome in the transfected cells with a few proportion of it being illegitimately incorporated into the host cell genome [23]. If the DNA is not randomly integrated, it is diluted over time after each cell division. Therefore, after several days under G418 selection, the surviving cells will be the ones that spontaneously integrated the plasmid in their genome.

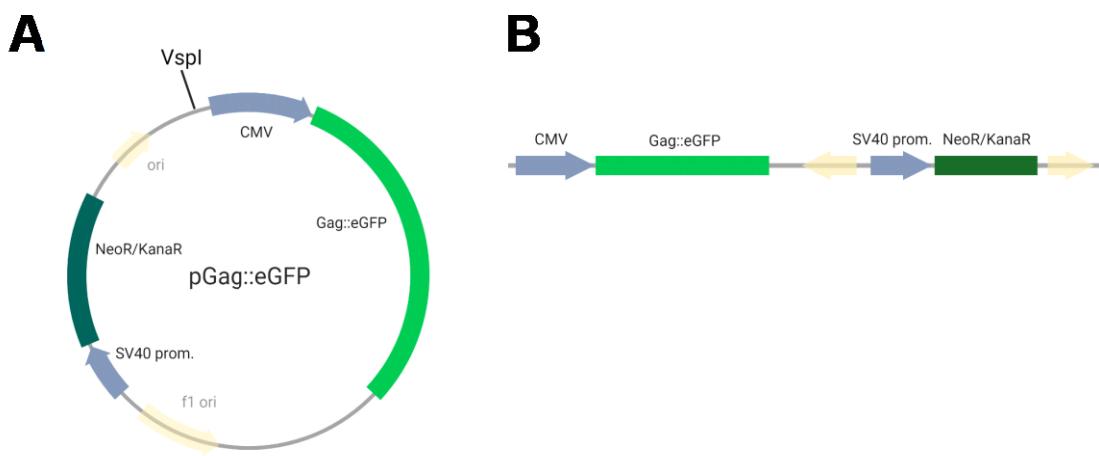


Figure 2. Schematic representation of the DNA used for the transfection. (A): pGag::eGFP circular plasmid. (B): pGag::eGFP vector linearized upon VspI restriction.

Initially, no selective pressure was added to the cell culture in order to allow the recovery of the transfected culture and the expression of the resistance gene by the cell machinery [24]. Transfection efficiency was assessed by flow cytometry, concluding that 30.7% of the circularly transfected and 24.4% of the linearly transfected cells were expressing the Gag::eGFP polyprotein at two days post-transfetion (dpt). We hypothesize that this difference may be due to cellular

exonuclease activity which degrades the DNA [25], diminishing transfection efficiency.

At 2 dpt, cells were diluted to $0.5 \cdot 10^6$ viable cells/mL and supplemented with G418 for a final concentration of 225 µg/mL (Figure 3). At 5 dpt, a significant drop in cellular viability due to the death of the non-transfected population was expected, but not observed. Contrary to this, high viabilities suggested a lack of antibiotic efficiency. For this reason, selective G418 concentration was increased to 450 µg/mL (Figure 3). After increasing the antibiotic concentration, cell growth stopped and a significant viability drop was observed.

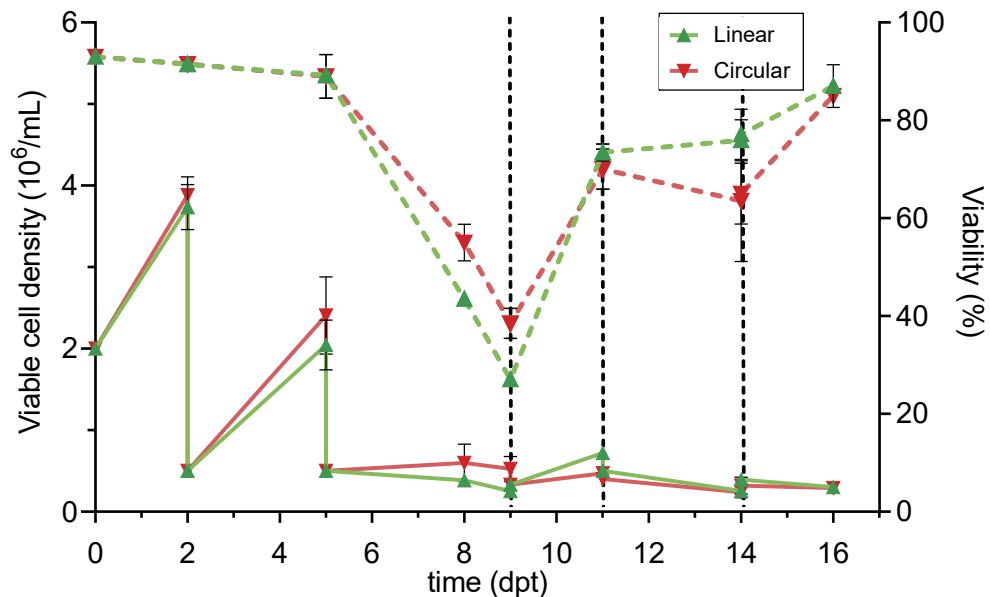


Figure 3. Cellular pool evolution: cell concentrations (solid lines) and viabilities (dotted lines) of HEK293 cells transfected with linear (green) or circular (red) plasmid. At 2dpt cells were diluted and supplemented with G418 at 225 µg/mL. At 5 dpt cells were diluted and supplemented with G418 at 450µg/mL. Medium exchanges are indicated by vertical dotted lines.

The viability drop experienced by cultures from 5 dps to 9 dps was caused by the death of the non-transfected population and transfected cells losing its expression by plasmid dilution and degradation. Only a very small part of the transfected population illegitimately integrated the construct on its genome, as it is a rare event [26]. From day 9 on, the positive selection of those cells was traduced in an increase of viability. However, pools were not showing cell growth. At 16 dpt, viability was totally recovered and therefore, cell pools were

clonally sorted. At this point, no significant differences were observed between linearly and circularly transfected pools. We can conclude that using a linear plasmid was as efficient as using a circularized plasmid for the generation of a stable pool by illegitimate integration.

3.3 Clonal selection: cell sorting

The surviving geneticin-resistant population was composed by cells that integrated the pGag::eGFP plasmid or at least its antibiotic resistance on their genome. In order to sort the polyclonal pool into individual cells with Gag::eGFP expression, 672 GFP+ clones were isolated using fluorescence-activated single cell sorting (FACS) in 96-well plates, as shown on Table 1. Two different gating criteria -coded as ‘medium’ or ‘high’ expressers- were used depending on fluorescein isothiocyanate (FITC) values, as indicated in Table 1 and represented in Figure 4. The more restrictive the FITC criteria were, the fewer cells from the pool matched it.

Table 1. Gating criteria: Transfection group, gating, FITC threshold, percentage of the cellular population meeting the criteria, number of isolated clones and survival.

| Transfected group | Gating criteria | FITC gating criteria | % of pool population | Isolated clones | Growing clones | Survival (%) |
|-------------------|-----------------|----------------------|----------------------|-----------------|----------------|--------------|
| Linear | Medium | >70 | 31.19% | 192 | 24 | 12.24% |
| | High | >220 | 9.94% | 192 | 30 | 15.63% |
| Circular | Medium | >80 | 23.88% | 192 | 32 | 16.67% |
| | High | >220 | 7.77% | 96 | 16 | 16.67% |

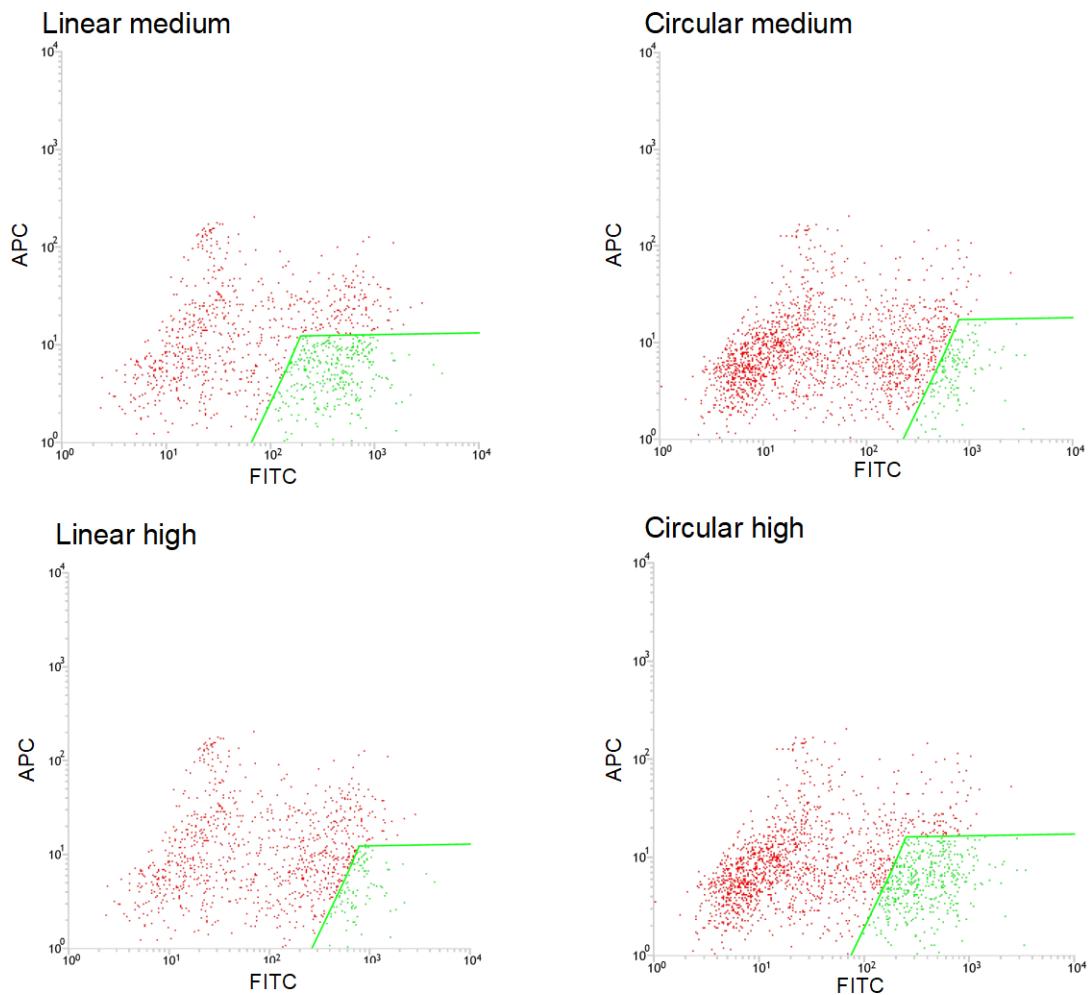


Figure 4. Gating criteria used in the FACS cell sorting of the linearly and circularly transfected cellular pools based on fluorescein isothiocyanate (FITC) values.

3.3 Clone expansion

In order to determine the monoclonal growth medium and conditions, preliminary tests were carried out with non-transfected HEK293. Monoclonal growth was only observed in plastic-adherent plates with FBS supplementation, as the HEK293 cellular growth from a unique isolated cell is a difficult event. Therefore, the fluorescent selected clones were grown in adherent 96-well plates with 10% fetal bovine serum (FBS) supplemented medium. During the following days, wells containing dead or non-dividing clones were discarded. At 7 days post sorting (dps) G418 was added at 450 µg/mL. Medium replacement of wells containing cells showing cellular division was carried out every seven days. From the 672 isolated clones, 102 showed cell growth and at 18 dps were subsequently amplified to 24 well plates. As shown in Table 1, no significant

differences in clone survival were observed between clones selected from the different pools.

In order to generate a cell library with the 102 different clones, cells were harvested and cryopreserved when reached confluence in 6-well plates. In order to verify its Gag::eGFP production, clones were analyzed at different stages of its amplification using fluorescent confocal microscopy (Figure 5).

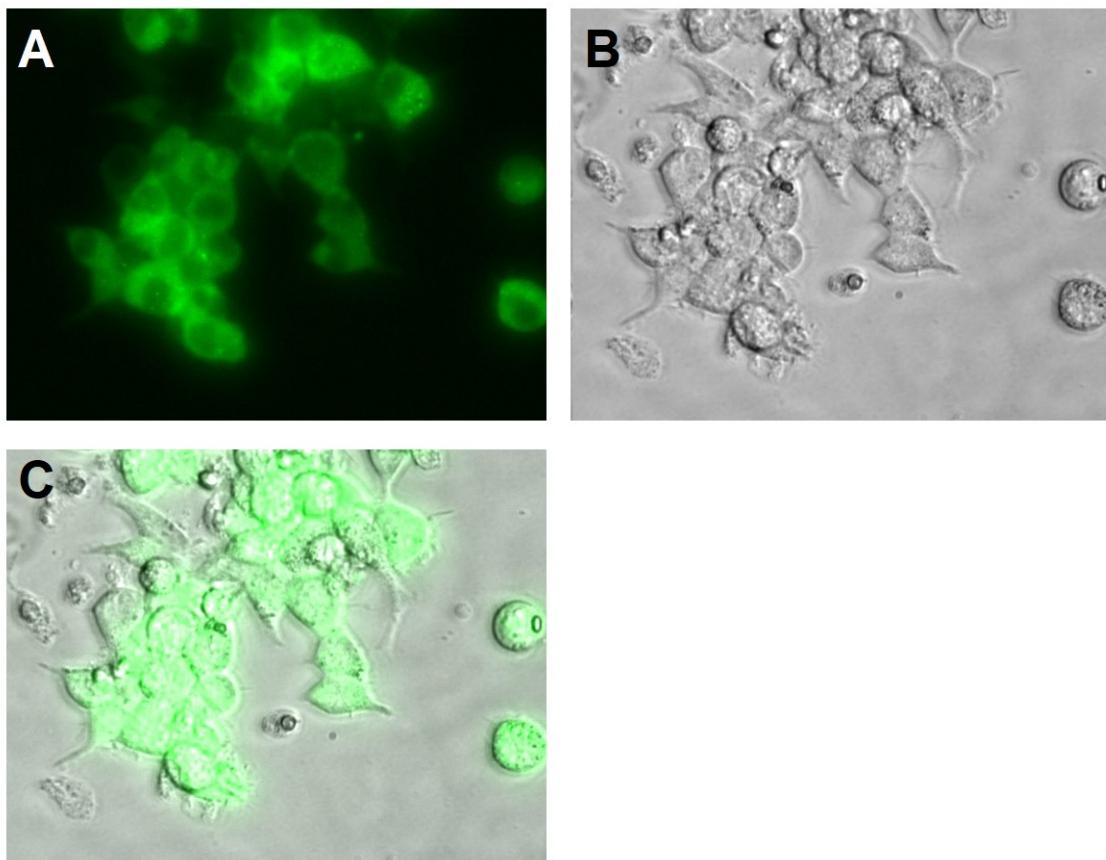


Figure 5. Confocal microscopy image (40X) of clone 9A7 at 24-well plate. All qualitatively observed clones showed positive fluorescence. (A): Green fluorescence channel (B): Bright-field (C): Green fluorescence and bright-field merge.

3.4 Gag::eGFP production quantification of the clones

The spectrofluorimetric quantitation analysis of the supernatants recovered from clones showing confluence at 6-well plates, allowed to generate an initial list of all the clones based on its Gag::eGFP production. The 34 clones exhibiting higher GFP fluorescence in the supernatant were subsequently analyzed by p24 enzyme-linked immunosorbent assay (ELISA). The results of the p24

ELISA allowed the classification of all the clones according to its Gag::eGFP production levels (Figure 6).

The ELISA test detects the HIV-1 Gag polyprotein without being able to discriminate if it is present in a monomeric form or as a part of a VLP. This can affect the reliability of the results in cultures presenting low viabilities, as dead cells release free monomers to the culture. Nevertheless, all the analyzed clones showed high viabilities; hence the mentioned phenomenon was not significantly affecting the analysis.

Figure 6 shows the selected 34 clones arranged by their Gag::eGFP production. As can be observed, clones derived from linear- or circular-DNA transfected pools are evenly distributed indicating that both methods are equally efficient for the generation of random integration stable cell lines. The 5 clones exhibiting higher production levels were 4A7, 10H9, 9A7, 9C1 and 3A1 (Figure 6).

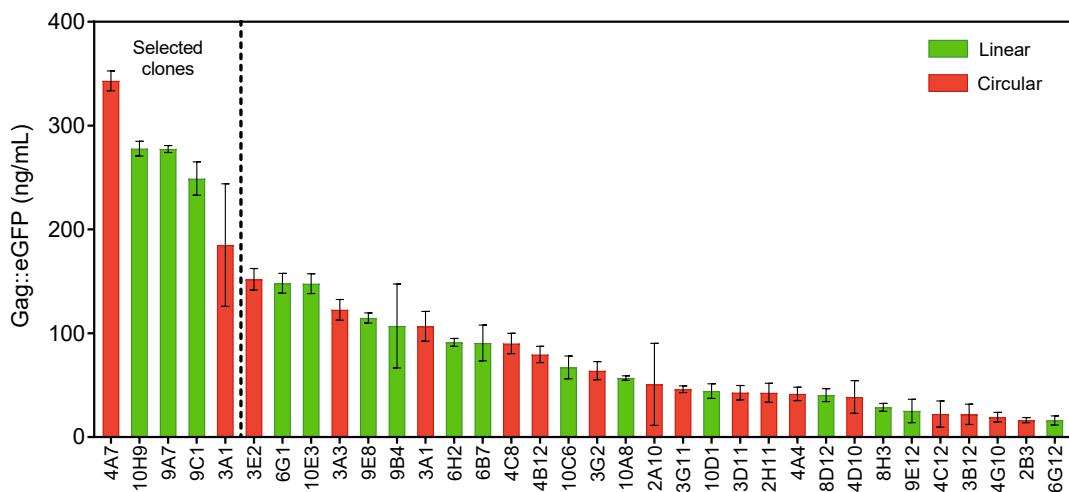


Figure 6. Selected 34 clones sorted by their Gag::eGFP supernatant concentration. Cell lines derived from linearly (green) or circularly (red) transfected pools are evenly distributed, indicating that both methods are equally efficient for the generation of random integration stable cell lines.

3.5 Suspension adaptation

As previously mentioned, to allow monoclonal growth clones were sorted and amplified in adherence. In order to achieve high density culture (hence, improve VLP production) and improve the scalability of the bioprocess by switching from

static to shaken conditions [27], the best cell lines were adapted to suspension-growth.

The 5 selected clones were defrosted in adherence. Adherent cultures showing more than 90% viability at 80-95% of confluence were seeded after a medium exchange at $0.5 \cdot 10^6$ viable cells/mL in 125 mL polycarbonate Erlenmeyer flasks, supplemented with 10% FBS and 450 µg/mL of G418. Cells were routinely diluted to $0.5 \cdot 10^6$ viable cells / mL when its concentration was higher than $1 \cdot 10^6$ viable cells. When no significant cell growth was observed, media was exchanged every 4 days.

As can be observed in Figure 7, clone suspension adaptation can be divided into three different phases: an initial cellular growth stop accompanied by a viability drop, a lag phase with no cellular growth in which the viability starts rising and a final cellular growth indicating the adaptation of the clone. A cryopreserved bank of the suspension adapted clones 10H9, 9A7, 3A1 and 4A7 was performed. Clone 9C1 failed to adapt to suspension, so it was discarded. Clones were considered adapted to suspension culture when attained a constant growth rate with viabilities higher than 95% after 3 passages.

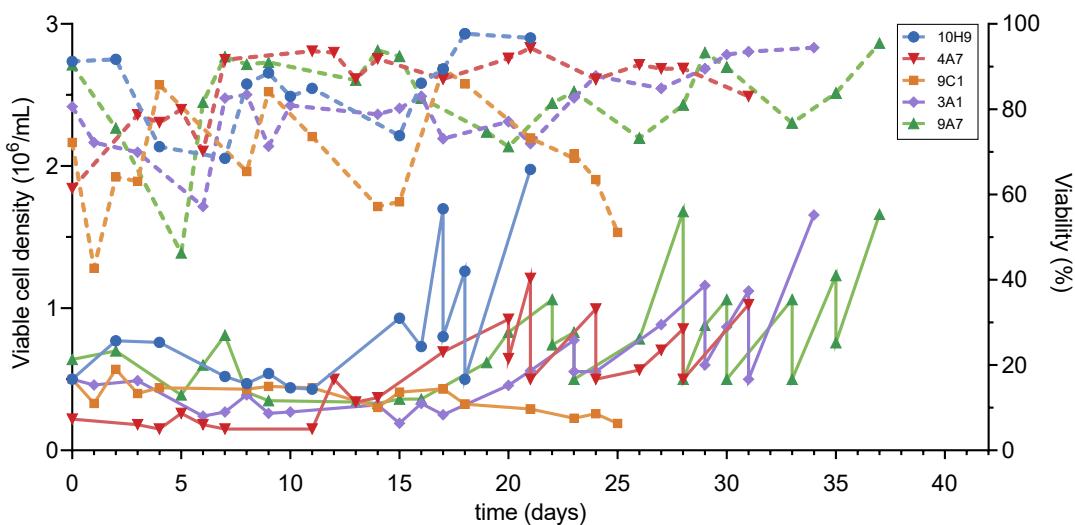


Figure 7. Suspension adaptation of the clones 10H9, 4A7, 9C1, 3A1 and 9A7. Full lines: viable cell density, dotted lines: viability. The viability of the different cell lines improves gradually over time, evidencing suspension adaptation. Clone 9C1 was discarded, as failed to adapt to suspension growth.

3.6 Serum-free media adaptation

Decades ago, FBS was widely used as a supplement in cell culture media. Even nowadays its use is necessary for specific cultures or procedures, as for the HEK293 clonal growth from single cells. Serum is an ill-defined animal-derived mixture with a variable composition which differs between lots, can present prions, growth inhibitors, endotoxins and non-attenuated viruses [28]. It's non-compatible with GMP procedures and interferes with downstream processes. For those reasons, clones were adapted to a chemically-defined serum-free media, which nowadays is the state of the art for drug production.

The adaptation of the adherent cell lines into FBS-free media was carried by gradual reduction of serum concentration from 10% to 0%. With this approach, suspension-adapted cell clones experienced a sequential adaptation. In contrast to the suspension adaptation, all cell lines successfully underwent the process with no significant changes in cell growth or viability (Figure 8). Clones were considered adapted to serum-free media after 3 passages without FBS and viabilities above 90%. A master cell bank for each cell line was generated at this point.

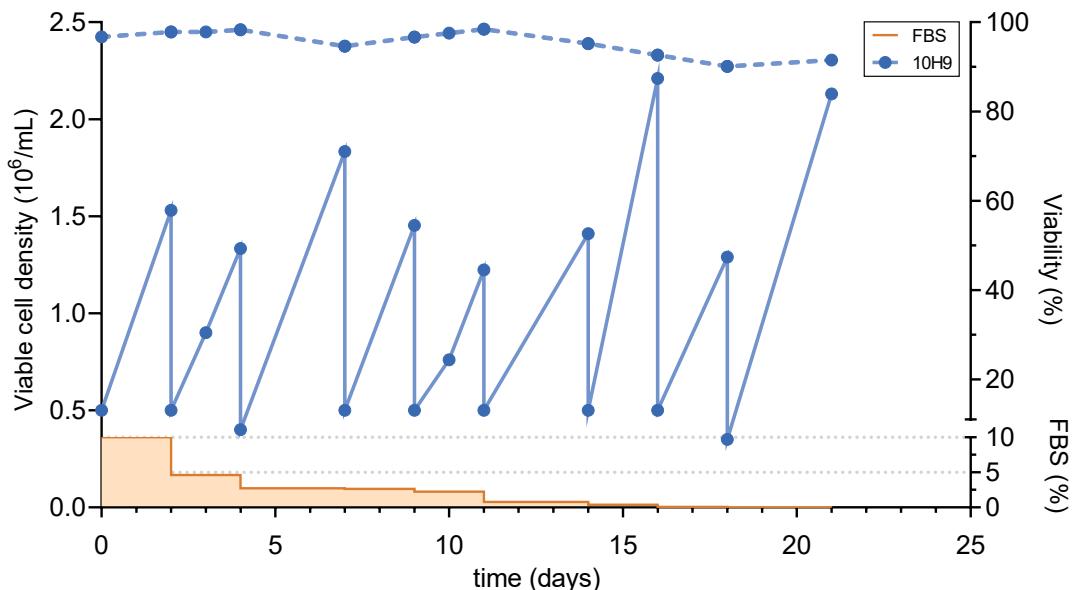


Figure 8. Example of the serum free adaptation of one clone, in this case 10H9. Full blue lines: viable cell density, dotted lines: viability, full orange lines: FBS presence in the media (in percent).

3.7 Selection of the best performing clone

Next, growth kinetic studies of the best four clones were performed (Figure 9), in order to determine its duplication time, maximum cell density, volumetric productivity and specific productivity (Figure 10).

As shown in Figure 9 and presented in Table 2, clones 10H9 and 3A1 reached the highest maximum cell density, of approximately $6 \cdot 10^6$ cells/mL, at 144 h and 192 h respectively. Clone 4A7 also reached its maximum cell density at 192 h, but together with 9A7, they reached approximately $5 \cdot 10^6$ cells/mL. Cell growth rate differed between cell lines: 10H9 showed the fastest duplication time (29 h), while 3A1 and 9A7 duplicated approximately every 38 h. 4A7 was the clone with the slowest growth, presenting a duplication time of 50 h. For comparison, wild-type HEK293 cells in the same conditions had duplication times of 26.2 h, peaking at $3.53 \cdot 10^6$ cells/mL (Table 2).

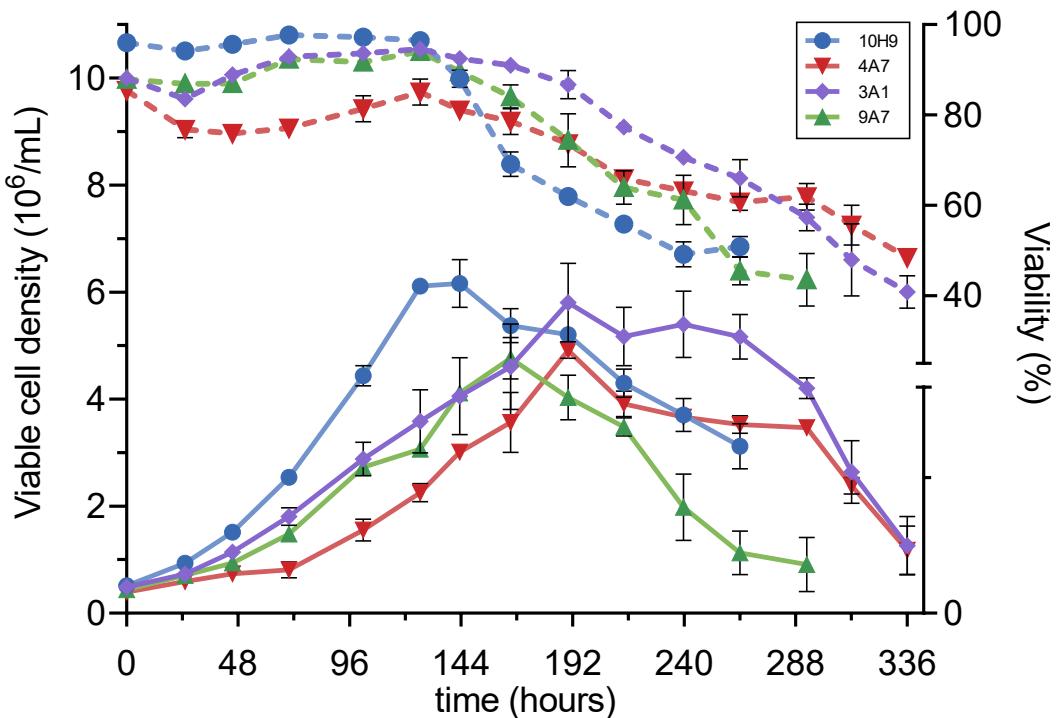


Figure 9. Growth kinetics of the suspension and serum-free adapted clones 10H9, 4A7, 3A1 and 9A7.

Table 2. Duplication time and maximum cell density of the suspension and serum-free adapted clones and HEK293 wild-type cells for reference.

| Clone name | Duplication time (hours) | Max. cell density (viable cells / mL) |
|------------|-----------------------------|--|
| 10H9 | 29.2 ± 0.3 | (6.16 ± 0.4)·10 ⁶ |
| 4A7 | 50.3 ± 2.1 | (4.92 ± 0.2)·10 ⁶ |
| 3A1 | 37.2 ± 2.1 | (5.80 ± 0.7)·10 ⁶ |
| 9A7 | 38.8 ± 2.3 | (4.77 ± 0.4)·10 ⁶ |
| HEK293 | 26.2 ± 1.0 | (3.53 ± 0.3)·10 ⁶ |

In order to determine VLP productions and productivity of the generated cell lines, their supernatants were analyzed at the cell density peak, 48 hours before and 24 hours later by ELISA.

VLP presence in the supernatant (Figure 10.A) and specific productivity (Figure 10.B) significantly differed between clones. This variability in production levels was concordant with what we might expect when comparing different clones generated by illegitimate integration: copies of the gene of interest (GOI) reside in different genomic loci between cell lines, so its transgenic expression will be differently affected by the genomic context. 10H9 production levels were 6.7 fold higher than those of the clone 4A7, the worst producer in terms of specific productivity (Table 3).

Overall, clone 10H9 not only showed the best growth profiles in terms of maximum cell density and growth rate, but also in terms of volumetric and specific VLP productivities (Table 3), so it was selected as the best cell line generated by illegitimate recombination in this work.

Table 3. Volumetric and specific productivity of the suspension serum-free adapted clones.

| Clone name | Volumetric productivity* (VLPs · mL ⁻¹ · day ⁻¹) | Specific productivity (VLPs · cell ⁻¹ · day ⁻¹) |
|------------|--|---|
| 10H9 | (1.61 ± 0.13)·10 ⁸ | 35.7 ± 2.9 |
| 4A7 | (2.4 ± 0.2)·10 ⁷ | 4.88 ± 0.4 |
| 3A1 | (9.2 ± 2.5)·10 ⁷ | 17.1 ± 4.6 |
| 9A7 | (1.3 ± 0.24)·10 ⁸ | 26.5 ± 4.9 |

* At cell density peak

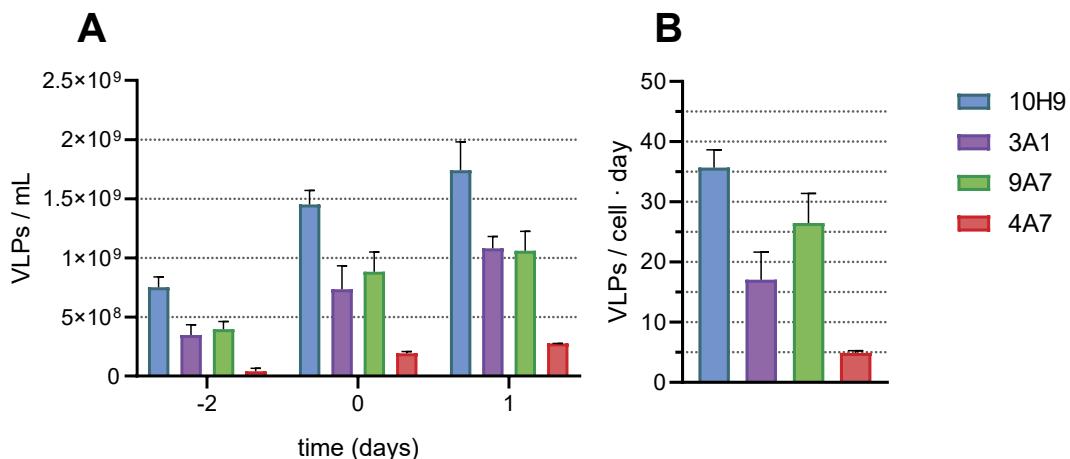


Figure 10. (A): VLP concentration of the four best clones 48 h prior, during, and 24 h after cell density peak. (B): Specific productivity of the best four clones.

3.10 Genomic analysis: number of integrated copies

Copy number variation (CNV) is a phenomenon in which the presence or number of copies of a particular DNA segment varies relative to a reference genome. In order to fully characterize the generated cell line, the 10H9 CNV was compared to its wild-type predecessor HEK293 for Gag::eGFP and a reference gene. This was assessed using digital droplet PCR (ddPCR), an ultrasensitive system capable of DNA absolute quantification overcoming inherent limitations of qPCR and microarray techniques when analyzing CNV. With this technique, different PCR samples for every condition containing the exact same amount of diluted genome were partitioned each into more than 20.000 nanoliter-sized independent droplets. Then, they were PCR amplified until the reaction end-point and every droplet was analyzed in order to determine if it was positive (contained the amplified DNA region) or negative (did not present the amplicon) (Figure 11.A). This resulted in the determination of the copy number for each studied construct. The Ribonuclease P Protein Subunit P30 (RPP30) gene from wild-type HEK293 was used as reference assay, as it has been reported to present two copies in human diploid genomes [29–31].

As shown in Figure 11.A, more than 30.000 unique end-point PCR events were analyzed for each gene and cell line. Knowing that RPP30 gene is present with two copies in the wild-type HEK293 genome, the number of copies per genome of the other studied DNA regions was assessed. As figure 11.B shows, clone

10H9 contained 4 copies of the RPP30 gene compared with the two copies present in the HEK293 parental cell line. 10H9 also presented four copies of the randomly integrated Gag::eGFP construct per genome, doubling its RPP30 copy number (Table 4).

It has been described that HEK293 can suffer chromosomal translocations or duplications involving significant parts of its genome [32]. The presented results suggest that, at some point, the whole or some part of the 10H9 cellular genome was duplicated. We hypothesize that this phenomenon may have occurred at the single 10H9 parental cell before its transfection, followed by the subsequent random integration of four Gag::eGFP constructs. Another theory is that shortly after the integration of two Gag::eGFP constructs, the whole genome was duplicated. In either case, this happened at very early stages of the cell line development, since the obtained copy number of the cell line is exactly four copies per genome. Contrarily, if the genome duplication had occurred in late clonal amplification phases, the population would present heterogeneity in the copy number of RPP30 gene, which would result in RPP30 CNV varying from two to four copies per genome when analyzed.

Table 4. Copies per microliter of the analyzed RPP30 and Gag::eGFP genes, its ratios and number of copies per genome.

| | HEK293 | 10H9 |
|--------------------------------|----------------|-----------------|
| RPP30 (copies/ μ l) | 54.3 \pm 4.0 | 108.9 \pm 3.9 |
| Gag::eGFP (copies/ μ l) | 0.0 | 108.0 \pm 6.6 |
| RPP30/HEK-RPP30 | 1.000 | 2.005 |
| Gag::eGFP/RPP30 | 0 | 0.991 |
| RPP30/HEK-RPP30 (nearest half) | 1 | 2 |
| Gag::eGFP/RPP30 (nearest half) | 0 | 1 |
| RPP30 copies/genome | 2 | 4 |
| Gag::eGFP copies/genome | 0 | 4 |

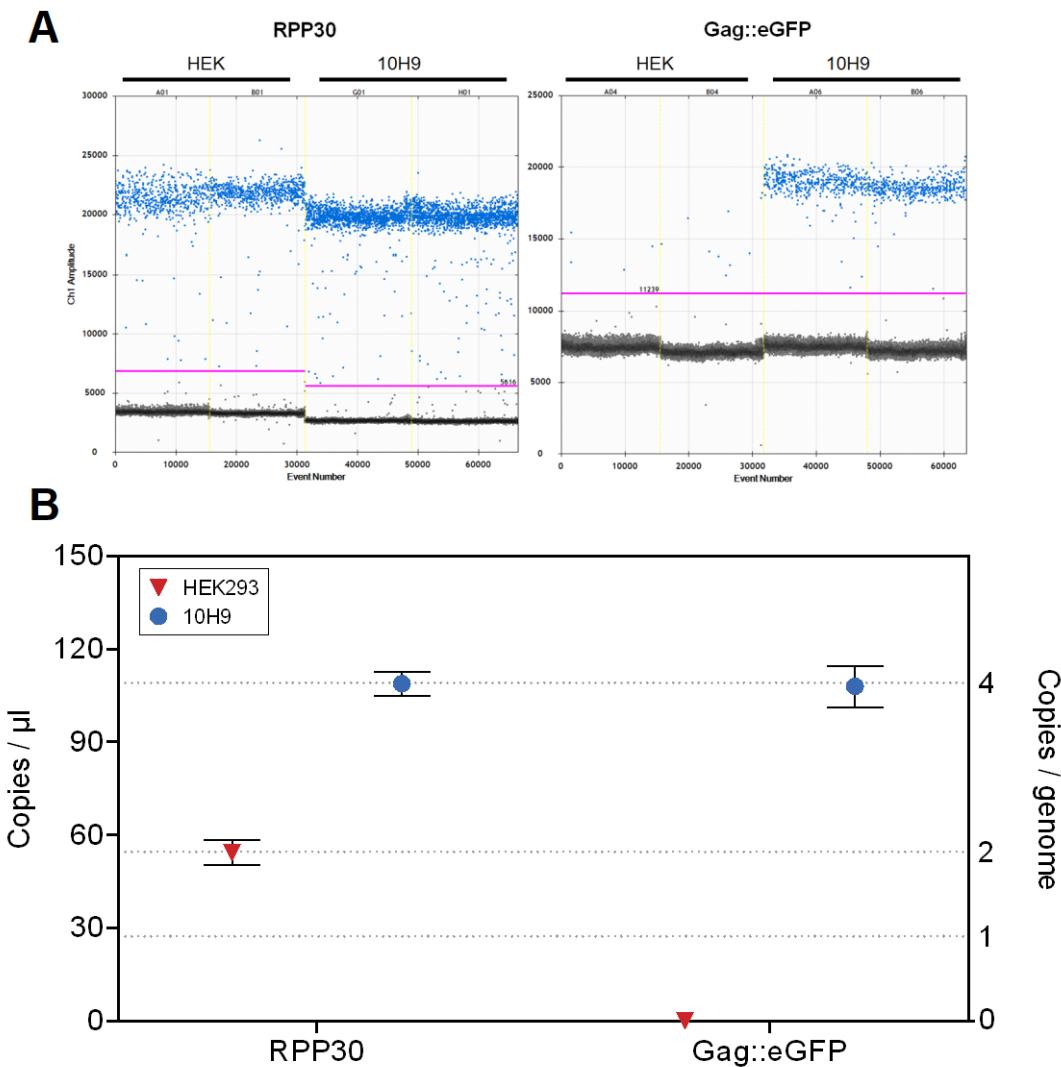


Figure 11. Data from absolute quantification of the 10H9 cell line for the RPP30 and Gag::eGFP genes. To estimate the average VCN values, the concentrations (copies/ μ l) of Gag::eGFP and the RPP30 reference gene were determined in 10H9 and HEK293 cell lines. Error bars indicate Poisson confidence limits.

Hoechst stained 10H9 cell line presented a majority of binucleated cells when observed under confocal microscopy (Figure 12), contrarily with what was observed in parental HEK293 cells. A transition from mono-nuclear to binuclear could explain the occurred genome duplication.

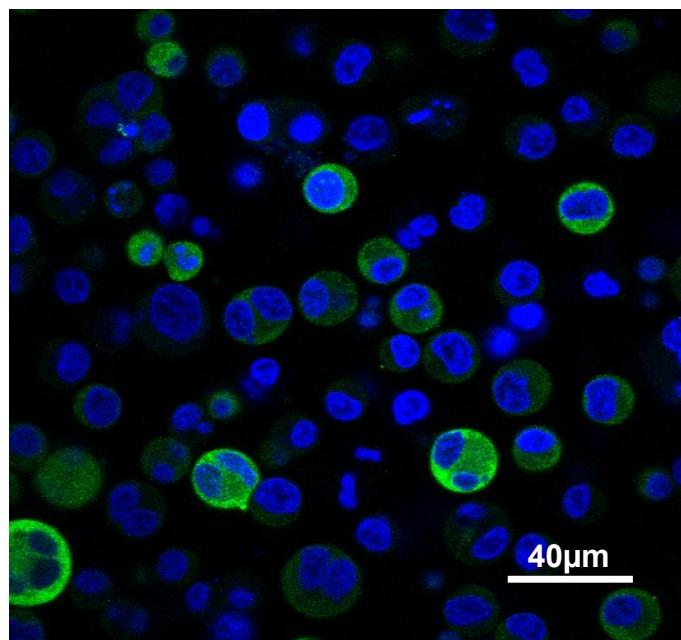


Figure 12. Confocal microscopy image of the binucleated 10H9 cell line. Gag::eGFP can be observed in green. Cell nuclei were stained with Hoechst (blue). If the 3D distribution of the two nuclei matched with the captured 2D plane, two nuclei could be clearly appreciated.

3.11 Stability analysis

In order to fully characterize the best clone, its stability in terms of VLP production was analyzed as it is a relevant parameter to determine its quality as a cell line. Clone 10H9 was passaged during 60 days, while analyzing its cellular growth and viability (Figure 13.A). Mean fluorescence, Gag::eGFP concentration in the supernatant and productivity was also assessed by flow virometry. As can be observed in Figure 13.B, the studied cell line experienced a slight tendency to decrease its mean GFP cellular fluorescence (analyzed by flow cytometry). However, this was not translated in a significant decrease in specific Gag::eGFP-VLP productivity levels, which were sustained at around 35 VLPs/(cell·day) throughout the experiment (Figure 13.B).

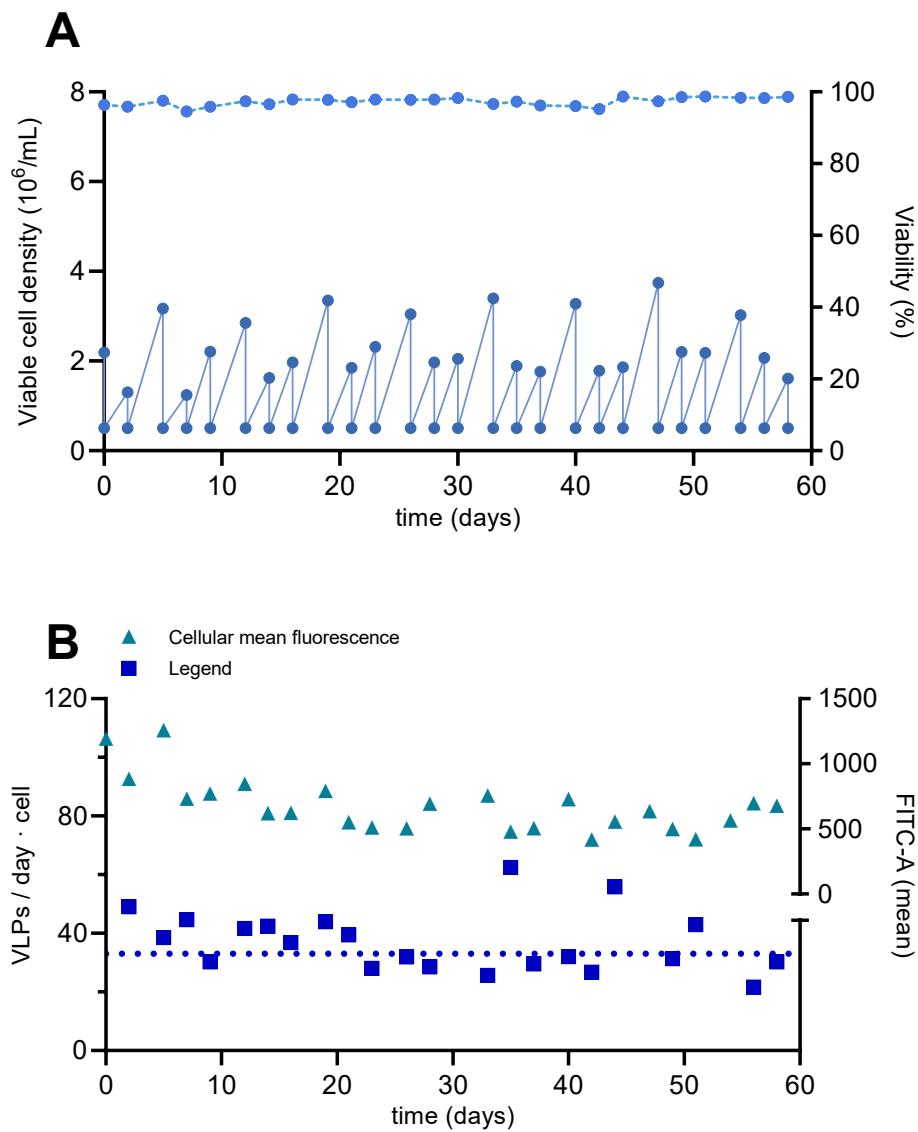


Figure 13. Stability analysis of the 10H9 cell line. **(A):** Cellular concentration (solid line) and viability (dotted line) during the studied 60 days. **(B):** Gag::eGFP-VLP production stability: mean cellular fluorescence was reduced during the studied 60 day period, while specific productivity level was maintained constant.

4. Conclusions

This work consisted in the generation of a stable cell line for the production of Gag::eGFP VLPs using illegitimate integration as approach. HEK293 cells were transfected with a construct coding for the Gag::eGFP transgene of interest and the G418 resistance gene in order to achieve its integration into the genome. The transfection was performed using both lineal and circular DNA in order to study the pros and cons of each approach. The transfection efficiency on the

transfected group with a lineal plasmid (24.4%) was moderately inferior comparing it with the cells transfected with a circular plasmid (30.7%), probably due to DNA degradation by exonucleases [25]. The resulting pools were challenged under G418 selection for 16 days in order to positively select the clones with genomic integration. Results showed no significant difference between cells transfected with linear or circular DNA during the generation of the pools, concluding that both methods were equally efficient for this purpose. When cell viability levels recovered, pools were sorted into individual clones in adherence, with 10% FBS supplementation. Clones were catalogued as: "high" or "medium" expressers depending on its GFP fluorescence levels. From the 672 isolated clones, in 102 it was observed cellular growth, with no significant difference in survival ratios between cells transfected with linear (\approx 14%) or circular DNA (\approx 16.6%). Clones were amplified and the study of its Gag::eGFP production levels by ELISA allowed the selection of the five best candidates for its suspension and serum-free adaptation. Four out of the five clones were successfully adapted to suspension and serum-free culture. Growth kinetics of the adapted clones were studied in order to determine its growth dynamics and Gag::eGFP production, which significantly varied between cell lines. 10H9 was selected as the best clonal cell line, since it presented the lowest duplication time (29.2 h), the highest maximum cell density ($6.16 \cdot 10^6$ viable cells/mL) and the best Gag::eGFP-VLP expression in terms of volumetric and specific productivities (35.7 VLPs/cell·day). The study of the copy number for the reference gene RPP30 suggested that the cell underwent whole genome duplication (probably at some point of its early clonal stage), at it presented four RPP30 copies per genome instead of the expected two. 10H9 presented 4 copies of the Gag::eGFP construct per genome. The clone stability study concluded that it maintained its VLP production unaltered throughout the studied 60 days.

