C3P3-G1: first generation of a eukaryotic artificial cytoplasmic expression system

Philippe H. Jaïs[1,*], Etienne Decroly [2], Eric Jacquet [3], Marine Le Boulch[1], Aurélien Jaïs[1], Olivier Jean-Jean[4], Heather Eaton[5], Prishila Ponien[3], Fréderique Verdier[6], Bruno Canard[2], Sergio Goncalves[1], Stéphane Chiron[1], Maude Le Gall[7], Patrick Mayeux[6] and Maya Shmulevitz[5]

[1]Eukarÿs SAS, Génopole Campus 3, 4 rue Pierre Fontaine, 91058 Evry Cedex, France, [2]Architecture et Fonction des Macromolécules Biologiques (AFMB) UMR 7257 CNRS/AMU, 163 Avenue de Luminy, 13288 Marseille Cedex 9, France, [3]Institut de Chimie des Substances Naturelles, CNRS UPR2301, Université Paris-Saclay, Avenue de la Terrasse, 91198 Gif-sur-Yvette, France, [4]Sorbonne Université, CNRS-UMR8256, Biological Adaptation and Ageing, Institut de Biologie Paris Seine (B2A-IBPS), F-75252 Paris, France, [5]Medical Microbiology and Immunology, University of Alberta, 6–142J Katz Group Centre for Pharmacy and Health Research, 114 Street NW, Edmonton, Alberta T6G 2E1, Canada, [6]INSERM Unit 1016, Institut Cochin, Bâtiment Gustave Roussy, 27 rue du faubourg Saint-Jacques, 75014 Paris, France and [7]Gastrointestinal and Metabolic Dysfunctions in Nutritional Pathologies, INSERM UMRS1149, 16 rue Henri Huchard, 75890 Paris Cedex 18, France

ABSTRACT

Most eukaryotic expression systems make use of host-cell nuclear transcriptional and post-transcriptional machineries. Here, we present the first generation of the chimeric cytoplasmic capping-prone phage polymerase (C3P3-G1) expression system developed by biological engineering, which generates capped and polyadenylated transcripts in host-cell cytoplasm by means of two components. First, an artificial single-unit chimeric enzyme made by fusing an mRNA capping enzyme and a DNA-dependent RNA polymerase. Second, specific DNA templates designed to operate with the C3P3-G1 enzyme, which encode for the transcripts and their artificial polyadenylation. This system, which can potentially be adapted to any *in cellulo* or *in vivo* eukaryotic expression applications, was optimized for transient expression in mammalian cells. C3P3-G1 shows promising results for protein production in Chinese Hamster Ovary (CHO-K1) cells. This work also provides avenues for enhancing the performances for next generation C3P3 systems.

INTRODUCTION

Despite their broad uses in life sciences, transient expression by plasmid-based expression systems has significant drawbacks. First, the transfer of plasmid DNA from the cytoplasm to the nucleus is a rate-limiting process in non-dividing cells. This limits efficient plasmid-based expression systems to dividing cells, in which this barrier is overcome by temporary disassembly of the nuclear membrane during mitosis (1,2). Such limited transfer to the nucleus of exogenous DNA in quiescent cells is a potential drawback for the efficacy of non-viral gene therapy and DNA vaccination. Second, plasmid-based expression depends on host cell nuclear RNA polymerase II (polII), a moderately processive enzyme with a rate of elongation of 25 and 6 nucleotides/second *in vitro* and *in cellulo*, respectively (3,4). Third, standard plasmid-based eukaryotic expression exploits the host cell transcriptional machinery and thereby competes with the host cell genome for transgene expression (4). Together, these challenges are thought to limit the efficacy of the standard eukaryotic expression systems.

To overcome these obstacles we introduce the C3P3-G1 expression system, which relies on two components: (i) an artificial chimeric enzyme having both 5′-capping and RNA synthesis activities and (ii) DNA templates that are specifically transcribed by the transduced enzyme and provide artificial polyadenylation. Once C3P3-G1 enzyme is expressed in the cytoplasm, capped and polyadenylated mRNAs are produced independently of the host cell transcription machinery. We present the first generation of the C3P3-G1 system (C3P3-G1), which was optimized for mammalian cells. This system shows promising results for transient expression in some mammalian cell lines including CHO-K1, while less efficient in other cell lines and therefore requiring further optimization.