5. Acknowledgements

The author wants to thank Dr. Javier Fuenmayor for his mentoring and help. We would like to thank Dr. Amine Kamen (McGill University, Montreal, Canada) for providing the HEK 293 SF-3SF6 cells. The pGag::eGFP (Cat#11468) reagent from Dr. Merilyn Resh was obtained through the National Institutes of Health

AIDS Reagent Program, Division AIDS, National Institute of Allergy and Infectious Diseases, National Institutes of Health. We would like to thank Dr Jose Amable Bernabé from the Institut de Ciència de Materials de Barcelona (ICMAB-CSIC, Bellaterra, Catalonia, Spain) for his support in nanoparticle tracking analysis techniques. The support of Núria Barba Bosch from the Servei de Microscòpia (Universitat Autònoma de Barcelona, Bellaterra, Catalonia, Spain) in the development of optical microscopy images and analysis is greatly appreciated. We would also like to thank Manuela Costa Garcia from the Servei de Cultius Cel·lulars, Producció d'Anticossos i Citometria (Universitat Autònoma de Barcelona, Bellaterra, Catalonia, Spain) for her help with flow cytometry and cell sorting; and Salvador Bartolomé Piñol from Laboratori de Luminescència i Espectroscòpia de Biomolècules (Universitat Autònoma de Barcelona, Bellaterra, Catalonia, Spain) for his help with qPCR and ddPCR analyses. Figures and Graphs were generated using GraphPad Prism version 8.0.2 for Windows (San Diego, CA, USA) and GIMP 2.10.28 (<http://www.gimp.org>).

References

- [1] Roldão, A., Mellado, M. C. M., Castilho, L. R., Carrondo, M. J. & Alves, P. M. Virus-like particles in vaccine development. *Expert Rev. Vaccines* **9**, 1149–1176 (2010)
- [2] Kushnir, N., Streatfield, S. J. & Yusibov, V. Virus-like particles as a highly efficient vaccine platform: Diversity of targets and production systems and advances in clinical development. *Vaccine* **31**, 58–83 (2012)
- [3] Venereo-Sánchez, A. *et al.* Hemagglutinin and neuraminidase containing virus-like particles produced in HEK-293 suspension culture: An effective influenza vaccine candidate. *Vaccine* **34**, 3371–3380 (2016)
- [4] Fuenmayor, J., Gòdia, F. & Cervera, L. Production of virus-like particles for vaccines. *N. Biotechnol.* **39**, 174–180 (2017)
- [5] Cervera, L. *et al.* Generation of HIV-1 Gag VLPs by transient transfection of HEK 293 suspension cell cultures using an optimized animal-derived component free medium. *J. Biotechnol.* **166**, 152–165 (2013)
- [6] Boix-Besora, A., Lorenzo, E., Lavado-García, J., Gòdia, F. & Cervera, L. Optimization, Production, Purification and Characterization of HIV-1 GAG-Based Virus-like Particles Functionalized with SARS-CoV-2. *Vaccines* **10**, 1–18 (2022)
- [7] Lavado-García, J. *et al.* Metabolic engineering of HEK293 cells to improve transient transfection and cell budding of HIV-1 virus-like particles. *Biotechnol. Bioeng.* **118**, 1649–1663 (2021)
- [8] Salimzadeh, L., Jaberipour, M., Hosseini, A. & Ghaderi, A. Non-viral transfection methods optimized for gene delivery to a lung cancer cell. *Avicenna J. Med. Biotechnol.* **5**, 68–77 (2013)
- [9] Cervera, L., Gutiérrez-Granados, S., Berrow, N. S., Segura, M. M. & Gòdia, F. Extended gene expression by medium exchange and repeated transient transfection for recombinant protein production enhancement. *Biotechnol. Bioeng.* **112**, 934–946 (2015)

- [10] Geisse, S. & Fux, C. *Chapter 15 Recombinant Protein Production by Transient Gene Transfer into Mammalian Cells. Methods in Enzymology* vol. 463 (Elsevier Inc., 2009).
- [11] Kim, T. K. & Eberwine, J. H. Mammalian cell transfection: The present and the future. *Anal. Bioanal. Chem.* **397**, 3173–3178 (2010)
- [12] Wurm, F. M. Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat. Biotechnol.* **22**, 1393–1398 (2004)
- [13] Hacker, D. L. & Balasubramanian, S. Recombinant protein production from stable mammalian cell lines and pools. *Curr. Opin. Struct. Biol.* **38**, 129–136 (2016)
- [14] González-Domínguez, I., Grimaldi, N., Cervera, L., Ventosa, N. & Gòdia, F. Impact of physicochemical properties of DNA/PEI complexes on transient transfection of mammalian cells. *N. Biotechnol.* **49**, 88–97 (2019)
- [15] Büssow, K. Stable mammalian producer cell lines for structural biology. *Curr. Opin. Struct. Biol.* **32**, 81–90 (2015)
- [16] Würtele, H., Little, K. C. E. & Chartrand, P. Illegitimate DNA integration in mammalian cells. *Gene Ther.* **10**, 1791–1799 (2003)
- [17] Mutskov, V. & Felsenfeld, G. Silencing of transgene transcription precedes methylation of promoter DNA and histone H3 lysine 9. *EMBO J.* **23**, 138–149 (2004)
- [18] Hermida-Matsumoto, L. & Resh, M. D. Localization of Human Immunodeficiency Virus Type 1 Gag and Env at the Plasma Membrane by Confocal Imaging. *J. Virol.* **74**, 8670–8679 (2000)
- [19] Schwartz, S. *et al.* Mutational inactivation of an inhibitory sequence in human immunodeficiency virus type 1 results in Rev-independent gag expression. *J. Virol.* **66**, 7176–7182 (1992)
- [20] Gutiérrez-Granados, S., Cervera, L., Gòdia, F., Carrillo, J. & Segura, M. M. Development and validation of a quantitation assay for fluorescently

tagged HIV-1 virus-like particles. *J. Virol. Methods* **193**, 85–95 (2013)

- [21] Kibbe, W. A. OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res.* **35**, W43–W46 (2007)
- [22] Vicens, Q. & Westhof, E. RNA as a drug target: The case of aminoglycosides. *ChemBioChem* **4**, 1018–1023 (2003)
- [23] Olsen, P. A., Gelazauskaite, M., Randøl, M. & Krauss, S. Analysis of illegitimate genomic integration mediated by zinc-finger nucleases: Implications for specificity of targeted gene correction. *BMC Mol. Biol.* **11**, (2010)
- [24] Lin, C. Y. *et al.* Enhancing protein expression in HEK-293 cells by lowering culture temperature. *PLoS One* **10**, 1–19 (2015)
- [25] Lejeune, F., Li, X. & Maquat, L. E. Nonsense-mediated mRNA decay in mammalian cells involves decapping, deadenyinating, and exonucleolytic activities. *Mol. Cell* **12**, 675–687 (2003)
- [26] Manivasakam, P., Aubrecht, J., Sidhom, S. & Schiestl, R. H. Restriction enzymes increase efficiencies of illegitimate DNA integration but decrease homologous integration in mammalian cells. *Nucleic Acids Res.* **29**, 4826–4833 (2001)
- [27] Shridhar, S. *et al.* Transcriptomic changes in CHO cells after adaptation to suspension growth in protein-free medium analysed by a species-specific microarray. *J. Biotechnol.* **257**, 13–21 (2017)
- [28] Brunner, D. Serum-free cell culture: the serum-free media interactive online database. *ALTEX* **4**, 53–62 (2010)
- [29] Lin, H. T. *et al.* Application of Droplet Digital PCR for Estimating Vector Copy Number States in Stem Cell Gene Therapy. *Hum. Gene Ther. Methods* **27**, 197–208 (2016)
- [30] Dyavar, S. R. *et al.* Normalization of cell associated antiretroviral drug concentrations with a novel RPP30 droplet digital PCR assay. *Sci. Rep.* **8**, 1–11 (2018)

- [31] De Oliveira, M. F. *et al.* Comparative analysis of cell-associated HIV DNA levels in cerebrospinal fluid and peripheral blood by droplet digital PCR. *PLoS One* **10**, 1–10 (2015)
- [32] Lin, Y. C. *et al.* Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. *Nat. Commun.* **5**, (2014)

Chapter four

**Stable HEK293 cell line generation for the production of
Gag::eGFP VLPs with a site-directed approach using
CRISPR / Cas9 technology**

Abstract: Gag-VLP production by transient gene expression (TGE) in HEK293 cells has been broadly studied, since they incorporate human-compatible post-translational modifications, which is crucial for their use as vaccines. Nevertheless, their production by stable gene expression (SGE) is an approach that can be of great impact. SGE is a more valuable strategy for its monoclonal nature, its prolonged expression over time and in comparison to TGE, SGE is not restricted to the use of transfection compatible media. Genomic Safe Harbors (GSHs) are useful targets for the integration of transgenes in the human genome as they can accommodate genetic constructs without their silencing and guaranteeing a solid and predictable long-term expression. There are currently three main well-described GSHs, from which good results in SGE have been previously reported for the adeno-associated virus site 1 (AAVS1). CRISPR/Cas9 gene editing methodology is a relatively recent developed tool which has facilitated cell line generation in human cell. This methodology relies on the Cas9 nuclease guided by a customized RNA molecule (sgRNA) to generate Double Stranded Breaks (DSBs) at the specific locus of interest in which the gene wants to be inserted. The DSB then triggers the eukaryotic cell repair mechanisms which will insert a donor sequence if this is flanked by homology arms. In this work we established a methodology for the generation of HEK293 cell lines to stably produce Gag-based VLPs via the locus-directed integration of the Gag::eGFP gene at AAVS1 GSH. This was achieved by co-transfection of plasmids encoding for the Cas9 nuclease, a sgRNA targeting the AAVS1 locus and the construct of interest flanked by AAVS1 homology arms. Clonal cell line candidates were isolated and the more promising ones were successfully adapted to suspension and serum-free media. From the eight adapted clones, analysis of the growth kinetics of the four best producers allowed the selection of the best clone which had a duplication time of 34.6 ± 1.4 hours, a maximum cell density peak of $3.96 \cdot 10^6 \pm 0.2$ viable cells/mL and a specific productivity of 222.9 VLPs/cell·day. Its expression remained stable during 33 days, the copy number of the integrated gene was assessed by ddPCR, and the incorporation of the construct at the AAVS1 locus with no undesired mutations was confirmed by PCR and DNA sequencing.

Keywords: VLPs, Stable gene expression, CRISPR, HEK293, Gag::eGFP, HIV-1

1. Introduction

There are two main approaches to develop a stable cell line: random gene integration and directed gene integration. The first method takes advantage of the spontaneous random integration into the host cell genome, while the second consists in a rationally-designed locus-specific genetic integration. The locus of integration plays a major role on the expression of the recombinant gene, known as the “position effect” phenomenon. However, the transgene can be integrated into inactive heterocromatin regions or be silenced, an inactivation phenomenon which has been associated with histone hipoacetylation, methylation variations at histone H3, and promoter CpG DNA methylation [1]. To overcome possible position effect drawbacks and not having to depend on a fortuitous optimal integration [2], targeting gene integration into specific pre-defined predictable transcriptionally active locus such as genomic safe harbors (GSHs), constitutes an attractive strategy [3].

GSHs are regions in the human genome that can accommodate the integration of functional transgenes without its silencing or undesired interactions with proximal chromosomal regions, and with a solid and predictable expression [4]. This predictability of outcome makes GSHs a useful target for the directed integration of transgenes. Three main GSHs that support long-term gene expression have been identified and reported: the adeno-associated virus site 1 (AAVS1), the chemokine receptor 5 (CCR5) gene locus and the human orthologue of the mouse ROSA26 locus (hRosa26) [5,6]. In this work, AAVS1 GSHs was chosen [5]. Integration at AAVS1 locus disrupts an intron of the gene PPP1R12C (Phosphatase 1 regulatory subunit 12C) from the human chromosome 19, which encodes for a protein with unknown function [4,7]. As AAVS1 disrupts PPP1R12C gene at 23 intron 1 (Figure 4), if an antibiotic resistance with no promoter is integrated in this locus, it will be expressed using the endogenous PPP1R12C promoter. This allows to select the cells with cassette integration at AAVS1 GSH. No gross abnormalities have been reported in human cells harboring transgenes in AAVS1 [4,7].

Integration of a transgene into a specific locus can be achieved via Homologous Recombination (HR) if the construct of interest is surrounded by homology arms

with the genomic region to be inserted at [8]. Nevertheless, relying in a spontaneous event such as HR -which takes place in less than one per 10^5 targeted cells- is tedious and inefficient [9]. However the induction of double strand breaks (DSB) in the desired locus greatly stimulates and increases HR occurrence [9]. Several targeted-location restriction enzymes have been developed for this purpose. Those technologies rely on the guide of homologous binding proteins or RNA. The most widely described ones are Zinc Finger Nucleases (ZFN), transcription activator-like effector nucleases (TALENs), and clustered regulatory interspaced short palindromic repeats associated with Cas9 (CRISPR/Cas9) [10].

When a DSB is created, it triggers two endogenous eukaryotic mechanisms of repair: non-homologous end joining (NHEJ) and the homology-directed repair (HDR) (Figure 1). The first one reseals the cut by joining the two broken DNA ends in a highly error-prone manner, causing common insertions or deletions (indels) of nucleic acids which disrupt the original sequence [11–13]. Alternatively, the repair can occur via HDR if a homologous DNA template (endogenous or donor) is present in the same cell. This precision mechanism uses the DNA template as a guide to make the repair [10,13]. Hence, HDR can be exploited to generate stable cell lines through directed gene integration.

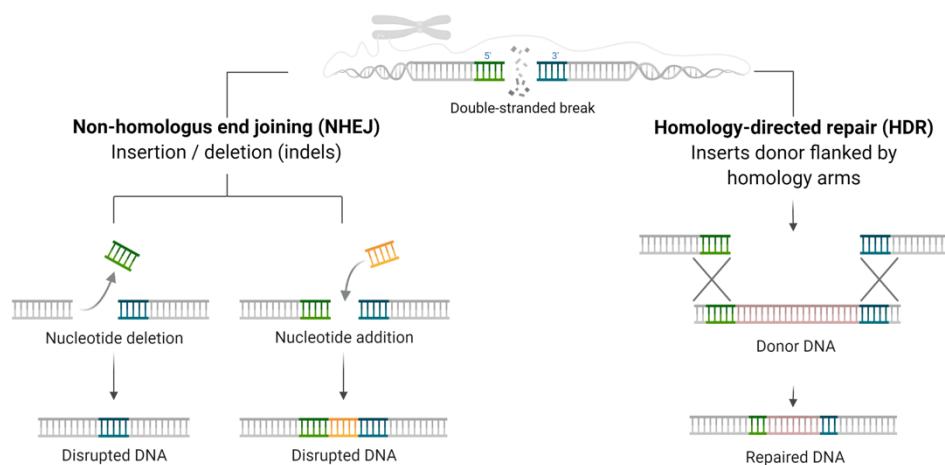


Figure 1. Endogenous eukaryotic double-strand break (DSB) repair mechanisms. The DSB can be repaired using the error-prone non-homologous end joining (NHEJ) mechanism which generates, or by homologous direct repair (HDR) if a complementary copy of the region is present.

CRISPR / Cas9 is an Archea and Bacteria defense mechanism that works as an immune system against viruses and plasmids (Figure 2). When a pathogen enters the cells, they incorporate segments of the invader's DNA into their own genome separated by interspaced repeats. The locus where the intruder DNA is placed is named CRISPR, an acronym for clustered regularly interspaced short palindromic repeats [13,14]. CRISPR locus will transcribe the so-called crRNAs if the cell is subsequently re-infected by the same pathogen. With about 20 RNA nucleotides, crRNAs will have a sequence perfectly complementary to the intruder's pathogen genome. Then, the crRNA will interact with a secondly expressed tracrRNA attracting the CRISPR-associated (Cas9) nuclease, which will bind and cut the foreign nucleic acids [13]. The nuclease Cas9 will bind and cleave the target DNA sequence only if there is a protospacer adjacent motif (PAM) sequence, consisting of any base followed by two guanines (NGG). PAM, distinguishes bacterial self from non-self DNA, thereby, preventing the cellular CRISPR locus from being self-destroyed by the Cas9 [5].

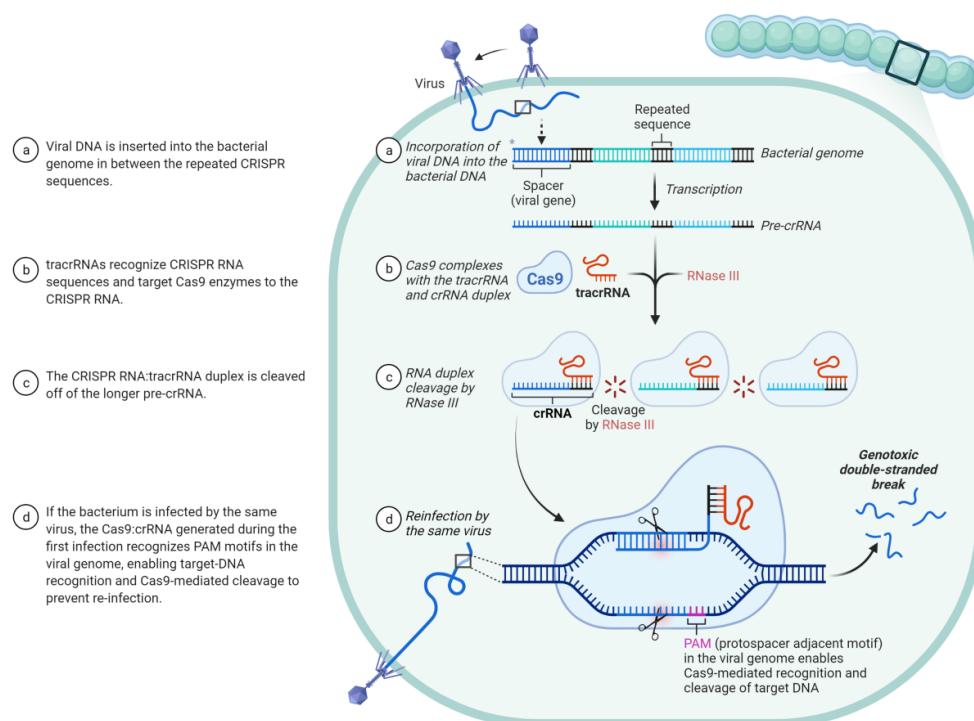


Figure 2. Schematic representation of CRISPR-Cas9 Adaptive Immune System of *Streptococcus pyogenes* against Viruses. Adapted from "Streptococcus' Adaptive Immune System against Viruses: CRISPR-Cas9", by BioRender (2022).

The discovery of the CRISPR / Cas9 system (from now on simply referred as CRISPR) has been one of the major milestones in the last decade, since Cas9 nuclease has been found to harbor a huge potential as a genetic engineering tool. Compared to previous nuclease technologies such as ZFN or TALENs, CRISPR adds a major fundamental benefit: the specificity is determined by the crRNA instead of the nuclease itself, which significantly simplifies its use requiring no protein engineering. In order to facilitate its design and application to different cell platforms -such as human cells-, tracrRNA and crRNA can be fused into a single guide RNA (sgRNA). sgRNA has approximately 100 nucleotides in length, of which 20 (the crRNA part) determine the 1:1 specificity with the DNA target sequence [13].

The simplicity, efficiency, flexibility and relative affordability of the system has made a huge impact in all biotechnology fields, and still generates high expectations about its clinical potential. Significant progress in this technology as well as delivery modalities for gene therapy applications are thought to trigger its clinical translation. In addition, CRISPR has become a very useful tool for the generation of locus-specific stable cell line generation.

To sum up, in order to achieve directed CRISPR genomic insertion in the AAVS1 locus, the presence of three elements is mandatory:

- Cas9 nuclease: is responsible for the generation of the DSB on the genomic DNA.
- sgRNA: binds and guides the Cas9 nuclease to the targeted locus of interest.
- Repair template: consists of the gene of interest surrounded by arms which present homology to the locus of interest. It is used as repair template by cellular HDR mechanisms, thus integrating the construct of interest.

In this work we generated a HEK293 cell line for the stable expression of Gag::eGFP via its locus-directed genetic integration at AAVS1 GSH using CRISPR/Cas9 technology (Figure 3). Cells were transfected with pT2 plasmid encoding for the Cas9 nuclease and a well described and validated sgRNA targeting AAVS1 locus [5], named T2. A second plasmid (pDonor) containing

the construct of interest flanked by AAVS1 homology arms allowed its integration via HDR (Figure 4). A cellular pool with puromycin-supplemented media was generated to select the clones with the donor construct (Puro-CMV-Gag::eGFP-pA) integration at the AAVS1 locus. Then, cells were clonally isolated, expanded and adapted to suspension and serum-free media. The best resulting clones were subsequently characterized and the best candidate was selected as the final CRISPR-generated stable cell line for the expression of Gag::eGFP VLPs.

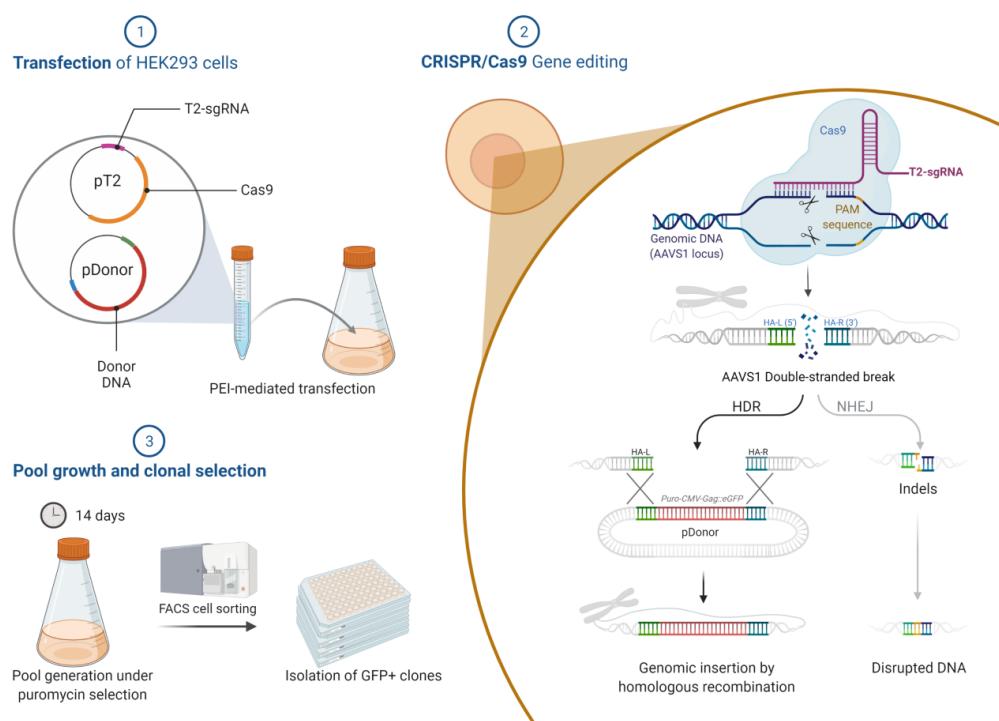


Figure 3. Schematic representation of the generation of a stable cell using CRISPR-Cas9. (1): Cell co-transfection with pT2 plasmid (encoding for the Cas9 nuclease and the T2-sgRNA targeting AAVS1 locus) and pDonor plasmid (containing the donor construct of interest flanked by AAVS1 homology arms). (2): Donor construct integration at the AAVS1 locus DSB generated by the Cas9 nuclease via HDR. (3): Pool generation under puromycin selection and clonal isolation for its subsequent expansion and characterization.

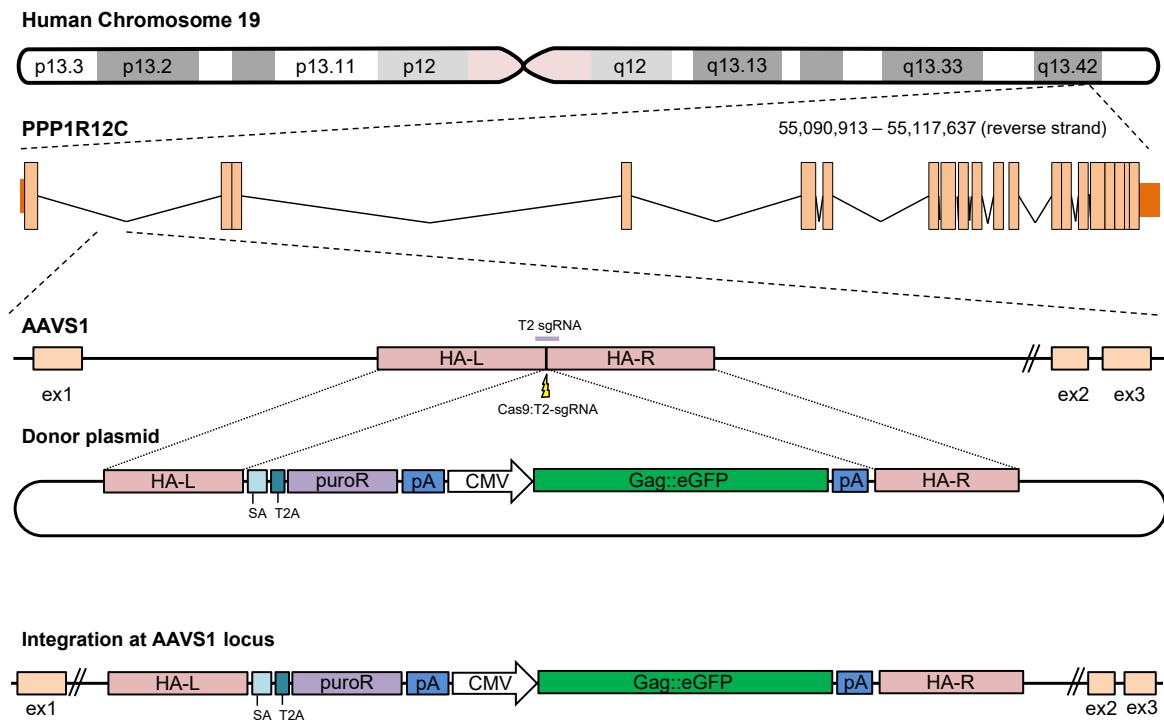


Figure 4. Integration at the genomic AAVS1 safe harbor using CRISPR-Cas9. From top to bottom: PPP1R12C gene location on human Chromosome 19. AAVS1 location between exons 1 and 2 of the PPP1R12C gene, where Cas9:T2-sgRNA complex will generate a DSB (yellow arrow). Donor plasmid carrying the construct to be integrated flanked by HA-L and HA-R homology arms. Construct of interest successfully integrated at the AAVS1 locus.

2. Materials and methods

2.1 Cell line, media and culture conditions

A serum-free suspension-adapted HEK293 cell line (HEK293SF-3F6) was used, kindly provided by Dr. Amine Kamen from the Biotechnology Research Institute at the National Research Council of Canada (Montreal, Canada) and McGill University. This cell line was derived from a Current Good Manufacturing Practice (cGMP) master cell bank available for manufacturing of clinical material.

The medium used for HEK293 cellular growth was the chemically defined and free from animal components FreeStyle™ 293 (ThermoFisher, Waltham, MA, USA) supplemented with 0.1% Pluronic (ThermoFisher, Waltham, MA, USA), 1.6 mg/L of r-transferrin (Merck Millipore, Kankakee, IL), 19.8 mg/L of r-insulin (Novo Nordisk Pharmatech, Køge, Denmark) and 0.9X of an in-house developed lipid mixture to maximize cell growth [15]. For cell selection, 450

$\mu\text{g/mL}$ of geneticin (ThermoFisher, Waltham, MA, USA) were used. After cell sorting the medium was supplemented with 20% FBS, puromycin ($1.7 \mu\text{g/mL}$) and antibiotic-antimycotic (ThermoFisher, Waltham, MA, USA).

Suspension cell cultures were maintained routinely in exponential growth phase in 125ml disposable polycarbonate Erlenmeyer flasks with vent cap (Corting, Tewksbury, NY, USA) or 6-well plates (ThermoFisher, Waltham, MA, USA), placed in a LT-X Kuhner shaker (LT-X Kuhner, Birsfelden, Switzerland) shaking at 130 rpm, at 37°C , 5% CO_2 and 85% RH.

Adherent cell cultures were maintained in exponential growth phase in 25 mm^2 T-flasks, 75 mm^2 T-flasks (Sigma-Aldrich, Saint Louis, MO, USA), 96-well, 24-well and 6-well plates (Sigma-Aldrich, Saint Louis, MO, USA) in a humidified incubator at 37°C and 5% CO_2 in air. Adherent cell monolayers were re-suspended by gently hitting the T-flask and pipetting, no trypsinization was needed.

Cell counts and viability determinations were performed using the NucleoCounter NC-3000 automatic cell counter (Chemometec, Lillerød, Denmark) following manufacturer's instructions. Adherent clones and cultures were observed under the Olympus CKX41 optical inverted microscope (Olympus, Tokyo, Japan) to qualitatively analyze cell growth, morphology and vitality.

Specific growth rate (μ) was calculated using the Equation 1. Specific productivity (q_p) and volumetric productivity (v_p) were calculated using Equation 2 and 3 respectively, where P is the product, X the viable cell density (VCD), t the time, and μ the specific growth rate of the cells during the selected time interval.

$$\mu = \frac{\ln(X_t - X_0)}{t_t - t_0} \quad (1)$$

$$q_p = \frac{\int_{t_0}^t P(t) \cdot dt}{\int_{t_0}^t X(t) \cdot dt} \cdot \mu \quad (2)$$

$$v_p = P/t \quad (3)$$

2.2 Plasmids and transfection

2.2.1 Initial plasmid expression vectors

The pGag::eGFP plasmid codes for a codon optimized Rev-independent HIV-1 Gag protein fused in frame to the enhanced GFP driven by CMV enhancer and promoter. The plasmid from the NIH AIDS Reagent Program (Cat 11468) [16] was constructed by cloning the Gag sequence from pCMV55M1-10 [17] into the pEGFP-N1 plasmid (Clontech, Palo Alto, CA, USA).

The lentiCRISPR-v2 plasmid was provided by Feng Zhang (Addgene plasmid #52961; [18]). This plasmid has a lentiviral backbone and codes for the human-codon optimized Cas9 preceded by EF-1 α core promoter. The Cas9 is fused at its C-terminus with the puromycin resistance protein through a P2A self-cleaving peptide. The plasmid also codes for the U6 RNA promoter and the sgRNA scaffold. The two elements are separated by a sequence of 1889bp flanked by Esp3I restriction sites.

The AAVS1-SA-2A-puro donor plasmid was gently provided by Rudolf Jaenisch (Addgene plasmid #22075; [19]). This plasmid has two short (800bp) homology arms that target the human AAVS1 safe harbor locus containing in between, the puromycin selection cassette without its promoter.

pMock plasmid does not have any mammalian promoter or Coding DNA Sequence (CDS). It was constructed by ligation of pGag::eGFP backbone.

2.2.2 Constructed plasmids

2.2.2.1 pT2 construction

The T2 dsDNA insert to be used in the Golden Gate reaction was generated by the hybridization (95°C, 7 minutes) of a pair of designed oligonucleotides (Sigma-Aldrich, Saint Louis, MO, USA) (Table 1). Four bases were added at the 5' of each oligonucleotide to generate compatible sticky ends with the Esp3I restricted lentiCRISPR-v2 backbone.

Table 1. Oligonucleotides used in this work. Sequences added to generate sticky ends or non-complementary compatible homologies for Gibson® or NEBuilder® HiFi DNA assembly are underlined. Added Stop mutations are showed in bold font. Primers were designed avoiding possibilities of self-dimerization and primer-dimer formation.

| Name | Objective | Sequence (5' -> 3') |
|-----------------|---------------------------|--|
| T2 forward | T2 dsDNA insert | <u>CACCGGGGCCACTAGGGACAGGAT</u> |
| T2 reverse | formation | <u>AAACATCCTGTCCCTAGTGGCCCC</u> |
| SEQ_U6+T2_fwd | T2-sgRNA and surroundings | TTGCATATACGATA <u>CAAGGCTG</u> |
| SEQ_Cas9_fwd | sequencing | AAAGTGATGTCGTGTACTGG |
| T2+CAS9_fwd | PCR cloning pT2 | <u>AGCCCATA<u>TATGGAGTCCGAGATCCAGTTGGTTAATTAAAG</u></u> |
| T2+CAS9_rev | insert | <u>AATATTAACGCTTACA<u>ATTTCA</u>CTTATCGTCATCGTCTTG</u> |
| BACKBONE_fwd | PCR cloning pT2 | AAATTGTAAGCGTTAATATTTGTTAAAATTC |
| BACKBONE_rev | backbone | CGGA <u>ACTCCATATATGGG</u> |
| SEQ_GIB_1_fwd | | TTTGCTGGCCTTTGCTCAC |
| SEQ_GIB_2_rev | pT2 sequencing | CCTAGCGCCCGCTCCTTCG |
| SEQ_GIB_3_fwd | | ACCTGTTACCCTGACCAAT |
| pGag-EGFP Fwd | PCR cloning | <u>GGATGCGGTGGGCTCTATGGCGTTACATAACTACGGTAATGGC</u> |
| pGag-EGFP Rev | pDonor insert | <u>TCACCAATCCTGTCCCTAGTACGCC</u> TTAAGATA <u>CATTGATGAGTTG</u> |
| pDonor Fwd | PCR cloning | <u>ATCAATGTATCTTAAGGCGTACTAGGGACAGGATTGGTGACA</u> |
| pDonor Rev | pDonor backbone | <u>TACCGTAAGTTATGTAACGCC</u> CATAGAGCCCACCGCATCC |
| SEQ_DON_fwd | pDonor | CATCGCATTGTCTGAGTAGG |
| SEQ_DON_rev | sequencing | GGGTGTGTCACCAGATAAGG |
| Fw Junct. 5' | 5' Junction PCR | CGGA <u>ACTCTGCCCTCTAACG</u> |
| Rv Junct. 5' | | GCACCGTGGGCTTGTACTC |
| Fw Junct. 3' | 3' Junction PCR | ACTGCATTCTAGTTGTGGTTG |
| Rv Junct. 3' | | ACCCCGAAGAGTGAGTTGC |
| Fw seq 5' | 5' Q5 PCR for | CACTTCAGGACAGCATGTTGC |
| Rv 5' Junction | Sequencing | GCACCGTGGGCTTGTACTC |
| Fw 3' Junction | 3' Q5 PCR for | ACTGCATTCTAGTTGTGGTTG |
| Rv seq 3' | Sequencing | AGAGATGGCTCCAGGAAATG |
| RPP30 ddPCR Fw | Gag::eGFP CNV | CCAACCTCATGCCACCCAGACCATC |
| RPP30 dd Rv | by ddPCR | CCCGCGCTAGGAATCAGACCAACAC |
| Gag::eGFP dd Fw | RPP30 CNV by | GTCCAGGAGCGCACCATCTCTTC |
| Gag::eGFP dd Rv | ddPCR | TTCAGCTCGATGCGGTTACCGAG |

Golden Gate T2 was performed as previously published by Cermak [20] and described in Table 2. FastDigest Esp3I (10U, ThermoFisher, Waltham, MA, USA), T4 DNA ligase (10U, ThermoFisher, Waltham, MA, USA) and 2x volume

of T4 DNA Ligase Buffer (ThermoFisher, Waltham, MA, USA) were used in this reaction. Golden Gate product was transformed into competent cells, positive colonies were obtained and Sanger sequencing (Servei de Genòmica, UAB, Bellaterra, Catalonia, Spain) confirmed the correct one-copy T2 sequence incorporation between U6 and sgRNA scaffold, confirming the generation of the lentiCRISPR-v2-T2 plasmid.

PCR clonings for the generation of the two fragments to be assembled were performed using Q5® High-Fidelity DNA polymerase (New England Biolabs, Ipswich, MA, USA) following manufacturer's instructions. Oligonucleotides (IDT-DNA, Leuven, Belgium) were designed to generate overlaps of 20 nucleotides between fragments, together with the incorporation of a stop codon at the 3' end of the Cas9 gene. Two-fragment assembly was performed using the NEBuilder® HiFi DNA Assembly mix (New England Biolabs, Ipswich, MA, USA) following manufacturer's instructions with 50 ng of vector and a vector-insert ratio of 1:2. Hundreds of positive colonies were obtained with no growth on re-ligation controls. The construction of pT2 plasmid was confirmed by sequencing.

2.2.2.2 pDonor construction

pDonor was constructed by two-fragment assembly of the insert of interest CMV-Gag::eGFP-pA into the backbone pAAVS1-SA-2A-puro. Two pairs of primers (Table 1) were designed (ThermoFisher, Waltham, MA, USA) both containing homology ends with each other. PCR clonings for the generation of the two fragments to be assembled were performed using Q5® High-Fidelity DNA polymerase (New England Biolabs, Ipswich, MA, USA) following manufacturer's instructions. The purified bands extracted from the gel were mixed with Gibson® Assembly Master Mix and were incubated as indicated at Table 2. A positive clone was identified by PCR colony and the construction of the pDonor plasmid was confirmed by Sanger sequencing (Servei de Genòmica, UAB, Bellaterra, Catalonia, Spain).

Table 2. Temperatures, cycles and steps used in the Golden Gate, PCRs, qPCRs, ddPCRs and assembly reactions presented in this work.

| Reaction | Cycles: Temperature and time |
|--|--|
| Golden Gate: T2 | |
| insertion into lentiCRISPR-v2 | 10x[37°C, 5 min + 16°C, 10 min] + 37°C, 15 min + 80°C, 5 min |
| PCR cloning T2+Cas9 | 98°C, 30 sec + 10x[98°C, 8 sec + 58°C, 20 sec + 72°C, 120 sec] + |
| Insert for pT2 | 25x[98°C, 8 sec + 70°C, 20 sec + 72°C, 121 sec] + 72°C, 120 sec |
| PCR cloning Backbone for pT2 | 98°C, 30 sec + 35x[98°C, 8 sec + 60°C, 20 sec + 72°C, 78 sec] + 72°C, 2 min |
| NEBuilder HiFi DNA assembly | 50°C, 60 min |
| PCR cloning Gag::eGFP | 98°C, 30 sec + 10x[98°C, 10 sec + 65°C, 30 sec + 72°C, 78 sec] + 25x |
| Insert for pDonor | [98°C, 10 sec + 72°C, 30 sec + 72°C, 80 sec] + 72°C, 2 min |
| PCR cloning Donor Backbone for pDonor | 98°C, 30 sec + 10x[98°C, 10 sec + 65°C, 30 sec + 72°C, 78 sec] + 25x [98°C, 10 sec + 68°C, 30 sec + 72°C, 5 min] + 72°C, 2 min |
| Gibson Assembly | 55°C, 1 hour |
| PCR colony for pDonor confirmation | 98°C, 30 sec + 33x[98°C, 10 sec + 50°C, 30 sec + 72°C, 180 sec] + 72°C, 5 min |
| PCR for integration at AAVS1 (5' Junction) | 95°C, 5 min + 72°C, 10 min + [95°C, 30 sec + 48.7°C, 45 sec + 72°C, 90 sec] + 72°C, 10 min |
| PCR for integration at AAVS1 (3' Junction) | 95°C, 5 min + 72°C, 10 min + [95°C, 30 sec + 46.9°C, 45 sec + 72°C, 90 sec] + 72°C, 10 min |
| High Fidelity PCR for 5' Junction sequencing | 98°C, 30 sec + 33x[98°C, 8 sec + 69°C, 20 sec + 72°C, 9 sec] + 72°C, 2 min |
| High Fidelity PCR for 3' Junction sequencing | 98°C, 30 sec + 33x[98°C, 8 sec + 64°C, 20 sec + 72°C, 9 sec] + 72°C, 2 min |
| qPCR for ddPCR annealing determination | 95°C, 5 min + 40x[95°C, 30 sec + 55.4-60.7°C, 1 min] + 4°C, 5 min + 90°C, 5 min + 4°C hold |
| ddPCR (Gag::eGFP and RPP30 amplicons) | 95°C, 5 min + 40x[95°C, 30 sec + 58°C, 1 min] + 4°C, 5 min + 90°C, 5 min + 4°C hold |

2.2.3 Plasmid amplification, DNA purification and Genomic DNA PCRs

Plasmids were amplified in *Escherichia coli* DH5α strain grown in LB medium (Conda, Madrid, Spain) supplemented with 10 µg/mL kanamycin (Sigma, St. Louis, MO, USA) or 100 µg/ml ampicillin (Sigma, St. Louis, MO, USA) depending on the *E. coli* antibiotic resistance present on each plasmid. Plasmid

purification was carried out using the Endofree Plasmid Mega kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions.

Purification of the PCR products for PCR cloning or sequencing from agarose gel bands was carried out using Gel Extraction Kit (Qiagen, Hilden, Germany).

Genomic DNA was purified using Wizard Genomic DNA Purification Kit (Promega, WI, Madison, USA) according to manufacturer's instructions. Purity and concentration of the extracted genomic DNA was assessed with a Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). 0.5-1 μ g of genomic DNA template was used per PCR reaction. Junction analytic PCRs were performed using DreamTaq Green PCR mix (ThermoFisher, MA, USA), while PCRs used for junction amplification, purification and its subsequent sequencing were performed using Q5® High-Fidelity DNA polymerase (New England Biolabs, Ipswich, MA, USA).

2.2.4 PEI-mediated Transient transfection

Exponentially growing HEK293 cells transfected at a density of 2·10⁶ cells/mL. A media exchange was performed prior to transfection by centrifugation of the cells at 300g for 5 minutes. PEIpro® (Polyplus-transfection SA, Illkirch-Graffenstaden, France) was used as transfection reagent. PEI-DNA complexes were formed under sterile conditions. Briefly, plasmid DNA was diluted in culture media (10% of the total volume of cell culture to be transfected) for a final total DNA concentration of 1 μ g/mL and vortexed for 10 s. Then, polyethylenimine (PEI), was added for a final concentration of 2 μ g/mL (a 2:1 PEI:DNA ratio (w/w)) and vortexed three times for 3 s. The mixture was incubated for 15 minutes at RT and then added to the culture. The ratio between plasmids for each transfection group is described at Table 4.

2.3 Toxicity assays

100 μ l of HEK293 cells were seeded in a 96-well plate at 0.5·10⁶ cells/ml together with 10 μ l of increasing antibiotic concentrations by triplicate. The plate was cultured under standard HEK293 culture conditions. After two days, 20 μ l of MTT reagent (Promega, WI, USA) was added to each well and the plate was incubated for one hour at 37°C with an agitation of 130 rpm. Before the MTT

addition, a standard curve of cell concentrations from 0 to $4 \cdot 10^6$ viable cells/ml was performed in duplicated columns. Absorbance at 490 nm was read on Multilabel Plater Reader VICTOR3® (Perkin Elmer, MA, USA).

2.4 Flow cytometry and cell sorting

The transfected cellular populations were assessed by flow cytometry using a BD FACS Canto flow cytometer (BD BioSciences, San Jose, CA, USA) using its FITC-A laser, at Servei de Cultius Cel·lulars, Producció d'Anticossos i Citometria (Universitat Autònoma de Barcelona, Bellaterra, Catalonia, Spain).

Cell sorting was carried out using a fluorescence activated BD FACSJazz® Cell Sorter (BD Biosciences/ThermoFisher, MA, USA).

2.5 Spectrofluorometry-based product quantitation

Green fluorescence intensity in Gag-GFP samples was measured using a Cary Eclipse Fluorescence Spectrophotometer (Agilent Technologies, Santa Clara, CA, USA). The instrument parameters were set as follows: $\lambda_{\text{ex}} = 488$ nm (slit 5 nm), $\lambda_{\text{em}} = 510$ nm (slit 10 nm). Readings were carried out at room temperature. Relative fluorescence units (RFU) values were calculated by subtracting fluorescence units (FU) of negative control samples from that given by the sample.

2.6 Confocal microscopy

Imaging of the clones was performed using Olympus FV1000 confocal fluorescence microscope (Olympus, Tokyo, Japan) at Servei de Microscòpia (Universitat Autònoma de Barcelona, Bellaterra, Catalonia, Spain). Prior to its visualization, cells were treated with 0.1% (v/v) of Hoechst 33342 (Thermo Fisher Scientific, Waltham, MA, USA) and 0.1% (v/v) of CellMask™ Deep Red (Thermo Fisher Scientific, Waltham, MA, USA) in order to stain cell nuclei and lipid membranes, respectively. Samples were placed in 35 mm glass bottom Petri dishes with 14 mm microwell (MatTek Corporation, Ashland, MA, USA) prior to its visualization under the microscope.

2.7 Nanoparticle Tracking Analysis

NTA-based Gag::eGFP VLP quantification and characterization was performed using a NanoSight®NS300 (Nanosight Ltd., Amesbury, UK) equipped with a blue filter module (488 nm) and a neutral filter at the Soft Material Service of the Institut de Ciència de Materials de Barcelona (ICMAB-CSIC, Bellaterra, Catalonia, Spain). Samples were previously diluted to a concentration of approximately 10^8 particles/mL. Sample injection was performed using a pump in order to improve the robustness of the measurement by continuous addition, and to minimize the photobleaching effect due to fluorescence depletion over time. 60 second videos were recorded at RT and analyzed with the NTA 3.4 software (Malvern Panalytical, Malvern, UK). Tracked particles size was determined from its Brownian motion. Three independent experimental replicas were carried out for each sample. Camera level and detection threshold were manually adjusted for each replica.

2.8 Flow virometry

This technique uses Flow cytometry with an ultraviolet side scatter laser to track particles down to 100nm, which combined with a FITC-A laser can discriminate Gag::eGFP VLPs from extracellular vesicles. VLP quantification by Flow Virometry was performed in a CytoFLEX (Beckman Coulter Inc., Brea, CA, USA) equipped with a violet side scatter (V-SSC) 405 nm filter configuration. The threshold of area trigger signal in V-SSC was set to 300 and the laser gains were set as follows: 72 for FSC, 135 for SSC, 9 for V-SSC, and 500 for FITC. Samples to be analyzed were diluted in 0.22 μm filtered HyClone™ PBS (GE Healthcare, Chicago, IL, USA) in order to be in the range of 500 - 5000 events/ μL with an abort rate below the 5%. 300.000 events were analyzed per sample at a flow rate of 10 $\mu\text{l}/\text{mL}$. V-SSC vs B525-FITC density plots were used to gate the VLPs (Figure 5). Results were analyzed with the CytExpert v.2.3 software (Beckman Coulter Inc., Brea, CA, USA) and VLP concentrations were calculated with Equation (4), where $C_{\text{NTA/FV}}$ is the standard for data harmonization between NTA and Flow Virometry instruments.

$$\text{VLP conc.} \left(\frac{\text{VLPs}}{\text{mL}} \right) = \frac{\text{events}}{\mu\text{L}} \cdot \frac{1000 \mu\text{L}}{\text{mL}} \cdot \text{Dilution} \cdot C_{\text{NTA/FV}} \quad (4)$$

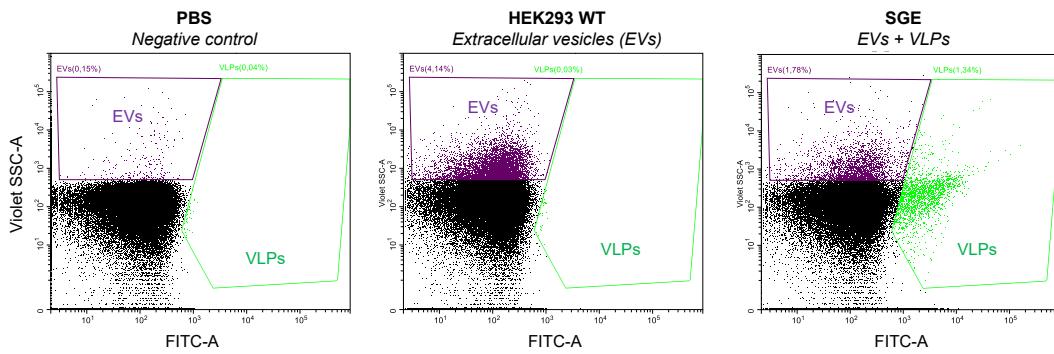


Figure 5. CytoFLEX Violet SSC-A vs FITC-A density plots showing VLP gating criteria for different samples: PBS, supernatant from a wild-type HEK293 and supernatant from a stable cell line expressing Gag::eGFP VLPs.

2.9 Digital droplet PCR (ddPCR)

2.9.1 Genomic DNA pre-digestion for ddPCR

A pre-digestion of the purified genomic DNA was performed in order to ensure that all potentially linked tandem gene copies would be randomly and independently distributed into the ddPCR droplets. For this purpose 2000 ng of each cell line gDNA were restricted with 14 units of HindIII (New England Biolabs, Ipswich, MA, USA) in a total digestion mixture of 50 µl for 1 hour at 37°C. HindIII was selected as it does not cut inside the RPP30 nor Gag::eGFP designed amplicons, is a high-fidelity enzyme with low star activity, has no methylation-sensitivity, and presents 100% activity in NEBuffer r2.1 (New England Biolabs, Ipswich, MA, USA) which does not interfere with ddPCR. No inactivation or further purification was performed.

2.9.2 Primer design for ddPCR

Two pairs of primers were designed to target and amplify fragments of 101 and 141 bp from Gag::eGFP and Ribonuclease P Protein Subunit P30 (RPP30) genes respectively, using the online PrimerQuest Tool from IDT (Integrated DNA Technologies IDT-DNA, Leuven, Belgium) website (<https://eu.idtdna.com/pages/tools>). The salt-adjusted Tm and primer self-complementarity were studied using the OligoCalc v3.27 online calculator [21] (<http://biotools.nubic.northwestern.edu/OligoCalc.html>) and primer-dimer

estimation was assessed by using Multiple Primer Analyzer tool from ThermoFisher (Waltham, MA, USA) website (<https://www.thermofisher.com/>). RPP30 sequence was retrieved from the NCBI GenBank database (Gene ID: 10556, National Centre for Biotechnology Information, Bethesda, MD, USA) website (www.ncbi.nlm.nih.gov/genbank/). RPP30 and Gag::eGFP assay primer sequences are shown in Table 1.

2.9.3 qPCR for ddPCR annealing determination

In order to optimize the annealing temperatures of the primers to be used in the ddPCR, a qPCR with a temperature gradient of 55.4 to 60.7°C was performed in a C1000 Touch thermal cycler (#1851197, Bio-Rad, Hercules, CA, USA). Each 10 µl PCR mixture contained 2x iTaq Universal SYBR Green (#1725151, Bio-Rad, Hercules, CA, USA), primers at a final concentration of 200 nM (IDT-DNA, Leuven, Belgium) and 80 ng of gDNA. The content was then transferred to a 96-well PCR plate, sealed and cycled using the following conditions: 95 °C for 5 min (1 cycle); 95 °C for 30 s and 55.4–60.7 °C for 1 min (40 cycles); 4 °C for 5 min, 90 °C for 5 min (1 cycle), 4 °C hold. On the basis of this assessment, the annealing temperature was determined to be 58°C for both RPP30 and Gag::eGFP primers. Additionally; its melting curves showed a single peak, indicating the presence of a single specific amplicon, as it is crucial to avoid unspecific amplifications such as primer dimers when performing ddPCRs. The tested gDNA samples had Cq values varying between 20.5 and 23.5 cycles at 58°C. As a consequence, samples were diluted for the subsequent ddPCRs in order to avoid saturation by an excessive number of positive droplets.

2.9.4 ddPCR for Copy Number Variation (CNV) determination

Each ddPCR mixture contained 2x EvaGreen ddPCR Supermix (#1864034, Bio-Rad, Hercules, CA, USA), the corresponding pair of primers at a final concentration of 200 nM (IDT-DNA, Leuven, Belgium) and 16 ng of gDNA (as the previous qPCR assay determined). The 20 µl PCR mixtures were loaded into DG8 Cartridges together with 70 µl of Droplet Generation Oil (#1863009 and #1863005 respectively, Bio-Rad, Hercules, CA, USA). After gasket attachment, the cartridges were placed into the QX200 Droplet Generator (#1864002 Bio-Rad, Hercules, CA, USA), which generated ~20.000 droplets

per sample. The content was then transferred to a 96-well PCR plate, sealed by means of a PX1 PCR Plate Sealer (#1814000, Bio-Rad, Hercules, CA, USA) and cycled using the conditions shown in Table 2. The QX200 Droplet Reader (#1864003, Bio-Rad, Hercules, CA, USA) was used to read the individual cycled droplets. The described process was carried on at Laboratori de Luminescència i Espectroscòpia de Biomolècules from the Universitat Autònoma de Barcelona.

The generated data was analyzed using QuantaSoft Software 1.7.4.0917, which determined the absolute quantity of target DNA molecules by fitting the number of positive and negative droplets after amplification (>31.000 accepted droplets per condition) into a Poisson distribution. The used statistical analysis took into account and compensated for the droplets presenting two or more target DNA copies due to the random nature of partitioning when droplets were generated.

2.10 Cryopreservation

The cryopreservation mixture consisted of 7.5% DMSO (Sigma-Aldrich, Saint Louis, MO, USA), in fresh medium. Each cryotube™ (ThermoFisher, MA, USA) contained 1 mL of cryopreservation mix and $5 \cdot 10^6$ cells. Cryotubes were placed in a Mr. Frosty™ (ThermoFisher, MA, USA) to reach -80°C with a rate of 1°C/minute. After 24 hours, cryotubes were stored in liquid nitrogen.

3. Results and discussion

3.1 pT2 and pDonor vector generation

pT2 plasmid codes for the human-codon optimized Cas9 downstream of EF-1 α core promoter and the T2-sgRNA (from [5]) under the U6 RNA promoter. The selected T2-sgRNA used in this study contains the 5'-GGGGCCACTAGGGACAGGAT [TGG] -3' specificity sequence to target the AAVS1 locus, which presents 100% identity with a segment of PPP1R12C intron 1 (Table 3). Immediately after the homology fragment, the genome presents the TGG PAM sequence, allowing Cas9 to produce the DSB.

Table 3. T2 sequence nucleotide BLAST on human genome. It presents 100% identity with a segment of PPP1R12C intron 1. Blast was carried out using NCBI BLAST online software (<https://blast.ncbi.nlm.nih.gov/BlastAlign.cgi>).

| Homo sapiens chromosome 19, alternate assembly CHM1_1.1 | | | |
|---|--|---------------------------------|--|
| Query 1 Sbjct 55620713 | GGGGCCACTAGGGACAGGAT GGGGCCACTAGGGACAGGAT | 20 55620694 | |
| Range 1: 55620694 to 55620713 (Genbank) | | | |
| Sequence ID: NC_018930.2 | | Length: 59121989 | |
| Score: 40.1 bits (20) | | Identities: 20/20 (100%) | |
| Strand: Plus/Minus | | Gaps: 0/20 (0%) | |
| Features: protein phosphatase 1 regulatory subunit 12C isoform a protein phosphatase 1 regulatory subunit 12C isoform b | | | |

pT2 construction was achieved by a two-step approach starting from the lentiCRISPR-v2 plasmid, as shown in Figure 6. The lentiCRISPR-v2 plasmid contains a region (flanked by Esp3I restriction enzyme sites) between the U6 RNA promoter and the sgRNA scaffold sequence. This region was substituted by the 20 nucleotide sequence complementary to the selected target locus, in this case the T2 sequence directed at the AAVS1 GSH. It was achieved with a Golden Gate reaction, which uses the property of Type IIS restriction endonucleases (like Esp3I) to cleave DNA outside of the recognition sequence. This generates non-compatible sticky ends, allowing cutting and ligating with high efficiencies in one-pot reaction (Figure 6.A) using an appropriated cycle of temperatures (Table 2).

The obtained intermediate plasmid (lentiCRISPR-v2-T2) presented the desired T2-sgRNA sequence but it needed to be further engineered in order to eliminate two remaining undesired features: its lentiviral backbone and a puromycin resistance gene fused at the C-terminus of the Cas9 protein through a P2A self-cleaving peptide. For this purpose, the construct of interest from the lentiCRISPR-v2-T2 (containing the T2-sgRNA, the Cas9 and its promoters), was cloned into the pGag::eGFP backbone using a two-fragment assembly reaction while adding a stop codon at the 3' end of the Cas9 protein, achieving the generation of pT2 (Figure 6.B).



Figure 6. Schematic representation of the pT2 plasmid construction. **(A)**: Generation of the intermediate *lentiCRISPR-v2-T2* vector from *lentiCRISPR-v2* plasmid. The addition of the T2 sgRNA sequence between U6 promoter and the gRNA scaffold was achieved using a Golden Gate reaction, using the ability of Esp3I nuclease to cleave DNA outside of the recognition sequence. This generated non-compatible sticky ends, which allowed cutting and ligating in one-pot two reactions cycle. **(B)**: Generation of the pT2 plasmid using NEBuilder HiFi DNA assembly of the indicated PCR cloning constructs (red and blue dotted lines, arrows indicate designed overlapping oligonucleotides). The size of the fragments is not representative to its actual proportional length of the genes.

pDonor was generated by the assembly of the CMV-Gag::eGFP-pA construct of interest into the AAVS1-SA-2A-puro plasmid. The insertion was directed downstream of the puromycin resistance gene poly-A signal and before the HA-R homology arm (Figure 7).

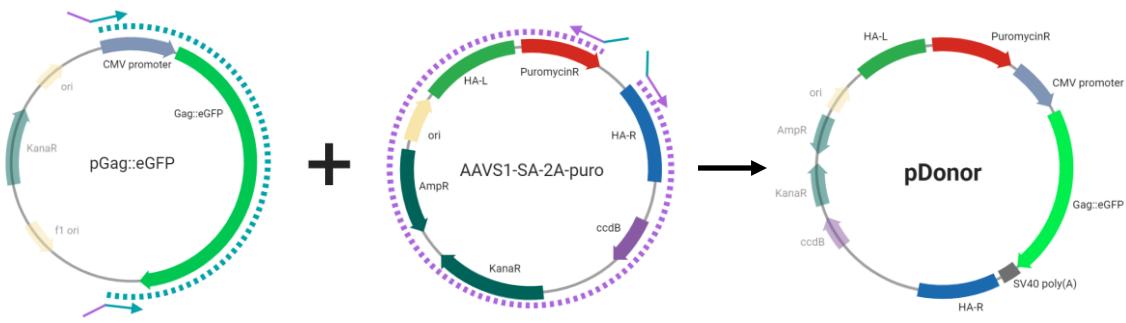


Figure 7. Schematic representation of the pDonor plasmid construction using NEBuilder HiFi DNA assembly of the indicated PCR cloning constructs (green and purple dotted lines, arrows indicate designed overlapping oligonucleotides). The size of the fragments is not representative to its actual proportional length of the genes.

3.2 Toxicity assay

In order to generate the stable HEK293 Gag::eGFP CRISPR/Cas9 cell line, a cellular pool with puromycin antibiotic selection was carried out. Puromycin causes premature translation termination at the ribosome, inhibiting protein expression [22]. The construct to be integrated at the AAVS1 safe harbor presents a puromycin resistance gene without an upstream promoter. With this approach, only the clones with the donor construct (Puro-CMV-Gag::eGFP-pA) integrated at the AAVS1 locus under its wild type promoter, will express the antibiotic resistance. Since antibiotic effect and tolerance differs between cell lines and culture conditions, a toxicity assay was performed with the aim to characterize the cellular response to puromycin and further determine its minimal selective concentration. The selected puromycin concentration range to perform the toxicity assay was from 0.3 µg/ml to 14.3 µg/ml [23].

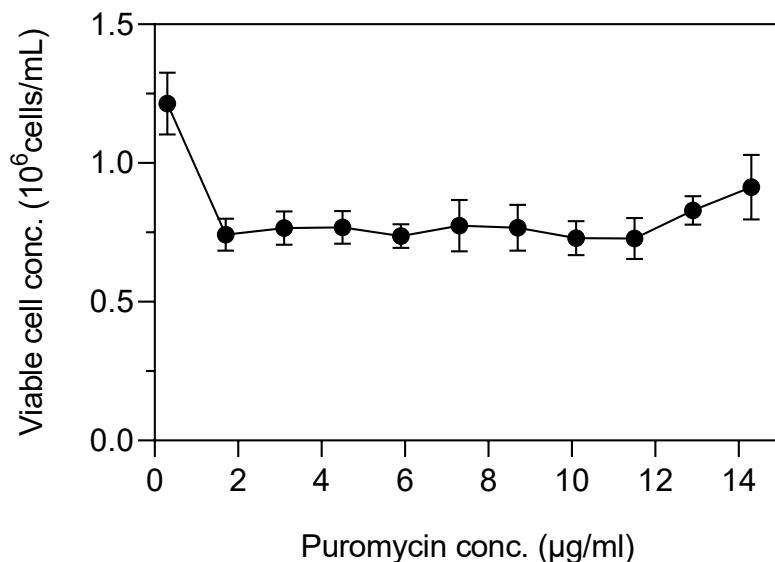


Figure 8. Puromycin toxicity assay. Viable cellular concentrations after being cultured for 48 hours under different puromycin concentrations.

The toxicity curve was carried out (Figure 8) and from the wide range of concentrations tested, only cells cultured with 0.3 $\mu\text{g}/\text{ml}$ of puromycin showed significant cellular growth. For this reason, 1.7 $\mu\text{g}/\text{ml}$ was chosen as the minimal puromycin concentration for selection experiments with this specific HEK293 cell line.

3.3 Transfection and pool generation

To generate the pool for the subsequent sorting of stable 293-Gag::eGFP CRISPR/Cas9 clones, pDonor and pT2 plasmids were co-transfected into HEK293 cells at a ratio 1:1, with PEIpro as a transfection reagent. Three control transfections -two with each individual plasmid and one with an empty pMock-were also performed, as shown at table 4.

Table 4. Pool and transfection control groups co-transfected plasmids.

| Group name | Plasmids | | Co-transfection ratio |
|-----------------------|----------|-------|-----------------------|
| CRISPR pool | pDonor | pT2 | 1:1 |
| pDonor control | pDonor | pMock | 1:1 |
| pT2 control | pT2 | pMock | 1:1 |
| pMock control | pMock | - | - |

At 2 days post transfection (dpt), the cell culture was diluted to $0.5 \cdot 10^6$ cells/ml in order to provide fresh nutrients and 1.7 µg/ml of puromycin were added, as previously determined, for cell selection (Figure 9). As mentioned above, only the clones that integrate the donor construct (Puro-CMV-Gag:eGFP-pA) at the AAVS1 locus should survive. A descent in viability and cellular growth was noticeable after 2 dpt. A media exchange was performed at 4 dpt and puromycin concentration was doubled up to 3.4 µg/ml to enhance the antibiotic selection effect. As a consequence, cells experienced a significant drop in cellular viability and growth (Figure 9). Then, at 7 dpt and for the rest of the experiment, puromycin concentration was reestablished to the initial 1.7 µg/ml. As can be observed, CRISPR pool engineered cells maintained its viability in higher levels (~40%) compared to control groups (~20%). From 9 dpt, the pool started recovering in terms of viability and viable cell density indicating the positive selection of the population with locus specific integration, while viability in all the control groups dropped to 0%. The pool continued growing until reaching a concentration of $4 \cdot 10^6$ viable cells / ml at 14 dpt, when it was clonally sorted (Figure 9).

The low viabilities (~40%) reached by the CRISPR pool are in line with what was expected, since not all the cells of the pool receive copies of the two co-transfected plasmids. In addition, even the double transfected cells need the Cas9 nuclease and the T2 gRNA to be expressed, interact with each other, generate a DSB at the AAVS1 locus, and this DSB needs to be repaired via cellular HDR using the pDonor plasmid as template. The subsequent rise in viability and cellular concentration is due to the growth of the clones that had undergone the AAVS1 integration of the construct of interest.

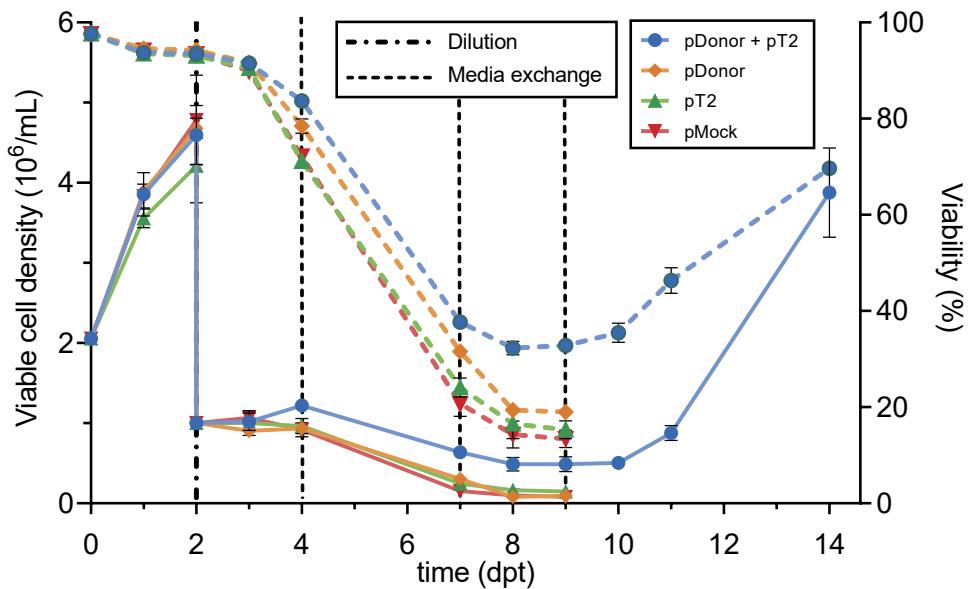


Figure 9. CRISPR transfected pool evolution. pDonor and pT2 plasmids were co-transfected into HEK293 cells, together with three control groups. At 2 dpt puromycin selection antibiotic was added. As a consequence, a drop in viability and cellular growth was observed. CRISPR pool (blue rounded dots and lines) maintained its viability in higher levels than control groups. From 9 dpt on, viability and viable cell density of the CRISPR pool recovered, indicating the positive selection of the population with locus specific integration, while viability in all the control groups dropped to 0%.

3.4 Clonal selection: cell sorting

At 14 dpt clonal selection was carried out in order to separate individual clones from the polyclonal pool. To select the Gag::eGFP producer clones, the cell sorting was based on GFP+ fluorescence, using three gating criteria depending on their fluorescent intensity (Figure 10). Sorting groups were coded as ‘GFP+’, ‘GFP++’ and ‘GFP+++’, containing clones with FITC values greater than 50, 500 and 2000, respectively. As Table 5 shows, different number of sorted clones of each group were isolated in 96-well plates, as the more restrictive the FITC criteria was, less cells from the pool matched it.

Table 5. Gating criteria: FITC threshold, percentage of the cellular population meeting the criteria, number of isolated clones, survival and nomenclature.

| Gating | FITC criteria | % of pool population | Isolated clones | Growing clones | Survival (%) | Nomenclature |
|----------------|---------------|----------------------|-----------------|----------------|--------------|--------------|
| GFP + | >50 | 80.65% | 480 | 59 | 12.3% | “Number” |
| GFP ++ | >500 | 10.33% | 288 | 16 | 5.6% | “Number”++ |
| GFP +++ | >2000 | 1.82% | 192 | 1 | 0.5% | “Number”+++ |

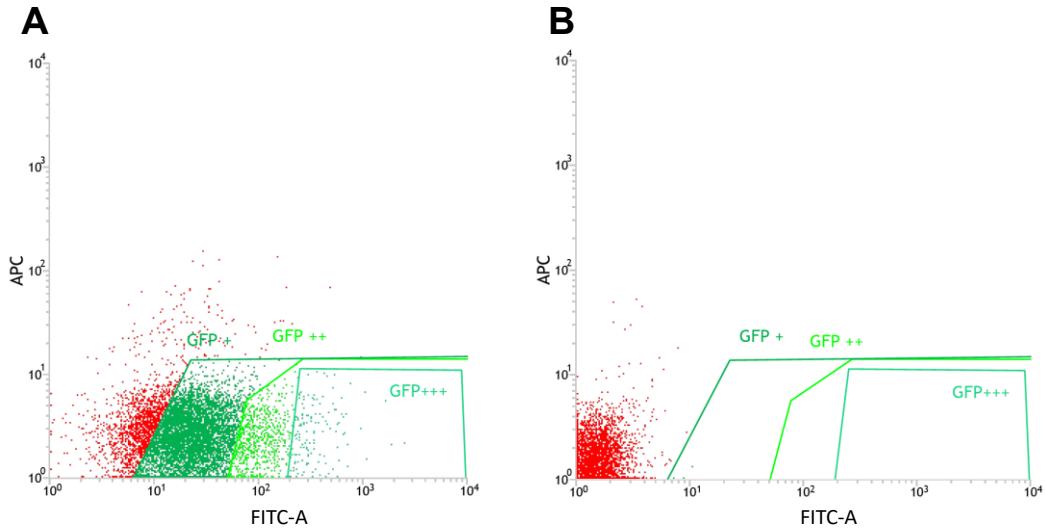


Figure 10. Gating criteria used in the cell sorting of the CRISPR cellular pool based on GFP+ fluorescence of the clones. (A): CRISPR Pool. (B): HEK293 WT (negative control).

3.5 Clone expansion

As it has been studied in preliminary tests and previous experiments, HEK293 monoclonal growth only occurs in plastic-adherent plates with FBS supplementation. For this reason, clones were seeded with Freestyle Mix supplemented with 20% FBS. From 7 days post-sorting (dps) on, puromycin was added at 1.7 µg/ml. Media supplemented with 15% FBS was exchanged periodically and when clones reached confluence they were progressively amplified to 24-well, 12-well, 6-well plates, 25 mm² and 75 mm² T-flasks. As shown in (Table 5), from all the isolated clones, a total of 76 survived all the amplification process. Remarkably, groups containing clones with greater FITC values, showed significantly lower survival ratios. We hypothesize that this may be a consequence of high Gag::eGFP polyprotein expression levels, since protein overexpression above some limits has been described to derive in cellular defects [24]. At this point, a cryopreserved stock of each of the 76 amplified clones was generated.

3.6 Gag::eGFP production quantification of the clones

When clones reached confluence at 25cm² T-flasks, they were analyzed by flow cytometry (Figure 11.A,B), determining their FITC-A mean levels (Figure 11.C,D) and supernatant was collected to assess its Gag::eGFP production by

fluorimetry. The fluorimetry analysis of the recovered supernatants allowed generating a range list of all the clones according to its Gag::eGFP production (Figure 11.F). This method cannot discriminate Gag::eGFP monomers from the Gag::eGFP VLPs, which can affect the reliability of the results especially in cultures with low viabilities, because the dead cells break releasing monomers into the medium. However, as all the clones showed similar high viabilities, it was assumed that viability was not significantly affecting the result.

After comparison, the 9 clones exhibiting higher GFP fluorescence levels in the supernatant were selected for further studies. From those 9 clones, 8 presented FITC-A mean values greater than 2000, analyzed by Flow Cytometry. From the unselected 62 remaining clones, only 6 had FITC values greater than 2000 (Figure 11.F, green bars), corroborating the coherence between the two methods used for the selection of the more promising clones: 1++, 7++, 9++, 13++, 15++, 8, 34, 35 and 43.

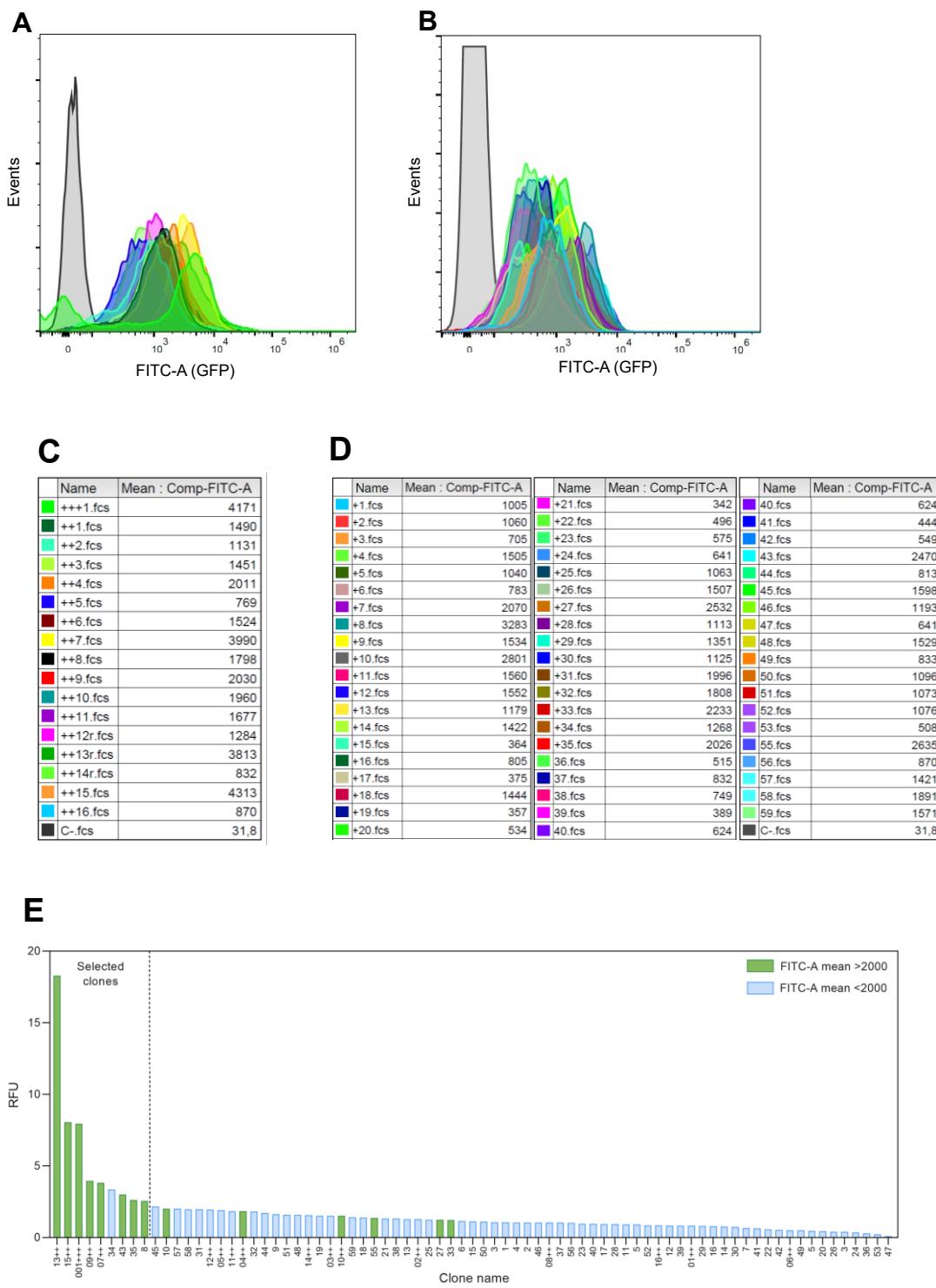


Figure 11. (A-B): GFP fluorescence histogram composition of all the adherent clones, by flow cytometry. **(A):** Histograms of clones from gating criteria “GFP+++” and “GFP++”. **(B):** Histograms from clones from gating criteria “GFP+”. **(C-D):** legend and FITC-A mean of the clones corresponding to graphs A and B **(E):** All clones sorted by its relative fluorescence. Color indicates if its mean FITC-A value analyzed by flow cytometry at confluent T-flasks is greater than 2000 relative units.

3.7 Suspension adaptation:

The adaptation of the adherent cell lines into suspension-growth was carried out with the goal to achieve high cell densities (and consequently good VLP production) and facilitate process scale-up.

The 9 selected clones were defrost and amplified in adherence. Cell count and viability assays were regularly performed to follow growth kinetics. When they achieved 85-90% confluence, they were seeded at $1 \cdot 10^6$ viable cells / mL in 125 mL polycarbonate Erlenmeyer flasks, with 7.5% FBS (15% FBS during the initial 48 hours) and 1.7 µg/mL of puromycin. Cells were routinely diluted to $0.5 \cdot 10^6$ viable cells/mL when their concentration was higher than $1.2 \cdot 10^6$ viable cells. If no significant growth was observed, media was exchanged each three days in order to provide fresh nutrients. Once the clones showed exponential growth together with viabilities higher than 95% during at least three passages, they were considered adapted to suspension culture. As it can be observed in Figure 12, the viability of the different cell lines improves gradually over time. Clone 15++ was unable to adapt to suspension, so it was discarded.

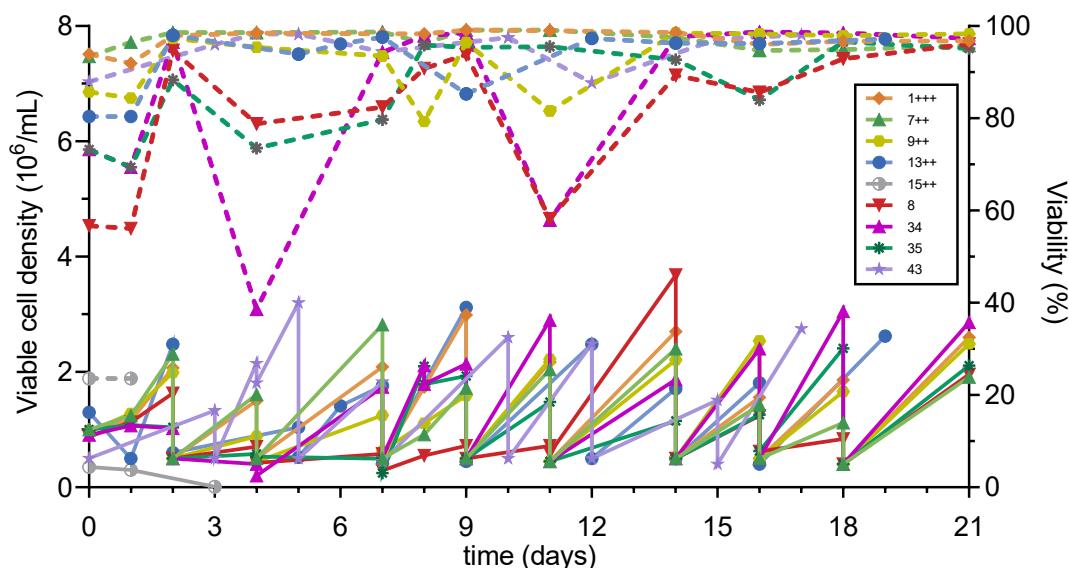


Figure 12. Suspension and serum-free media adaptation of the clones. Full lines: viable cell density, dotted lines: viability.

3.8 Serum-free adaptation

As mentioned before, medium was supplemented with FBS to facilitate the monoclonal cellular growth post cell sorting. The adaptation of the suspension cell lines into FBS-free media was carried by gradual reduction of serum concentration from 7.5% to 0%. With this methodology, clones achieved a sequential adaptation. Unlike the suspension adaptation, no significant changes in cell viabilities were observed during the process (Figure 12). Clones were considered successfully adapted to serum-free media after 3 passages without FBS and viabilities above 95%. Adapted clones attained a constant growth rate. A cryopreserved cell bank for each adapted clone was generated.

3.9 Selection of the best performing clone:

The supernatant from the adapted clones was collected and analyzed for Gag::eGFP VLPs by flow virometry (Figure 13.A). This allowed selecting the potential best four candidates from the 8 adapted candidates in terms of VLP production, named 1++, 7++, 13++, and 8.

Growth kinetic studies of the top four clones allowed the evaluation of its performance in terms of duplication time, maximum cell density, volumetric productivity and specific productivity.

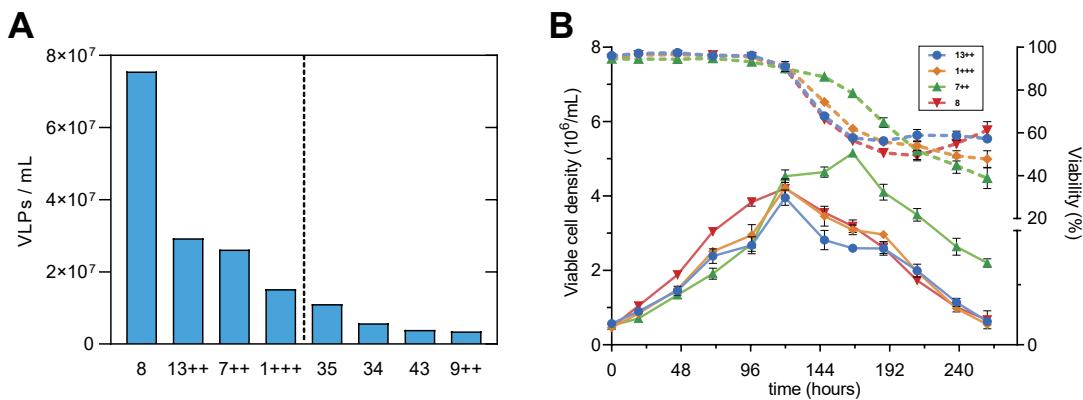


Figure 13. (A): VLP concentrations at the supernatant of the adapted CRISPR clones, 48 hours post seeding. (B): Growth kinetics of the suspension and serum-free adapted clones 13++, 1++, 7++ and 8.

Clones 13++, 1++ and 8 showed maximum cell density peaks of approximately $4 \cdot 10^6$ cells/mL at 120 h with duplication times between 26.9 and 34.6 h, while 7++ showed a different cell growth profile, reaching its maximum

of $5.16 \cdot 10^6$ cells/mL at 168 h, together with a duplication time of 39.1 h (Figure 13.B). For comparison, wild-type HEK293 cells in the same media and conditions showed duplication times of 26.2 h, peaking at $3.53 \cdot 10^6$ cells/mL (Table 6).

Table 6. Duplication time and maximum cell density of the suspension and serum-free adapted clones and HEK293 wild-type cells for reference.

| Clone name | Duplication time (hours) | Max. cell density (viable cells / mL) |
|------------|--------------------------|---------------------------------------|
| 13++ | 34.6 ± 1.4 | $(3.96 \pm 0.2) \cdot 10^6$ |
| 1+++ | 29.3 ± 0.6 | $(4.22 \pm 0.2) \cdot 10^6$ |
| 7++ | 39.1 ± 0.3 | $(5.16 \pm 0.1) \cdot 10^6$ |
| 8 | 26.9 ± 0.2 | $(4.20 \pm 0.3) \cdot 10^6$ |
| HEK293 | 26.2 ± 1.0 | $(3.53 \pm 0.3) \cdot 10^6$ |

In order to assess and compare VLP productions, samples prior to, during and after the cell density peak for each clone were analyzed. Remarkably, the viability drop experienced by the clones after 120 h of cell culture, is well known to cause a significant release of Gag::eGFP monomers, which was traduced in a sudden supernatant fluorescence increase after 120 h (data not shown). Therefore, to avoid this interference which affects the reliability of fluorimetric analysis in low viability cultures, samples were analyzed using Flow Virometry, which quantified the Gag::eGFP-VLPs.

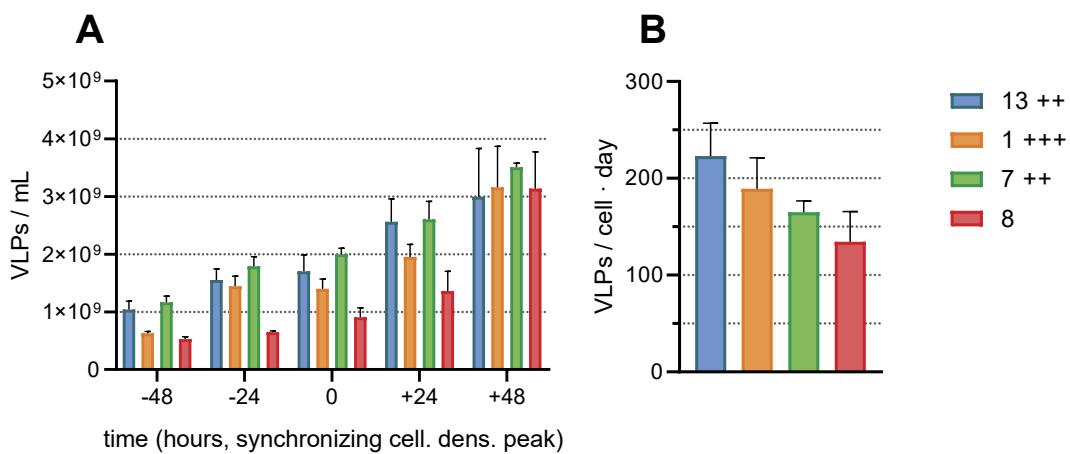


Figure 14. (A): VLP concentration of the four best CRISPR clones the days prior and after cell density peak. All clones showed a similar VLP production at 48 hours post cell density peak with no significant differences. (B): Specific productivity of the four best clones.

In contrast to what was observed in previous studies on cell line generation by illegitimate integration, VLP quantification showed that all four clones had similar VLP productions at 48 hours post cell density peak with no significant differences (Figure 14.A). This behavior is coherent with the expected when generating parallel cell lines directing the transgene into the same GSH. While clones may show small differences in production, in general terms, as the construct of interest is located under the same promoter and genetic context, the expression levels between clonal cell lines should be similar. Nevertheless, clone 13++ showed the highest specific productivity (Figure 14.B) (Table 7), therefore it was selected as the best clone. It was also observed under confocal microscopy, where it showed normal cellular morphology (Figure 15).

Table 7. Volumetric and specific productivity of the suspension serum-free adapted clones at cell density peak.

| Clone name | Volumetric productivity (VLPs · mL ⁻¹ · day ⁻¹) | Specific productivity (VLPs · cell ⁻¹ · day ⁻¹) |
|------------|---|---|
| 13++ | $(3.4 \pm 0.6) \cdot 10^8$ | 222.8 ± 33.9 |
| 1+++ | $(2.8 \pm 0.3) \cdot 10^8$ | 189.1 ± 32.2 |
| 7++ | $(2.9 \pm 0.2) \cdot 10^8$ | 165.0 ± 11.5 |
| 8 | $(1.8 \pm 0.3) \cdot 10^8$ | 134.4 ± 31.3 |

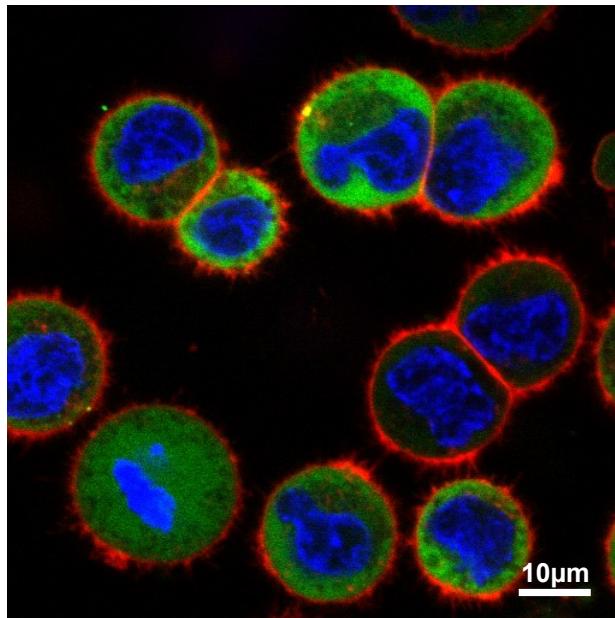


Figure 15. Confocal microscopy image of the suspension and serum-free adapted clone 13++. Gag::eGFP can be observed in green, lipid membranes were stained with CellMask (red) and nuclei with Hoechst (blue).

3.10 Integration site analysis

The genome of the clone 13++ was analyzed in order to confirm the integration of the construct of interest in the AAVS1 locus via HDR. Two pairs of oligonucleotides were designed to amplify each junction between the construct of interest and the AAVS1 locus outside the homology arms. PCRs showed the expected bands for 5' and 3' junctions (Figure 16), confirming the integration of the locus of interest at the desired AAVS1 safe harbor locus. Additionally, the junctions from the genomic DNA were sequenced in order to analyze if undesired mutations occurred at the surroundings of the Cas9 restriction point. No insertion or deletions were observed at 5' or 3' junction surroundings, obtaining the expected sequence consequence of the construct integration via HDR of the homology arms (Figure 16).

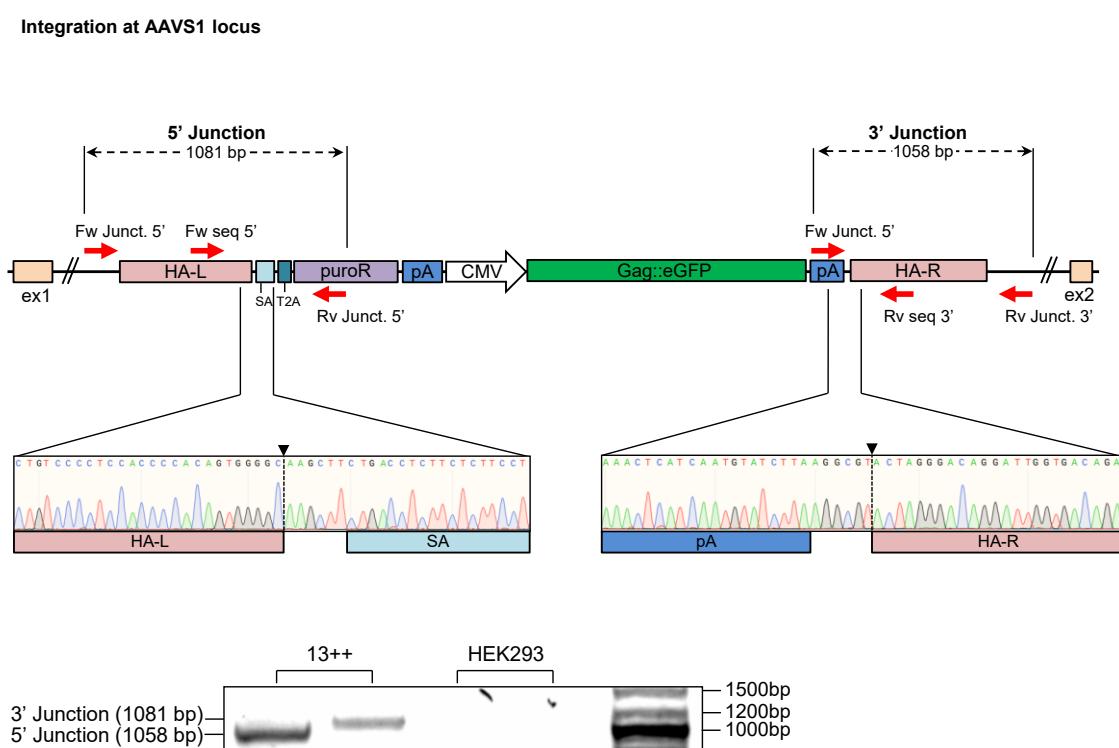


Figure 16. Construct integration analysis at the AAVS1 genomic locus of the 13++ clone. Two PCR reactions were designed to amplify each junction between the construct of interest and the AAVS1 locus outside the homology arms. PCRs showed the expected bands for 5' and 3' junctions and sequencing showed no insertion or deletions, confirming the construct integration at the AAVS1 human safe harbor without undesired mutations.

3.11 Genomic analysis: number of integrated copies

In order to characterize the 13++ cell line in terms of copy number of RPP30 and the integrated Gag::eGFP, it was studied using digital-droplet PCR (ddPCR) and compared with wild-type HEK293 cells. As explained in Chapter 3, RPP30 was used as a reference gene since it is well known that it presents two copies at the human genome [25–27].

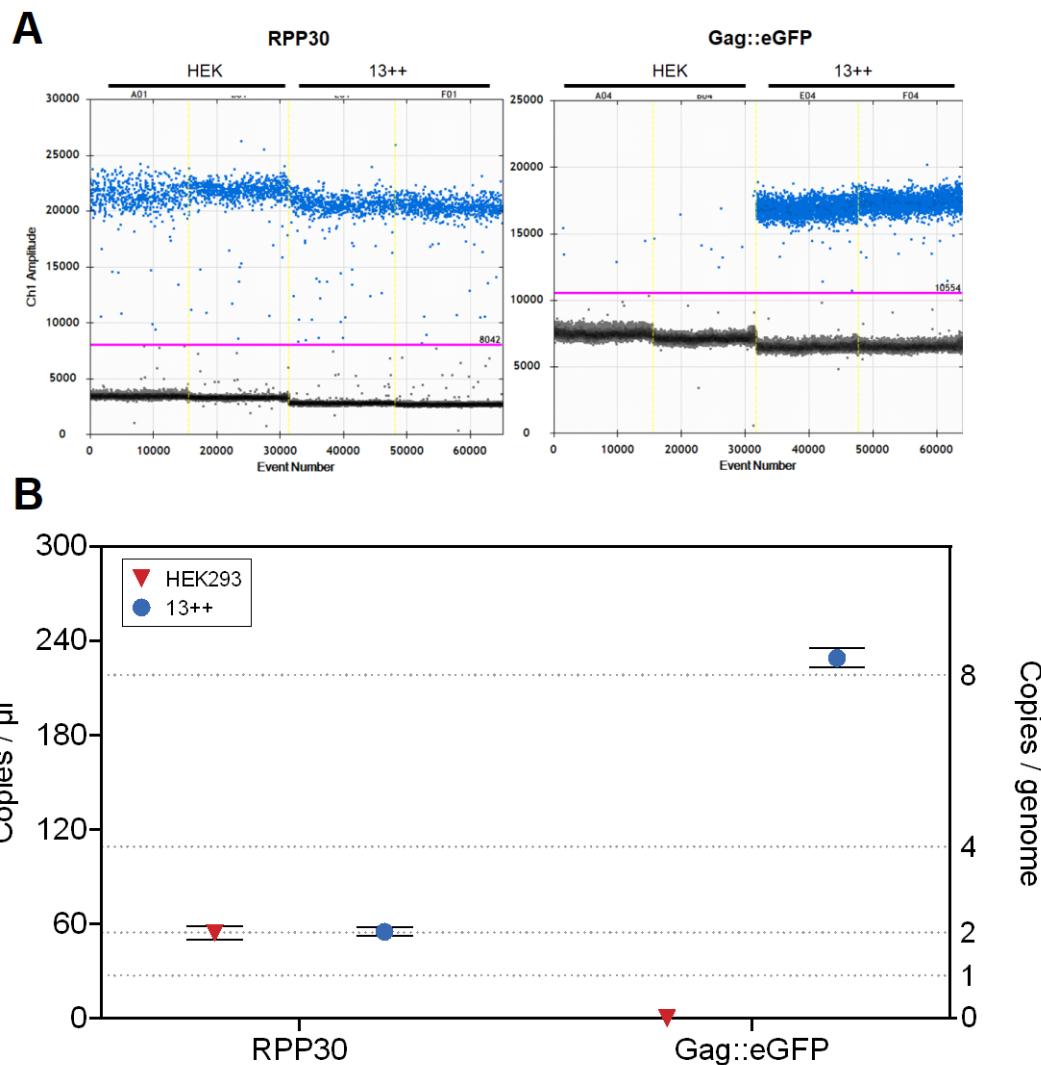


Figure 17. Data from absolute quantification of the 13++ cell line for the RPP30 and Gag::eGFP genes. To estimate the average VCN values, the concentrations (copies/µl) of Gag::eGFP and the RPP30 reference gene were determined in 10H9 and HEK293 cell lines. Error bars indicate Poisson confidence limits.

ddPCR allowed the analysis of more than 30.000 PCR events (Figure 17.A) for the studied genes. RPP30 gene in the wild-type HEK293 genome allowed the calculation of the number of copies per genome of each analyzed DNA region.