MATERIALS AND METHODS

Plasmids

Artificial gene sequences were synthesized and assembled from stepwise PCR using oligonucleotides, cloned and fully sequence verified by GeneArt AG (Regensburg, Germany). The coding sequences of the C3P3-G1 enzymes and all other constructions were optimized for expression in human cells with respect to codon adaptation index using the GeneOptimizer algorithm (5).

The C3P3-G1 enzyme sequences were subcloned into the pCMVScript plasmid backbone (Stratagene, La Jolla, CA, USA), following the removal of the T7 φ10 promoter sequence (i.e. the promoter commonly used in E. coli exogenous expression vectors, such as the pET series). All pCMV-C3P3-G1 plasmids and derivatives had the same design: IE1 promoter/enhancer from the human cytomegalovirus (CMV), 5′-untranslated region (5′-UTR), Kozak consensus sequence followed by the ORFs of the C3P3-G1 enzymes, 3′-untranslated region (3′-UTR), and SV40 polyadenylation signal. Substitution of the coding sequence components from the C3P3-G1 enzymes was enabled by digestion at endonuclease restriction enzyme sites of six nucleotides located between each domain of the coding sequences of the C3P3-G1 enzyme (i.e. NH_2 terminus, upstream and downstream to the linker) and eight nucleotides immediately downstream to stop codon. The C3P3-G1 enzyme plasmids are identified by the different modules in the ORF of the C3P3-G1 enzyme downstream to the IE1 human CMV promoter/enhancer. For instance, NP868R-$(G_4S)_4$-K1ERNAP(R551S) designates the fusion of the NP868R capping enzyme with the mutant R551S K1E RNA polymerase through a $(G_4S)_4$ flexible linker.

Plasmids containing the Firefly Luciferase gene were used to optimize C3P3-G1 enzymes and DNA template sequences. They consist of a phage RNA polymerase promoter, variable 5′-UTR, Kozak consensus sequence followed by the ORF of the Luciferase gene from *Photinus pyralis* and stop codon, variable 3′-UTR, poly[A] track that was routinely of 40 adenosine residues, followed by a self-cleavage RNA sequence that was generally the genomic ribozyme sequence from the hepatitis D virus, and terminated by the bacteriophage T7 φ10 transcription stop. Restriction enzymatic sites were inserted between each motif of the luciferase plasmids to allow easy swapping of each motif by subcloning. The plasmids are identified by the corresponding ORF (e.g. Luciferase) preceded by the phage promoter (e.g. pT7φ10-Luciferase).

Plasmids used for comparison with the standard transient expression system consisted of the ORF under consideration subcloned in the commercial pCMVScript plasmid, e.g. pCMVScript-Luciferase. The resulting construction therefore contained the IE1 human CMV promoter/enhancer, Kozak consensus sequence followed by the ORF, and late SV40 polyadenylation signal.

Cell culture and transfection

For standard experiments, the Human Embryonic Kidney 293 (HEK-293, ATCC CRL 1573) and Chinese Hamster Ovary K1 (CHO-K1, ATCC CCL-61) were routinely grown at 37°C in 5% CO_2 atmosphere at 100% relative humidity. Cells were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 4 mM L-alanyl-L-glutamine, 10% fetal bovine serum (FBS), 1% non-essential amino-acids, 1% sodium pyruvate, 1% penicillin and streptomycin and 0.25% fungizone.

Cells were routinely plated in 24-well plates at 1×10^5 cells per well the day before transfection and transfected at 80% cell confluence. Transient transfection was performed with Lipofectamine 2000 reagent (Invitrogen, Carlsbad, CA, USA) according to manufacturer's recommendations, except when otherwise stated. For standard luciferase and hSEAP gene reporter expression assays, cells were analyzed 24 h after transfection.