Cell line 13++ contained the same copy number of the RPP30 gene compared with the HEK293 parental cell line (Figure 17.B). This suggested that no major translocations or duplication events occurred in the generated cell line genome [28]. Interestingly, 13++ presented 8 copies of the Gag::eGFP construct per genome (Table 8, Figure 17.B). 8 copies exceeded the expected copy number for a transgene rationally-integrated into the AAVS1 locus, as diploid human genomes present two AAVS1 loci. Previously, the integration into the AAVS1 was confirmed using junction PCR, so we hypothesize that this elevated number of copies can be due to tandem integration of multiple donor constructs into one or two AAVS1 locus. A plausible explanation for the tandem integration is that it may have been caused by the HRD cellular mechanisms joining not only the homology arms of the donor with its equivalent sequence regions of the AAVS1 locus, but also ligating different donor constructs from different donor plasmids between them.

Table 8. Copies per microliter of the analyzed RPP30 and Gag::eGFP genes, its ratios and number of copies per genome.

| | HEK293 | 13++ |
|--------------------------------|----------------|-----------------|
| RPP30 (copies/ μ l) | 54.3 \pm 4.0 | 55.1 \pm 2.8 |
| Gag::eGFP (copies/ μ l) | 0.0 | 229.0 \pm 6.0 |
| RPP30/HEK-RPP30 | 1.000 | 1.014 |
| Gag::eGFP/RPP30 | 0 | 4.151 |
| RPP30/HEK-RPP30 (nearest half) | 1 | 1 |
| Gag::eGFP/RPP30 (nearest half) | 0 | 4 |
| RPP30 copies/genome | 2 | 2 |
| Gag::eGFP copies/genome | 0 | 8 |

3.12 Stability analysis

One of the critical issues to analyze when characterizing a cell line is its stability expressing a recombinant protein or product of interest over time. Instability or diminution of expression can affect process yield, protein quality, and significantly reduces the capability of the clone to produce in continuous or prolonged operations. It is also one of the parameters analyzed by regulatory agencies in order to approve clonal cell lines for the production of

pharmaceuticals [29]. Therefore, Gag-VLP production stability of the generated 13++ cell line constitutes a key factor to determine its quality as a clone and its potential uses. Stability was studied by analyzing the cellular density, viability, mean fluorescence and its Gag-VLP production during 33 consecutive days. Results show that upon different passages, cell line 13++ does not significantly reduce viability, growth (data not shown) or production since its specific productivity remains constant around 218 VLPs / (day · cell) (Figure 18).

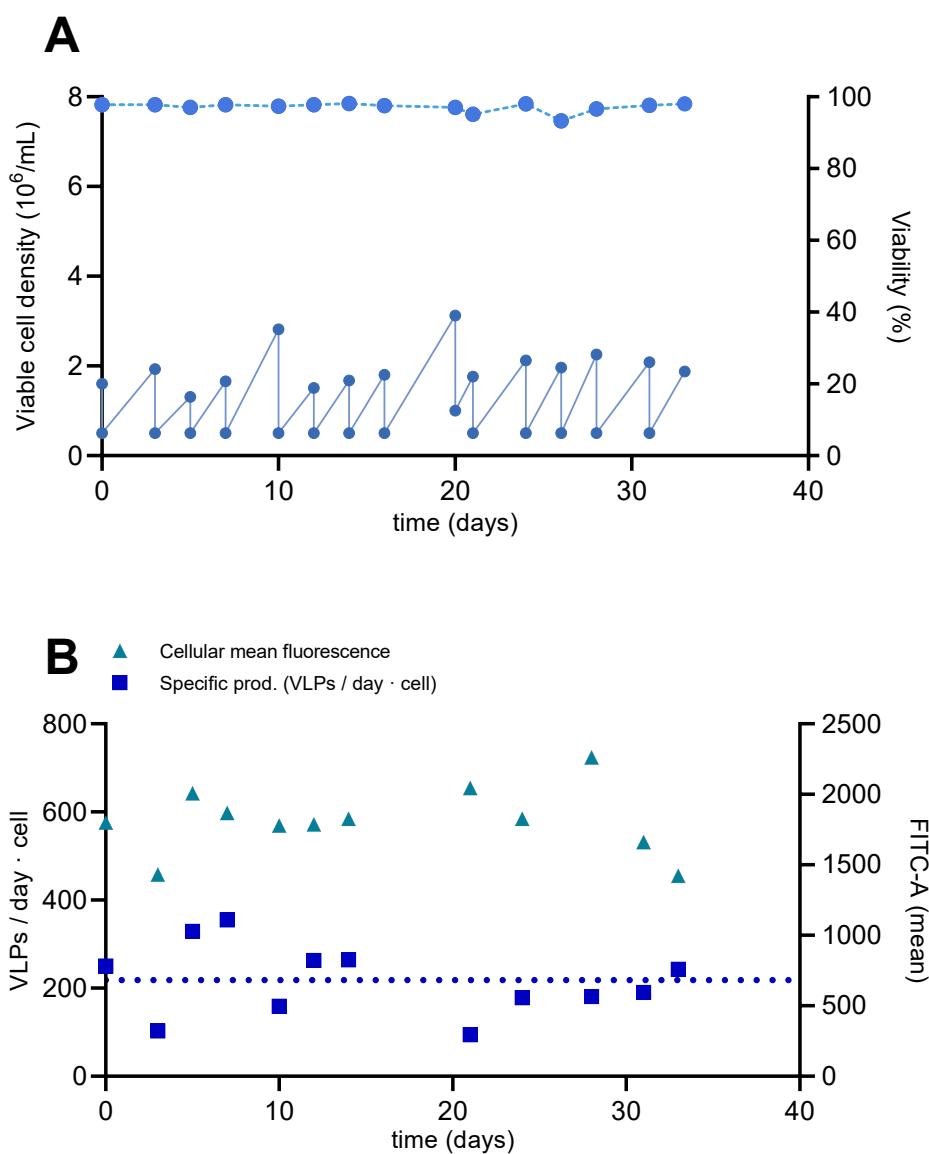


Figure 18. Stability analysis of the 13++ cell line. (A): Cellular concentration (solid line) and viability (dotted line) during the studied 60 days. (B): Gag::eGFP-VLP production stability: mean cellular fluorescence was reduced during the studied 60 day period, while specific productivity level was maintained constant.

4. Conclusions

The development of a stable cell line expressing Gag-VLPs was carried out successfully. Two plasmids coding for the necessary CRISPR / Cas9 elements for the integration of the Gag::eGFP and puromycin resistance genes in the AAVS1 locus were generated and used to co-transfect HEK293 cells. Unlike the co-transfected cell pool, control groups did not survive in puromycin containing medium, indicating that the integration into AAVS1 was only caused by the HDR repair mechanism responding the CRISPR-generated DSB. Among the 960 unicellular seeded clones, 76 showed cellular growth, with a survival rate range depending on the gating used in the cell sorting: the less restrictive the criteria was, more clones survived. The study of the Gag::eGFP presence in the adherent clones supernatant, allowed the selection of the best candidates for its suspension and serum-free adaptation. The selected clones were successfully adapted to suspension and FBS free culture conditions and growth kinetic studies of the best four potential candidates were performed, which allowed its comparison and study in terms of growth, viability and specific VLP productivity. 7++ was the clone that reached a higher maximum cell density ($5.16 \cdot 10^6$ viable cells/mL) and the longest duplication time, of 39.1 hours. Cell lines 13++, 1+++ and 8 showed maximum densities of approximately $4 \cdot 10^6$ cells/mL, with duplication times between 26.9 and 34.6 hours. Clone 13++ was selected as the best CRISPR generated clone based on its specific productivity, of 222.8 VLPs/(cell·day). Analysis of its genome allowed us to conclude that the construct of interest was added at the AAVS1 locus, while sequencing showed no insertion or deletion mutations at the 5' nor 3' junctions. Gene copy number of the gene RPP30 remained invariable compared to its HEK293 wild-type predecessor, indicating that the clone did not suffer major genomic alterations during its generation. 13++ presented 8 copies of the Gag::eGFP gene per genome. We hypothesize that this was caused by tandem integration of multiple donor constructs into the AAVS1 locus. Clone 13++ also showed stable VLP expression during 33 days, maintaining its specific productivity unaltered.

5. Acknowledgements

The authors thank Dr. Amine Kamen (McGill University, Montreal, Canada) for providing the HEK 293 SF-3SF6 cells. The pGag::eGFP (Cat#11468) reagent from Dr. Merilyn Resh was obtained through the National Institutes of Health AIDS Reagent Program, Division AIDS, National Institute of Allergy and Infectious Diseases, National Institutes of Health. We would also like to thank Laia Bosch Molist for her help with unpublished experiments and Dr. Carlo Carolis from Centre de Regulació Genòmica de Barcelona for his help designing the pDonor plasmid assembly as well as providing the necessary Gibson® reagents. We would like to thank Dr Jose Amable Bernabé from the Institut de Ciència de Materials de Barcelona (ICMAB-CSIC, Bellaterra, Catalonia, Spain) for his support in nanoparticle tracking analysis techniques. The support of Núria Barba Bosch and Meritxell Vendrell from the Servei de Microscòpia (Universitat Autònoma de Barcelona, Bellaterra, Catalonia, Spain) in the development of optical microscopy images and analysis is greatly appreciated. We would also like to thank Manuela Costa Garcia from the Servei de Cultius Cel·lulars, Producció d'Anticossos i Citometria (Universitat Autònoma de Barcelona, Bellaterra, Catalonia, Spain) for her precious help with flow cytometry, the support of Sílvia Torrents Zapata from the Banc de Sang i Teixits (Barcelona) and Marc García Trujillo with flow virometry; and Salvador Bartolomé Piñol from Laboratori de Luminescència i Espectroscòpia de Biomolècules (Universitat Autònoma de Barcelona, Bellaterra, Catalonia, Spain) for his help with qPCR and ddPCR analyses. Figures (1-3, 6 and 7) were created with Biorender.com, and figures 1 and 2 were adapted from templates retrieved from <https://app.biorender.com/biorender-templates>.

References

- [1] Mutskov, V. & Felsenfeld, G. Silencing of transgene transcription precedes methylation of promoter DNA and histone H3 lysine 9. *EMBO J.* **23**, 138–149 (2004)
- [2] Hansen, J. *et al.* A large-scale, gene-driven mutagenesis approach for the functional analysis of the mouse genome. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 9918–9922 (2003)
- [3] Wurm, F. M. Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat. Biotechnol.* **22**, 1393–1398 (2004)
- [4] Sadelain, M., Papapetrou, E. P. & Bushman, F. D. Safe harbours for the integration of new DNA in the human genome. *Nat. Rev. Cancer* **12**, 51–58 (2012)
- [5] Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823–6 (2013)
- [6] Aznauryan, E. *et al.* Discovery and validation of human genomic safe harbor sites for gene and cell therapies. *Cell Reports Methods* **2**, (2022)
- [7] Oceguera-Yanez, F. *et al.* Engineering the AAVS1 locus for consistent and scalable transgene expression in human iPSCs and their differentiated derivatives. *Methods* **101**, 43–55 (2016)
- [8] Kandavelou, K. *et al.* Targeted manipulation of mammalian genomes using designed zinc finger nucleases. *Biochem. Biophys. Res. Commun.* **388**, 56–61 (2009)
- [9] Krämer, O., Klausing, S. & Noll, T. Methods in mammalian cell line engineering: From random mutagenesis to sequence-specific approaches. *Appl. Microbiol. Biotechnol.* **88**, 425–436 (2010)
- [10] LaFountaine, J. S., Fathe, K. & Smyth, H. D. C. Delivery and therapeutic applications of gene editing technologies ZFNs, TALENs, and CRISPR/Cas9. *Int. J. Pharm.* **494**, 180–194 (2015)

- [11] Lieber, M. R., Ma, Y., Pannicke, U. & Schwarz, K. Mechanism and regulation of human non-homologous DNA end-joining. *Nat. Rev. Mol. Cell Biol.* **4**, 712–720 (2003)
- [12] Van Gent, D. C. & Van Der Burg, M. Non-homologous end-joining, a sticky affair. *Oncogene* **26**, 7731–7740 (2007)
- [13] Schmidt, F. & Grimm, D. CRISPR genome engineering and viral gene delivery: A case of mutual attraction. *Biotechnol. J.* **10**, 258–272 (2015)
- [14] Pennisi, E. The CRISPR craze. *Science/AAAS* **341**, 833–836 (2013)
- [15] Cervera, L. *et al.* Generation of HIV-1 Gag VLPs by transient transfection of HEK 293 suspension cell cultures using an optimized animal-derived component free medium. *J. Biotechnol.* **166**, 152–165 (2013)
- [16] Hermida-Matsumoto, L. & Resh, M. D. Localization of Human Immunodeficiency Virus Type 1 Gag and Env at the Plasma Membrane by Confocal Imaging. *J. Virol.* **74**, 8670–8679 (2000)
- [17] Schwartz, S. *et al.* Mutational inactivation of an inhibitory sequence in human immunodeficiency virus type 1 results in Rev-independent gag expression. *J. Virol.* **66**, 7176–7182 (1992)
- [18] Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014)
- [19] Hockemeyer, D. *et al.* NIH Public Access. **27**, 851–857 (2014)
- [20] Cermak, T. *et al.* Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res.* **39**, 1–11 (2011)
- [21] Kibbe, W. A. OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res.* **35**, W43–W46 (2007)
- [22] Aviner, R. The science of puromycin: From studies of ribosome function to applications in biotechnology. *Comput. Struct. Biotechnol. J.* **18**, 1074–1083 (2020)

- [23] McCall, C. M. *et al.* Human Immunodeficiency Virus Type 1 Vpr-Binding Protein VprBP, a WD40 Protein Associated with the DDB1-CUL4 E3 Ubiquitin Ligase, Is Essential for DNA Replication and Embryonic Development. *Mol. Cell. Biol.* **28**, 5621–5633 (2008)
- [24] Mori, Y., Yoshida, Y., Satoh, A. & Moriya, H. Development of an experimental method of systematically estimating protein expression limits in HEK293 cells. *Sci. Rep.* **10**, 1–11 (2020)
- [25] Lin, H. T. *et al.* Application of Droplet Digital PCR for Estimating Vector Copy Number States in Stem Cell Gene Therapy. *Hum. Gene Ther. Methods* **27**, 197–208 (2016)
- [26] Dyavar, S. R. *et al.* Normalization of cell associated antiretroviral drug concentrations with a novel RPP30 droplet digital PCR assay. *Sci. Rep.* **8**, 1–11 (2018)
- [27] De Oliveira, M. F. *et al.* Comparative analysis of cell-associated HIV DNA levels in cerebrospinal fluid and peripheral blood by droplet digital PCR. *PLoS One* **10**, 1–10 (2015)
- [28] Lin, Y. C. *et al.* Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. *Nat. Commun.* **5**, (2014)
- [29] Bailey, L. A., Hatton, D., Field, R. & Dickson, A. J. Determination of Chinese hamster ovary cell line stability and recombinant antibody expression during long-term culture. *Biotechnol. Bioeng.* **109**, 2093–2103 (2012)

Chapter five

**Stable HEK293 RMCE cell line generation for the production of
Gag::eGFP VLPs by lentiviral infection**

Abstract: The production of Gag-based VLPs in HEK293 cell lines is of great interest, as they can correctly assemble complex VLPs with appropriate post-translational modifications (PTMs). It can be performed in transient gene expression (TGE) and stable gene expression (SGE). SGE is a valuable strategy due to its homogeneous nature. Additionally, in comparison to TGE, SGE is not restricted to the use of transfection compatible media and its expression is prolonged over time. Randomly generated cell lines rely on the fortuitous integration of the gene of interest (GOI), which is not an ideal approach as the locus of integration plays a major role on the expression of a protein. However there had been few defined genomic safe harbors (GSH) in the human genome. Recombinase-Mediated Cassette Exchange (RMCE) is a two-step molecular mechanism which relies on the random integration of a cassette expressing a reporter protein into the cellular genome in a process known as tagging. In this work, the HIV-1 Gag polyprotein was fused with the green fluorescent protein (GFP), to act simultaneously as a reporter protein for the tagging of the cells, while generating Gag-based VLPs, the product of interest. When a high expresser tagged cell line is identified and characterized, the integrated cassette acts as a molecular “landing pad” for the integration by RMCE of a gene of interest (GOI), known as targeting. The tagged cell line must be a good stable producer with good growth and it is of great importance the presence of a single copy of the integrated cassette on its genome. This was achieved by lentiviral infection using low multiplicity of infection (MOI). Infected line candidates were enriched in pools based on their fluorescence and more than 2000 clones were isolated. The 13 more promising clones were adapted to suspension and serum-free media. Growth kinetic studies of the 5 clones exhibiting higher Gag::eGFP concentrations in their supernatant were performed and cell lines were tested by their G418 sensitivity. Two clones showing G418 resistance were discarded. The three remaining clones were characterized in terms of targeting capabilities using DsRed as a targeting GOI and G418 for the selection of targeted clones. One clone was discarded due to its inability to perform a correct targeting and from the remaining two the best clone was selected in terms of specific productivity. Digital droplet PCR (ddPCR) analysis confirmed that the clone presented a single integration of the RMCE cassette on its genome. The selected clone had a duplication time of

29.0 ± 2.5 h, reached a maximum cell density of $(4.58 \pm 0.13) \cdot 10^6$ cells/mL and a specific productivity of 110 VLPs/(cell·day), maintaining its expression stable during the studied 45 days.

Keywords: VLPs, Stable gene expression, RMCE, lentivirus, HEK293, Gag::eGFP, HIV-1, Targeting.

1. Introduction

Stable cell lines can be developed following two main approaches: random gene integration or directed gene integration. The first approach is based on the spontaneous random integration of a transgene construct of interest into the cell's genome, while the second approach consists in a rationally-designed locus-specific integration. Random integration can result in high expression cell lines if a high enough number of clones is screened, as the integration is randomly distributed [1] and the locus of integration plays a major role on the expression of a recombinant construct [2], so the directed gene integration is usually the recommended way to go. However, aiming the integration into specific pre-defined high transcriptionally active locus in HEK293 cells can be sometimes difficult as few genomic safe harbors (GSHs) have been identified in the human genome [3]. Recombinase-Mediated Cassette Exchange (RMCE) constitutes an attractive strategy as it combines the advantages from the two previously explored methodologies. It initially relies on a fortuitous integration of a cassette into the cellular genome in order to identify a promising high-expression clone and generate a high expression characterized cell line [4]. Then, this cell line carrying a tagged cassette in a high transcriptionally active locus can be repeatedly modified in a "locus-specific" way in order to quickly and easily generate high producer cell lines [4,5].

RMCE is a molecular mechanism relying on site-specific recombination. Briefly, site-specific recombination is catalyzed by recombinases and differs from homologous recombination as DNA is cut and conservatively rearranged without any degradation, synthesis process or need of co-factor addition [6,7]. In mammalian cell culture, the most prominent Site-specific recombinase (SSR) proteins are Cre (acronym for "Causes recombination") and Flp ("Flippase") which act on their recombinase target sites (RTs) *LoxP* ("locus of crossover in phage P1") and *FRT* (Flp-recombinase target), respectively [8,9]. Genomic RMCE is a process driven by SSRs where a tagging cassette flanked by heterospecific RTs anchored in the host genome is exchanged for a targeting cassette flanked by identical sites, which ends integrated into the host genome (Figure 1) [8,10].

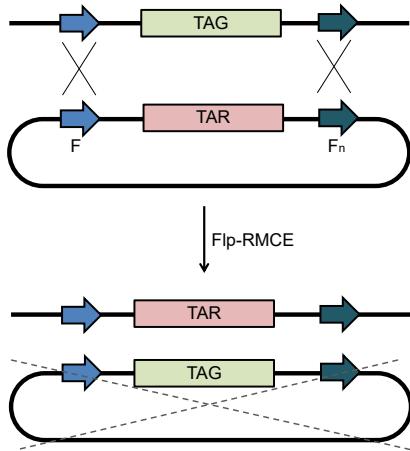


Figure 1. Genomic RMCE is a process driven by SSRs where a tagging (TAG) cassette flanked by heterospecific RTs anchored in the host genome is exchanged for a targeting (TAR) cassette flanked by identical sites, which ends integrated into the host genome.

The general outline of the RMCE strategy for the generation of stable cell lines (otherwise known as “tag-and-exchange” approach) is based on two main steps, usually referred as tagging and targeting [8]. First, a cell line is generated by random integration of one single copy of the tagging cassette coding for a reporter protein, creating a “landing pad” flanked by heterospecific RTs [11]. The use of a reporter protein facilitates the screening for the selection of clones showing high and stable expressions. Then, the tagging cassette can be cleanly replaced by a targeting cassette presenting the same set of RTs, causing the non-disruptive insertion of a gene of interest (GOI) into the pre-characterized genomic hot-spot [6,9]. For the site-specific cassette exchange to occur, the targeting cassette must be provided to the “tagged” recipient cell line in presence of a recombinase (Flp for *FRT* RTs or Cre for *LoxP* RTs) which catalyzes the site-specific genome integration. This process can be repeated at will, and as it is based on a highly-specific event.

When generating a tagged master cell line, it is important to select the best clone in terms of cellular growth, stability and productivity, combined with a single integration site of the tagging cassette [4,11]. These parameters ensure that the integration site is a “hot-spot” with the potential of harboring a “landing pad”, enabling the production of targeted clones presenting the expression characteristics of the tagged master cell line [11,12]. Otherwise, if a RMCE cell line presents two or more tagging constructs in its genome, when targeted, a

portion of its cellular population will partially recombine, leading to undesired partially-targeted resistant clones expressing simultaneously the tagging and targeting transgenes. The delivery of DNA into mammalian cells for the generation of stable cell lines can be performed using different techniques, however most of them lack the control over the copy number of the integrated DNA constructs. Different DNA transduction protocols have been reviewed and evaluated in the literature, concluding that lentiviral vectors are the best option for the generation of tagged cell lines with single copy integration events [12].

Lentiviral vector (LV) systems are one of the most used and studied gene transfer vectors, and have a great potential for the delivery of genes to cells [13]. They can be diluted to transfect cells with very low multiplicities of infection (MOI) ensuring the delivery of only one copy per infected cell. They are mainly derived from HIV-1 and over the last decades, this vector system has been improved for example with its pseudotyping with heterologous envelopes such as the vesicular stomatitis virus G protein (VSV-G), which produce high-titer and stable LV able to transduce a large spectrum of cell types [13,14]. Due to HIV-1 pathogenicity in humans, different generations of LV vector systems have been developed for safety reasons (Figure 2) [13,15]. First generation LV system consisted in the expression of all the HIV-1 proteins except Vpu and Env. Second generation LV additionally removed the genes *Vif*, *vpr*, *vpu* and *nef* from the first generation, as they were not necessary for the production of LV [15]. Third generation LV is characterized by the removal of the *tat* regulatory gene as its function was to transactivate genes under U3 5' long terminal repeat (LTR), which was also replaced by the constitutively active Rous Sarcoma Virus (RSV) promoter, or cytomegalovirus (CMV) promoter (Figure 2) [15,16]. Nowadays, third generation LVs are the most commonly used in research. They involve the co-transfection of four plasmids with a highly improved biosafety as they take advantage of split non-overlapping packaging constructs reducing the potential undesired recombination events which could derive in generation of replication-competent viruses, the main concern regarding the use of any HIV-1 vector.

The four plasmids for third generation production of LVs used in this work are:

- Transfer vector: Self-inactivating (SIN) lentiviral backbone containing the cassette of interest (in this work the Gag::eGFP RMCE tagging cassette), flanked by the *cis*-acting elements required for reverse transcription, encapsidation and integration: RSV promoter, 5'LTR (truncated), HIV-1 ψ , RRE, cPPT/CTS and the 3' LTR- Δ U3 with a deletion on its U3 element (self-inactivating) for biosafety reasons, as well as the WPRE (W) element for transgene expression enhancement [15,16].
- Gag-pol: Codes for viral enzymes and structural proteins.
- Rev: Codes for the post-translational regulator, under RSV promoter.
- Env: Codes for the VSV-G glycoprotein.

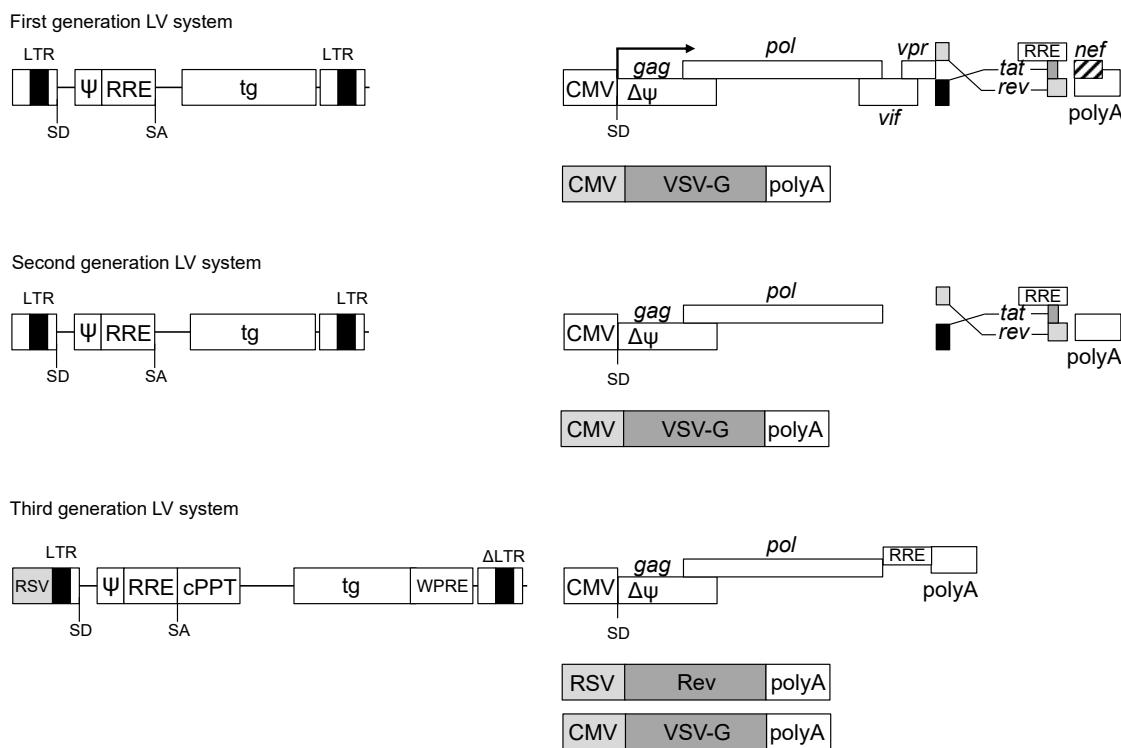


Figure 2. Schematic representation of different lentiviral HIV-1 based LV from each generation. First generation vectors expressed all the HIV-1 proteins except for Vpu and Env, which was substituted by the VSV-G protein to generate pseudotyped particles. Second generation LV additionally removed the genes Vif, vpr, vpu and nef. Third generation LV removed the tat regulatory gene, truncated the 5'LTR by addition of a the constitutive RSV promoter, and deleted the U3 element of the 3'LTR, generating self-inactivated (SIN)-LV. Adapted from [16].

The RMCE methodology presented in this work uses the Flp-FRT system and it is based on the differential property of the properly “targeted” cells to resist selection under G418 antibiotic compared to their un-recombined counterparts (Figure 3). This is caused by the presence of a truncated neomycin resistance gene (Δ neo) lacking the initial ATG codon situated downstream of the recombinase target site FRT of the tagged cell line, combined with the presence of an ATG triplet at the 3' end of the targeting cassette (Figure 3). When RMCE takes place, the initial ATG codon of the targeting cassette activates the expression of the neomycin resistance gene allowing the selection of the “targeted” clones by G418 selection [17] (Figure 3).

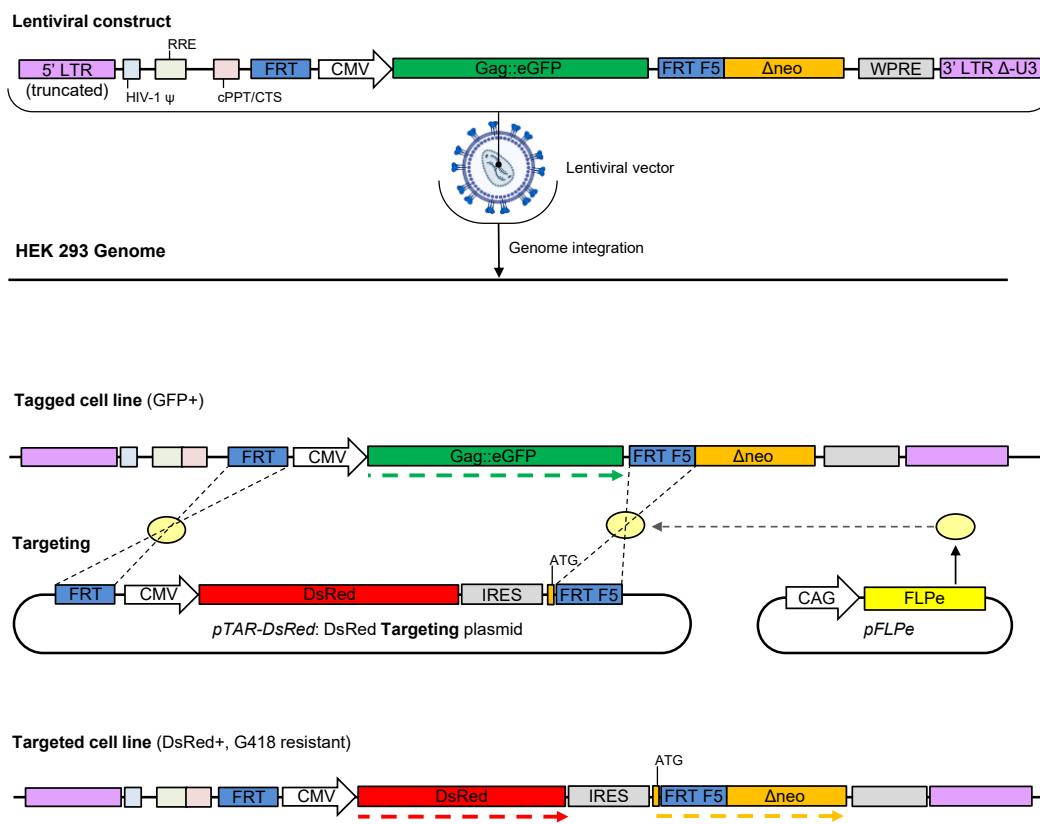


Figure 3. The RMCE methodology presented in this work is based on the Flp-FRT system and takes advantage of the differential ability of properly “targeted” cells to resist selection under G418 antibiotic compared to their un-recombined counterparts. From top to bottom: Lentiviral vector carrying the tagging cassette was generated and used to transduce its DNA into a host cell genome. The resulting tagged cell line was then transfected with pTAR-DsRed and pFLPe plasmids in order to perform RMCE. Genomic construction after targeting switches the GFP by DsRed expression while cells become G418 resistant.

Here, we generated a stable cell line presenting a tagging cassette coding for the Gag::eGFP protein flanked by heterospecific FRT RTs, as well as a

neomycin truncated gene (Figure 3). In order to achieve the presence of a single copy of the cassette into the cell genome, a third generation lentiviral vector carrying the tagging cassette was generated, titrated and used to infect the parental cell lines with a low MOI. Infected cells were sorted into enriched pools according to its GFP fluorescence. Then, cells were clonally isolated, expanded and adapted to suspension and serum-free media. The best resulting “tagged” clones were subsequently characterized and selected in terms of Gag::eGFP production, G418 sensitivity and ability to perform RMCE tagging reaction. The tagging vector used in this work coded for the Gag::eGFP gene, which consists of the HIV-1 Gag polyprotein fused in frame with the green fluorescent protein (GFP). With this approach, the gene acted simultaneously as a reporter protein which facilitated the tagging process, while generating Gag::eGFP VLPs, a valuable product of interest [18]. Once generated, the obtained cell line harbors the potential to have its Gag::eGFP gene exchanged by any other recombinant protein of interest.

2. Materials and methods

2.1 Cell lines, media and culture conditions

2.1.1 Parental cell line

A serum-free suspension-adapted HEK293 cell line (HEK293SF-3F6) was used, kindly provided by Dr. Amine Kamen from the Biotechnology Research Institute at the National Research Council of Canada (Montreal, Canada) and McGill University. This cell line was derived from a Current Good Manufacturing Practice (cGMP) master cell bank available for manufacturing of clinical material.

The medium used for HEK293 cellular growth was the chemically defined and free from animal components FreeStyle™ 293 (ThermoFisher, Waltham, MA, USA) supplemented with 0.1% Pluronic (ThermoFisher, Waltham, MA, USA), 1.6 mg/L of r-transferrin (Merck, Kankakee, IL), 19.8 mg/L of r-insulin (Novo Nordisk Pharmatech, Køge, Denmark) and 0.9X of an in-house developed lipid mixture to maximize cell growth [18]. After cell sorting the medium was supplemented with 10% heat-inactivated Fetal Bovine Serum (FBS) (26140079,

ThermoFisher, Waltham, MA, USA) and 50 µg/mL gentamicin (G1272, Merck, Kankakee, IL). For RMCE targeting, selection with G418 (J62671-03, ThermoFisher, Waltham, MA, USA) was used.

Suspension cell cultures were maintained routinely in exponential growth phase in 125ml disposable polycarbonate Erlenmeyer flasks with vent cap (Corning, Tewksbury, NY, USA) or 6-well plates (ThermoFisher, Waltham, MA, USA), placed in a LT-X Kuhner shaker (LT-X Kuhner, Birsfelden, Switzerland) shaking at 130 rpm, at 37°C, 5% CO₂ and 85% RH.

Adherent cell cultures were maintained in exponential grown phase in 25 mm² T-flasks, 75 mm² T-flasks (Sigma-Aldrich, Saint Louis, MO, USA), 96-well, 24-well and 6-well plates (Sigma-Aldrich, Saint Louis, MO, USA) in a humidified incubator at 37°C and 5% CO₂ in air. Adherent cell monolayers were washed with PBS and re-suspended by 4 min trypsinization treatment.

Cell counts and viability determinations were performed using the NucleoCounter®NC-3000 automatic cell counter (Chemometec, Lillerød, Denmark) following manufacturer's instructions. Adherent clones and cultures were observed under the Olympus CKX41 optical inverted microscope (Olympus, Tokyo, Japan) to qualitatively analyze cell growth, morphology and vitality.

Specific growth rate (μ) was calculated using the Equation 1. Specific productivity (q_p) and volumetric productivity (v_p) were calculated using Equation 2 and 3 respectively, where P is the product, X the viable cell density (VCD), t the time, and μ the specific growth rate of the cells during the selected time interval.

$$\mu = \frac{\ln(X_t - X_0)}{t_t - t_0} \quad (1)$$

$$q_p = \frac{\int_{t_0}^t P(t) \cdot dt}{\int_{t_0}^t X(t) \cdot dt} \cdot \mu \quad (2)$$

$$v_p = P/t \quad (3)$$

2.1.2 Lentiviral producer cell line

HEK-293T cell line (CRL-11268, American Type Culture Collection, Manassas, VA, USA) constitutively expressing the simian virus 40 (SV40) large T antigen, was used for lentiviral production.

The medium used for HEK-293T growth was DMEM-HG (4.5 g/L) (61965, Gibco, Carlsbad, CA, USA) supplemented with 10 mM HEPES, 0.1mM β -Mercaptoethanol (31350, Gibco, Carlsbad, CA, USA), 1mM sodium pyruvate solution (S8636, Sigma Aldrich, St. Louis, MO, USA), 1X Penicillin-Streptomycin-L-Glutamine (10378016, Gibco, Carlsbad, CA, USA), 1X Non-Essential Amino Acids Solution (11140, Gibco, Carlsbad, CA, USA) and 12.5% heat-inactivated FBS (F7524, Thermo Fisher Scientific, Waltham, MA, USA).

Cultures were maintained in exponential growth phase in 25 mm² T-flasks, 75 mm² T-flasks and 100mm culture dishes in a humidified incubator at 37°C and 5% CO₂ in air. Adherent cell monolayers were washed with PBS and then detached from adherent culture by 4 min trypsinization treatment.

2.2 Plasmids and transfection

2.2.1 Plasmid expression vectors

The pGag::eGFP plasmid codes for a codon optimized Rev-independent HIV-1 Gag protein fused in frame to the enhanced GFP driven by CMV enhancer and promoter. The plasmid from the NIH AIDS Reagent Program (Cat 11468) [19] was constructed by cloning the Gag sequence from pCMV55M1-10 [20] into the pEGFP-N1 plasmid (Clontech, Palo Alto, CA, USA).

The pLenti-GFP is a self-inactivating (SIN) lentiviral transfer vector. It encodes the eGFP gene controlled by the CMV promoter, both sequences flanked by FRT and FRT-F5 regions. A neomycin phosphotransferase gene lacking the initial ATG codon (Δ neo) is located downstream of FRT-F5 sequence. The whole mentioned construct, together with a Woodchuck Hepatitis Virus (WHV) Posttranscriptional Regulatory Element (WPRE), a Rev response element (RRE) and a HIV-1 packaging signal (HIV-1 ψ) are flanked by a 5' Long

Terminal Repeat (LTR) and a self-inactivating 3' ΔU3 LTR. Rous Sarcoma Virus (RSV) promoter is located at 5' of the 5' LTR.

The pLenti-Gag::eGFP is a self-inactivating (SIN) lentiviral transfer vector constructed from the pLenti-GFP plasmid as later explained and represented in Figure 4. It encodes the Gag::eGFP gene controlled by the CMV promoter, both flanked by the Flp Recombination Targets FRT and FRT-F5 regions. A neomycin phosphotransferase gene lacking the initial ATG codon (Δ neo) is located downstream of FRT-F5 sequence. The whole construct, together with a WPRE element, a RRE and a HIV-1 packaging signal (HIV-1 ψ) are flanked by a 5' LTR and a self-inactivating 3' ΔU3 LTR. RSV promoter is located at 5' of the 5' LTR.

The pGag-Pol, pRsv-Rev and pVSV-G plasmids encode for HIV-1-derived Gag-pol (structural proteins and viral enzymes), HIV-1 Rev (post-transcriptional regulator) and Vesicular Stomatitis Virus Glycoprotein (VSV-G), respectively. Those proteins provide the packaging and envelope functions needed for the generation of recombinant lentiviral particles containing the genetic constructs coded in pLenti- plasmids between its LTR sequences.

The targeting pTAR-DsRed vector comprises the DsRed reporter protein driven by CMV promoter and harbors an encephalomyocarditis virus (EMCV) internal ribosome entry site (IRES) element followed by an ATG translation initiation codon. The mentioned construct is flanked by Flipase Recognition Targets (FRT) FRT-WT and F5. Upon targeting the cassette of a tagged clone, the ATG ends up located in frame with the neomycin gene, conferring resistance towards G418 antibiotic (Figure 3).

The pFlp plasmid codes for the nuclear-targeted site-specific recombinase Flp driven by the CAG promoter.

pMock plasmid does not have any mammalian promoter or coding DNA sequence (CDS). It was constructed by the ligation of the pGag::eGFP backbone.

2.2.2 Competent Dcm methylation negative *E. Coli* generation

SCS110 is a Dam and Dcm methylation deficient *E. coli* strain kindly provided by Dr. Ramon Alemany from Catalan Institute of Oncology (IDIBELL, Barcelona). SCS110 chemically competent cells were generated following the rubidium chloride salts protocol. Briefly, the exponentially growing SCS110 culture was chilled in ice 5 minutes when its 600 nm absorbance reached 0.5 AU, centrifuged 5 min 4000 x g and its supernatant was discarded. Then, cells were resuspended in ice cold TFB1 buffer (100 mM RbCl, 50 mM MnCl₂, 30 mM Potassium Acetate, 10 mM CaCl₂, 15% Glycerol in milli-Q water at pH=5.8 adjusted with HCl) in a ratio of 3mL of buffer to 10mL of initial culture and incubated 90 min in ice. Finally, supernatant was removed as previously mentioned and cells were resuspended in ice cold TFB2 buffer (10 mM MOPS, 10 mM RbCl, 75 mM CaCl₂, 15% Glycerol in milli-Q water at pH=6.8 adjusted with 1 M KOH) in a ratio of 2mL of buffer to 10mL of initial culture. Chemically competent SCS110 aliquots were stored at -80°C.

2.2.3 Lentiviral plasmid coding for Gag::eGFP construction

For the generation of pLenti-Gag::eGFP, pGag::eGFP and pLenti-GFP plasmids were amplified in methylation-deficient SCS110-competent cells grown in LB medium (Conda, Madrid, Spain) at 32°C supplemented with 10 µg/mL kanamycin (Sigma, St. Louis, MO, USA) or 100 µg/ml ampicillin (Sigma, St. Louis, MO, USA) depending on the antibiotic resistance present on each plasmid. Plasmid purification was carried out using the plasmid Miniprep kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. pGag::eGFP and pLenti-GFP vectors were restricted using NdeI and the Dcm methylation sensitive PfoI restriction enzymes (Thermo Fisher Scientific, Waltham, MA, USA). Purification of the PCR products from agarose gel bands was carried out using Gel Extraction Kit (Qiagen, Hilden, Germany). The resulting CMV-Gag::eGFP construct fragment extracted from pGag::eGFP plasmid was cloned between the NdeI and PfoI restriction sites of the opened pLenti-CMV vector using T4 DNA ligase (Thermo Fisher Scientific, Waltham, MA, USA) (Figure 4). The obtained pLenti-Gag::eGFP plasmid was verified by sequencing.

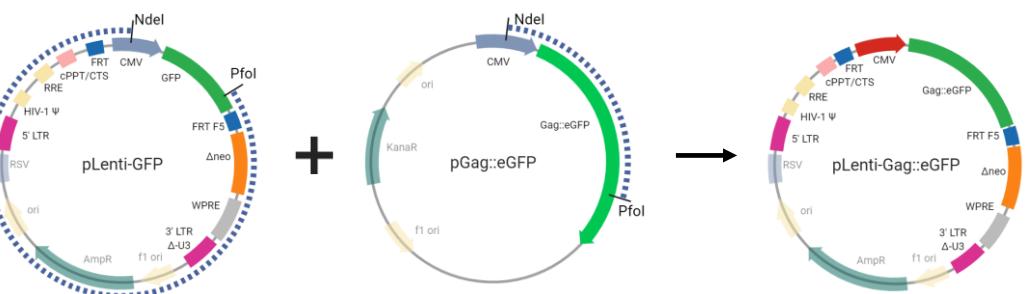


Figure 4. Schematic representation of the pLenti-Gag::eGFP plasmid generation from pLenti-Gag and pGag::eGFP. Two fragments (blue dotted lines) generated by NdeI and Pfol restriction of pLenti-GFP and pGag::eGFP vectors were purified and ligated to obtain pLenti-Gag::eGFP plasmid. The size of the fragments is not representative to its actual proportional length of the genes.

2.2.4 Plasmid amplification and Genomic DNA extraction

Plasmids were amplified in *Escherichia coli* Top10 strain grown in LB medium at 32°C supplemented with 100 µg/ml ampicillin (Sigma, St. Louis, MO, USA). Plasmid purification was carried out using the NucleoBond Xtra Maxi kit (Macherey-Nagel, Düren, Germany) according to the manufacturer's instructions.

Genomic DNA was purified using Wizard Genomic DNA Purification Kit (Promega, WI, Madison, USA) according to manufacturer's instructions. Purity and concentration was assessed with a Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). 0.5-1µg of genomic DNA template was used per PCR reaction. Junction analytic PCRs were performed using DreamTaq Green PCR mix (ThermoFisher, MA, USA), while PCRs used for junction amplification, purification and its subsequent sequencing were performed using Q5® High-Fidelity DNA polymerase (New England Biolabs, Ipswich, MA, USA).

2.2.5 Calcium-phosphate transfection for lentiviral vector generation

Exponentially growing 293T cells were seeded in 10 cm dishes 24 hours prior to transfection in order to have a homogeneous monolayer with 60-70% confluence at transfection time. One hour before the transfection, chloroquine was added to the culture for a final concentration of 25 µM. At transfection time, plasmid DNAs encoding for the lentiviral backbone and packaging proteins were mixed in 8 mL of milliQ water as follows: 10 µg of lenti backbone, 10 µg of

pGag-Pol, 5 µg of pRsv-Rev and 3 µg pVSV-G (Figure 5). Then, 400 µl of 1.25M CaCl₂ were added and mixed with a serologic pipette. After one minute, 2 mL of HEPES Buffered Salt solution (HBS) (280 mM NaCl, 50 mM HEPES, 1.5 mM Sodium phosphate dibasic heptahydrate in milli-Q water at pH=6.79 adjusted with NaOH) were added and quickly mixed by vigorously pipetting while creating bubbles. Then 4 mL of the transfection mixture were homogeny distributed per dish. After 4 to 6 hours, when sediment precipitates were observable under microscope, media was exchanged. Finally, transfected cells were incubated for 48 hours, the medium containing the produced lentiviral particles collected, centrifuged 15 minutes at 6000 x g supernatant was recovered, sterile filtrated at 0.45 µm, and aliquots were stored at -80°C.

2.2.6 PEI-mediated transient transfection

Exponentially growing HEK293 cells were passaged to have a cell density of 2·10⁶ cells/mL at transfection time. A media exchange was performed prior to transfection by centrifugation of the cells at 300g for 5 minutes. PEIpro® (Polyplus-transfection SA, Illkirch-Graffenstaden, France) was used as transfection reagent. PEI-DNA complexes were formed under sterile conditions. Briefly, plasmid DNA was diluted in culture media (10% of the total volume of cell culture to be transfected) for a final total DNA concentration of 1 µg/mL and vortexed for 10 s. Then, polyethylenimine (PEI), was added for a final concentration of 2 µg/mL (a 2:1 PEI:DNA ratio (w/w)) and vortexed three times for 3 s. The mixture was incubated for 15 minutes at RT and then added to the culture. The ratio between plasmids is described at Table 6.

2.3 Tagged cell line generation by lentiviral infection

One day prior to the infection, parental HEK293 cells were seeded in order to reach cell densities of ~60.000 cells / cm² at the adherent 6-well plates or 0.5·10⁶ cells/mL at the suspension 125 mL Erlenmeyer flask. Lentiviral infections were performed using a MOI of 0.01, diluting the tittered viral stock with culture media supplemented with 8 µg/mL polybrene (TR-1003-G, Merck, Kenilworth, NJ, USA). 6 hours post infection (hpi) residual virus was removed with a medium exchange. 7 days post infection (dpi) single transduced clones

expressing Gag::eGFP were isolated by fluorescent activated cell sorting (FACS).

2.4 Flow cytometry and cell sorting

For flow cytometry, adherent cells were detached by 4 minute trypsin treatment and resuspended in FACS buffer (2% FBS in PBS). Cells were analyzed using a Becton-Dickinson FACScalibur (BD Biosciences, San Jose, CA, USA) at Helmholtz Centre for Infection Research (Braunschweig, Germany) and a BD FACS Canto flow cytometer (BD BioSciences, San Jose, CA, USA), at Servei de Cultius Cel·lulars, Producció d'Anticossos i Citometria (Universitat Autònoma de Barcelona, Bellaterra, Catalonia, Spain). Data was processed using FlowJo V10 (BD Biosciences, San Jose, CA, USA). In order to remove cell aggregates before cell sorting, adherent and suspension pools were trypsinized 4 min.

2.5 PCR Analysis of Targeted Clones

Genomic DNA of G418-resistant targeted and un-targeted control clones was analyzed for correct integration of the DsRed targeting cassette by PCR (Figure 17). For this purpose, three primers were designed to bind to the integrated targeting and tagging cassettes, shown in Table 1. The PCR was carried on using the 2X Red PCR Master Mix (302004, PJK, Kleinblittersdorf, Germany) with the temperatures shown at Table 2.

Table 1. Oligonucleotides used in this work. Primers were designed avoiding possibilities of self-dimerization and primer-dimer formation.

| Name | Objective | Sequence (5' -> 3') |
|-------------|--------------------------------|------------------------|
| Fw GFP Tag | Tagging analysis | CTCGTGACCACCCCTGACCTA |
| Rv Neo | Tagging and Targeting analysis | CGTCAAGAAGGCGATAGAAGGC |
| Fw IRES Tar | Targeting analysis | GCCACGTTGTGAGTTGGATA |

Table 2. Reaction temperatures, cycles and steps.

| Reaction | Cycles: Temperature and time |
|-------------------------------------|---|
| Tagging and Targeting analysis PCRs | 95°C, 5 min + 72°C, 10 min + 34x[95°C, 30 sec + 55.1°C, 45 sec + 72°C, 90 sec] + 72°C, 10 min |

2.6 Toxicity assays

100 µl of HEK293 cells were seeded in a 96-well plate at $0.5 \cdot 10^6$ cells/ml together with 10 µl of increasing antibiotic concentrations, all points triplicated. The plate was cultured under standard HEK293 culture conditions. After two days 20 µl of MTT reagent (Promega, WI, USA) was added to each well and the plate was incubated for one hour at 37°C with an agitation of 130 rpm. Before the MTT addition, a standard curve of cell concentrations from 0 to $4 \cdot 10^6$ viable cells/ml was performed in duplicated columns. Absorbance at 490 nm was read on Multilabel Plater Reader VICTOR3® (Perkin Elmer, MA, USA).

2.7 Spectrofluorometry-based product quantitation

Green fluorescence intensity in Gag-GFP samples was measured using a Cary Eclipse Fluorescence Spectrophotometer (Agilent Technologies, Santa Clara, CA, USA). The instrument parameters were set as follows: $\lambda_{\text{ex}} = 488$ nm (slit 5 nm), $\lambda_{\text{em}} = 510$ nm (slit 10 nm). Readings were carried out at room temperature. Relative fluorescence units (RFU) values were calculated by subtracting fluorescence units (FU) of negative control samples from that given by the sample.

2.8 Confocal microscopy

The imaging of the clones was performed using Leica TCS SP5 confocal fluorescence microscope (Leica Microsystems, Wetzlar, Germany) and Olympus FV1000 confocal fluorescence microscope (Olympus, Tokyo, Japan) at Servei de Microscòpia (Universitat Autònoma de Barcelona, Bellaterra, Catalonia, Spain). Prior to visualization, cells were treated with 0.1% (v/v) of Hoechst 33342 (Thermo Fisher Scientific, Waltham, MA, USA) and 0.1% (v/v) of CellMask Deep Red (Thermo Fisher Scientific, Waltham, MA, USA) in order to stain cell nuclei and lipid membranes, respectively. Samples were placed in 35 mm glass bottom Petri dishes with 14 mm microwells (MatTek Corporation, Ashland, MA, USA) prior to its visualization under the microscope.

2.9 Nanoparticle Tracking Analysis

NTA-based Gag::eGFP VLP quantification and characterization was performed with a NanoSight®NS300 (Nanosight Ltd., Amesbury, UK) equipped with a blue filter module (488 nm) and a neutral filter at the Soft Material Service of the Institut de Ciència de Materials de Barcelona (ICMAB-CSIC, Bellaterra, Catalonia, Spain). Samples were previously diluted to a concentration of approximately 10^8 particles/mL. Sample injection was performed using a pump in order to improve the robustness of the measurement by continuous addition, and to minimize the photobleaching effect due to fluorescence depletion over time. 60 second videos were recorded at RT and analyzed with the NTA 3.4 software (Malvern Panalytical, Malvern, UK). Tracked particles size was determined from its Brownian motion. Three independent experimental replicas were carried out for each sample. Camera level and detection threshold were manually adjusted for each replica.

2.10 Flow virometry

VLP quantification by Flow Virometry was performed in a CytoFLEX (Beckman Coulter Inc., Brea, CA, USA) equipped with a violet side scatter (V-SSC) 405 nm filter configuration. The threshold of area trigger signal in V-SSC was set to 300 and the laser gains were set as follows: 72 for FSC, 135 for SSC, 9 for V-SSC, and 500 for FITC. Samples to be analyzed were diluted in 0.22 μm filtered HyClone™ PBS (GE Healthcare, Chicago, IL, USA) in order to be in the range of 500 - 5000 events/ μL with an abort rate below the 5%. 300.000 events were analyzed per sample at a flow rate of 10 $\mu\text{l}/\text{mL}$. V-SSC vs B525-FITC density plots were used to gate the VLPs. Results were analyzed with the CytExpert v.2.3 software (Beckman Coulter Inc., Brea, CA, USA) and VLP concentrations were calculated with Equation (4), where $C_{\text{NTA/FV}}$ is the standard for data harmonization between NTA and Flow Virometry instruments.

$$\text{VLP conc.} \left(\frac{\text{VLPs}}{\text{mL}} \right) = \frac{\text{events}}{\mu\text{L}} \cdot \frac{1000 \mu\text{L}}{\text{mL}} \cdot \text{Dilution} \cdot C_{\text{NTA/FV}} \quad (4)$$

2.11 Digital droplet PCR (ddPCR)

2.11.1 Genomic DNA pre-digestion for ddPCR

A pre-digestion of the purified genomic DNA was performed in order to ensure that all potentially linked tandem gene copies would be randomly and independently distributed into the ddPCR droplets. For this purpose 2000 ng of each cell line gDNA were restricted with 14 units of HindIII (New England Biolabs, Ipswich, MA, USA) in a total digestion mixture of 50 µl for 1 hour at 37°C. HindIII was selected as it does not cut inside the RPP30 nor Gag::eGFP designed amplicons, is a high-fidelity enzyme with low star activity, has no methylation-sensitivity, and presents 100% activity in NEBuffer r2.1 (New England Biolabs, Ipswich, MA, USA) which does not interfere with ddPCR. No inactivation or further purification was performed.

2.11.2 Primer design for ddPCR

Two pairs of primers were designed to target and amplify fragments of 101 and 141 bp from Gag::eGFP and Ribonuclease P Protein Subunit P30 (RPP30) genes respectively, using the online PrimerQuest Tool from IDT (Integrated DNA Technologies IDT-DNA, Leuven, Belgium) website (<https://eu.idtdna.com/pages/tools>). The salt-adjusted Tm and primer self-complementarity were studied using the OligoCalc v3.27 online calculator [21] (<http://biotools.nubic.northwestern.edu/OligoCalc.html>) and primer-dimer estimation was assessed by using Multiple Primer Analyzer tool from ThermoFisher (Waltham, MA, USA) website (<https://www.thermofisher.com/>). RPP30 sequence was retrieved from the NCBI GenBank database (Gene ID: 10556, National Centre for Biotechnology Information, Bethesda, MD, USA) website (www.ncbi.nlm.nih.gov/genbank/). RPP30 assay primer sequences were (forward primer) 5'-CCAACCTCATGCCACCCAGACCATC-3' and (reverse primer) 5'-CCCGCGCTAGGAATCAGACCAACAC-3'. Gag::eGFP assay primer sequences were (forward primer) 5'-GTCCAGGAGCGCACCATCTTCTTC-3' and (reverse primer) 5'-TTCAGCTCGATGCGGTTACCCAG-3'.

2.11.3 qPCR for ddPCR annealing determination

In order to optimize the annealing temperatures of the primers to be used in the ddPCR, a qPCR with a temperature gradient of 55.4 to 60.7°C was performed in a C1000 Touch thermal cycler (#1851197, Bio-Rad, Hercules, CA, USA). Each 10 µl PCR mixture contained 2x iTaq Universal SYBR Green (#1725151, Bio-Rad, Hercules, CA, USA), primers at a final concentration of 200 nM (IDT-DNA, Leuven, Belgium) and 80 ng of gDNA. The content was then transferred to a 96-well PCR plate, sealed and cycled using the following conditions: 95 °C for 5 min (1 cycle); 95 °C for 30 s and 55.4–60.7 °C for 1 min (40 cycles); 4 °C for 5 min, 90 °C for 5 min (1 cycle), 4 °C hold. On the basis of this assessment, the annealing temperature was determined to be 58°C for both RPP30 and Gag::eGFP primers. Additionally; its melting curves showed a single peak, indicating the presence of a single specific amplicon, as it is crucial to avoid unspecific amplifications such as primer dimers when performing ddPCRs. The tested gDNA samples had Cq values varying between 20.5 and 23.5 cycles at 58°C. As a consequence, samples were diluted for the subsequent ddPCRs in order to avoid saturation by an excessive number of positive droplets.

2.11.4 ddPCR for Copy Number Variation (CNV) determination

Each ddPCR mixture contained 2x EvaGreen ddPCR Supermix (#1864034, Bio-Rad, Hercules, CA, USA), the corresponding pair of primers at a final concentration of 200 nM (IDT-DNA, Leuven, Belgium) and 16 ng of gDNA (as the previous qPCR assay determined). The 20 µl PCR mixtures were loaded into DG8 Cartridges together with 70 µl of Droplet Generation Oil (#1863009 and #1863005 respectively, Bio-Rad, Hercules, CA, USA). After gasket attachment, the cartridges were placed into the QX200 Droplet Generator (#1864002 Bio-Rad, Hercules, CA, USA), which generated ~20.000 droplets per sample. The content was then transferred to a 96-well PCR plate, sealed by means of a PX1 PCR Plate Sealer (#1814000, Bio-Rad, Hercules, CA, USA) and cycled using the following conditions: 95 °C for 5 min (1 cycle); 95 °C for 30 s and 58°C for 1 min (40 cycles); 4 °C for 5 min, 90 °C for 5 min (1 cycle), 4 °C hold. The QX200 Droplet Reader (#1864003, Bio-Rad, Hercules, CA, USA) was used to read the individual cycled droplets. The described process was carried

on at Laboratori de Luminescència i Espectroscòpia de Biomolècules from the Universitat Autònoma de Barcelona.

The generated data was analyzed using QuantaSoft Software 1.7.4.0917, which determined the absolute quantity of target DNA molecules by fitting the number of positive and negative droplets after amplification (>31.000 accepted droplets per condition) into a Poisson distribution. The used statistical analysis took into account and compensated for the droplets presenting two or more target DNA copies due to the random nature of partitioning when droplets were generated.

2.12 Cryopreservation

The cryopreservation mixture consisted of 7.5% DMSO (Sigma-Aldrich, Saint Louis, MO, USA), in fresh medium. Each cryotube™ (ThermoFisher, MA, USA) contained 1 mL of cryopreservation mix and $5 \cdot 10^6$ cells. Cryotubes were placed in a Mr. Frosty™ (ThermoFisher, MA, USA) to reach -80°C with a rate of 1°C/minute. After 24 hours, cryotubes were stored in liquid nitrogen.

3. Results and discussion

3.1 Lentivirus generation and titration

The generation of a lentivirus carrying the tagging construct of interest expressing the Gag::eGFP gene under CMV promoter, was achieved by co-transfection of HEK293T viral producer cell line with the pLenti-backbone (Gag::eGFP or GFP for positive control) transfer vector, the packaging helper plasmids pGag-Pol, pRsv-Rev and the envelope plasmid pVSV-G (Figure 5) using the calcium phosphate precipitation method. HEK293T cells were used as it has been shown that the SV-40 T-antigen present in the cell line, increases its vector production efficiency [15].

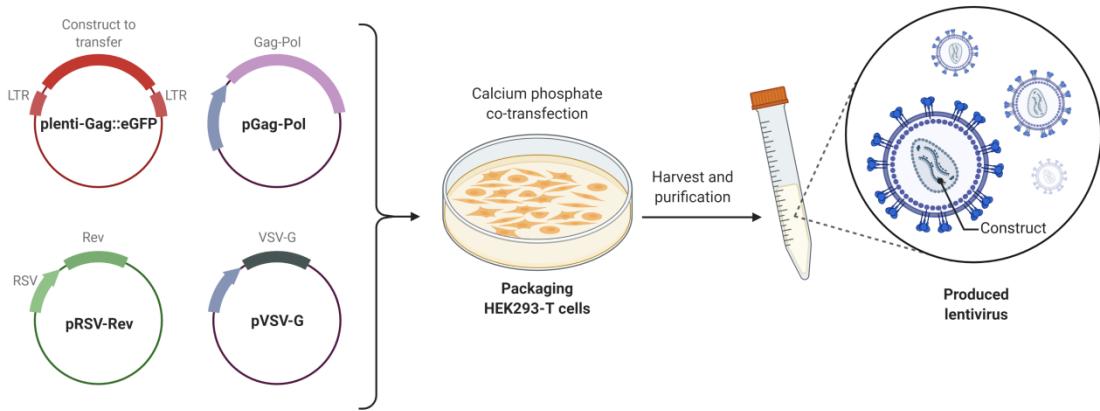


Figure 5. Schematic representation of the four-plasmid co-transfection for the generation of third generation lentiviral vectors. HEK293-T cells were transfected with a pLenti-backbone transfer vector, the packaging helper plasmids pGag-Pol, pRsv-Rev and the envelope plasmid pVSV-G using the calcium phosphate precipitation method.

The obtained viral stock was then tittered by infecting a known concentration of adherent HEK293 cells with different viral stock dilutions and analyzing its fluorescent population at 48 hours post-infection (hpi) (Figure 6). The viral titter was determined using Equation (5) below.

$$\text{Titer } \left(\frac{\text{viruses}}{\mu\text{l}} \right) = \frac{\text{cells} \cdot \text{Fluorescent cells (\%)} \cdot \text{dilution}}{\text{culture volume (\mu l)}} \quad (5)$$

This lead us to determine that the positive control lenti-GFP had a concentration of 436.5 virions/ μl while the lenti-Gag::eGFP tagging concentrate had 56.16 virions/ μl . The difference in titer is coherent with the bibliography as the increase of the vector genome size has been correlated with a significant reduction in functional lentiviral titer [13], and lenti-Gag::eGFP construct is greater than lenti-GFP by 1.5 kbp.

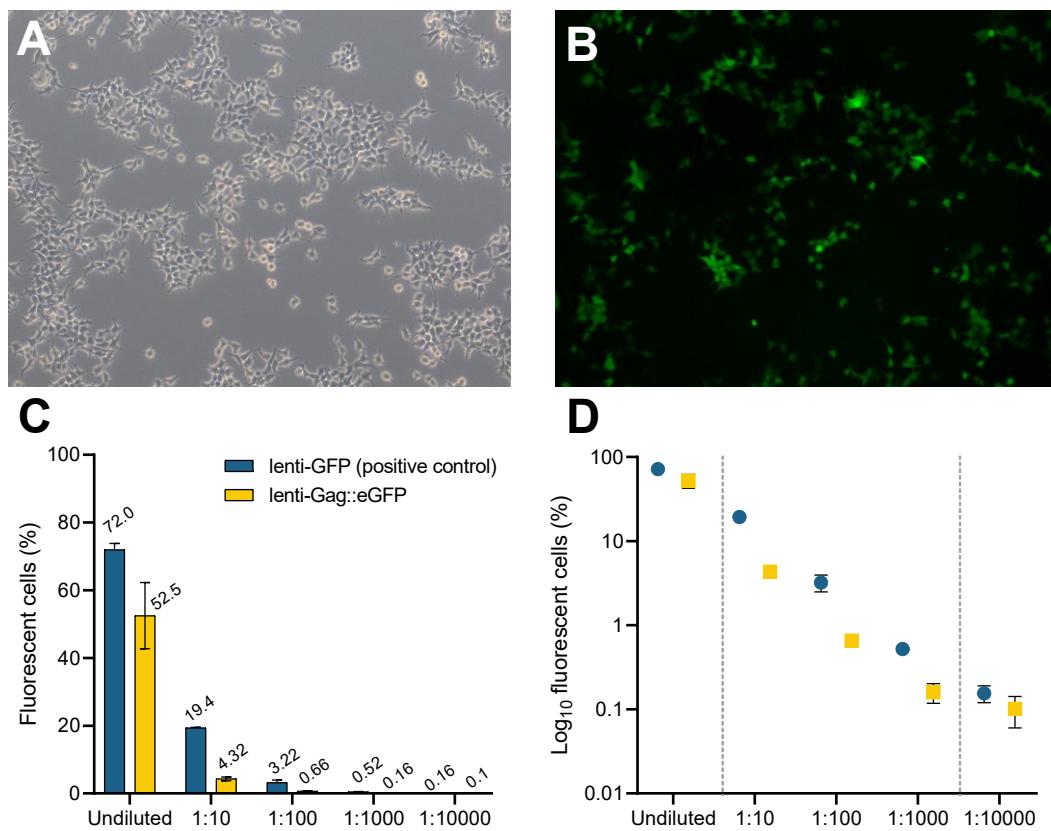


Figure 6. Titering of the generated lentiviral stocks. **(A-B):** Adherent cells infected with undiluted lenti-GFP lentivirus at 48 hpi: **(A):** Bright-field, **(B):** GFP channel. **(C-D):** Fluorescent cellular populations upon lentiviral infection with different stock dilutions. **(D):** An exponential fashion of positive fluorescent population correlating with the dilutions was observed.

3.2 Cell pool generation by lentiviral infection: tagging

As mentioned earlier, to generate a RMCE stable cell line, the construct of interest containing the transgene, recombination sequences and the truncated antibiotic resistance, needs to be integrated into the cellular genome in the tagging step. When tagging cells for the subsequent generation of a stable cell line, the number of integrated copies in the cellular genome is relevant as it will affect the targeting recombination efficiency of the clones. To achieve a viral transduction of one copy of the RMCE tagging construct per clone, HEK293 cells growing in suspension and in adherence were infected in parallel using a MOI of 0.01 viral particles per cell. At 3 dpi, suspension cells were subjected to a media exchange and adherent cells were amplified in T-flasks. At 6 dpi, both groups showed normal cellular growth and tagged Gag::eGFP producer cells were observable as GFP fluorescent by confocal microscopy (Figure 7).

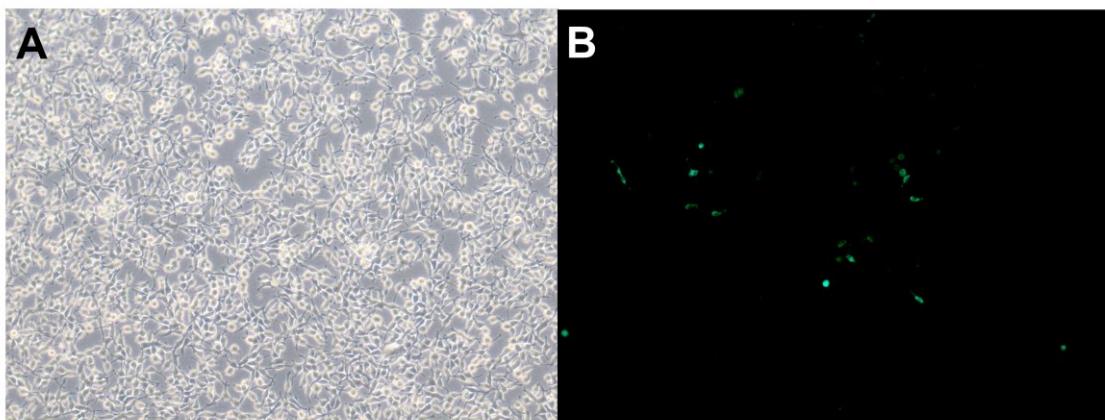


Figure 7. Adherent 293 cells at 6 days post-infection observed by confocal microscopy. **(A):** Bright-field image shows cells forming an adherent monolayer close to confluence. **(B):** Tagged Gag::eGFP producer cells can be observed at the GFP channel (green).

3.3 Pool generation by cell sorting enrichment

At 6 dpi, positive cell enrichment was performed in order to prevent the dilution of the expression tagged cells, as they have no survival advantage over non-infected cells. To select the Gag::eGFP producer clones, the cell sorting was based on GFP+ fluorescence, using two gating criteria for each group, depending on fluorescence (Figure 8). Sorting pools were coded as ‘high’ or ‘medium’ expression, as presented in Table 3. The more restrictive the FITC criteria was, the less cells from the pool matched it.

Table 3. Pool enrichment gating criteria: Parental infected cells group, gating criteria, FITC-A threshold, percentage of the cellular population meeting the criteria and number of sorted clones.

| Parental infected cells | Gating criteria | FITC-A gating criteria | % of parent population | Sorted cells | Pool name |
|-------------------------|-----------------|------------------------|------------------------|----------------|-------------------|
| Adherent | Medium | [200-1300] | 3.86% | $7 \cdot 10^5$ | Adherent Medium |
| | High | >1300 | 1.05% | $2 \cdot 10^5$ | Adherent High |
| Suspension | Medium | [700-5000] | 10.83% | $6 \cdot 10^5$ | Suspension Medium |
| | High | >6000 | 4.7% | $3 \cdot 10^5$ | Suspension High |

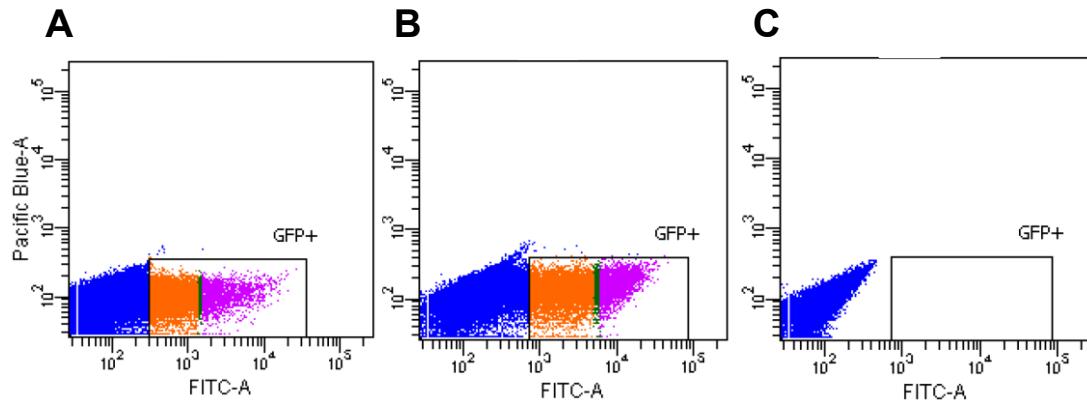


Figure 8. Gating criteria used in the cell sorting for the pool enrichment based on GFP+ fluorescence of the clones. High GFP expressers gating criteria is shown in purple, medium expressers is shown in orange. **(A):** Adherent pools. **(B):** Suspension pools. **(C):** Negative control: Wild-type HEK293

3.4 Clonal selection: cell sorting

14 days after pool enrichment, cells were sorted in adherence into 96-well plates to separate individual clones from each polyclonal tagged pool. Sorting criteria is shown in Figure 9. Adherent pool was showing good growth and viabilities while suspension pool presented low viabilities and abnormal cellular growth, hence the low number of isolated clones from suspension pools, as shown in Table 4.

Table 4. Gating criteria: FITC threshold, percentage of the cellular population meeting the criteria, number of isolated clones, survival and nomenclature.

| Pool | FITC-A gating criteria | % of pool population | Isolated clones | Growing clones | Survival (%) | Clone nomenclature |
|-------------------|------------------------|----------------------|-----------------|----------------|--------------|--------------------|
| Adherent Medium | >600 | 20.08% | 672 | 22 | 3.27% | AM"number" |
| Adherent High | >1300 | 19.72% | 768 | 28 | 3.64% | AH"number" |
| Suspension Medium | >600 | 84.67% | 288 | 3 | 1.04% | SM"number" |
| Suspension High | >1300 | 62.72% | 384 | 9 | 2.34% | SH"number" |

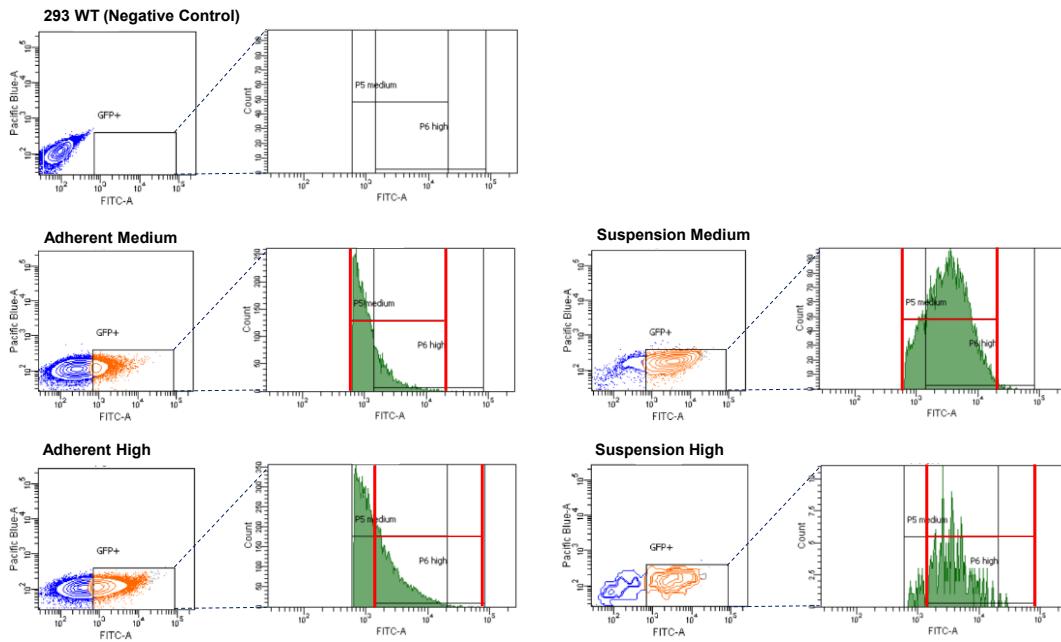


Figure 9. Gating criteria for the clonal sorting based on GFP+ fluorescence. Cells were sorted into 96-well plates to separate individual clones from each polyclonal tagged pool.

3.4 Clonal expansion

Clones were seeded with Freestyle Mix supplemented with 10% FBS and media was exchanged periodically. Clones were sequentially amplified to 24-well, 12-well, 6-well plates, 25 mm² and 75 mm² T-flasks when confluent. From the 2112 isolated clones, 62 successfully went through the clonal growth and amplification process, which lasted four weeks. Adherent clones had a slightly superior survival rate (~3.5%) than suspension ones (~1.7%). A cryopreserved stock of each amplified clone was generated at this point.

3.5 Gag::eGFP production quantification of the clones

Clones presenting confluence and high viabilities at 25 cm² T-flasks were analyzed using flow cytometry (Figure 10). Previous studies on the generation of stable cell lines showed a direct correlation between flow cytometry relative GFP mean values and Gag::eGFP expression. The 13 clones exhibiting greater GFP fluorescence intensity mean values (Figure 10.E) were selected as promising for further studies. All of them showed >93% positive GFP population. From the best 13 clones, 9 were derived from the “adherent high” pool, two from the “suspension high” pool, while “suspension medium” and “adherent medium” were represented with one clone in each pool.

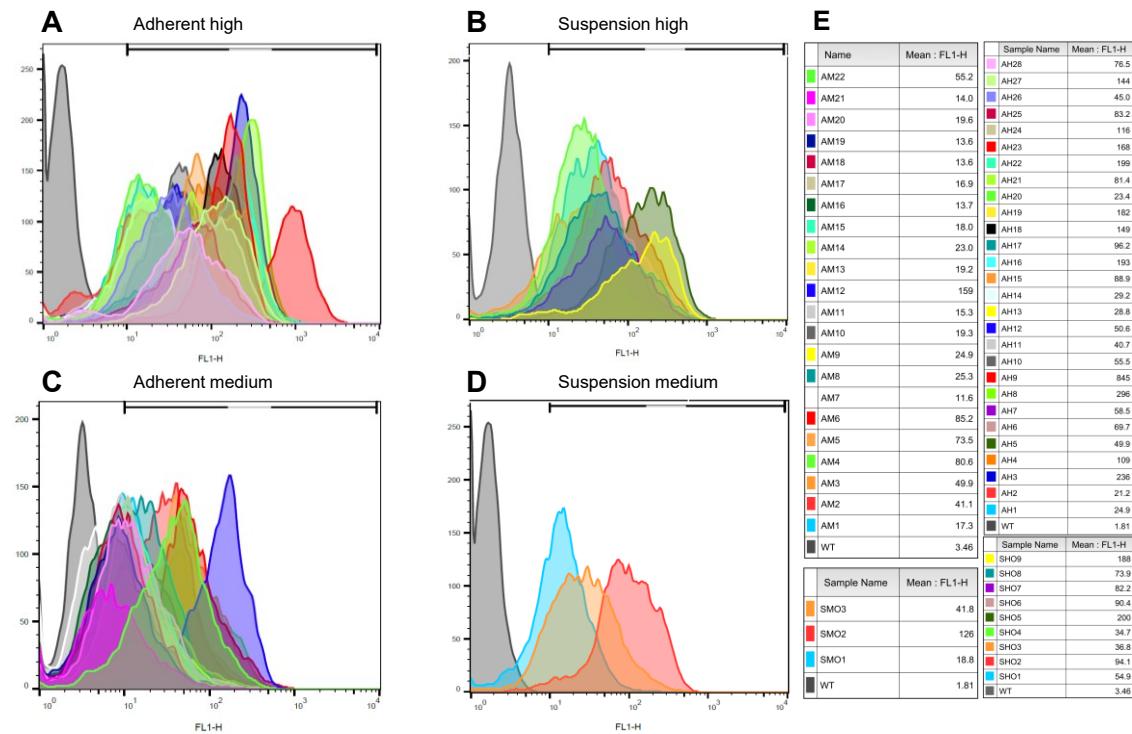


Figure 10. Clones analysis by flow cytometry. The 13 clones exhibiting greater GFP fluorescence intensity mean values were selected as promising for further studies. (A): Overlay of the clones derived from the “adherent high” pool. (B): Overlay of the clones derived from the “suspension high” pool. (C): Overlay of the clones derived from “adherent medium” pool. (D): Overlay of the clones derived from “suspension medium” pool. (E): Clone color legend and its mean GFP fluorescence intensity.

3.5 Suspension and serum-free media adaptation

Production cell lines need to achieve high cell densities and allow an easy scale up of the bioprocess. Suspension cell growth facilitates the achievement of those goals. Additionally, cells are growing in presence of FBS, previously introduced to facilitate the clonal growth from single cells after cell sorting. The adaptation of the suspension cell lines into FBS-free media was carried by gradual reduction of serum concentration.

The 13 selected clones were defrost and amplified in T-Flasks. Viability and cell density were analyzed to monitor adherent growth. When clones exhibited 85-90% confluence, the cells of two parallel 75cm² T-flasks were detached and seeded at 2·10⁶ viable cells/mL in 125 mL polycarbonate Erlenmeyer flasks, with 10 mL of media with its FBS supplementation reduced at 5%. Cells were maintained in exponential growth by routinely diluting them when reaching densities higher than 1.5·10⁶ viable cells/mL. When clones showed exponential

growth in addition with viabilities higher than 90% during at least 2 passages, they were considered adapted to suspension culture. As can be observed in Figure 11, the viability of the different cell lines improves gradually over time, evidencing suspension adaptation of all the clones (except for the AH19) during the first five days. Although AH19 initially struggled to grow, it achieved suspension adaptation at day 11.

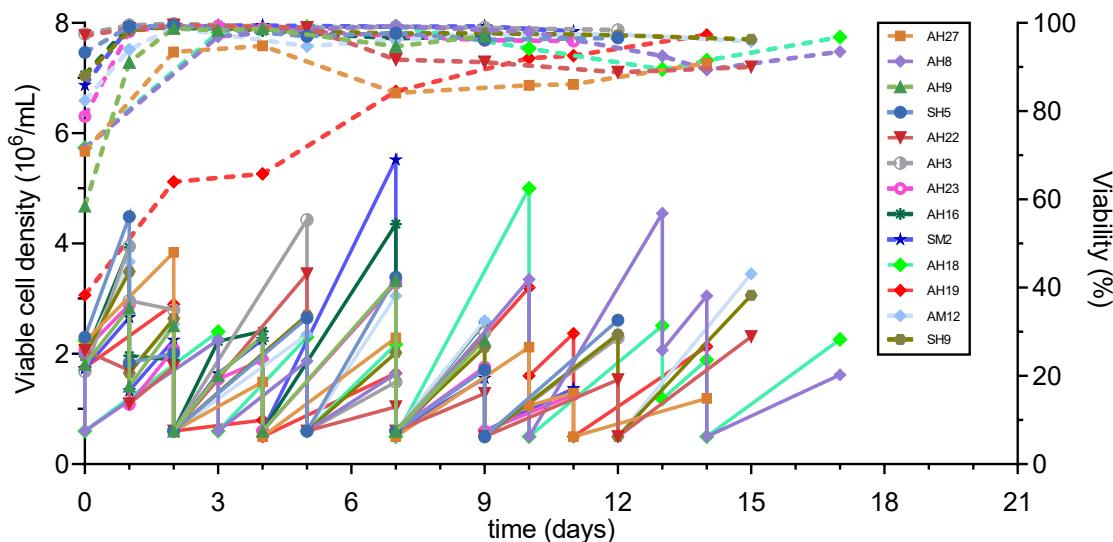


Figure 11. Suspension and serum-free media adaptation. The viability of the different cell lines improves gradually over time, evidencing suspension adaptation of the majority of the clones between days 4 and 5. Then, cell lines were adapted to serum-free media: a part from the clones AH22, AH27 and AM12 -which showed a slight reduction in viability during its adaptation process- no significant changes in cell viabilities were observed. Solid lines: viable cell density, dashed lines: viability.

The adaptation of the clones to FBS-free media was carried out after suspension adaptation by gradual reduction of serum concentration from 5% to 0%. Cells were seeded by reducing its FBS concentration each passage achieving a sequential adaptation. Apart from the clones AH22, AH27 and AM12 -which showed a slight reduction in viability during its adaptation process- no significant changes in cell viabilities were observed (Figure 11). Clones were considered successfully adapted to serum-free media after 3 passages without serum. Adapted clones exhibited a constant growth rate and viabilities above 93%. A cryopreserved cell bank for each adapted clone was generated.

3.6. Growth kinetics of the best producers

When clones were adapted to suspension and serum-free media growth, a preliminary study of their Gag::eGFP specific productivity was performed. For this purpose supernatant fluorescence was analyzed (Figure 12.A). This allowed the selection of the best four candidates from the 13 adapted clones, named AH9, AH8, AH22, SH5 and AH27.

Growth kinetic studies of the top five clones were performed to evaluate their duplication time and maximum cell density (Figure 12.B).

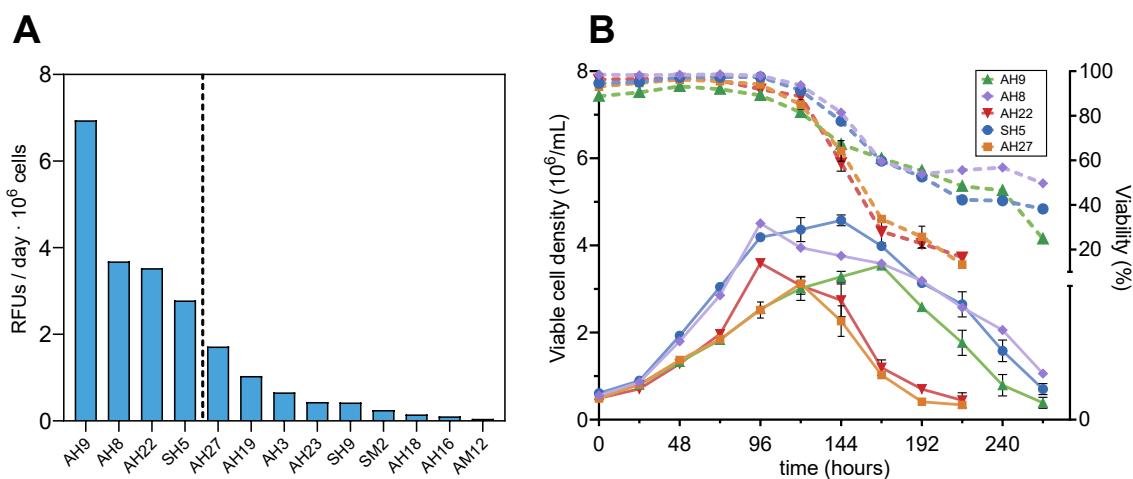


Figure 12. (A): Relative fluorescence units produced per day and million cells, analyzed from the supernatant of the 13 adapted RMCE clones, 48 hours post seeding. (B): Growth kinetics of the suspension and serum-free adapted clones AH9, AH8, AH22, SH5 and AH27.

As shown in Table 5, clones AH9 and AH22 showed maximum cell densities of approximately $3.5 \cdot 10^6$ cells/mL (similar to that of wild-type HEK293 cultured in the same media and conditions), while AH8 and SH5 reached $4.5 \cdot 10^6$ cells/mL approximately. The time where the maximum took place also differs between clones, being 96 h for clones AH22 and AH8; 144 h for SH5 and 168 h for AH9. All clones presented similar duplication times between 29 and 33.2 h, while AH9 showed a much slower growth profile, with a duplication time of 43.2 h. For comparison, HEK293 cells showed duplication times of 26.2 h (Table 5).

Table 5. Duplication time and maximum cell density of the suspension and serum-free adapted clones and HEK293 wild-type cells for reference.

| Clone name | Duplication time (hours) | Max. cell density (viable cells / mL) |
|------------|-----------------------------|--|
| AH9 | 43.2 ± 0.9 | (3.54 ± 0.01)·10 ⁶ |
| AH8 | 30.3 ± 1.2 | (4.50 ± 0.20)·10 ⁶ |
| AH22 | 31.4 ± 2.9 | (3.59 ± 0.06)·10 ⁶ |
| SH5 | 29.0 ± 2.5 | (4.58 ± 0.13)·10 ⁶ |
| AH27 | 33.2 ± 2.0 | (3.12 ± 0.04)·10 ⁶ |
| HEK293 | 26.2 ± 1.0 | (3.53 ± 0.3)·10 ⁶ |

3.7 G418 sensitivity assay

The next step was the selection of the best candidates in terms of its targeting capability. As previously mentioned, to allow the selection of the “targeted” sub-clones emerged from the RMCE reaction, it is relevant that the parental tagged cell line presents a completely inactivated neomycin resistance gene with no leaky genetic expression.

To test its antibiotic sensitivity, suspension clones were cultured in presence and absence of G418. Wild-type HEK293 cells were used as a G418-sensitive control cell line. When cell lines treated with G418 showed no cellular growth together with viabilities lower than 60% (horizontal grey dotted line), they were considered G418-sensitive. Otherwise they were catalogued as G418-resistants. From the 5 tested clones, AH22, AH27 and SH5 were G418-sensitive (Figure 13). Cell lines AH8 and AH9 -which had previously shown the highest Gag::eGFP production levels-, were discarded for being G418-resistants.

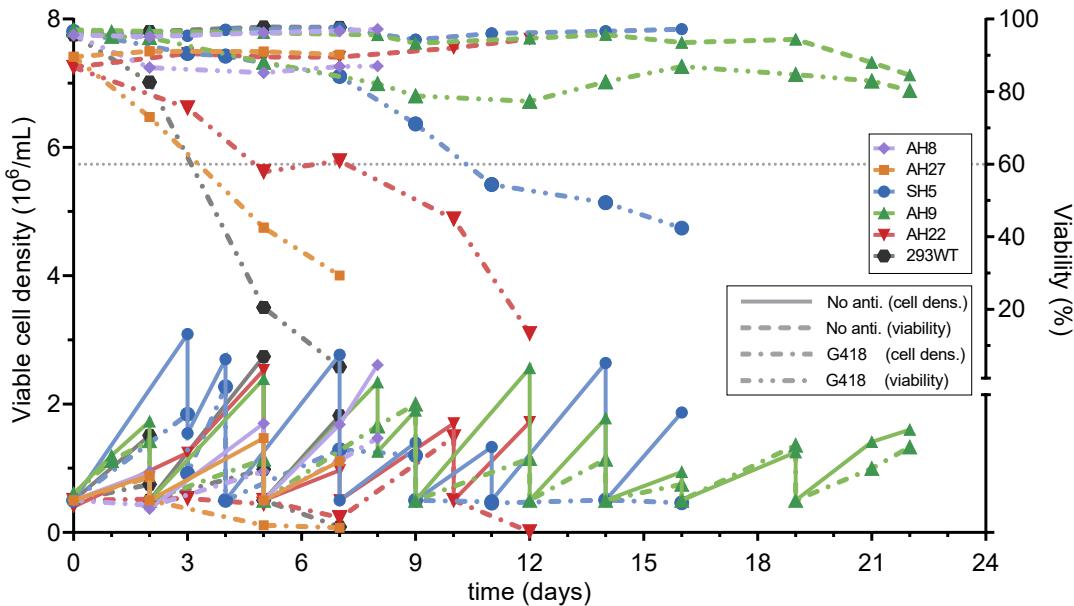


Figure 13. G418 sensitivity assay of the five potential best clones. When cell lines treated with antibiotic showed no cellular growth together with viabilities lower than 60% (horizontal grey dotted line), they were considered G418-sensitive. Otherwise they were catalogued as G418-resistants. From the 5 tested clones, AH22, AH27 and SH5 resulted G418 sensitive. Solid lines: untreated clones viable cell density; dashed lines: untreated clones viability; dash-dotted lines: G418-treated clones viable cell density; dash-double-dotted lines: G418-treated clones viability.

3.8 Targeting analysis

To test the targeting capability of the three candidate cell lines, they were co-transfected using a targeting vector (pTAR-DsRed) together with a plasmid coding for the flipase responsible for the FRT-FLP reaction (pFIP). The targeting vector pTAR-DsRed codes for the DsRed reporter protein and its promoter, flanked by FRT sequences. With this targeting strategy, cell lines undergoing adequate targeting shift from single GFP expressers to become single DsRed expressers as RMCE proceeds upon G418 selection.

Table 6. RMCE transfection groups and its plasmid composition.

| Transfection group | Plasmid A | Plasmid B | Ratio |
|--------------------|------------|-----------|-------|
| TAR | pTAR-DsRed | pFIP | 1:1 |
| FIP | pFIP | pMock | 1:1 |
| DsRed | pTAR-DsRed | pMock | 1:1 |

Suspension exponentially growing cells were co-transfected using PEI as transfection reagent. Control transfections were also performed (Table 6). Two days post-transfection (dpt), all groups reached high cell densities and were diluted in antibiotic-containing media to have a final concentration of 3 mg/mL of G418. After antibiotic addition, all groups experienced a continuous decline in viability and cell growth (data not shown). To provide fresh nutrients to the cells, media was exchanged every two or three days. At 8 dpt cells were cultured in adherence with G418 concentration decreased to 1.5mg/ml. During the subsequent passages, the single-transfected Flp and TAR-DsRed control groups did not show adherent growth, experiencing apoptosis until the death of all the cells (Figure 14.B,C). Otherwise, the targeted cell lines showed adherent growth of clustered single DsRed + cells, as observed by confocal microscopy (Figure 14.A,D).

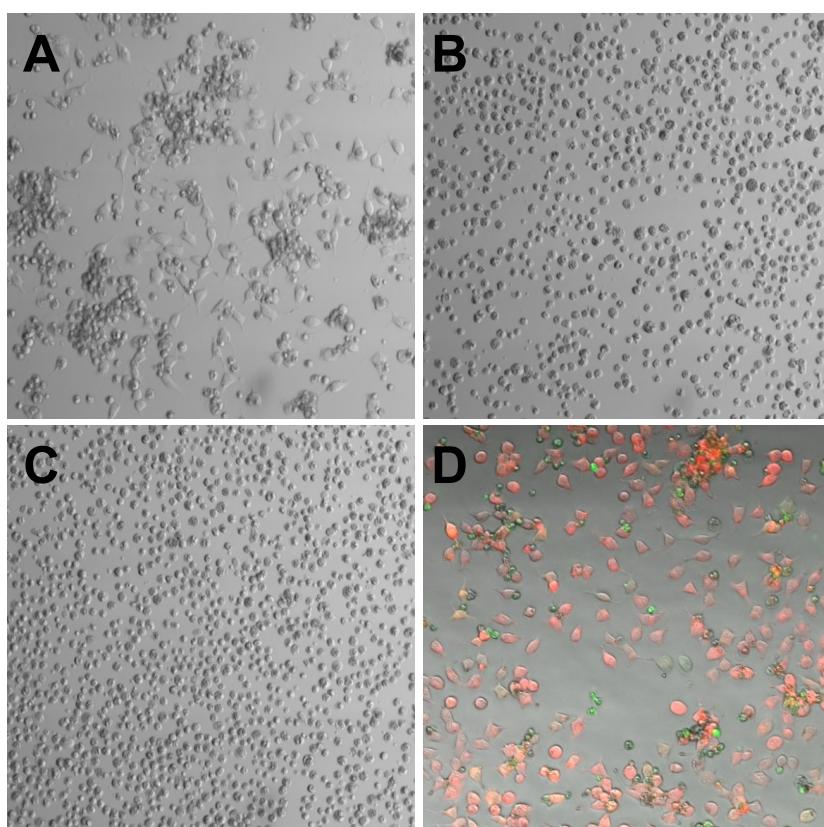


Figure 14. Microscopy images of the three transfected groups (as detailed in Table 6) of the clone AH22 growing in adherence at 15 dpt. (**A-C**): Bright-field images of TAR (**A**), DsRed (**B**) and Flp (**C**) transfection groups. TAR RMCE cells (**A**) show adherent growth under G418 selection, unlike DsRed (**B**) and Flp (**C**) controls. Apoptotic cells (non-adhered round-shaped) were easily removed by media Exchange. (**D**): Merge of DsRed, GFP and bright-field channels. It can be easily appreciated that only single DsRed positive cells showed good adherent growth.