Firefly luciferase and eSEAP gene reporter assays

Luciferase luminescence was assayed by the Luciferase Assay System (Promega, Madison, WI, USA) according to the manufacturer's recommendations. In brief, cells were lysed in Cell Culture Lysis Reagent buffer (CLR), and then centrifuged at 12 000 × g for 2 min at 4°C. Luciferase Assay Reagent (Promega; 100 μl/well) diluted at 1:10 for HEK-293 cells and 1:50 for CHO-K1 cells was added to supernatant (20 μl/well). Luminescence readout was taken on a Tristar 2 microplate reader (Berthold, Bad Wildbad, Germany) with a read time of one second per well for HEK-293 cells and 0.1 s for CHO-K1 cells.

In order to normalize for transfection efficacy, cells were transfected with the pORF-eSEAP plasmid (InvivoGen, San Diego, CA, USA), which encodes for the human secreted embryonic alkaline phosphatase driven by the EF-1α/HTLV composite promoter. Enzymatic activity was assayed in cell culture medium using the Quanti-Blue colorimetric enzyme assay kit (InvivoGen). Gene reporter expression was expressed as the ratio of luciferase luminescence (RLU; relative light units) to eSEAP absorbance (OD, optic density).

Semi-quantitative assessment of mRNA capping rate by tethered capping enzyme assay

For the semi-quantitative assessment of mRNA capping efficiency, we took advantage of the λ-phage N protein-boxB RNA interaction, which normally regulates antitermination during transcription of λ-phage mRNAs (6). The short N-terminal peptide of the λN protein mediates its binding to the 17 nucleotides boxB RNA hairpins at nanomolar affinity (7). The λN peptide was fused the N-terminus of the NP868R African swine fever virus capping enzyme, resulting in a tethered capping enzyme (i.e. pCMV-λN-NP868R), while four BoxBr hairpins were introduced to the 3′UTR of the Firefly Luciferase gene (i.e. pT7φ10-Luciferase-4xBoxBr). The effects of this tethered capping system were tested on C3P3-G1 transcripts, together with various controls. HEK-293 cells were transfected as described above with the appropriate combination of plasmid using an empty dummy plasmid to transfect the same amount of DNA to all conditions. Luciferase reporter expression was monitored by conventional luciferin oxidation assays and normalized by hSEAP expression as described above.

NP868R protein production

The full-length ORF from the NP868R capping enzyme was optimized for codon usage in Sf9 cells (*Spodoptera frugiperda* 9). The resulting sequence was synthesized into the F8 donor plasmid (GenScript, Piscataway, NJ), then transferred into the Bacmid shuttle DNA through site-specific recombination using the *Escherichia coli* DH10Bac strain (Invitrogen). The recombinant Bacmid DNA was transfected into Sf9 cells and incubated in serum-free culture medium Sf-900 II SFM (Life Technologies) for 5 days at 27°C before harvest of the budded virus particles released into the medium. The recombinant baculovirus was further amplified by propagation for three days post-infection in Sf9 cells. Cells were then lysed by sonication in presence of protease inhibitor. The HISx10-EK-NP868R recombinant protein was purified from cell extracts on nickel-nitrilotriacetic acid columns using a standard imidazole stepwise elution protocol. The samples were analyzed by SDS-PAGE and western-blot using an anti-His antibody, then stored at –20°C in 50% glycerol buffer. The yields of the baculovirus/Sf9 expression assay were of 4 mg/l of cell culture with a 72% purity rate, as estimated by Coomassie gel staining.

NP868R biochemical enzymatic assay

NTPase assay. NP868R recombinant protein (0.5 μM) was incubated with 0.5 μCi [α-^{32}P]-GTP or [α-^{32}P]-ATP in 50 mM Tris–HCl (pH 8.5), 5 mM DTT, 5 mM KCl supplemented or not with 2.5 mM MgCl$_2$ and incubated at 37°C. Reaction was stopped by adding an equal volume of formamide/EDTA gel-loading buffer, and hydrolysis products were separated over a 20% polyacrylamide/8 M urea gel prior to phosphorimaging using FLA3000 instrument (Fuji, Tokyo, Japan).