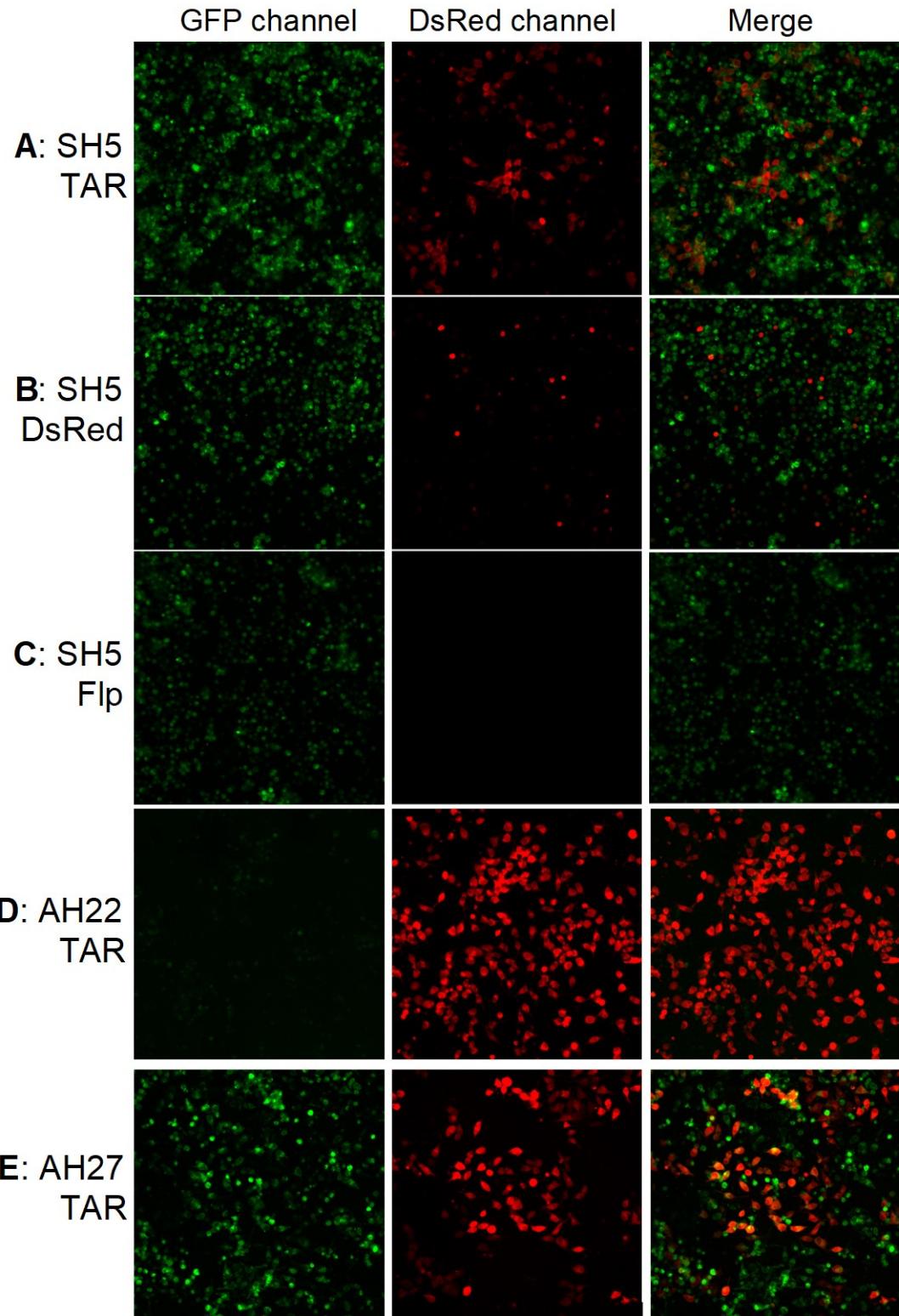


Figure 15. Transfected groups under fluorescent confocal microscopy. SH5 Flp and DsRed control groups did not show adherent growth, experiencing apoptosis (round shaped cells) even of the DsRed transiently transfected positive cells. As can be observed in DsRed and merge channels, the SH5 and AH22 targeted group showed adherent growth of clustered DsRed + cells. AH27 targeted group was not able to derive into a single DsRed population, showing also adherent growth of GFP positive cells.

Some of the TAR DsRed+ adherent cells presented Gag::eGFP leftovers on its cytoplasm, presumably from expression occurred before targeting took place. As expected, upon different passages and RMCE taking place, Gag::eGFP presence on the viable cells disappeared. On the contrary, dead and apoptotic cell populations were either Gag::eGFP expressers (Figure 15) or DsRed transfected cells which had not undergone targeting (Figure 15.B).

Cell population analysis by flow cytometry showed similar profiles between clones (Figure 16). They went through three different main phases. In the first phase, after transfection, part of the cell population raised as double positives as a result of the GFP positive nature of the cell lines plus its DsRed expression from the transfected (and not necessarily integrated) targeting construct. Secondly, the double positive population diminished due to the dilution and degradation of the transfected non-integrated DNA. Finally, the positive G418 selection of the clones that successfully performed cassette exchange was translated in a progressive increase of the single DsRed+ population in clones SH5 and AH22. AH27 cell line failed to become a single DsRed+ expresser population (Figure 15.E and 16). This may be indicative of the presence of two copies of the tagging construct on AH27 genome, which resulted in partially targeted cells harboring together the targeting and tagging cassettes.

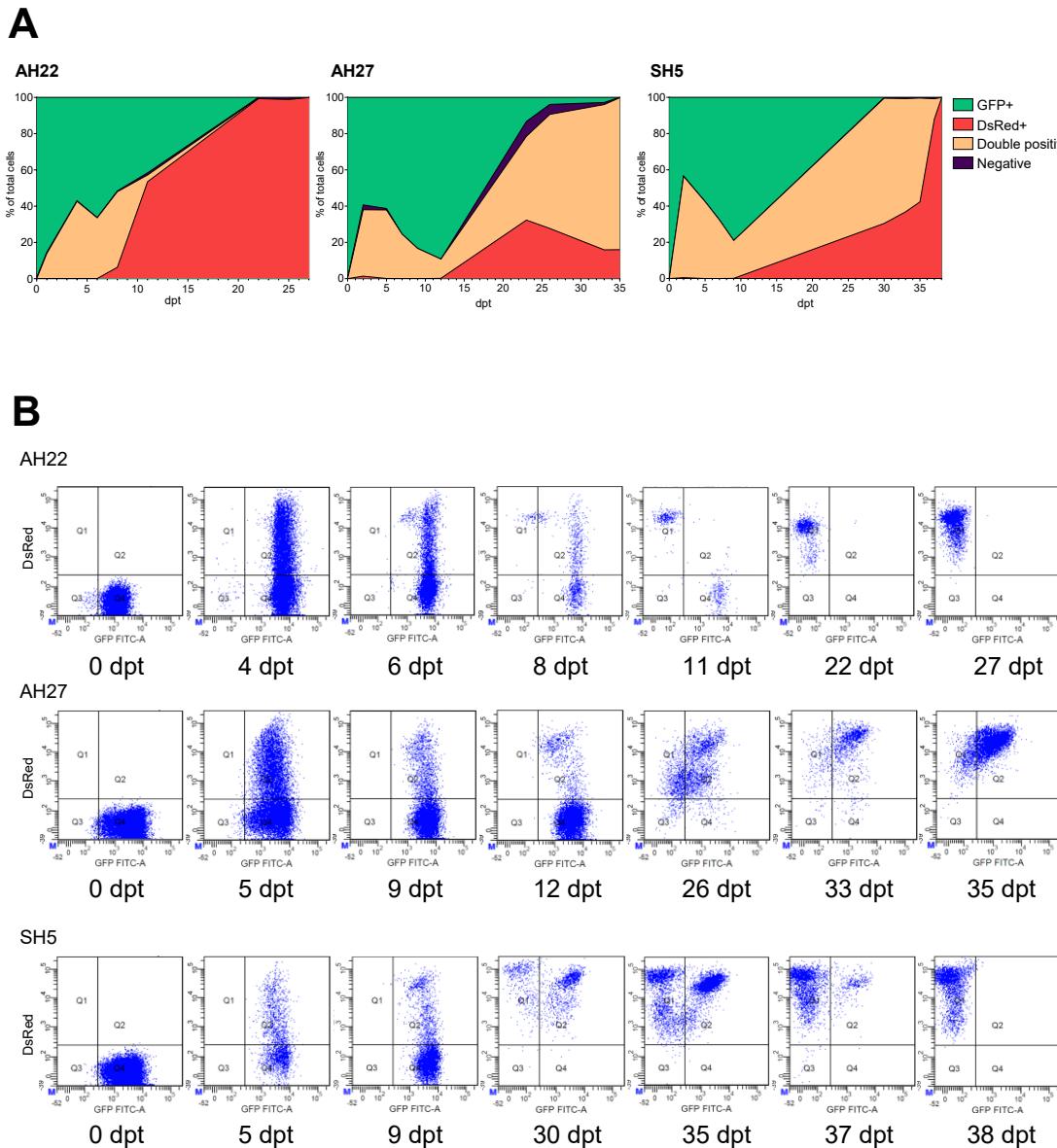


Figure 16. Flow cytometry analysis of SH5, AH22 and AH27 showed different population migrations during the RMCE cassette exchange procedure. Three different phases can be differentiated. Firstly, after transfection, a double positive population raised as a result of the DsRed expression from the transfected (and not necessarily integrated) targeting construct by the already GFP positive clones. Secondly, double positive population diminished due to the dilution and degradation of the transfected non-integrated DNA. Finally, the positive selection of the clones harboring the targeting cassette resulted in an increase of double positive population which switched into a homogeneous single DsRed+ population in clones SH5 and AH22. Clone AH27 failed to become a single DsRed+ expresser population. (**A**): Population migration over time. (**B**): DsRed versus GFP fluorescence intensity population distribution at different time points.

The presence of the tagging cassette on the genome of the generated cell lines and the precise integration of the targeting construct was analyzed by PCR. For this purpose, three primers (Table 1) were designed to bind to the tagging and targeting cassettes (Figure 17.A). As shown in Figure 17.B, at the moment of

targeting (0 hpt) the three cell lines show the PCR bands corresponding to the tagging cassettes, and as expected, none showed bands for the targeting cassette. The three analyzed cell lines showed the PCR band corresponding to the targeting cassette at the last day of the targeting procedure. SH5 and AH22 pools showed no trace of the tagging construct in their genome at 38 and 27 dpt, respectively. However, AH27 exhibited a loose band indicating presence of untargeted cells at 35 dpt (Figure 17.B). Those results were consistent with flow cytometry and optical microscopy analysis, which showed single isolated DsRed populations for the clones SH5 and AH22 at the end of targeting procedure (Figures 15.A, D and 16), while AH27 showed a double positive population (Figure 15.E and 16), failing to become a single DsRed expresser. Those results derived in the decision of discarding clone AH27.

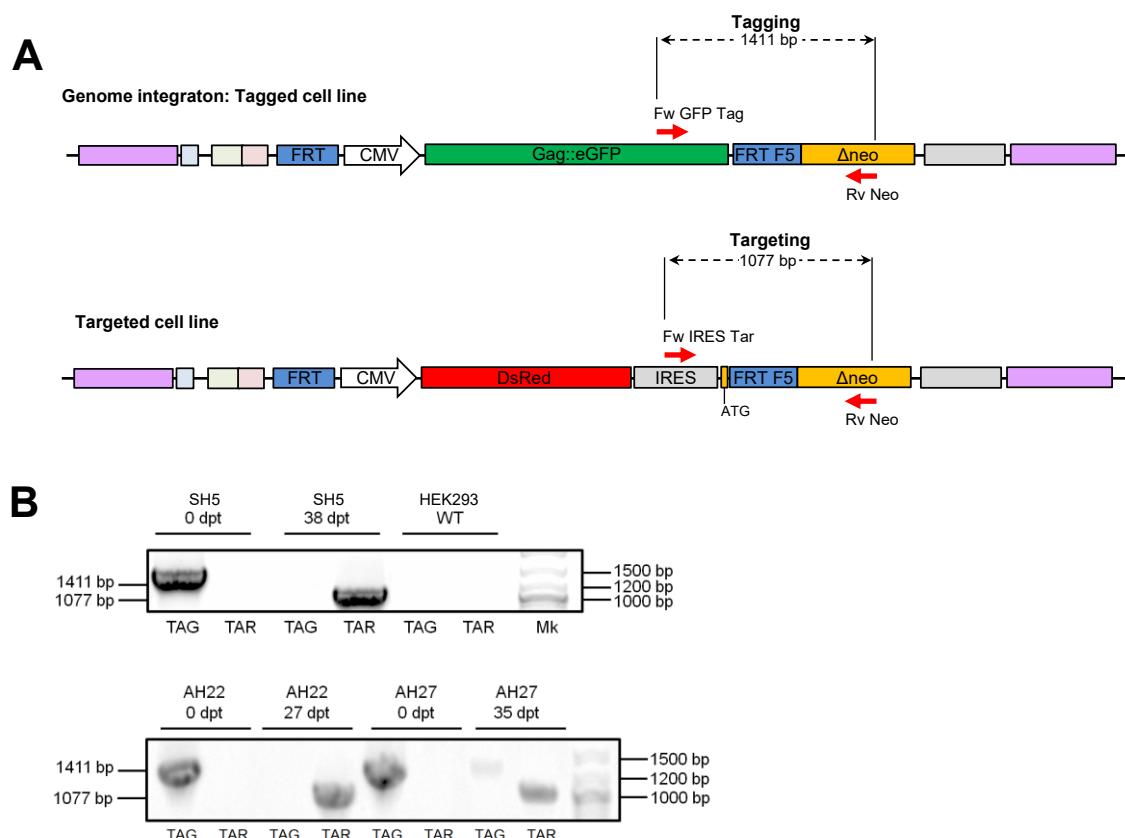


Figure 17. Targeting analysis of the clones SH5, AH22 and AH27. **(A)**: Two PCR reactions were designed to amplify the surroundings of the 3' FRT F5, for the identification of the tagging and targeting cassettes. **(B)**: The three analyzed cell lines showed the PCR bands corresponding to the tagging cassettes at the moment of targeting (0 hpt), and were negative for the targeting cassette band. After targeting and G418 selection, all the clones showed the PCR band corresponding to the targeting cassette. At this point, SH5 and AH22 pools show no trace of the tagging construct on its genome, while AH27 shows a loose band indicating presence of untargeted cells.

3.10 Specific productivity analysis: selection of the best clone

The targeting-competent SH5 and AH22 cell lines were imaged by means of confocal microscopy (Figures 18.A, B) and analyzed in terms of Gag::eGFP VLP production and cell productivity (Figures 18.C, D). For this purpose, both cell lines were seeded at $0.5 \cdot 10^6$ cells/mL and its harvest at 72 hours was analyzed by NTA. They showed similar growth profiles and high cell viabilities (data not shown). Fluorescence microscopy images of both clones showed Gag::eGFP presence in the cellular cytoplasm, especially in clone AH22 (Figure 18.A). Surprisingly, AH22 cell line presented extremely low Gag::eGFP VLP presence in its supernatant compared to SH5 (Figure 18.C). We hypothesize that clone AH22, despite having a stable Gag::eGFP expression, may have experienced mutations in protein or groups of proteins associated with pathways playing a major role in the Gag-VLPs generation or budding [22], such as the endosomal sorting complexes required for transport (ESCRT) machinery. SH5 was selected as the best clone, presenting a VLP concentration of $9 \cdot 10^8$ VLPs/mL and a specific productivity of 110 VLPs/cell·day.

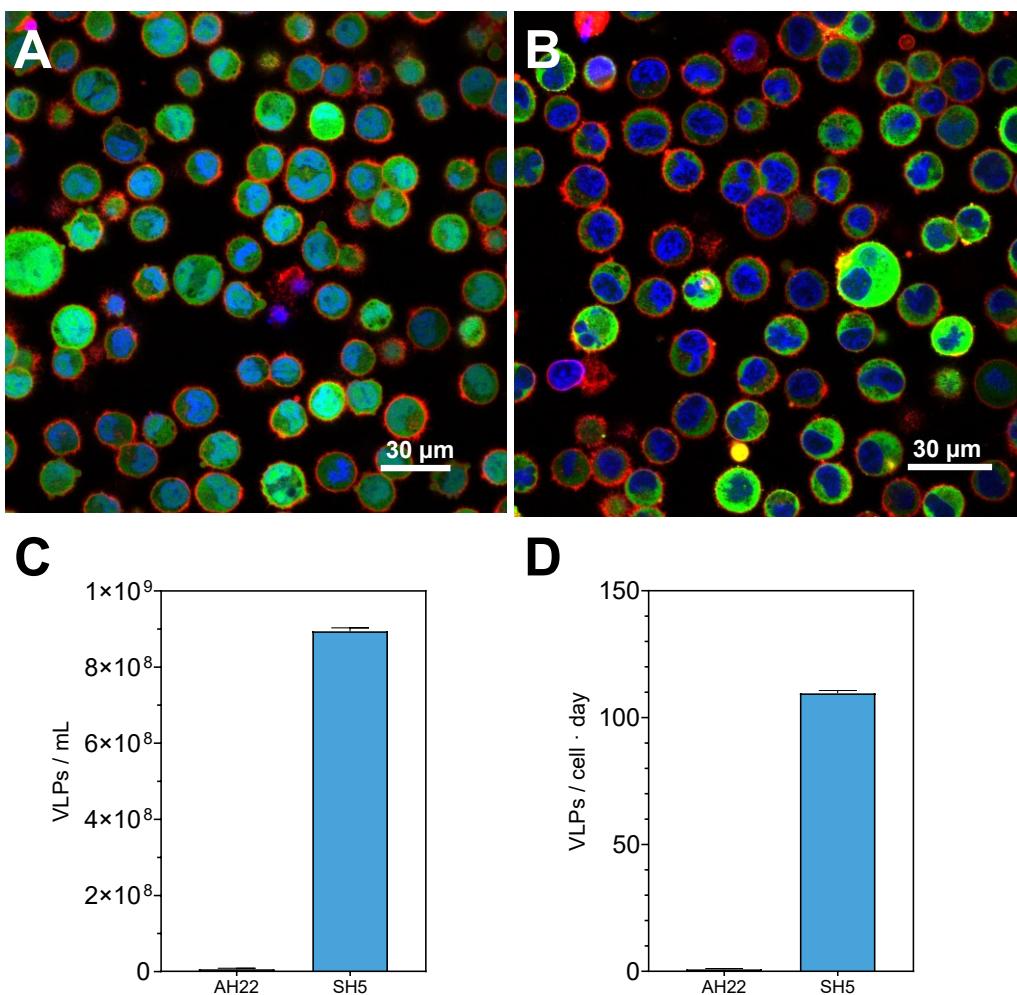


Figure 18. Characterization of the targeting-competent clones AH22 and SH5 in suspension. **(A,B):** Confocal microscopy images of the clones AH22 **(A)** and SH5 **(B)**. Lipid membranes (red) and nuclei (blue) were stained. Gag::eGFP can be observed as green. **(C):** VLP concentration of the cultured clones. SH5 presented a VLP concentration of $9 \cdot 10^8$ VLPs/mL at 72 hours post seeding while AH22 VLP concentration was negligible. **(D):** Specific productivity of the clones SH5 and AH22. SH5 had a specific productivity of 110 VLPs/(cell·day).

3.11 Genomic analysis: number of integrated copies

The SH5 cell line was also characterized to confirm that only one copy of the Gag::eGFP GOI was integrated. This was studied using digital-droplet PCR (ddPCR) and compared with wild-type HEK293 cells. As explained in the Chapter 3, RPP30 was used as a reference gene since it is a well-known reference gene that is present at the human genome with two copies [23–25].

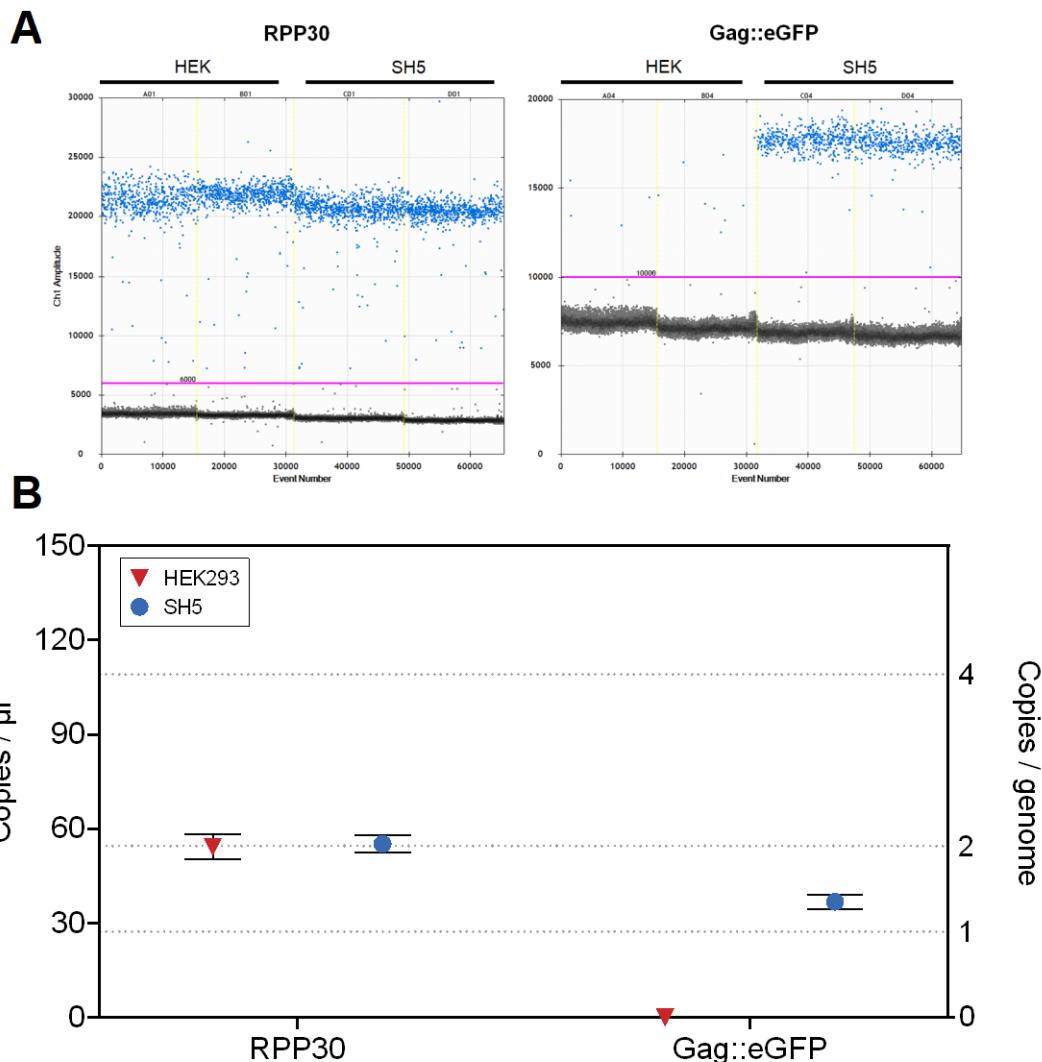


Figure 19. Absolute quantification of the SH5 cell line for the RPP30 and Gag::eGFP genes. (A): Plot of ddPCR reactions. Fluorescent amplitude is indicated on the Y-axis and number of read events (droplets) is indicated on the X-axis. Positive droplets (blue) contained one or more copies of the target DNA. The threshold is indicated as a pink line. Negative droplets without any target DNA are shown in grey. (B): Copies of the RPP30 and Gag::eGFP genes for each cell line. To estimate the copy number per genome, the obtained concentrations for the same studied genomic mass were compared to the RPP30 reference gene were of HEK293 cell lines. Error bars indicate Poisson confidence limits.

ddPCR allowed the analysis of more than 30.000 events (Figure 19.A) for each studied group. The comparison with the RPP30 gene of the wild-type HEK293 cells allowed the assessment of the copy number of each analyzed DNA region per genome. SH5 presented two copies of the reference RPP30 gene on its genome, exactly the same as the parental HEK293 cell line from which SH5 derives (Figure 19.B). This indicates that probably no major translocations or duplication events occurred in the SH5 genome during its generation process

[26]. As previously explained, it is important for cell lines to be used in RMCE, to present only one integrated cassette per genome. That's the reason why the cell line was generated by lentiviral transduction with a MOI of 0.01. ddPCR analysis showed that, as desired, SH5 presented approximately half the number of copies of Gag::eGFP compared to RPP30, meaning that one single copy of the Gag::eGFP gene (thus one copy of the RMCE tagging cassette) was present on SH5 genome (Figure 19.B, Table 7). This also confirmed that the use of lentiviral vectors at low MOIs is an appropriate approach for the generation of stable cell lines with a single integration of the delivered construct into the host cell's genome.

Table 7. Copies per microliter of the analyzed RPP30 and Gag::eGFP genes, its ratios and number of copies per genome.

| | HEK293 | SH5 |
|--------------------------------|----------|----------|
| RPP30 (copies/μl) | 54.3±4.0 | 55.2±2.8 |
| Gag::eGFP (copies/μl) | 0.0 | 36.8±2.3 |
| RPP30/HEK-RPP30 | 1.000 | 1.016 |
| Gag::eGFP/RPP30 | 0 | 0.667 |
| RPP30/HEK-RPP30 (nearest half) | 1 | 1 |
| Gag::eGFP/RPP30 (nearest half) | 0 | 0.5 |
| RPP30 copies/genome | 2 | 2 |
| Gag::eGFP copies/genome | 0 | 1 |

3.12 Stability analysis

VLP production stability of the generated SH5 cell line is relevant as determine its potential uses and quality as a clone. Stability was studied by analyzing the cellular density, viability, mean fluorescence and its Gag-VLP production for a prolonged time period. Results showed that upon 45 days of passages, SH5 did not significantly reduced its viability, growth (data not shown) or production since its specific productivity remained constant at approximately 96 VLPs / (day·cell) (Figure 20).

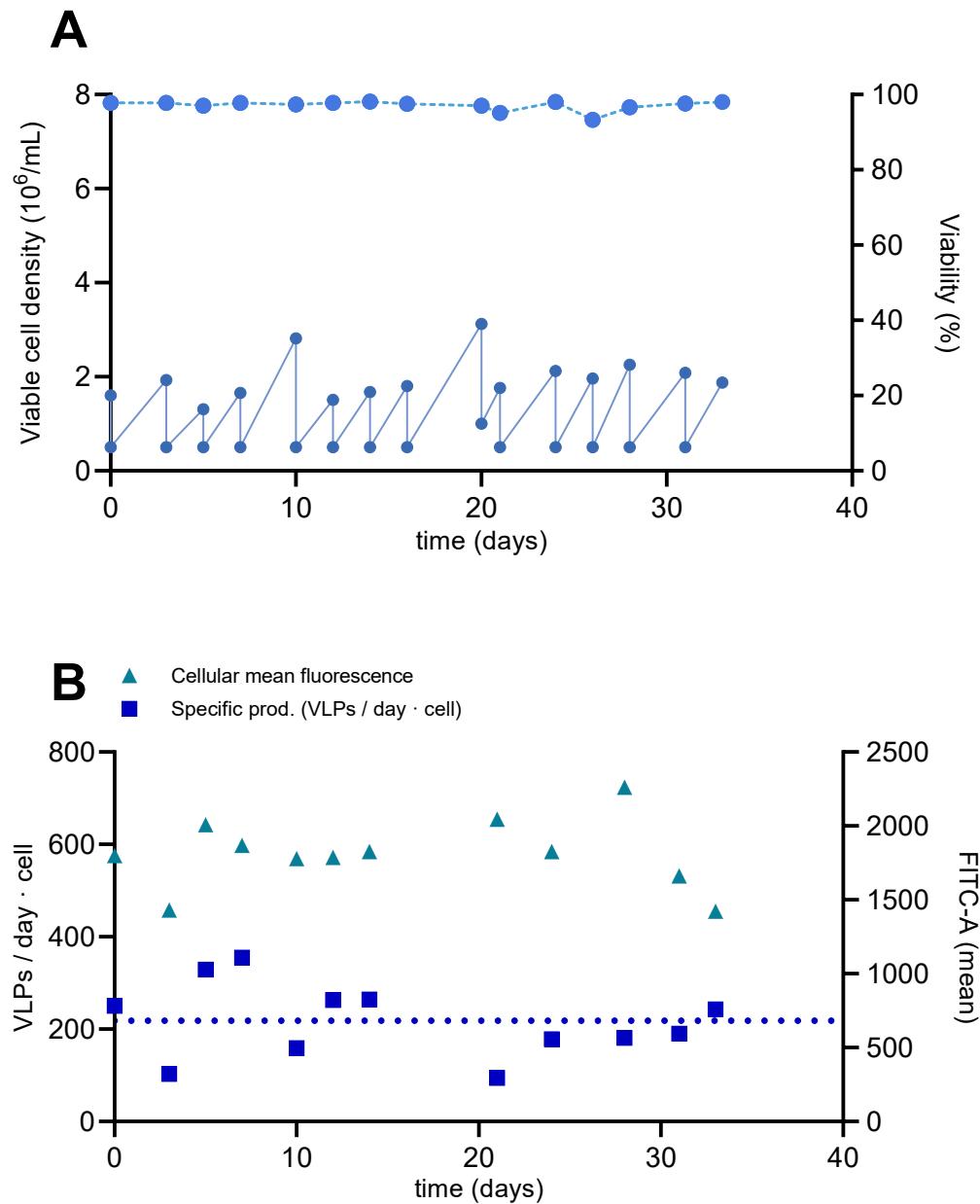


Figure 20. Gag::eGFP-VLP production stability of the SH5 cell line. Mean cellular fluorescence and specific productivity were not significantly reduced during the studied 45 day period. Mean cellular fluorescence and VLP concentration were quantified using flow cytometry and flow virometry, respectively.

4. Conclusions

In this work we successfully generated a RMCE Gag::eGFP VLP producer cell line with the potential of switching its expression to any protein of interest via RMCE Tagging. When generating a RMCE cell line, the number of integrated copies in the cellular genome is relevant as affects the targeting recombination efficiency of the subsequent produced clones. In order to achieve the integration of one copy of the cassette into the cellular genome, a third generation lentiviral vector carrying the tagging cassette was generated by calcium phosphate co-transfection of four plasmids. The lentiviral stock titration was then carried out, obtaining a concentration of 56.16 virions/ μ l. Suspension and adherent parental HEK293 cells were tagged by lentiviral transduction using a MOI of 0.01. Six days post infection (dpi), a GFP-positive cell enrichment was performed to prevent the dilution of the tagged cells. Suspension and adherent infected cells were sorted using two gating criteria in terms of GFP fluorescence. The pools resulting from each group were catalogued as: “high” and “medium” expression. Two weeks after the pool enrichment, more than 2000 clones were isolated in 96-well plates. From all the seeded clones, 62 showed cell growth, resulting in survival ratios of \approx 1.7% and \approx 3.5% for the suspension and adherent infected cells, respectively. Clones were amplified and the study of their mean cellular fluorescence using flow cytometry, allowed the selection of the 13 best candidates for their suspension and serum-free adaptation. The adaptation of the cells was achieved successfully and their Gag::eGFP productivity was studied, allowing the selection of the five potential best candidates. Growth kinetic studies of the selected cell lines were performed to analyze cell growth dynamics: cell density, duplication time and time until reaching maximum cell density peak. As the targeting RMCE methodology presented in this work is based on the differential ability of the targeted sub-clones to resist G418 antibiotic selection compared to their un-recombined counterparts, a G418 sensitivity assay of the clones was performed. From the 5 tested cell lines, two showed antibiotic resistance, so they were discarded. The targeting of the remaining 3 clones was tested by transient co-transfections of plasmids encoding for the Flp flipase and a targeting cassette encoding the expression of the DsRed reporter protein. With

this methodology, cell lines undergoing correct targeting shifted from single GFP expressers to single DsRed expressers upon G418 selection. GFP and DsRed expression of the targeted clones and control groups was analyzed by confocal microscopy and flow cytometry. The presence of the targeting or tagging cassettes in the cellular genomes was also analyzed by PCR. Cell lines SH5 and AH22 were catalogued as targeting-competent as they showed excellent targeting capabilities, while clone AH27 was unable to derive into a single DsRed positive population. The specific productivity of the clones showed that SH5 was the best clone in terms of specific productivity, generating 110 VLPs/(cell·day). It presented the same copies per genome of the RPP30 reference gene compared to HEK293 cells, indicating that SH5 genome did not suffer major alterations during its generation. As desired for an RMCE-compatible cell line, SH5 presented a single integration of the tagging cassette per genome. The stability analysis showed that upon 45 days, SH5 does not significantly reduce its growth, viability or production, and maintains its specific productivity constant.

5. Acknowledgements

Part of this research was conducted at Helmholtz Centre for Infection Research (Braunschweig, Germany) during an international doctoral stay. The author wants to thank Natascha Gödecke and Dagmar Wirth (Model Systems for Infection and Immunity, Helmholtz Centre for Infection Research, Braunschweig, Germany) for their precious support mentoring this work as well as their help in the contribution of the HEK 293-T cell line, RMCE plasmids, reagents and protocols. The authors also thank Dr. Amine Kamen (McGill University, Montreal, Canada) for providing the HEK 293 SF-3SF6 cells. The pGag::eGFP (Cat#11468) reagent from Dr. Merilyn Resh was obtained through the National Institutes of Health AIDS Reagent Program, Division AIDS, National Institute of Allergy and Infectious Diseases, National Institutes of Health. We would also like to thank Dr. Diego Fontana (Universidad Nacional del Litoral, Santa Fe, Argentina) for its help conceptualizing the design of the pLenti-Gag::eGFP plasmid. We would like to thank Dr Jose Amable Bernabé from the Institut de Ciència de Materials de Barcelona (ICMAB-CSIC, Bellaterra, Catalonia, Spain) for his support in nanoparticle tracking analysis techniques.

The support of Núria Barba Bosch and Meritxell Vendrell from the Servei de Microscòpia (Universitat Autònoma de Barcelona, Bellaterra, Catalonia, Spain) in the development of optical microscopy images and analysis is greatly appreciated. We would also like to thank Manuela Costa Garcia from the Servei de Cultius Cel·lulars, Producció d'Anticossos i Citometria (Universitat Autònoma de Barcelona, Bellaterra, Catalonia, Spain) for her help with flow cytometry and cell sorting; and Salvador Bartolomé Piñol from Laboratori de Luminescència i Espectroscòpia de Biomolècules (Universitat Autònoma de Barcelona, Bellaterra, Catalonia, Spain) for his help with qPCR and ddPCR analyses. Figures and Graphs were generated using GraphPad Prism version 8.0.2 for Windows (San Diego, CA, USA) and GIMP 2.10.28 (<http://www.gimp.org>).

References

- [1] Hansen, J. *et al.* A large-scale, gene-driven mutagenesis approach for the functional analysis of the mouse genome. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 9918–9922 (2003)
- [2] Mutskov, V. & Felsenfeld, G. Silencing of transgene transcription precedes methylation of promoter DNA and histone H3 lysine 9. *EMBO J.* **23**, 138–149 (2004)
- [3] Wurm, F. M. Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat. Biotechnol.* **22**, 1393–1398 (2004)
- [4] Wirth, D. *et al.* Road to precision: recombinase-based targeting technologies for genome engineering. *Curr. Opin. Biotechnol.* **18**, 411–419 (2007)
- [5] Turan, S. *et al.* Recombinase-mediated cassette exchange (RMCE): Traditional concepts and current challenges. *J. Mol. Biol.* **407**, 193–221 (2011)
- [6] Prorocic, M. M. & Stark, W. M. *Site-specific Recombination. Brenner's Encyclopedia of Genetics: Second Edition* vol. 6 (Elsevier Inc., 2013).
- [7] Blakely, G. W. *Mechanisms of Horizontal Gene Transfer and DNA Recombination. Molecular Medical Microbiology: Second Edition* vols 1–3 (Elsevier Ltd, 2014).
- [8] Turan, S., Kuehle, J., Schambach, A., Baum, C. & Bode, J. Multiplexing RMCE: Versatile Extensions of the Flp-Recombinase-Mediated Cassette-Exchange Technology. *J. Mol. Biol.* **402**, 52–69 (2010)
- [9] Schucht, R., Wirth, D. & May, T. Precise regulation of transgene expression level and control of cell physiology. *Cell Biol. Toxicol.* **26**, 29–42 (2010)
- [10] Baer, A. & Bode, J. Coping with kinetic and thermodynamic barriers: RMCE, an efficient strategy for the targeted integration of transgenes. *Curr. Opin. Biotechnol.* **12**, 473–480 (2001)

- [11] Zhang, L. *et al.* Recombinase-mediated cassette exchange (RMCE) for monoclonal antibody expression in the commercially relevant CHOK1SV cell line. *Biotechnol. Prog.* **31**, 1645–1656 (2015)
- [12] Nehlsén, K. *et al.* Recombinant protein expression by targeting pre-selected chromosomal loci. *BMC Biotechnol.* **9**, 1–12 (2009)
- [13] Kumar, M., Keller, B., Makalou, N. & Sutton, R. E. Systematic determination of the packaging limit of lentiviral vectors. *Hum. Gene Ther.* **12**, 1893–1905 (2001)
- [14] Akkina, R. K. *et al.* High-efficiency gene transfer into CD34+ cells with a human immunodeficiency virus type 1-based retroviral vector pseudotyped with vesicular stomatitis virus envelope glycoprotein G. *J. Virol.* **70**, 2581–2585 (1996)
- [15] Merten, O. W., Hebben, M. & Bovolenta, C. Production of lentiviral vectors. *Mol. Ther. - Methods Clin. Dev.* **3**, 16017 (2016)
- [16] Hawley, R. G. Overview of the HIV-1 Lentiviral Vector. *Mol. Biol.* 1–15 (2001)
- [17] Pieters, T. *et al.* Structure-function studies in mouse embryonic stem cells using recombinase-mediated cassette exchange. *J. Vis. Exp.* **2017**, 1–10 (2017)
- [18] Cervera, L. *et al.* Generation of HIV-1 Gag VLPs by transient transfection of HEK 293 suspension cell cultures using an optimized animal-derived component free medium. *J. Biotechnol.* **166**, 152–165 (2013)
- [19] Hermida-Matsumoto, L. & Resh, M. D. Localization of Human Immunodeficiency Virus Type 1 Gag and Env at the Plasma Membrane by Confocal Imaging. *J. Virol.* **74**, 8670–8679 (2000)
- [20] Schwartz, S. *et al.* Mutational inactivation of an inhibitory sequence in human immunodeficiency virus type 1 results in Rev-independent gag expression. *J. Virol.* **66**, 7176–7182 (1992)
- [21] Kibbe, W. A. OligoCalc: an online oligonucleotide properties calculator.

Nucleic Acids Res. **35**, W43–W46 (2007)

- [22] Lavado-García, J., Jorge, I., Cervera, L., Vázquez, J. & Gòdia, F. Multiplexed Quantitative Proteomic Analysis of HEK293 Provides Insights into Molecular Changes Associated with the Cell Density Effect, Transient Transfection, and Virus-Like Particle Production. *J. Proteome Res.* **19**, 1085–1099 (2020)
- [23] Lin, H. T. *et al.* Application of Droplet Digital PCR for Estimating Vector Copy Number States in Stem Cell Gene Therapy. *Hum. Gene Ther. Methods* **27**, 197–208 (2016)
- [24] Dyavar, S. R. *et al.* Normalization of cell associated antiretroviral drug concentrations with a novel RPP30 droplet digital PCR assay. *Sci. Rep.* **8**, 1–11 (2018)
- [25] De Oliveira, M. F. *et al.* Comparative analysis of cell-associated HIV DNA levels in cerebrospinal fluid and peripheral blood by droplet digital PCR. *PLoS One* **10**, 1–10 (2015)
- [26] Lin, Y. C. *et al.* Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. *Nat. Commun.* **5**, (2014)

Chapter six

**Study and process optimization of stable HEK293 production of
Gag::eGFP VLPs by different strategies**

Abstract: The production of Gag-based virus like particles (Gag-VLPs) by stable gene expression (SGE) in mammalian platforms represents a valuable strategy due to its robust and prolonged expression, homogeneous nature and ease bioprocess scale-up. Additionally, SGE cell lines constitutively expressing Gag-VLPs can be functionalized with epitopes from pathogenic agents, simplifying a process that would otherwise result in lower gene delivery efficiencies. In this study we evaluated the production of Gag VLPs using different stable cell lines, generated by plasmid random integration, lentiviral transduction and locus-specific genomic safe harbor (GSH) integration with CRISPR/Cas9; by comparing them with a typical transient gene expression (TGE) production approach. The studied variables were cellular growth, culture viability, VLP concentration, VLP productivity and VLP quality. Additionally, the improvement of the process yield was performed by cell media optimization, which increased the viable cell density and allowed the elongation of the production process, resulting in increased volumetric productivities. The best platform resulted to be the CRISPR-generated cell line cultured with HyCell media and harvested 10 days after seeding. With this approach, $2.50 \cdot 10^{10}$ VLPs/mL were obtained at harvest time, improving the production achieved by TGE by 2.7 fold.

Keywords: VLPs, Stable gene expression, random integration, Recombinase-mediated cassette exchange, RMCE, targeting, CRISPR, GSH, AAVS1, HEK293, Gag::eGFP, HIV-1, media optimization.

1. Introduction

Although transient transfection is a high yield small-scale interesting approach for the production of VLPs [1], its large-scale production should be based on the use of stable gene expression (SGE) cell lines [2]. SGE represents a valuable strategy due to its robust and prolonged expression, the ability to produce in high densities, its monoclonal nature, low batch-to-batch variability and its easy scale-up for large-scale bioprocesses. SGE cell lines constitutively expressing Gag-VLPs can be transiently transfected with expression vectors coding for epitopes to produce Gag-based vaccine candidates for a broad range of pathogens or variants simplifying a process that would otherwise result in lower gene delivery efficiencies. Additionally, the constitutive production of VLPs by SGE avoids the use of transfection reagents, which not only increase the economic cost of the process, but also negatively affects its cell growth and viability. Unlike standard TGE protocols, SGE VLP production is not limited to be harvested at 72 hpt [3,4]. The high viabilities and cell densities obtained with SGE, combined with the optimization of culture media, can be exploited to extend the production process, resulting in an increase of VLP concentrations at harvest without altering its quality.

In this work, three platforms constitutively expressing Gag-VLPs were analyzed in terms of cellular growth, culture viability, VLP concentration, VLP productivity and VLP quality, in order to select the most promising expression platform.

2. Materials and methods

2.1 Cell lines, media and culture conditions

The serum-free suspension-adapted 10H9, SH5 and 13++ stable gene expression (SGE) cell lines were used. They are derived from Current Good Manufacturing Practice (cGMP) HEK293 cell bank. These cell lines stably express the Gag::eGFP gene under CMV promoter which derives in the constitutive expression of Gag::eGFP-VLPs. 10H9 was generated by random insertion of the linearized vector coding for the CMV-Gag::eGFP construct together with a neomycin resistance gene, which confers G418 resistance. SH5 was generated by lentiviral transduction using a multiplicity of infection (MOI) of

0.01. This resulted in a single integration into the cellular genome of a transfer vector containing a RMCE tagging cassette. The integrated cassette codes for the CMV-Gag::eGFP construct flanked by two heterospecific Flp and recombination targets (FRT and FRT F5 regions) and a neomycin phosphotransferase gene lacking of the initial ATG codon (Δ neo) which will activate its expression after RMCE with an appropriate targeting cassette. 13++ cell line was generated by the integration of the CMV-Gag::eGFP construct into the adeno-associated virus site 1 (AAVS1) genomic safe harbor (GSH) using CRISPR/Cas9 technology. It is also puromycin resistant.

The serum-free suspension-adapted HEK293 cell line (HEK293SF-3F6) was used as non VLP-producer control cell line, kindly provided by Dr. Amine Kamen from the Biotechnology Research Institute at the National Research Council of Canada (Montreal, Canada) and McGill University.

Two media were used for cellular growth. The chemically defined and free from animal components FreeStyle 293 (ThermoFisher, Waltham, MA, USA) supplemented with 0.1% Pluronic (ThermoFisher, Waltham, MA, USA), 1.6 mg/L of r-transferrin (Merck, Kankakee, IL), 19.8 mg/L of r-insulin (Novo Nordisk Pharmatech, Køge, Denmark) and 0.9X of an in-house developed lipid mixture to maximize cell growth [5], and the chemically defined animal-component free HyCell TransFx-H from HyClone (GE Healthcare, Chicago, IL, USA) supplemented with 4 mM GlutaMAX (Gibco, Thermo Fisher Scientific, Waltham, MA, USA) and 0.1% Pluronic F-68 Non-ionic Surfactant (Gibco, Thermo Fisher Scientific, Waltham, MA, USA). From now on, they will be referred as Freestyle Mix and HyCell, respectively.

To adapt FreestyleMix growing cell lines to HyCell growth, they were seeded gradually increasing the ratio of HyCell/FreestyleMix in each passage. When cells attained at a constant growth rate and viabilities higher than 95% during two or more passages in 100% HyCell media, they were considered adapted to HyCell media.

Suspension cell cultures were maintained routinely in exponential growth phase in 125ml disposable polycarbonate Erlenmeyer flasks with vent cap (Corning, Tewksbury, NY, USA), placed in an orbital shaker (orbit diameter 16mm, Stuart,

Stone, UK) at 130 rpm in a humidified Steri-Cult CO₂ incubator (3310, ThermoFisher, Waltham, MA, USA) at 37°C, 85% RH and 5% CO₂ in air.

Cell counts and viability determinations were performed using the NucleoCounter NC-3000 automatic cell counter (Chemometec, Lillerød, Denmark) following manufacturer's instructions. Specific growth rate (μ) was calculated using the Equation 1.

$$\mu = \frac{\ln(X_t - X_0)}{t_t - t_0} \quad (1)$$

2.2 Plasmids and transfection

2.2.1 Plasmid expression vectors and amplification

The pGag::eGFP plasmid codes for a codon optimized Rev-independent HIV-1 Gag protein fused in frame to the enhanced GFP driven by CMV enhancer and promoter. The plasmid from the NIH AIDS Reagent Program (Cat 11468) [6] was constructed by cloning the Gag sequence from pCMV55M1-10 [7] into the pEGFP-N1 plasmid (Clontech, Palo Alto, CA, USA).

Plasmid DNA was amplified in *Escherichia coli* Top10 strain grown in LB medium at 37°C supplemented with 10 µg/ml of kanamycin (Sigma, St. Louis, MO, USA). Plasmid purification was carried out using the Endofree Plasmid Mega kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions.

2.2.2 PEI-mediated transient transfection

Exponentially growing HEK293 suspension cells were passaged to have a cell density of 2·10⁶ cells/mL at transfection time. PEIpro (Polyplus-transfection SA, Illkirch-Graffenstaden, France) was used as transfection reagent. PEI-DNA complexes were formed under sterile conditions. Briefly, DNA was diluted in culture media (10% of the total volume of cell culture to be transfected) for a final total DNA concentration of 1 µg/mL and vortexed for 10 s. Then, polyethylenimine (PEI), was added for a final concentration of 2 µg/mL (a 2:1

PEI:DNA ratio (w/w)) and vortexed three times for 3 s. The mixture was incubated for 15 minutes at RT and then added to the culture.

HEK293 cells growing in Freestyle Mix media, were subjected to a media exchange prior to its transfection by 5 minutes 300 x g centrifugation.

2.4 Flow cytometry

For flow cytometry, cells were analyzed using a BD FACS Canto flow cytometer (BD BioSciences, San Jose, CA, USA), at Servei de Cultius Cel·lulars, Producció d'Anticossos i Citometria (Universitat Autònoma de Barcelona, Bellaterra, Catalonia, Spain).

2.5 Spectrofluorometry-based product quantitation

Green fluorescence intensity in Gag-GFP samples was measured using a Cary Eclipse Fluorescence Spectrophotometer (Agilent Technologies, Santa Clara, CA, USA). The instrument parameters were set as follows: $\lambda_{\text{ex}} = 488 \text{ nm}$ (slit 5 nm), $\lambda_{\text{em}} = 510 \text{ nm}$ (slit 10 nm). Readings were carried out at room temperature. Relative fluorescence units (RFU) values were calculated by subtracting fluorescence units (FU) of negative control samples from that given by the sample.

2.6 Nanoparticle Tracking Analysis

NTA-based Gag::eGFP VLP quantification and characterization was performed using a NanoSight NS300 (Nanosight Ltd., Amesbury, UK) equipped with a blue filter module (488 nm) and a neutral filter at the Soft Material services of the Institut de Ciència de Materials de Barcelona (ICMAB-CSIC, Bellaterra, Catalonia, Spain). Samples were previously diluted to a concentration of approximately 10^8 particles/mL. Sample injection was performed using a pump in order to improve the robustness of the measurement by continuous addition, and to minimize the photobleaching effect due to fluorescence depletion over time. 60 second videos were recorded at RT and analyzed with the NTA 3.4 software (Malvern Panalytical, Malvern, UK). Tracked particles size was determined from its Brownian motion. Three independent experimental replicas

were carried out for each sample. Camera level and detection threshold were manually adjusted for each replica.

3. Results and discussion

3.1 Cell line comparison, product characterization and media optimization

In the previous chapters, three stable gene expression (SGE) cell lines constitutively producing Gag::eGFP VLPs have been generated using random integration (10H9), lentiviral transduction of an RMCE cassette (SH5) and locus-specific integration into the AAVS1 GSH using CRISPR/Cas9 (13++) (Figure 1).

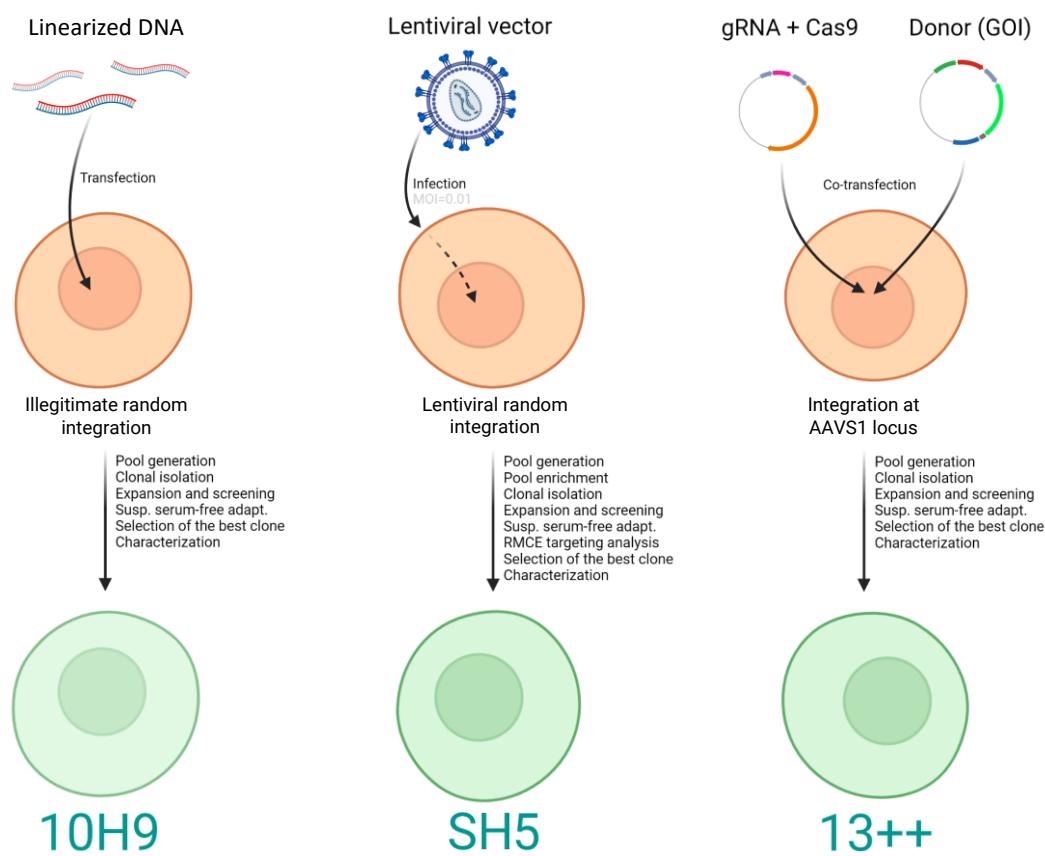


Figure 1. Schematic representation of the generation of the three studied cell lines constitutively expressing Gag::eGFP-VLPs. 10H9 was generated by illegitimate random integration of a linearized plasmid coding for the Gag::eGFP gene. SH5 was generated using lentiviral DNA transduction by infection at a multiplicity of infection (MOI=0.01) of a recombinase-mediated cassette exchange (RMCE) cassette coding for the Gag::eGFP gene. 13++ was generated by cotransfection of two plasmids coding for the Cas9 nuclease, a sgRNA for AAVS1 integration and a Donor plasmid coding for Gag::eGFP flanked by AAVS1 homology arms. After an extensive screening process, the best cell lines emerged from each methodology were selected and characterized.

Each cell line has been selected as the best candidate of its class, after an extensive screening and characterization process. Here, a comparison experiment is performed in order to determine which cell line is the best in terms of cell growth and specific productivity. To put the performance of the three cell lines in context, they are compared with a transiently-transfected HEK293 cell line, from now on, referred as TGE.

The media in which cells are cultured plays a major role when it comes to protein production as it can affect cell growth, the duration of the bioprocess, the specific productivity of the cell lines and the product quality. For that reason, a comparison of each platform growth and production between Freestyle Mix (the media in which the clones were generated) and HyCell (the commercial media that has shown the best Gag-VLP production in HEK293 cells in previous studies [8]) was performed.

3.1.1 Growth kinetics

The three cell lines were seeded at $0.5 \cdot 10^6$ viable cells/mL in Erlenmeyer flasks to study their growth and VLP production dynamics. When TGE groups reached $2 \cdot 10^6$ cells/mL (day two), they were transfected with the pGag::eGFP vector.

In Freestyle Mix, the three SGE clones showed similar growth kinetics, reaching a maximum of approximately $4.5 \cdot 10^6$ cells/mL between the fourth and fifth day after seed (Figure 2). TGE reached a maximum of $3 \cdot 10^6$ cells/mL, due to transfection reagent (PEI) toxic effect, a remarkable drawback of TGE strategies when compared to SGE. After reaching the maximum, all cell lines showed a rapid decrease in viability.

Table 1. Duplication times and maximum cell density peaks of the studied production platforms with FreestyleMix and HyCell media.

| Platform | FreestyleMix | | HyCell | |
|----------|--------------------------|---------------------------------------|--------------------------|---------------------------------------|
| | Duplication time (hours) | Max. cell density (viable cells / mL) | Duplication time (hours) | Max. cell density (viable cells / mL) |
| 13++ | 31.2 ± 0.3 | $(4.2 \pm 0.3) \cdot 10^6$ | 31.0 ± 0.8 | $(11.0 \pm 0.9) \cdot 10^6$ |
| SH5 | 30.8 ± 0.5 | $(4.4 \pm 0.1) \cdot 10^6$ | 29.8 ± 1.0 | $(14.2 \pm 0.9) \cdot 10^6$ |
| 10H9 | 30.6 ± 0.8 | $(4.8 \pm 0.3) \cdot 10^6$ | 28.0 ± 0.7 | $(13.3 \pm 1.4) \cdot 10^6$ |
| TGE | 33.5 ± 0.9 | $(3.3 \pm 0.3) \cdot 10^6$ | 27.7 ± 2.6 | $(3.1 \pm 0.2) \cdot 10^6$ |

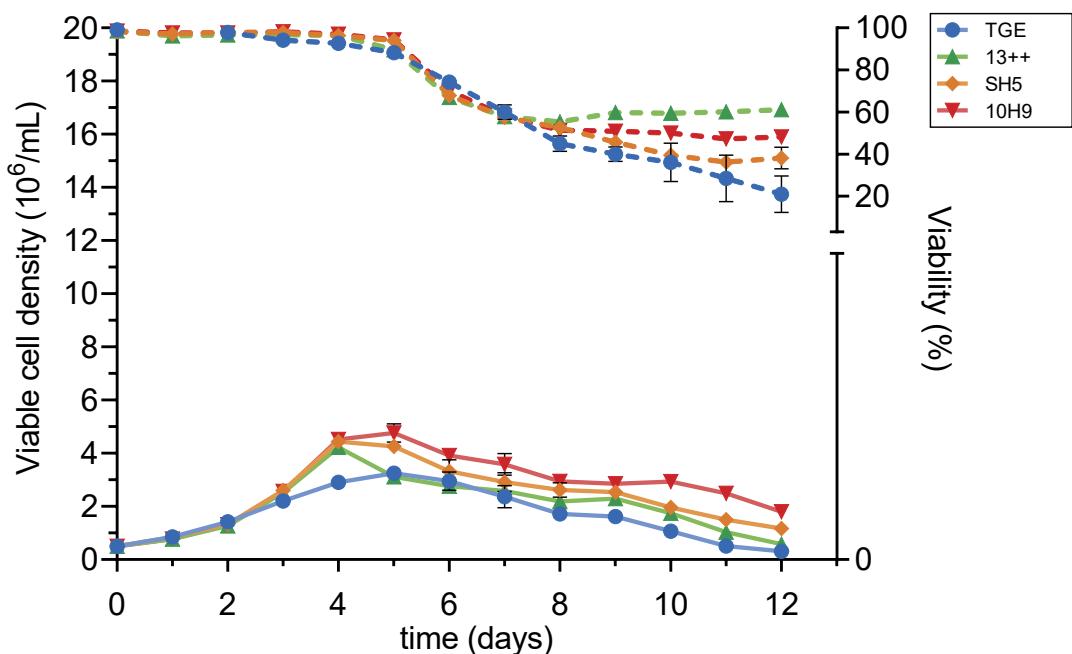


Figure 2. Growth kinetics in FreestyleMix media of the stable cell lines 13++, SH5 and 13++ compared with HEK293 cells transiently transfected (TGE) at day 2. A significant drop in cellular viability can be observed after day 5.

When cultured in HyCell, cell lines did not experience a significant change in terms of duplication times as shown in Table 1. However, SGE cell lines reached much higher cellular concentrations (Figure 3). They all showed similar growth profiles, showing an initial maximum around day 6, followed by a period with a staggered phase while maintaining high viabilities. This was followed by a second increase in concentration until they reached a second maximum peak at day 10, of more than 10^7 viable cells/mL. After this, cultures experienced a quick viability drop at day 10 (Figure 3), probably due to some media components depletion. In any case, the cell concentration reached is significant for batch culture of HEK 293 cells [9]. The positive effect observed in SGE cultured in HyCell was not reproduced by the TGE cells, which reached significantly lower cell densities and experienced low viabilities especially from day 5. This was caused by the toxic effect of the transfection reagents introduced at day 2.

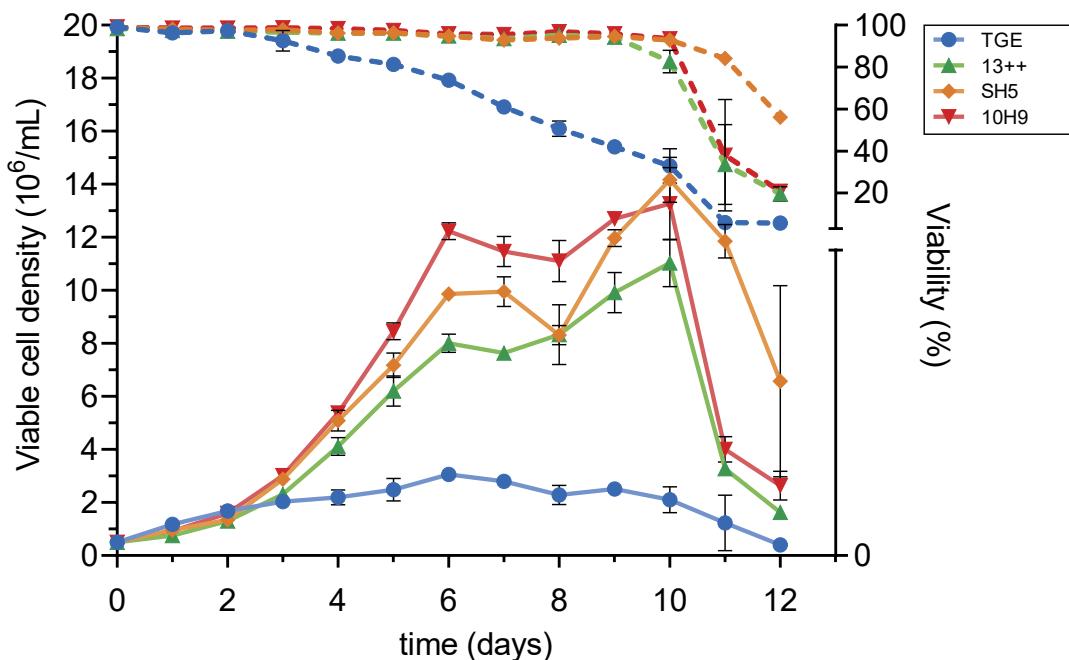


Figure 3. Growth kinetics in HyCell media of the stable cell lines 13++, SH5 and 13++ compared with HEK293 cells transiently transfected (TGE) at day 2.

The growth behavior showed by all SGE groups in HyCell media harbored the potential for the production processes to be extended maintaining prolonged high viabilities compatible with the production of properly assembled VLPs and avoiding the release of Gag monomers or contaminants derived from apoptotic cells. For that reason, the SGE batch production with HyCell was established to be harvested at day 10, compared to day 5 in FreestyleMix. TGE productions for both media were harvested at day 5 (72 hours-post transfection (hpt)) as it was determined in previous studies [3,4]. The elongation of the production process was hypothesized to be translated in a production improvement.

3.1.2 VLP production and specific productivity

NTA and fluorescence analyses were conducted to analyze VLP production kinetics, specific productivity and volumetric productivity from the start of the bioprocess until harvest.

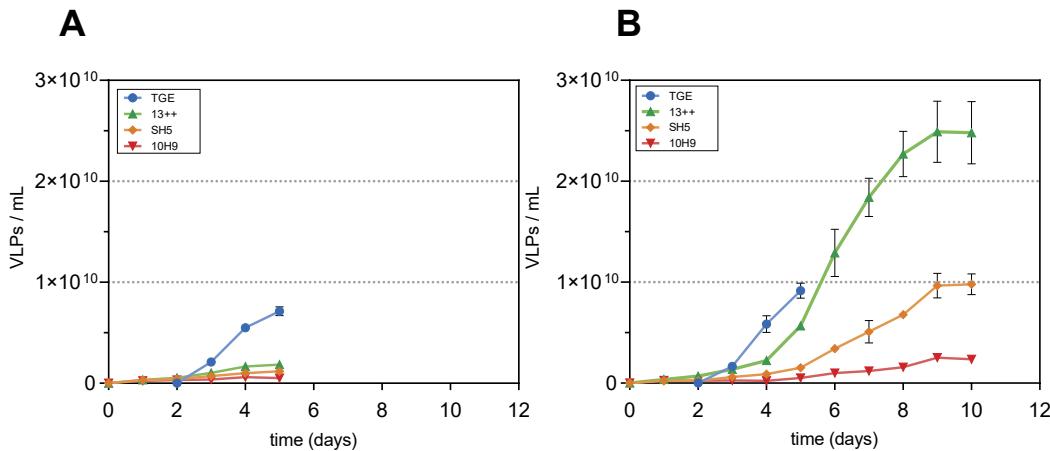


Figure 4. VLP production kinetics of the studied platforms, cultured in (A) FreestyleMix and (B) HyCell.

Among conditions at day 5 (or 72 hpt), TGE presented the highest VLP concentration (Figure 4). When comparing VLP concentrations at this time point, HyCell culture resulted in a significant improvement in TGE and 13++ conditions ($p<0.05$ and $p<0.01$ respectively) over FreestyleMix, while SH5 and 10H9 were not significantly affected ($p>0.05$). However, the extension of the production process thanks to the high prolonged viabilities in HyCell media resulted in higher VLP concentrations for all the platforms at harvest time (day 10) (Figure 4). Remarkably, 13++ surpassed TGE at day 6, and reached a final VLP concentration of $2.5 \cdot 10^{10}$ particles/mL, improving the production achieved by TGE by 2.7 fold. Cell line SH5 reached similar VLP concentrations as TGE on its harvest (Figure 4).

To understand the dynamics that resulted in higher VLP productions in HyCell compared to FreestyleMix, specific and volumetric productivities were studied. Specific productivity (q_p) and volumetric productivity (v_p) were calculated using Eq. 2 and Eq. 3 respectively, where P is the product, X the viable cell density (VCD), t the time, and μ the specific growth rate of the cells during the selected time interval, calculated using Eq. 1.

$$q_p = \frac{\int_{t_0}^t P(t) \cdot dt}{\int_{t_0}^t X(t) \cdot dt} \cdot \mu \quad (2)$$

$$v_p = P/t \quad (3)$$

In FreestyleMix, SGE cell lines presented specific productivities between 2.5 and 5 fold less than TGE (Table 2). Culture in HyCell only had a positive impact in the specific productivity of the clone 13++ compared to FreestyleMix. TGE specific productivity in HyCell remained significantly higher than all of SGE cell lines.

Table 2. Volumetric and specific productivities at day 5 of the studied production platforms with FreestyleMix and HyCell media.

| Platform | FreestyleMix | | HyCell | |
|----------|--|--|--|--|
| | Specific Productivity (VLPs·cell ⁻¹ ·day ⁻¹) | Volumetric Productivity (VLPs·mL ⁻¹ ·day ⁻¹) | Specific Productivity (VLPs·cell ⁻¹ ·day ⁻¹) | Volumetric productivity (VLPs·mL ⁻¹ ·day ⁻¹) |
| 13++ | 258.32 | $(3.7 \pm 0.5) \cdot 10^8$ | 368.68 | $(2.5 \pm 0.3) \cdot 10^9$ |
| SH5 | 145.98 | $(2.4 \pm 0.5) \cdot 10^8$ | 114.65 | $(1.6 \pm 0.1) \cdot 10^9$ |
| 10H9 | 89.52 | $(1.0 \pm 0.4) \cdot 10^8$ | 35.68* | $(2.4 \pm 0.1) \cdot 10^8$ |
| TGE | 704.92 | $(1.4 \pm 0.1) \cdot 10^9$ | 551.86 | $(1.8 \pm 0.2) \cdot 10^9$ |

*Days 1-6. After day 6, 10H9 specific productivity was reduced to 8.22 (VLPs·cell⁻¹·day⁻¹) until harvest time.

However, in HyCell, the low specific productivities of the 13++ and SH5 cell lines compared to TGE, were compensated by the process extension and the high cell densities reached, resulting in a remarkable improvement of the production titers at harvest time. This was manifested in volumetric productivities equaling (SH5) or exceeding (13++) the one shown by TGE ($1.8 \cdot 10^9$ VLPs·mL⁻¹·day⁻¹) (Table 2). These results support HyCell as the best media for the production of functionalized Gag-based VLPs with the SGE platforms, from which 13++ resulted to be the best producer.

3.1.3 VLP characterization

VLP characterization was carried on using NTA. The average distribution and particle diameter varied among the different production platforms and media used. As expected, all the conditions showed main populations of fluorescent particles between 50 and 200 nm (Figure 5), corresponding to Gag::eGFP VLPs [4]. Except for 10H9, all platforms improved VLP homogeneity and reduced their >200 nm subpopulations when the bioprocess was carried on in HyCell instead of FreestyleMix (Figure 5). Particle size distributions were not substantially affected between the different media and SGE cell lines (118-128 nm), however VLPs produced in TGE presented larger diameters (144 nm) (Table 3).

When producing VLPs, it is important to distinguish them from extracellular vesicles (EVs): similar nano-particles produced by the host cells [10], which are process-related impurities [11]. For this purpose, the VLP/EV ratio was defined as the percentage of VLPs among the total particles (EVs + VLPs). Again, all platforms except for 10H9 showed remarkable purity improvements in HyCell (Table 3). In this media 13++ reported a VLP ratio of 34.8% over the total EVs, the same as VLPs produced by TGE.

Table 3. VLP characterization by nano-tracking analysis (NTA) of the studied platforms at time of harvest cultured with FreestyleMix or HyCell media. Mean particle diameter of the main population, percentage of the main population and VLP ratio respect the total number of extracellular vesicles are shown.

| Platform | FreestyleMix (day 5) | | | HyCell (day 10) | | |
|----------|-----------------------------|---------------|--------------|-----------------------------|---------------|--------------|
| | Average part. diameter (nm) | Main pop. (%) | VLPs/EVs (%) | Average part. diameter (nm) | Main pop. (%) | VLPs/EVs (%) |
| 13++ | 122 | 92.8 | 10.3 | 125 | 94.9 | 34.8 |
| SH5 | 118 | 90.4 | 5.2 | 118 | 97.8 | 26.3 |
| 10H9 | 118 | 87.7 | 4.6 | 128 | 85.4 | 4.9 |
| TGE | 128 | 96.6 | 26.1 | 144* | 99.3* | 34.8* |

*Harvested at day 5

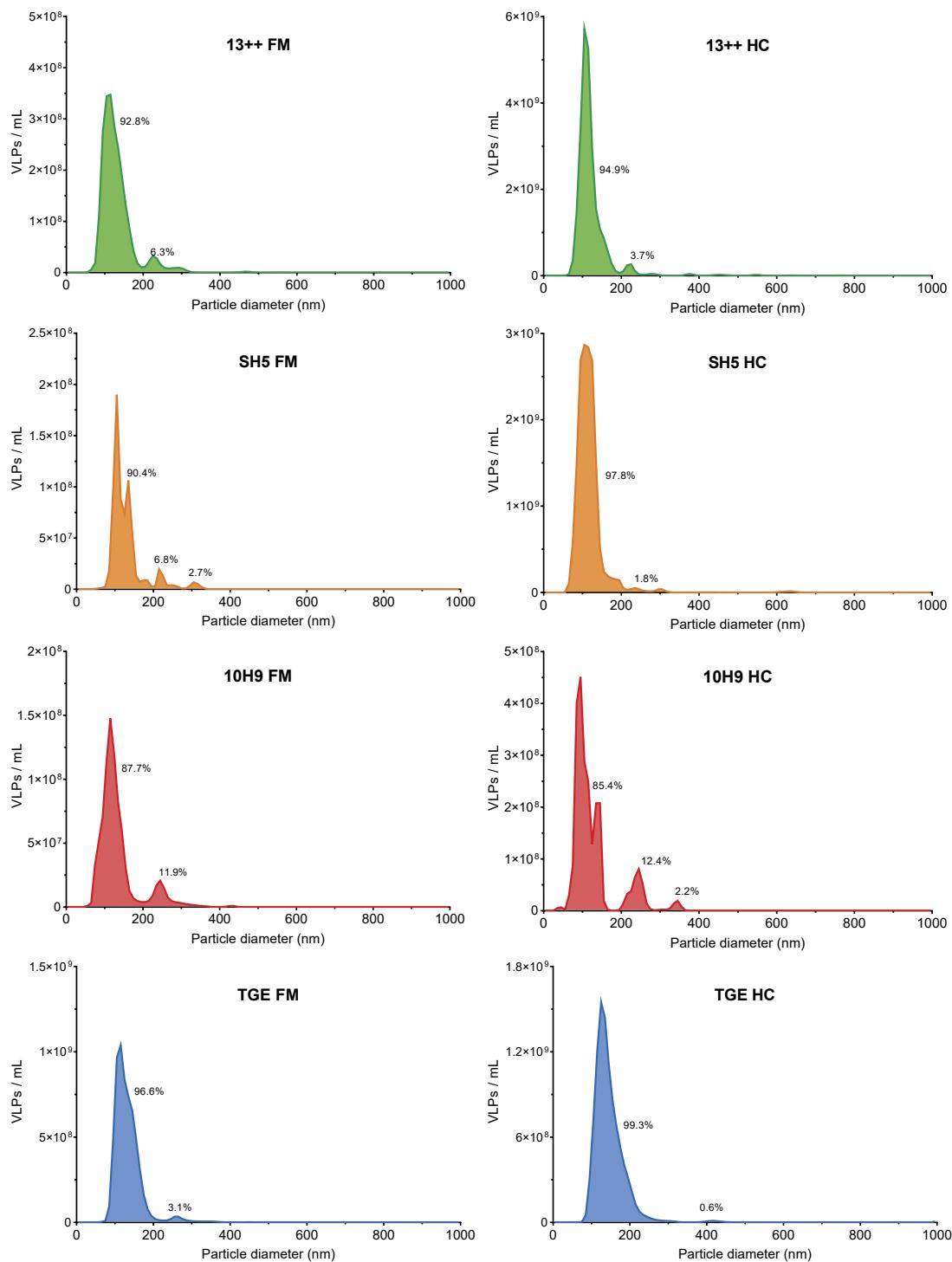


Figure 5. Fluorescent particle size distribution of the supernatants from each studied platform, in FreestyleMix (FM) or HyCell (HC) media at time of harvest, analyzed by nano-tracking analysis (NTA).

4. Conclusions

Three different previously generated SGE platforms constitutively producing Gag::eGFP VLPs were analyzed and compared with TGE as a reference. Cell line 10H9 was generated by random integration, SH5 is a RMCE-competent cell line generated by lentiviral transduction of a RMCE cassette and 13++ was generated by locus-specific CRISPR/Cas-mediated integration of the Gag::eGFP gene at the AAVS1 GSH. Growth analysis showed that no major differences were observed between cell lines in terms of duplication time. Cell lines growing in FreestyleMix media showed similar growth kinetics with a maximum of approximately $4.5 \cdot 10^6$ vc/mL around day four. In HyCell, all SGE platforms behaved similarly experiencing a first maximum at day 6, followed by a lower growth rate phase and a second maximum at day 10, moment until which they maintained high viabilities. Remarkably, SGE cell lines reached much higher cellular concentrations ($>10^7$ viable cells/mL) in HyCell culture. The characteristics of HyCell culture, sustained a longer culture time span and harvest could be performed at day 10.

During the first 5 days of the production process TGE resulted to be the platform with higher production levels. In fact SGE cell lines showed specific productivities lower than those of TGE in both FreestyleMix and HyCell media. However, the elongation of the production process and the high cellular densities reached thanks to the use of HyCell media, compensated the low specific productivities of the SGE cell lines, considerably increasing its VLP concentrations at harvest (day 10). Clone SH5 reached similar volumetric productivities compared to the TGE reference group, while 13++ improved the production levels of transient transfection by 2.7 fold.

VLP characterization and purity analysis also confirmed 13++ in HyCell media as the best strategy for stable Gag-based VLP production, as the platform presented a main VLP subpopulation (94.9%) of an average of 125 nm in diameter (Table 3), with no significant contaminant heterogeneous subpopulations, unlike clone 10H9 (Figure 5). The ratio of VLPs among the total extracellular vesicles (EV) was determined to be 34.8%, exactly the same as

the obtained by TGE, and improving the VLP/EV ratios obtained in SH5 and 10H9 (26.3% and 4.9%, respectively).

This study concludes that the 13++ cell line cultured in HyCell media and harvested at day 10 after seed, is an ideal approach for the generation of Gag-based VLPs. This SGE platform obtains $2.50 \cdot 10^{10}$ VLPs/mL at harvest time, improving the production achieved by TGE by 2.7 fold. SGE cell lines constitute a valuable strategy for the generation of recombinant products due to their monoclonal nature and consequently a more homogeneous product, with low batch-to-batch variations. Additionally, SGE avoids the need to generate and use large amounts of DNA or transfection reagents, simplifying the bioprocess while making it more economic and easy to implement at large scale.

Although not reaching the volumetric productivities of the 13++ cell line, SH5 presents similar production levels as TGE (Table 3). Taking into account the mentioned advantages of SGE in front of TGE, this makes SH5 an interesting alternative to the classic transient transfection approach for the production of VLPs. Additionally, SH5 is a RMCE-competent cell line, with the ability to switch its Gag::eGFP expression by any gene of interest after a rapid and easy cassette-exchange [12].

5. Acknowledgements

The author wants to thank Dr. Amine Kamen (McGill University, Montreal, Canada) for providing the HEK 293 SF-3SF6 cells. The pGag::eGFP (Cat#11468) reagent from Dr. Merilyn Resh was obtained through the National Institutes of Health AIDS Reagent Program, Division AIDS, National Institute of Allergy and Infectious Diseases, National Institutes of Health. We would like to thank Dr Jose Amable Bernabé from the Institut de Ciència de Materials de Barcelona (ICMAB-CSIC, Bellaterra, Catalonia, Spain) for his support in nanoparticle tracking analysis techniques. Figures and Graphs were generated using GraphPad Prism version 8.0.2 for Windows (San Diego, CA, USA).

References

- [1] Cervera, L., Gutiérrez-Granados, S., Berrow, N. S., Segura, M. M. & Gòdia, F. Extended gene expression by medium exchange and repeated transient transfection for recombinant protein production enhancement. *Biotechnol. Bioeng.* **112**, 934–946 (2015)
- [2] Kim, T. K. & Eberwine, J. H. Mammalian cell transfection: The present and the future. *Anal. Bioanal. Chem.* **397**, 3173–3178 (2010)
- [3] Cervera, L., González-Domínguez, I., Segura, M. M. & Gòdia, F. Intracellular characterization of Gag VLP production by transient transfection of HEK 293 cells. *Biotechnol. Bioeng.* **114**, 2507–2517 (2017)
- [4] Gutiérrez-Granados, S., Cervera, L., Gòdia, F., Carrillo, J. & Segura, M. M. Development and validation of a quantitation assay for fluorescently tagged HIV-1 virus-like particles. *J. Virol. Methods* **193**, 85–95 (2013)
- [5] Cervera, L. *et al.* Generation of HIV-1 Gag VLPs by transient transfection of HEK 293 suspension cell cultures using an optimized animal-derived component free medium. *J. Biotechnol.* **166**, 152–165 (2013)
- [6] Hermida-Matsumoto, L. & Resh, M. D. Localization of Human Immunodeficiency Virus Type 1 Gag and Env at the Plasma Membrane by Confocal Imaging. *J. Virol.* **74**, 8670–8679 (2000)
- [7] Schwartz, S. *et al.* Mutational inactivation of an inhibitory sequence in human immunodeficiency virus type 1 results in Rev-independent gag expression. *J. Virol.* **66**, 7176–7182 (1992)
- [8] García, J. L. Bioprocess engineering for HIV-1 Gag VLP production in HEK293 cells. (Universitat Autònoma de Barcelona, 2021).
- [9] Lavado-García, J. Bioprocess engineering for HIV-1 Gag VLP production in HEK293 cells. (Universitat Autònoma de Barcelona, 2021).
- [10] Akers, J. C., Gonda, D., Kim, R., Carter, B. S. & Chen, C. C. Biogenesis of extracellular vesicles (EV): exosomes, microvesicles, retrovirus-like vesicles, and apoptotic bodies. *J. Neurooncol.* **113**, 1–11 (2013)

- [11] González-Domínguez, I., Puente-Massaguer, E., Cervera, L. & Gòdia, F. Quality assessment of virus-like particles at single particle level: A comparative study. *Viruses* **12**, 1–23 (2020)
- [12] Turan, S. *et al.* Recombinase-mediated cassette exchange (RMCE): Traditional concepts and current challenges. *J. Mol. Biol.* **407**, 193–221 (2011)

Chapter seven

**Production of SARS-CoV-2 pseudotyped VLPs in HEK293 cell
lines with stable Gag::eGFP expression**

Abstract:

Gag-based virus-like particles (Gag-VLPs) constitute an attractive approach to produce potential vaccine candidates against COVID-19, by their functionalization with the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) Spike (S) protein. The production of SARS-CoV-2 functionalized VLPs can be performed by transient co-transfection. This approach, can lead to low yields when two proteins have to be co-expressed, as the optimal total amount of DNA delivered in a transient transfection is fixed. Stable gene expression (SGE) systems constitutively expressing Gag represent a valuable strategy to facilitate the functionalization of Gag-VLPs. They only need to be transfected with the DNA coding for the epitope to be incorporated to the outer part of the VLP. Apart from simplifying a process that would otherwise result in low delivery efficiencies, SGE is a valuable approach due to its homogeneous nature and ease to scale-up. However, this approach presents two main drawbacks: the presence of non-functionalized VLPs generated the days prior to the transfection, and the low cell densities reached due to the toxicity of transfection reagents which negatively impacts in the productivity of the process. Here, we report the bench-scale bioreactor for production of functionalized VLPs based on the transfection of a stable cell line constitutively producing Gag::eGFP-VLPs with the S antigen. The obtained results allow the evaluation of the advantages and drawbacks of this approach.

Keywords: Stable gene expression, stirred-tank bioreactor, Gag-VLPs, HIV-1, SARS-CoV-2, Spike

Introduction

COVID-19 is a disease caused by SARS-CoV-2 which generated a worldwide pandemic with huge sanitary, social and economic consequences [1,2], making more evident than ever the need to develop vaccine platforms to fight against emerging infectious diseases. In this regard, virus-like particles (VLPs), constitute a promising approach in order to prevent and fight existing or new emerging diseases or pathogen variants [3]. VLPs are highly immunogenic nanostructures that resemble native virus without carrying its genetic information, making them a safer alternative to conventional vaccines [4]. HIV-1 based VLPs are generated by the recombinant expression of the Gag polyprotein, a process that can be performed by transient transfection (TGE) or stable gene expression (SGE) [5].

The potential of Gag-VLPs to present the spike (S) protein from SARS-CoV-2 by TGE has been previously studied in Chapters 1 and 2 [6]. SARS-CoV-2 functionalized VLPs (from now on, S-VLPs) were generated by the co-transfection of two plasmids, coding for the Gag::eGFP core protein and the S protein to be present at the outer part of the VLP. This approach is relatively easy and rapid to perform, but co-transfection results in low copy delivery efficiencies for each individual gene, as the total amount of delivered DNA is limited by the toxicity effects on the cells and the maximum has been determined by previously optimized transfection protocols [7,8]. Previous studies optimized as well the ratio between the plasmids used in the transfection in order to maximize the number of cells co-expressing simultaneously Gag and S proteins [6]. This limitation inherent to multiple-gene co-transfection could, in principle, be resolved by the use of SGE cell lines constitutively expressing the Gag polyprotein. With this approach, all the DNA delivered to the cells in the transfection is coding for the S protein, resulting in all transfected cells generating S-VLPs. This is also beneficial, as the S protein to be presented in the VLPs can be easily changed, allowing its continuous improvement to present the most recent circulating variants [9], while the structural Gag core of the vaccine candidate remains unchanged.

In this work, the stable cell line 13++ expressing Gag::eGFP-VLPs was cultured in a bench-scale bioreactor and transfected for the S protein in order to study an alternative approach for the generation of pseudotyped VLPs other than the co-transfection of approach.

2. Materials and methods

2.1 Cell lines, media and culture conditions

The serum-free suspension-adapted 13++ cell line is derived from Current Good Manufacturing Practice (cGMP) HEK293 cells. It stably express the Gag::eGFP gene under CMV promoter which derives in the constitutive expression of Gag::eGFP VLPs. 13++ cell line was generated by the integration of the CMV-Gag::eGFP construct into the adeno-associated virus site 1 (AAVS1) genomic safe harbor (GSH) using CRISPR/Cas9. It is also puromycin resistant. For further details check Chapter 4.

The chemically defined and animal-component free HyCell TransFx-H from HyClone (GE Healthcare, Chicago, IL, USA) supplemented with 4 mM GlutaMAX (Gibco, Thermo Fisher Scientific, Waltham, MA, USA) and 0.1% Pluronic F-68 Non-ionic Surfactant (Gibco, Thermo Fisher Scientific, Waltham, MA, USA).

Suspension cell sub-cultures were maintained routinely in exponential growth phase in 125ml disposable polycarbonate Erlenmeyer flasks with vent cap (Corning, Tewksbury, NY, USA), placed in an orbital shaker (orbit diameter 16mm, Stuart, Stone, UK) at 130 rpm in a humidified Steri-Cult CO₂ incubator (3310, ThermoFisher, Waltham, MA, USA) at 37°C, 85% RH and 5% CO₂ in air.

Cell counts and viability determinations were performed using the NucleoCounter NC-3000 automatic cell counter (Chemometec, Lillerød, Denmark) following manufacturer's instructions. Specific growth rate (μ) Specific productivity (q_p) and volumetric productivity (v_p) were calculated using Equation 1, 2 and 3 respectively, where P is the product, X the viable cell density (VCD), t the time, and μ the specific growth rate of the cells during the selected time interval.

$$\mu = \frac{\ln(X_t - X_0)}{t_t - t_0} \quad (1)$$

$$q_p = \frac{\int_{t_0}^t P(t) \cdot dt}{\int_{t_0}^t X(t) \cdot dt} \cdot \mu \quad (2)$$

$$v_p = P/t \quad (3)$$

2.2 Stirred Tank Reactor (STR)

The bioprocess was performed with a BioStat B Plus bioreactor (Sartorius Stedim Biotech, Göttingen, Germany) equipped with a 3-blade segment dual impeller with UP-DP configuration [10]. The agitation was set at 200 rpm, the temperature at 37 °C, and the pH at 7.1, controlled with CO₂ and NaHCO₃ (7.5% w/v). Dissolved oxygen was controlled at 40% of air saturation by supplementing air through a sparger at a constant flow of 0.1 L/min, enriched with pure oxygen when needed. 13++ growing exponentially in disposable polycarbonate 1 L shake flasks (Corning, Tewksbury, MA, USA) were used to seed the bioreactor at 0.5·10⁶ cells/mL in 1 L working volume.

2.3 Plasmids and transfection

2.3.1 Plasmid expression vectors and amplification

The pSpike plasmid codes for a mammalian cell codon optimized nucleotide sequence coding for the spike protein of SARS-CoV-2 driven by the CAG enhancer and β-actin promoter. It was produced under HHSN272201400008C and obtained through BEI Resources, NIAID, NIH: Vector pCAGGS Containing the SARS-Related Coronavirus 2, Wuhan-Hu-1 spike Glycoprotein Gene, NR-52310.

Plasmid DNA was amplified in *Escherichia coli* Top10 strain grown in LB medium at 37°C supplemented with 100 µg/ml of ampicillin (Sigma, St. Louis, MO, USA). Plasmid purification was carried out using the Endofree Plasmid Mega kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions.

2.3.2 PEI-mediated transient transfection

Exponentially growing HEK293 suspension cells were passaged to have a cell density of $2 \cdot 10^6$ cells/mL at transfection time. PEIpro (Polyplus-transfection SA, Illkirch-Graffenstaden, France) was used as transfection reagent. PEI-DNA complexes were formed under sterile conditions. Briefly, DNA was diluted in culture media (10% of the total volume of cell culture to be transfected) for a final total DNA concentration of 1 $\mu\text{g}/\text{mL}$ and vortexed for 10 s. Then, polyethyleneimine (PEI), was added for a final concentration of 2 $\mu\text{g}/\text{mL}$ (a 2:1 PEI:DNA ratio (w/w)) and vortexed three times for 3 s. The mixture was incubated for 15 minutes at RT and then added to the culture.

2.4 Immunocytochemistry staining for flow cytometry

For IF-ICC staining, cells were centrifuged 5 min at $300 \times g$ and rinsed with staining solution (1.5% (v/v) fetal bovine serum (FBS) 1X phosphate-buffered saline (PBS)) before primary antibody incubation for 20 min at 4 °C in the dark. After rinsing twice, cells were incubated with the corresponding secondary antibody for 20 min at 4 °C. After IF-ICC staining, fixation was performed using 2% (v/v) formaldehyde 1X PBS for 10 min at RT. Cells were resuspended in staining solution and stored at 4 °C prior to analysis.

Primary human anti-SARS-CoV-2 spike glycoprotein RBD domain antibody (ab272854, AbCam, Cambridge, UK) was diluted 1:1000. The secondary antibody was an anti-human IgG (H+L) coupled with Cy™5, produced in donkey (709-175-149, Jackson ImmunoResearch, West Grove, PA, USA), diluted 1:400. All antibodies were diluted using staining solution.

Cells were analyzed using a BD FACS Canto flow cytometer (BD BioSciences, San Jose, CA, USA), at Servei de Cultius Cel·lulars, Producció d'Anticossos i Citometria (Universitat Autònoma de Barcelona, Bellaterra, Catalonia, Spain).

2.5 Spectrofluorometry-based product quantitation

Green fluorescence intensity in Gag-GFP samples was measured using a Cary Eclipse Fluorescence Spectrophotometer (Agilent Technologies, Santa Clara, CA, USA). The instrument parameters were set as follows: $\lambda_{\text{ex}} = 488$ nm (slit 5 nm), $\lambda_{\text{em}} = 510$ nm (slit 10 nm). Readings were carried out at room temperature.

Relative fluorescence units (RFU) values were calculated by subtracting fluorescence units (FU) of negative control samples from that given by the sample.

2.6 Nanoparticle Tracking Analysis

NTA-based Gag::eGFP VLP quantification and characterization was performed using a NanoSight NS300 (Nanosight Ltd., Amesbury, UK) equipped with a blue filter module (488 nm) and a neutral filter at the Soft Material services of the Institut de Ciència de Materials de Barcelona (ICMAB-CSIC, Bellaterra, Catalonia, Spain). Samples were previously diluted to a concentration of approximately 10^8 particles/mL. Sample injection was performed using a pump in order to improve the robustness of the measurement by continuous addition, and to minimize the photobleaching effect due to fluorescence depletion over time. 60 second videos were recorded at RT and analyzed with the NTA 3.4 software (Malvern Panalytical, Malvern, UK). Tracked particles size was determined from its Brownian motion. Three independent experimental replicas were carried out for each sample. Camera level and detection threshold were manually adjusted for each replica.

2.7 Dot Blot

Samples were charged into Bio-Dot Apparatus (#1706545, Bio-Rad, Hercules, CA, USA) while low vacuum was applied. Nitrocellulose membrane (#88018, Thermo Fisher Scientific, Waltham, MA, USA) was placed on top of a humidified filter paper. Once samples were transferred, membrane was incubated with anti-SARS-CoV-2 spike glycoprotein S2 monoclonal antibody (Ab281312, AbCam, Cambridge, UK) and an anti-rabbit secondary antibody (A9919, Merck, Kenilworth, NJ, USA) following the same procedure previously described for Western blot. Once dried, membranes were scanned, and the pixel density for each loaded sample was analyzed using the software ImageJ2 Fiji (National Institutes of Health, Bethesda, MD, USA). The standard used for quantification was a recombinant human coronavirus SARS-CoV-2 spike glycoprotein S2 subunit (Ab272106, AbCam, Cambridge, UK).

3. Results

3.1 SARS-CoV-2 VLP production by transient transfection of a Gag-VLP SGE cell line in a STR

In Chapter 6, it was determined that the best platform for stable Gag-VLP expression was the 13++ cell line cultured with HyCell media. The SARS-CoV-2 VLP production at bioreactor scale was performed in a 1L stirred-tank bioreactor (STR) using this clone and transfecting with a plasmid coding for the SARS-CoV-2 Spike protein. Exponentially growing 13++ cells were inoculated at $0.5 \cdot 10^6$ viable cells/mL (vc/mL). The process was set to operate at 200 rpm, with temperature, pH and dissolved oxygen controlled at 37°C, 7.1 and 40%, respectively. The bioreactor was transfected two days after its inoculation with the plasmid DNA coding for the SARS-CoV-2 Spike protein, at a cell density of $2 \cdot 10^6$ vc/mL. As a consequence of the transfection, cell growth in the STR experienced a punctual drop, reflected in the evolution of viable cell density, probably due to a negative effect caused by the transfection in STR conditions, since the same behavior was not observed in the parallel Erlenmeyer flasks (20 mL, n=3). From 24 hpt on, cells resumed its growth, reaching a final density of $2.46 \cdot 10^6$ vc/mL, with a viability of 84.9% similar to the values obtained at the parallel Erlenmeyer flasks, of $2.73 \cdot 10^6$ vc/mL and 89.3% viability (Figure 1). Bioprocess was ended and supernatant was harvested at 72 hpt, as previously established for TGE in HEK293 cells [11].

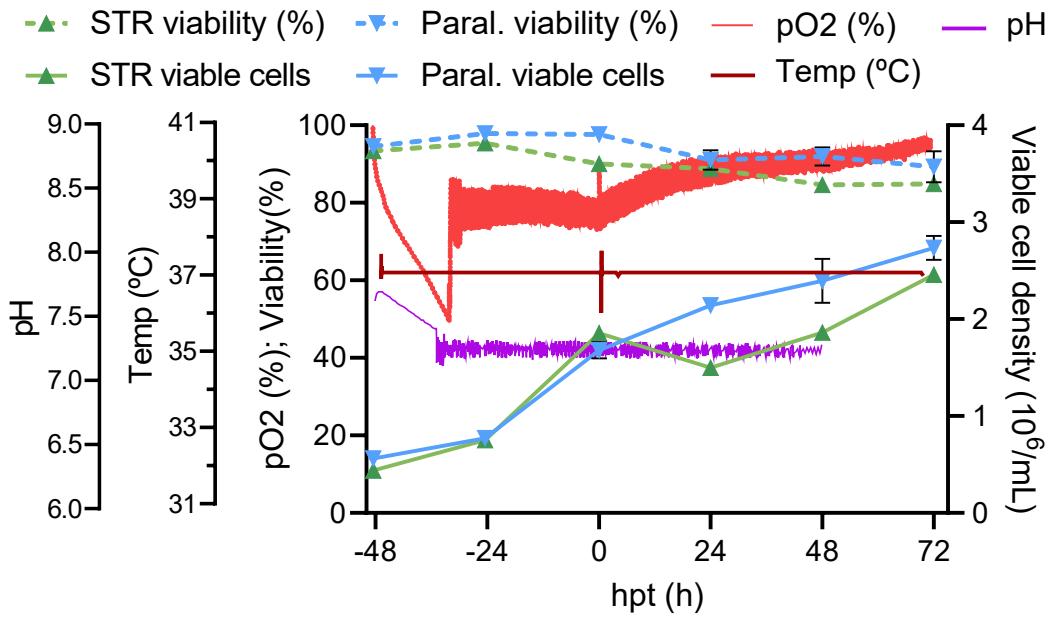


Figure 1. Production of S-VLPs by TGE of 13++ cell line in a stirred 1 L bioreactor. On-line profile measurements of different process parameters: temperature, pH and dissolved oxygen concentration (pO_2). Cell density and cell viability values are also shown.

ICC-staining and flow cytometry analysis determined the S-transfected population. Reactor presented a 99% population constitutively expressing Gag::eGFP as expected, and 45% population co-expressing S and Gag::eGFP proteins, slightly less than the parallel Erlenmeyer flasks, with 50.6% (Figure 2.A). This represents low transfection percentages compared with ~80% obtained for standard single plasmid transfection procedures [5]. Moreover it does not improve the S and Gag::eGFP expresser populations obtained with two plasmid co-transfection (~55%) [6], a method with, *a priori*, less gene deliver efficiency (Table 1).

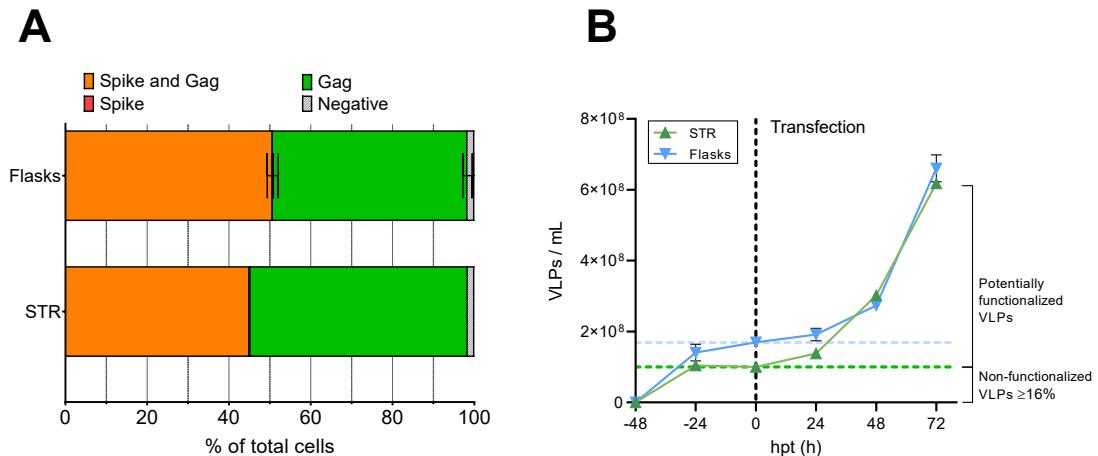


Figure 2. (A): Transfected populations, analyzed by ICC at 72 hpt. (B): Gag-VLP production kinetics: particle concentration at different time-points pre and after transfection. Dashed vertical line indicates transfection. Dashed horizontal lines indicate the VLP concentration at transfection time, meaning that the VLPs below the line are non-functionalized with the Spike protein.

VLP concentration was monitored during the whole process, where reactor and parallel flasks behaved similarly (Figure 2.B). 13++ cell line generated Gag VLPs since its inoculation, experiencing a remarkable production boost from 48 to 72 hpt. However, $\sim 6 \cdot 10^8$ VLPs/mL were obtained at 72 hpt harvest (Figure 3.A). This represents a 5.8-fold less concentration compared to the VLPs generated with transient co-transfection of S and Gag::eGFP in a STR [6]; and 50-fold decrease compared with the SGE VLP production by 13++ cell line harvested at day 10 (Table 1). As it was determined in Chapter 6, 13++ cell line has a lower VLP specific productivity compared to TGE; however, this can be compensated by its ability to reach high cell densities during a prolonged period, resulting in greater volumetric productivities than TGE. Therefore, the low cellular densities of the 13++ cell line after transfection in the bioreactor were translated into low VLP production levels.

Additionally, a drawback that needs to be considered when using the presented approach for the production of pseudotyped VLPs is the fact that the cells not expressing the S protein produce non-functionalized VLPs. Non-functionalized VLP population was constituted by all the VLPs produced before the transfection took place (Figure 2 and Figure 3.A), together with the VLPs produced by the non-transfected 13++ cells, a ~55% of the total cells.

Table 1: Comparison between different VLP and S-VLP production approaches, in terms of transfected populations, S-VLP expresser populations, VLP concentrations, VLP/EV ratio and Spike concentrations.

| Approach | HEK293 Gag::eGFP TGE (Chapter 6) | 13++ SGE (Chapter 6) | HEK293 Gag::eGFP + Spike TGE (Chapter 1) [6] | 13++ SGE + Spike by TGE |
|---|----------------------------------|--------------------------|--|-------------------------|
| Total transfected population | ~80% | 0% | ~80% | ~45% |
| Gag-VLP producer population | 80% (TGE) | ~100% (SGE) | ~70% (TGE) | ~99% (SGE) |
| Pseudotyped VLP producer population | 0% | 0% | ~55% (TGE) | ~45% (TGE) |
| Max. cell density (viable cells/mL) | $3.1 \cdot 10^6$ | $11 \cdot 10^6$ | $3.95 \cdot 10^6$ | $2.46 \cdot 10^6$ |
| VLP concentration (VLPs/mL) | $\sim 1 \cdot 10^{10}$ | $\sim 2.5 \cdot 10^{10}$ | $\sim 3.5 \cdot 10^9$ | $\sim 6 \cdot 10^8$ |
| VLP/EV ratio | 34.8% | 34.8% | 16.8% | 6.8% |
| Spike concentration ($\mu\text{g}/\text{mL}$) | - | - | 1.46 | 0.192 |

Reactor and Erlenmeyer flasks presented an almost identical ratio of VLP to extracellular vesicles as evaluated by NTA fluorescent particle analysis (Figure 3.B) with no significant difference in mean particle diameter at harvest (147.5 vs 164.2 nm). Quantitative dot blot analysis allowed the determination of the SARS-CoV-2 spike concentration in the harvested supernatants (Figure 3.C). The reactor showed a concentration of ~0.192 μg spike/mL, while flasks supernatant had a slightly lower concentration, of 0.173 μg spike/mL. This represented a 7.6-fold decrease compared to the S and Gag::eGFP co-transfection approach in a STR (1.46 μg spike/mL) [6] (Table 1).

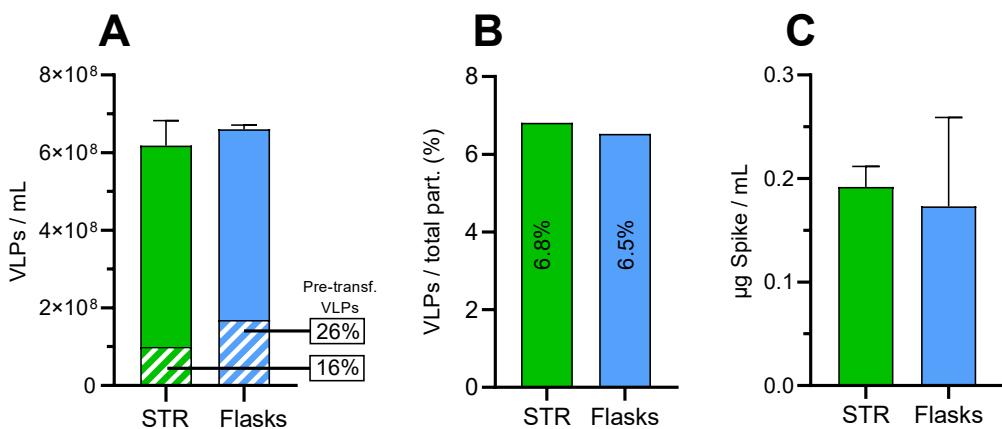


Figure 3: Analysis of the harvested VLPs. (A) Gag-VLP concentrations at harvest measured by NTA fluorescent particle analysis. The pre- transfection VLPs (therefore, non-functionalized) are represented as dashed bars and its percentage among the total harvested VLPs of each condition is indicated. (B): Percentage of VLPs from the total particles present at the harvest, by NTA fluorescent and non-fluorescent particle analysis. (C): SARS-CoV-2 spike concentrations at harvest, analyzed by dot blot.

4. Conclusions

The production of Gag-based VLPs using the SGE 13++ cell platform transfected with a plasmid coding for the SARS-CoV-2 Spike glycoprotein was explored. The bioprocess was carried out successfully, observing good cellular growth and viabilities. Erlenmeyer flasks showed a very similar behavior compared to the STR in all analyzed aspects. ICC analysis determined that 45% of the 13++ population was transfected, a surprisingly low transfection efficiency compared with other VLP and S-VLP production approaches with theoretically less transfection efficiency (Table 1). For this reason, further studies on the TGE of 13++ and other SGE cell lines need to be performed in order to determine if the observed low transfection efficiency is caused by the nature of the cell line itself or exclusive for this approach or bioprocess. The production kinetic of the process was assessed, observing the generation of VLPs from the inoculation of the bioreactor 48 hours pre-transfection until its harvest at 72 hpt. This behavior is not ideal for the production of pseudotyped VLPs, as the VLPs produced before transfection are non-functionalized, in this case accounting for 16% of the total VLPs. The use of inducible promoters would avoid this issue, and should be considered for the development of future VLP-expressing SGE cell lines. A media exchange at the moment of the

transfection would also allow removing the already produced VLPs but it is challenging and not ideal. The harvest of the produced VLPs was performed at 72 hpt, obtaining 5.8-fold less VLPs compared to alternative co-transfection approaches for the production of S-VLPs (Table 1), while Spike concentrations were reduced in 7.6-fold. This can be explained as not only few VLPs were generated, but also a greater proportion of them were non-functionalized. Moreover, the low harvested VLP concentration was caused by the low densities reached by 13++ after its transfection, probably due to the toxic effect of PEI transfection reagent. The cell density effect (a not well-known phenomena limiting the transfection of the cells to low cell densities) needs to be studied in order to implement protocols allowing efficient transfection of cells at cell densities greater than $2 \cdot 10^6$ vc/mL [12]. Another factor that determined the low VLP production of the cells in this study was the process length. As determined in Chapter 6, the specific productivity of the 13++ does not reach the one shown by TGE cells, however this was compensated by the combination of high cell densities together with the process elongation until day 10. With this approach, volumetric productivity of the 13++ ended up exceeding by 2.7 fold the ones obtained with TGE. For this reason, approaches like continuous operation using extended gene expression (EGE) should be considered [11]. EGE is based on multiple cell re-transfections to avoid the dilution of the transfected DNA and it has been successfully used for Gag VLP production at bioreactor scale [13]. Without this approach, the S protein transfection of the 13++ cell line would derive in the production of non-functionalized VLPs after 72 hpt.

Two more approaches can be considered for the production of S-VLPs at reactor scale. One is the generation of Gag-VLPs by the 13++ cell line without epitope transfection, reaching high cellular densities with the possibility to elongate operations using fed-batch or continuous processes for the generated VLPs to be later functionalized by means of click-chemistry or other chemical approaches [14]. Another option would be the generation of a new SGE cell line for the co-expression of Gag and S proteins. However, this would be a long, tedious, and time-consuming process, also accompanied by the lack of flexibility for the change of the epitope in case of need.

Overall this work evidenced the possibility of operating and transfecting a SGE for the production of pseudotyped VLPs at bioreactor scale, while demonstrating its scalability obtaining the same results as in Erlenmeyer flasks. The results allowed the analysis of the advantages and drawbacks of the presented approach which helped to determine the future steps needed to follow in order to further optimize the production of pseudotyped VLPs using SGE cell lines.

5. Acknowledgements

We would like to thank Dr Jose Amable Bernabé from the Institut de Ciència de Materials de Barcelona (ICMAB-CSIC, Bellaterra, Catalonia, Spain) for his support in nanoparticle tracking analysis techniques. We also would like to thank Paula Grau and Laia Miranda for their help setting up the bioreactor. Figures and Graphs were generated using GraphPad Prism version 8.0.2 for Windows (San Diego, CA, USA).

6. References

- [1] Guan, D. *et al.* Global supply-chain effects of COVID-19 control measures. *Nat. Hum. Behav.* **4**, 577–587 (2020)
- [2] Hu, B., Guo, H., Zhou, P. & Shi, Z. L. Characteristics of SARS-CoV-2 and COVID-19. *Nat. Rev. Microbiol.* **19**, 141–154 (2021)
- [3] Fuenmayor, J., Gòdia, F. & Cervera, L. Production of virus-like particles for vaccines. *N. Biotechnol.* **39**, 174–180 (2017)
- [4] Lavado-García, J. *et al.* Characterization of HIV-1 virus-like particles and determination of Gag stoichiometry for different production platforms. *Biotechnol. Bioeng.* **118**, 2660–2675 (2021)
- [5] Lavado-García, J. *et al.* Metabolic engineering of HEK293 cells to improve transient transfection and cell budding of HIV-1 virus-like particles. *Biotechnol. Bioeng.* **118**, 1649–1663 (2021)
- [6] Boix-Besora, A., Lorenzo, E., Lavado-García, J., Gòdia, F. & Cervera, L. Optimization, Production, Purification and Characterization of HIV-1 GAG-Based Virus-like Particles Functionalized with SARS-CoV-2. *Vaccines* **10**, 1–18 (2022)
- [7] Kreiss, P. Plasmid DNA size does not affect the physicochemical properties of lipoplexes but modulates gene transfer efficiency. *Nucleic Acids Res.* **27**, 3792–3798 (1999)
- [8] Pezzoli, D., Giupponi, E., Mantovani, D. & Candiani, G. Size matters for in vitro gene delivery: investigating the relationships among complexation protocol, transfection medium, size and sedimentation. *Sci. Rep.* **7**, 44134 (2017)
- [9] Vasireddy, D., Vanaparthy, R., Mohan, G., Malayala, S. V. & Atluri, P. Review of COVID-19 Variants and COVID-19 Vaccine Efficacy: What the Clinician Should Know? *J. Clin. Med. Res.* **13**, 317–325 (2021)
- [10] Buffo, M. M. *et al.* Influence of dual-impeller type and configuration on oxygen transfer, power consumption, and shear rate in a stirred tank

- bioreactor. *Biochem. Eng. J.* **114**, 130–139 (2016)
- [11] Cervera, L., Gutiérrez-Granados, S., Berrow, N. S., Segura, M. M. & Gòdia, F. Extended gene expression by medium exchange and repeated transient transfection for recombinant protein production enhancement. *Biotechnol. Bioeng.* **112**, 934–946 (2015)
- [12] Backliwal, G., Hildinger, M., Hasija, V. & Wurm, F. M. High-density transfection with HEK-293 cells allows doubling of transient titers and removes need for a priori DNA complex formation with PEI. *Biotechnol. Bioeng.* **99**, 721–727 (2008)
- [13] Fuenmayor, J., Cervera, L., Gòdia, F. & Kamen, A. Extended gene expression for Gag VLP production achieved at bioreactor scale. *J. Chem. Technol. Biotechnol.* **94**, 302–308 (2019)
- [14] Moses, J. E. & Moorhouse, A. D. The growing applications of click chemistry. *Chem. Soc. Rev.* **36**, 1249–1262 (2007)

Discussion and future work

This PhD thesis studied the functionalization of Gag::eGFP VLPs with the SARS-CoV-2 spike (S) protein and tested their recognition by convalescent COVID-19 human sera. It also developed different HEK 293 cell lines for the production of Gag::eGFP VLPs using three different strategies. Then, the best cell line was selected, and its pseudotyping with the S protein was studied at bioreactor scale.

Chapter one consisted in the study of the Gag::eGFP VLPs functionalization using the S protein from the SARS-CoV-2 virus. Previously published studies were focused on the generation SARS-CoV and SARS-CoV-2 VLPs by the co-expression of the coronavirus M, S and E proteins [1]. In this chapter, a potential COVID-19 vaccine candidate, based on HIV-1 Gag-based S VLPs (from now on S-VLPs) was studied a novel approach to the best of our knowledge. HIV-1 Gag VLPs can be used to generate immunity against different diseases, since they are composed by a core of Gag molecules surrounded by a lipid bilayer, a membrane that can be functionalized with epitopes of interest [2]. No deleterious effects on cell viability were observed upon the recombinant co-expression of the S and Gag::eGFP proteins by transient transfection of HEK 293 cells. Otherwise, low viabilities would indicate cytotoxicity caused by the S protein incomplete maturation through the secretory pathway [3]. Confocal images showed co-localization of the S and Gag::eGFP proteins at the plasmatic membrane. These results confirmed that after its recombinant expression, S protein travels to membrane, where Gag::eGFP-VLPs are generated via budding [4]. Purified VLPs were analyzed by DotBlot and electron microscopy (EM). Dot Blot led us to conclude that S protein was present in the produced VLPs while EM immunogold labeling confirmed the incorporation of S antigens in the outer membrane of the VLPs. S-VLPs presented a mode diameter of 134.9 ± 1.2 nm with no significant structural differences with the Gag::eGFP-VLPs.

Further, the transfection protocol was optimized using Design of Experiments to increase the S-VLP population. The obtained model fitted the data correctly, and allowed us to determine the optimal pSpike, pGag::eGFP and PEI concentrations that resulted in the highest double-positive transfected cell population at 72 hpt. The optimal concentrations found were 0.308 µg/mL of

pGag::eGFP, 1.058 µg/mL of pSpike, and 2.045 µg/mL of PEI. With these parameters a 57.5±2.3% of double expresser population at 72 hpt was predicted. This was then corroborated experimentally, obtaining a 58.9±0.4% of double expresser population. Then, the optimized bioprocess was satisfactorily performed at 1L bioreactor, showing a comparable behavior with the parallel flasks in all the analyzed parameters and obtaining a 55.1% of cells expressing S and Gag::eGFP proteins, which was in agreement with the model prediction.

A scalable and cGMP-compatible downstream process (DSP) was used to purify the bioreactor harvest. The DSP consisted of two clarification steps, an anion-exchange (IEX) capture step and a size-exclusion (SEC) final polishing step. Primary clarification with a deep filter successfully removed cell debris, intact cells, impurities, and aggregates among other undesired contaminant particulate materials. Secondary clarification reduced bioburden, undesired particles and precipitates, as required for the proper performance of the subsequent IEX. Overall, clarifications reduced the turbidity of the harvested supernatant from 22.5 NTU to 2.13 NTU. Capture and polishing steps greatly succeeded in the elimination of dsDNA (98.97% reduction), host cell proteins (99.78% reduction) and VLP aggregates while increasing the VLP ratio among total particles. Globally, the DSP was very efficient in the reduction of contaminants while it improved the VLP concentration, VLP ratio and S protein concentration. Unfortunately, the VLP recovery yield of the whole process was 2.18%, much lower than the one obtained in previously reported Gag-VLP DSP (23%) [5]. Analyzing DSP by steps, clarifications showed a recovery of 79.6%, IEX a 10% and SEC a 38%. The recovery percentages presented by IEX and SEC were lower than expected (36% for IEX and 61% for SEC, according to previous work) [5]. IEX fraction analysis showed that 33% of the VLPs were lost at the Flow Through (Ft), while 14% were lost at the E1 fraction.

Globally, Chapter one demonstrated the possibility to generate S-VLPs while defining a cGMP-compatible bioprocess, which facilitates its potential transfer to industrial scale. The defined Gag-based functionalization approach can be used to produce new chimeric VLPs against different COVID-19 variants or any pathogen of interest, becoming a methodology with a great potential for the generation of vaccines in case of future pandemics.

In **Chapter two**, the functionalization approach defined in Chapter one was used to generate different VLP structures, named S_{WT} -VLP, S_{mut2} -VLP and S_{mut3} , considering several desired traits in their design. S_{WT} codes for the SARS-CoV-2 Wuhan spike sequence. S_{mut2} incorporates stabilizing substitutions [6,7], the mutation D614G from the early dominant pandemic form [8], the elimination of the furin cleavage site to improve homogeneity [9,10]; and two cysteine substitutions to increase immunogenicity and to avoid potential undesired antibody-dependent enhancement (ADE) responses [10]. S_{mut3} additionally presents the suppression of the C-terminal ERGIC retention motif [11], with the aim to promote its membrane localization and thereby its incorporation to the generated VLPs.

The stabilizing mutations present in S_{mut2} and S_{mut3} variants facilitated their recombinant expression, which was translated in a ~10% viability improvement at harvest compared to S_{WT} . Additionally, the population expressing both the Gag::eGFP and the S proteins was superior in S_{mut2} and S_{mut3} expresser cells (60.6% and 62.2%) than in S_{WT} (55.2%). When purified, S_{mut2} and S_{mut3} VLPs presented 2.2 and 1.6-fold concentration improvements compared with the unmodified S_{WT} -VLPs, duplicating their VLP percentage among total extracellular vesicles up to (14%-16.2%). Dot Blot analysis showed that the S_{mut3} -VLP was the candidate presenting a greater S concentration, exceeding by 2.2 fold the other variants. Additionally, S_{mut3} -VLPs presented significantly greater S functionalization levels than S_{mut2} -VLPs, indicating the detrimental effect of the ERGIC retention signal in the incorporation of the S protein at the VLP surface. These results show evidence that the introduced mutations improved S_{mut2} and S_{mut3} recombinant expression while enhancing VLP production and their functionalization with the S protein.

The three produced S-VLP immunogens were then tested for their recognition by human sera from patients recovered from COVID-19. All the SARS-CoV-2 convalescent sera showed recognition to the functionalized VLPs, varying on intensity depending on their antibody positivity against SARS-CoV-2. Remarkably, in the majority of the cases, the S-VLP variants were better recognized than the soluble version of the S protein used as positive control. This was probably due to a much more advantageous antigenic presentation of

the S protein when incorporated at the surface of the VLPs, facilitating its proper conformation and emulating the natural viral context surrounding the S protein [12]. Unexpectedly, the immunogens showing the best recognition were the S_{WT} -VLPs and the ones showing the worst levels were the S_{mut3} -VLPs. In conclusion, even though the introduction of the studied mutations resulted beneficial in the production process and S-VLP quality, mutated variants recognition by convalescent sera did not reach the levels shown by S_{WT} -VLPs, which were selected as the best immunogen for further studies.

The COVID-19-convalescent human sera assay gave us very valuable insights on the immunogenic potential of the tested S-VLPs. Next steps should be done to assess the immunoresponse of the S_{WT} -VLPs when inoculated in animal models subsequently challenged with the SARS-CoV-2 virus. With this approach the real protective effect of the VLPs could be evaluated. In this regard, S_{mut2} and S_{mut3} -VLPs should not totally be left behind. Even though sera assays clearly demonstrated that S_{mut2} and S_{mut3} -VLPs were less recognized, it is possible that this was a consequence of the concealment of ADE epitopes, while in S_{WT} -VLPs these undesired epitopes were recognized by non-neutralizing low-quality antibodies [13]. In-vitro assays to study ADE-triggering potential of the produced variants should also be considered as future work.

Chapters 1 and 2 demonstrated the potential of the S protein to be incorporated at Gag::eGFP-VLPs for their use as immunogens. In order to transfer the production of S-VLPs at large scale and to overcome the drawbacks associated with transient gene expression (TGE) [14], Chapters 3, 4 and 5 developed different stable gene expression (SGE) cell lines for the constitutive production of Gag::eGFP-VLPs using different methodologies.

In **Chapter three**, a SGE cell line constitutively producing Gag::eGFP-VLPs was generated through illegitimate random integration. A pool with selective antibiotic pressure was generated to select the clones expressing the construct of interest, and then 672 individual Gag::eGFP+ clones were sorted by fluorescence-activated single cell sorting (FACS), grown and screened. The best producers were adapted to serum-free media and suspension culture. In order to select the best cell line, growth kinetics of the more promising clones

were studied. The best cell line, named 10H9, presented a duplication time of 29.2 hours, a maximum viable cell density (VCD) of $6.16 \cdot 10^6$ vc/mL and a specific productivity of 35.7 VLPs/cell·day which remained stable during the studied 60 days. 10H9 presented 4 copies of the integrated genetic construct on its genome.

Although the 10H9 clone showed good cellular growth, homogeneity, and prolonged stable expression, its VLP production was low compared with TGE approaches. For that reason the generation of a new cell line with improved VLP production was explored in **Chapter four**. The illegitimate integration approach used at Chapter three relies on the random integration of a construct of interest into the host cell genome [15]. The spontaneous integration of the transgene into a high transcriptional locus is a very unlike event, and can be inactivated or silenced over time [16]. For that reason, the high-throughput screening of a large number of clones would be required in order to generate a high-producer cell line using random integration. Another option is to target the integration of the transgene into a specific pre-defined locus [17]. In this regard, genomic safe harbors (GSH) such as AAVS1 are transcriptionally active regions in the human genome with a solid and predictable expression that can accommodate the integration of transgenes without their silencing [18,19]. There are several methodologies that can accomplish this purpose, from which CRISPR/Cas9 was selected as the most promising approach [20].

In Chapter four, a SGE cell line constitutively producing Gag::eGFP-VLPs was generated by site-directed integration at AAVS1 GSH using CRISPR/Cas9. For this purpose two plasmids were constructed: pT2 coding for the Cas9 nuclease and a sgRNA targeting the AAVS1 locus; and the pDonor containing the construct of interest flanked by AAVS1 homology arms. A pool with selective antibiotic pressure was generated to select the clones presenting integration at the AAVS1 locus, as the integration in this region activates the antibiotic resistance gene present in the construct. FACS allowed the isolation of 960 individual Gag::eGFP+ clones. The more promising ones were adapted to serum-free media and suspension culture. The analysis of the growth kinetics allowed the selection of the best cell line, named 13++. It presented a duplication time of 34.6 hours, a maximum VCD of $3.96 \cdot 10^6$ vc/mL and a

specific productivity of 222.9 VLPs/cell·day, which remained stable during the studied 33 days. 13++ greatly improved the VLP production levels presented by the 10H9 cell line and showed potential to overcome TGE upon bioprocess intensification. ddPCR analysis determined that 13++ presented 8 integrated copies of the construct of interest per cell, while their incorporation into the AAVS1 locus was confirmed by PCR and sequencing. This large number of integrated copies can be explained by the tandem integration of different donor constructs in the same locus by homologous recombination of their compatible homology arms.

Chapter four successfully established a methodology for the generation of SGE HEK293 cell lines. Additionally, the pT2 plasmid generated can be used for the production of virtually any desired SGE HEK293, as long as it is co-transfected with a plasmid coding for the construct of interest to be integrated flanked by AAVS1 homology arms. This is a relevant tool that will certainly be used in future work of the group in the development os SGE HEK 293 cell lines for VLPs production.

In **Chapter five**, a Recombinase-Mediated Cassette Exchange (RMCE) stable cell line producing Gag::eGFP-VLPs was generated by lentiviral transduction. This methodology relies in the random integration of a tagging (TAG) cassette into a cellular pool, followed by the selection of a high-expresser clone for the generation of a stable cell line [21]. Then, the TAG cassette incorporated into a highly transcriptional region of the cell line genome, can be swapped for any construct of interest in a simple, rapid and economic approach known as targeting (TAR) [22,23]. Importantly, the initially generated cell line needs to present a single copy of the TAG cassette in order to be able to correctly perform subsequent TAR procedures [23]. A lentiviral infection using low multiplicity of infection was used to achieve the integration of a single cassette per cell. More than 2000 individual Gag::eGFP+ clones were isolated. The more promising ones were adapted to suspension and serum-free media. Growth kinetic studies of the 5 best producers were performed and the best clone with ability to correctly perform RMCE was selected. The selected clone, named SH5, presented a duplication time of 29.0 hours, a maximum VCD of $4.58 \cdot 10^6$ cells/mL and a stable specific productivity of 110 VLPs/cell·day during the

studied 45 days. SH5 presented a single cassette copy on its genome and it was RMCE-competent. SH5 has the potential to generate new high producer stable cell lines for the expression of virtually any gene of interest.

SH5 and 10H9 were both generated using random integration approaches. Interestingly, the higher number of screened clones for the generation of SH5 compared to 10H9, resulted in a much more productive cell line, evidencing the importance of high throughput screening for the generation of cell lines.

In **Chapter six**, the 10H9, SH5 and 13++ cell lines generated at Chapters 3, 4 and 5 were studied and compared with a standard TGE Gag-VLP production approach to assess their potential as Gag::eGFP VLP production platforms. Culture media is a crucial parameter when it comes to bioprocess optimization as it can affect cell growth, cellular specific productivity and VLP quality. For this purpose, growth kinetics, production levels and VLP quality of the different mentioned platforms were studied when cultured in FreestyleMix (the media in which cell lines were initially generated) and HyCell (a commercial media which showed the best Gag-VLP production levels in other studies [24]).

When cultured in FreestyleMix, all the groups presented duplication times between 30.6 and 33.5 hours and SGE cell lines reached their maximum cell density of approximately $4.5 \cdot 10^6$ cells/mL between the fourth and fifth day. In the same conditions, TGE reached a maximum of $3 \cdot 10^6$ cells/mL at 72 hpt (equivalent to 5 days of culture). The determination of the time of harvest for Gag-VLP production is mainly driven by three parameters. First, the production of VLPs has been observed to be predominantly growth-associated; therefore, extending the bioprocess after cells stopped growing is not beneficial. Second, the release of host-cell proteins and other undesired elements by apoptotic cells results in the presence of contaminants in the harvested product, making subsequent DSP more difficult [25]. Finally, when producing VLPs with conventional TGE approaches, the transfected plasmid DNA is progressively degraded and diluted in each cell division. Therefore, even if cells continue to grow with optimal viabilities at 72 hpt (which is improbable due to the cytotoxic effect of the transfection reagents), the dilution of the plasmid DNA at this point is translated in a higher cellular population percentage not producing Gag-VLPs

but releasing undesired EVs [26,27]. The low viabilities experienced by SGE cell lines after day 5 in FreestyleMix, determined that the harvest should be performed at this time. Productions in FreestyleMix resulted in $7.13 \cdot 10^9$ VLPs/mL for TGE and significantly lower values for all the SGE candidates, being 13++ the maximum producer with $1.84 \cdot 10^9$ VLPs/mL at time of harvest.

When cultured in HyCell media, cell lines reached significantly higher cell densities and for a prolonged time period before experiencing sudden apoptosis. This allowed the VLP harvest to be performed at day 10 for SGE cell lines cultured in HyCell. TGE cell lines reached much lower concentrations and viabilities due to the PEI's cytotoxic effect. In any case, TGE cells were harvested at day 5 (72 hpt) because of the previously mentioned DNA dilution phenomena. This resulted in the TGE approach producing $9.16 \cdot 10^9$ VLPs/mL. Interestingly, the high VCD reached, together with the extension of the production process, resulted extremely beneficial for clones SH5 and 13++. SH5 produced $1.57 \cdot 10^{10}$ VLPs/mL while 13++ harvest contained $2.48 \cdot 10^{10}$ VLPs/mL (2.7-fold improvement compared to TGE). VLPs were characterized and their quality assessed. Results showed that when 13++ and SH5 were cultured in HyCell particle size distributions, VLP homogeneity, and VLPs among total particles ratio significantly improved, reaching levels comparable to those of TGE.

Overall, 13++ resulted to be the best SGE cell line in terms of specific productivity, volumetric productivity and VLP quality. For that reason 13++ was selected as the best candidate to produce Gag-based VLPs. Moreover, the good results showed by SH5 cell line, equaling in production the TGE, also makes it a very interesting approach for the production of VLPs not only because all the benefits of SGE in front of TGE, but also because it harbors the potential to express any protein of interest as it is an RMCE-competent cell line.

The VLP production kinetics showed no significant VLP production increase from day 9 to day 10 in either 13++ or SH5 cell lines, therefore we suggest that batch productions for 13++ or SH5 in HyCell should be harvested at day 9 to avoid the possible presence of EVs or undesired cellular by-products.

Shortening the batch operation by one day will ease the bioprocess, while obtaining the same VLP production levels.

In Chapter six, the change from FreestyleMix to HyCell resulted in a successful production enhancement. However, future work should be considered in order to further optimize 13++ and SH5 VLP production. This can be achieved mainly by three strategies: increasing VCD (more cells producing VLPs), modulating the bioprocess operation (more time to produce VLPs) or enhancing specific productivity (more VLPs produced per cell) [28]. As previously described, the use of HyCell media resulted in a significant increase of the VCD and allowed the extension of the bioprocess. Additionally, it also enhanced the specific productivity of the 13++ cell line. However, a cell-line-specific media optimization should be done to further improve 13++ and SH5 productivities by using Design of Experiments. In this regard, non-animal derived additives such as tocopherol, fatty acids, synthetic cholesterol, polysorbate 80, r-transferrin, albumin and insulin should be considered as they had been reported to enhance HIV-1 Gag-VLP production [29].

Furthermore, the implementation of fed-batch or perfusion operations should also be considered in future work. These operations allow reaching high VCDs and substantially extend the bioprocess duration. They have been successfully used for the production of Gag-VLPs by TGE [30]. TGE in perfusion processes need to be constantly re-transfected with repeated medium exchanges (method known as Extended Gene Expression [EGE] [31]) to counteract the plasmid dilution effect and prolong the production phase; and even though, expresser population is usually below 80% [30]. Conversely, perfusion approaches using SGE cell lines ensures that during the bioprocess the whole cellular population is expressing Gag-VLPs which greatly simplifies the operation, avoiding the use of transfection reagents, plasmid DNA, its production, reduces the process economic cost, manipulation and decreases the bioreactor contamination possibility.

In Chapter two, S_{WT} was determined to be the most promising S variant in terms immunogenicity, as it showed the highest recognition levels by COVID-19 human convalescent sera. In Chapter six, 13++ SGE was determined to be the

best cell line for the production of Gag::eGFP-VLPs. Given this context, in **Chapter 7**, 13++ was cultured in a bench-scale 1L bioreactor and transfected with the S_{WT} protein to study the production of pseudotyped S_{WT}-VLPs. The bioreactor showed similar behavior as the parallel shake flasks, indicating that the scale up of the process was achieved successfully. At 72 hpt, STR presented a 99% of population constitutively expressing Gag::eGFP and a 45% of transfected cells co-expressing the S_{WT} and Gag::eGFP proteins, a surprisingly low value given that single-plasmid TGE procedures typically reach transfection values of ~80% [4]. Moreover, the obtained 45% of double expresser population is smaller than the ~55% observed at Chapter one with the co-transfection of two plasmids also coding for S_{WT} and Gag::eGFP [12], an approach with theoretically less gene delivery efficiency [32]. More research needs to be done in order to determine if the observed low double expresser population is caused by the nature of the bioprocess itself or 13++ cell-line-specific.

A limitation derived from the TGE functionalization of SGE cell lines stably-expressing Gag::eGFP-VLPs, is that during the period prior to the transfection, cells are already expressing non-functionalized VLPs. In this case, the study of the production kinetics allowed us to determine that 16% of the total harvested VLPs were produced before transfection. In order to solve this issue, the generation of SGE cell lines expressing Gag-VLPs under inducible promoter systems should be considered. A much more simple approach would be to perform a media exchange immediately before transfection, although this method would be more laborious and challenging.

The bioprocess harvest contained $\sim 6 \cdot 10^8$ VLPs/mL, a 5.8-fold less VLPs and 7.6-fold less S protein concentrations, compared with Chapter one. The low S protein concentrations can be attributed to the low concentration of VLPs, added to the fact that a significant part of them were non-functionalized. The low VLP production of the 13++ cell line when transfected was caused by the low VCD reached together with the short duration of the production process, which was harvested at 72 hpt to avoid the plasmid dilution phenomena. In Chapter six, when these parameters were improved, 13++ reached productions of $2.5 \cdot 10^{10}$ particles/mL, 50-fold superior than the obtained here (Chapter 7). In

this perspective, two approaches should be tested. First, to extend the length of the process, perfusion or continuous approaches using the previously described EGE methodology should be considered. Second, the low cellular densities reached after transfection (probably due to the toxic effect of PEI transfection reagent) should be increased. New protocols enabling the efficient transfection of mammalian cells at VCDs greater than $2 \cdot 10^6$ vc/mL need to be developed. In this regard, future work should be done to fully comprehend the cell density effect (CDE), a not well-known phenomena which limits the transfection of the cells to low VCDs.

Future work should also explore the use of 13++ SGE cell line without any antigen transfection for the production of Gag::eGFP-VLPs, to be subsequently functionalized by means of click chemistry or other chemical techniques for antigen display [33]. These techniques, based on the modification of chemical groups present in the surface of the VLPs to couple proteins or molecules of interest [34], are rapid, biocompatible and highly specific [35]. This would allow the generation of large amounts of Gag::eGFP-VLPs using the 13++ cell line at high VCD by perfusion or continuous processes. The generated VLPs would be purified and then functionalized in a cell-free step. Additionally this would allow the functionalization of the generated Gag::eGFP-VLPs with any S variant or antigen of interest in a cell-independent more flexible approach.

Finally, a strategy to be explored is the generation of a stable cell line co-expressing the S and Gag proteins. With this approach, high VCDs could be reached and operation elongated using fed-batch, perfusion or continuous operations. Additionally, the whole cellular population would be double-expressers, minimizing the presence of non-functionalized VLPs at the harvest which are difficult to separate during DSP. This stable cell line could be created by site-directed insertion at the AAVS1 locus using CRISPR/Cas9, the methodology used for the generation of the 13++ cell line (Chapter four) which resulted to be the best producer among all the generated SGE cell lines (Chapter six). This would be a promising but tedious and time-consuming process with the main drawback of its lack of flexibility for antigen modification in case of need. The SH5 cell line (Chapter five) can solve the presented drawbacks, as it has the potential to be targeted in order to integrate and

express any desired construct of interest via a relatively easy, economic and rapid cassette-exchange approach while it shows the same production levels as a TGE (Chapter six).

Overall, this work represents a contribution to the development of a HEK 293 based platform for the production of Gag VLPs and their functionalization with antigens of interest and opens several additional aspects to be addressed in order to continue this effort.

References

- [1] Mortola, E. & Roy, P. Efficient assembly and release of SARS coronavirus-like particles by a heterologous expression system. *FEBS Lett.* **576**, 174–178 (2004)
- [2] Lavado-García, J. *et al.* Characterization of HIV-1 virus-like particles and determination of Gag stoichiometry for different production platforms. *Biotechnol. Bioeng.* **118**, 2660–2675 (2021)
- [3] Esposito, D. *et al.* Optimizing high-yield production of SARS-CoV-2 soluble spike trimers for serology assays. *Protein Expr. Purif.* **174**, 105686 (2020)
- [4] Lavado-García, J. *et al.* Metabolic engineering of HEK293 cells to improve transient transfection and cell budding of HIV-1 virus-like particles. *Biotechnol. Bioeng.* **118**, 1649–1663 (2021)
- [5] González-Domínguez, I. *et al.* A Four-Step Purification Process for Gag VLPs: From Culture Supernatant to High-Purity Lyophilized Particles. *Vaccines* **9**, 1154 (2021)
- [6] Pallesen, J. *et al.* Immunogenicity and structures of a rationally designed prefusion MERS-CoV spike antigen. *Proc. Natl. Acad. Sci.* **114**, E7348–E7357 (2017)
- [7] Kirchdoerfer, R. N. *et al.* Stabilized coronavirus spikes are resistant to conformational changes induced by receptor recognition or proteolysis. *Sci. Rep.* **8**, 1–11 (2018)
- [8] Korber, B. *et al.* Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell* **182**, 812–827.e19 (2020)
- [9] Millet, J. K. & Whittaker, G. R. Host cell proteases: Critical determinants of coronavirus tropism and pathogenesis. *Virus Res.* **202**, 120–134 (2015)
- [10] Henderson, R. *et al.* Controlling the SARS-CoV-2 spike glycoprotein conformation. *Nat. Struct. Mol. Biol.* **27**, 925–933 (2020)

-
- [11] Lontok, E., Corse, E. & Machamer, C. E. Intracellular Targeting Signals Contribute to Localization of Coronavirus Spike Proteins near the Virus Assembly Site. *J. Virol.* **78**, 5913–5922 (2004)
 - [12] Boix-Besora, A., Lorenzo, E., Lavado-García, J., Gòdia, F. & Cervera, L. Optimization, Production, Purification and Characterization of HIV-1 GAG-Based Virus-like Particles Functionalized with SARS-CoV-2. *Vaccines* **10**, 1–18 (2022)
 - [13] Iwasaki, A. & Yang, Y. The potential danger of suboptimal antibody responses in COVID-19. *Nat. Rev. Immunol.* **20**, 339–341 (2020)
 - [14] Hacker, D. L. & Balasubramanian, S. Recombinant protein production from stable mammalian cell lines and pools. *Curr. Opin. Struct. Biol.* **38**, 129–136 (2016)
 - [15] Hansen, J. *et al.* A large-scale, gene-driven mutagenesis approach for the functional analysis of the mouse genome. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 9918–9922 (2003)
 - [16] Mutskov, V. & Felsenfeld, G. Silencing of transgene transcription precedes methylation of promoter DNA and histone H3 lysine 9. *EMBO J.* **23**, 138–149 (2004)
 - [17] Wurm, F. M. Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat. Biotechnol.* **22**, 1393–1398 (2004)
 - [18] Sadelain, M., Papapetrou, E. P. & Bushman, F. D. Safe harbours for the integration of new DNA in the human genome. *Nat. Rev. Cancer* **12**, 51–58 (2012)
 - [19] Oceguera-Yanez, F. *et al.* Engineering the AAVS1 locus for consistent and scalable transgene expression in human iPSCs and their differentiated derivatives. *Methods* **101**, 43–55 (2016)
 - [20] LaFountaine, J. S., Fathe, K. & Smyth, H. D. C. Delivery and therapeutic applications of gene editing technologies ZFNs, TALENs, and CRISPR/Cas9. *Int. J. Pharm.* **494**, 180–194 (2015)

- [21] Wirth, D. *et al.* Road to precision: recombinase-based targeting technologies for genome engineering. *Curr. Opin. Biotechnol.* **18**, 411–419 (2007)
- [22] Nehlsén, K. *et al.* Recombinant protein expression by targeting pre-selected chromosomal loci. *BMC Biotechnol.* **9**, 1–12 (2009)
- [23] Zhang, L. *et al.* Recombinase-mediated cassette exchange (RMCE) for monoclonal antibody expression in the commercially relevant CHOK1SV cell line. *Biotechnol. Prog.* **31**, 1645–1656 (2015)
- [24] García, J. L. Bioprocess engineering for HIV-1 Gag VLP production in HEK293 cells. (Universitat Autònoma de Barcelona, 2021).
- [25] González-Domínguez, I., Puente-Massaguer, E., Cervera, L. & Gòdia, F. Quality assessment of virus-like particles at single particle level: A comparative study. *Viruses* **12**, 1–23 (2020)
- [26] Cervera, L., González-Domínguez, I., Segura, M. M. & Gòdia, F. Intracellular characterization of Gag VLP production by transient transfection of HEK 293 cells. *Biotechnol. Bioeng.* **114**, 2507–2517 (2017)
- [27] Gutiérrez-Granados, S., Cervera, L., Gòdia, F., Carrillo, J. & Segura, M. M. Development and validation of a quantitation assay for fluorescently tagged HIV-1 virus-like particles. *J. Virol. Methods* **193**, 85–95 (2013)
- [28] Xu, J. *et al.* Development of an intensified fed-batch production platform with doubled titers using N-1 perfusion seed for cell culture manufacturing. *Bioresour. Bioprocess.* **7**, 17 (2020)
- [29] Cervera, L. *et al.* Generation of HIV-1 Gag VLPs by transient transfection of HEK 293 suspension cell cultures using an optimized animal-derived component free medium. *J. Biotechnol.* **166**, 152–165 (2013)
- [30] Lavado-García, J., Cervera, L. & Gòdia, F. An Alternative Perfusion Approach for the Intensification of Virus-Like Particle Production in HEK293 Cultures. *Front. Bioeng. Biotechnol.* **8**, (2020)
- [31] Cervera, L., Gutiérrez-Granados, S., Berrow, N. S., Segura, M. M. &

Gòdia, F. Extended gene expression by medium exchange and repeated transient transfection for recombinant protein production enhancement. *Biotechnol. Bioeng.* **112**, 934–946 (2015)

- [32] Pezzoli, D., Giupponi, E., Mantovani, D. & Candiani, G. Size matters for in vitro gene delivery: investigating the relationships among complexation protocol, transfection medium, size and sedimentation. *Sci. Rep.* **7**, 44134 (2017)
- [33] Moses, J. E. & Moorhouse, A. D. The growing applications of click chemistry. *Chem. Soc. Rev.* **36**, 1249–1262 (2007)
- [34] Mohsen, M. O., Zha, L., Cabral-Miranda, G. & Bachmann, M. F. Major findings and recent advances in virus–like particle (VLP)-based vaccines. *Semin. Immunol.* **34**, 123–132 (2017)
- [35] Zhang, X. & Zhang, Y. Applications of Azide-Based Bioorthogonal Click Chemistry in Glycobiology. *Molecules* **18**, 7145–7159 (2013)

Conclusions

Conclusions

1. When co-expressed together, HIV-1 Gag::eGFP and SARS-CoV-2 spike (S) proteins generate S functionalized Gag::eGFP-based VLPs (S-VLPs).
2. The co-transfection approach for the production of S-VLPs was optimized to maximize the cell population expressing both Gag::eGFP and S proteins to 57.5% by using Design of Experiments. Optimal concentrations obtained were: 0.308 µg/mL of pGag::eGFP, 1.058 µg/mL of pSpike, and 2.045 µg/mL of PEI.
3. S-VLP production was successfully performed at 1L stirred-tank bioreactor (STR), showing the same behavior, production levels and S-VLP quality as in erlenmeyer flasks.
4. A scalable and cGMP-compatible downstream process consisting of two clarification steps, an anion-exchange capture step and a size-exclusion final polishing step was implemented to successfully purify the S-VLPs harvested from STR production.
5. Three S-VLP variants harboring different rationally-engineered spike mutations to improve their stability, homogeneity, production and immunogenicity were designed, produced and purified.
6. The expression of the mutated S_{mut2} and S_{mut3} variants contributed to increase cellular viability, double-positive expressing population, VLP production (by 2.2 and 1.6-fold respectively) and VLP ratio among total extracellular particles from 8.3% to 16.2% and 14%. S_{mut3} -VLPs presented remarkably greater S functionalization levels than S_{WT} -VLPs and S_{mut2} -VLPs.

-
7. The produced S-VLP immunogens were specifically recognized by COVID-19 convalescent human sera. The S_{WT} -VLPs showed the highest recognition levels and the S_{mut3} -VLPs the lowest immunogenic potential.
 8. A SGE cell line for the constitutive production of Gag::eGFP-VLPs was generated by illegitimate integration. The best clone, named 10H9, presented a maximum cell density of $6.16 \cdot 10^6$ vc/mL, a duplication time of 29.2 h and a specific productivity of 35.7 VLPs/cell·day.
 9. A rationally-designed SGE cell line producing Gag::eGFP-VLPs was successfully generated by directed integration into the AAVS1 Genomic Safe Harbor using CRISPR/Cas9 technology. The locus-specific integration was confirmed by PCR and sequencing. The best clone, named 13++, presented a maximum cell density of $3.96 \cdot 10^6$ vc/mL, a duplication time of 34.6 h and a specific productivity of 222.8 VLPs/cell·day.
 10. An RMCE-competent SGE cell line producing Gag::eGFP-VLPs was successfully generated by lentiviral infection. The best clone in terms of productivity, cell growth and ability to perform correct cassette-exchange was named SH5. It showed a maximum cell density of 4.58 vc/mL, a duplication time of 29 h and a specific productivity of 110 VLPs/cell·day. The presence of a single-copy genomic integration of the RMCE cassette was confirmed by ddPCR.
 11. 13++ was determined to be the best SGE cell line for the production of Gag::eGFP-VLPs. When cultured in HyCell media it improved the production levels of HEK293 TGE by 2.7 fold reaching $2.48 \cdot 10^{10}$ VLPs/mL. It presented a specific productivity of 368.68 VLPs/cell·day and a volumetric productivity of $(2.5 \pm 0.3) \cdot 10^9$ VLPs/mL·day.

12. 13++ SGE cell line was transiently transfected for the production of pseudotyped S-VLPs. Operation was successfully scaled to 1 L bioreactor showing similar behavior as parallel flasks. This approach obtained 5.8-fold less VLPs and 7.6-fold less S concentration compared to standard co-transfection approaches, due to the its low VCDs and short operation time.