GTase assay. The pppAC$_4$ RNA was synthetized using T7 DNA primase as described by Peyrane *et al.* (8). The RNA (pppAC$_4$) were then radiolabelled at their 3′ end by ligation of [α-^{32}P]-pCp using T4 RNA ligase (New England Biolabs, Ipswich, MA, USA) according to the manufacturer's instructions (9). NP868R recombinant protein (0.5 μM) was incubated 1 h at 30°C with 10 mM GTP and 5 μM pppAC4pCp* in 50 mM Tris–HCl (pH 8.5), 5 mM DTT, 5 mM KCl supplemented or not with 2.5 mM MgCl$_2$. After incubation, the reactions were stopped by addition of an equal volume of formamide/EDTA gel-loading buffer and hydrolysis products were separated over a 20% polyacrylamide/8 M urea gel, prior to phosphorimaging.

MTase assays. MTase activity assays were performed by combining 0.5 μM of enzymes with 1 μM of GpppAC$_4$ or mGpppAC$_4$ RNAs, 2 μM SAM and 0.33 μM [methyl-^{3}H]-SAM (Perkin Elmer, Waltham, MA, USA) in 40 mM Tris–HCl (pH 8.5), 1 mM DTT. Reactions at 30°C were stopped by a 10-fold dilution in 100 μM ice-cold SAH and the samples transferred to DEAE (Perkin Elmer) using a Filter-mats Harvester (Packard Instruments, Downers Grove, IL, USA). The RNA-retaining mats were washed twice with 10 mM ammonium formate (pH 8.0), twice with water, and once with ethanol. They were then soaked with liquid scintillation fluid, allowing the measurement of [methyl-^{3}H] transfer to the RNA substrates using a Wallac MicroBeta TriLux Liquid Scintillation Counter (Perkin Elmer).

Thin-layer chromatography (TLC) analysis of cap structures. small RNA pppAC$_4$ (10 μM) was incubated 1 h at 30°C with recombinant NP868R protein (0.5 μM) or Vaccina virus capping Enzyme (New England Biolabs) in the presence of 5 μCi [α-^{32}P]-GTP (Perkin Elmer). The capped RNA was purified by precipitation in 3 M sodium acetate supplemented with 1 μg/μl of glycogen (Thermo Scientific, Waltham, MA), and digested with 1 U of Nuclease P1 (US Biologicals, Salem, MA, USA) in 30 mM sodium acetate (pH 5.3), 5 mM ZnCl$_2$ and 50 mM NaCl for 4 h at 37°C. The products were spotted onto polyethylenimine cellulose TLC plates (Macherey Nagel, Düren, Germany), and resolved using 0.65 M LiCl as mobile phase. The radio-labeled caps released by nuclease P1 were visualized using a Fluorescent Image Analyzer FLA3000 phosphorimager.

Western-blot analysis

For C3P3 immunoblotting, HEK-293 and CHO-K1 cells were transfected with the pCMV-FLAGx3-NP868R-(G$_4$S)$_2$-K1ERNAP(R551S) plasmid, which encodes for the FLAGx3 tag fused in frame to the amino-terminal ends of the NP868R-(G$_4$S)$_2$-K1ERNAP(R551S) C3P3 enzyme. Cells were lysed in 200 μl of CLR buffer and lysate was clarified by spinning for 15 s at 12 000 × g at room temperature. Twenty milligrams of total protein were resolved on 4–12% NuPAGE SDS-polyacrylamide gradient gel (Life Technologies, Carlsbad, CA, USA), and subjected to western blotting onto nitrocellulose Hybond membrane (GE Healthcare, Pittsburgh, PA, USA) overnight at +4°C.

Membranes with transferred proteins were blocked with 5% skim milk powder in PBS, then incubated with the rabbit polyclonal F7425 anti-FLAG primary antibody (1:1000; Sigma-Aldrich, Saint-Louis, MO, USA), then with anti-rabbit IgG-conjugated horseradish peroxidase NA9340V antibody (1:10000; GE Healthcare). Bands were visualized using the SuperSignal West Pico Chemiluminescent Substrate solution (Thermo Scientific) and scanned with the Fusion XPRESS gel imager (Vilber Lourmat, Marne-la-Vallée, France). Molecular weights were determined using the Novex Sharp Pre-stained Protein Standard color markers (Thermo Fisher).

Cell viability and death assay

To investigate the effects of C3P3-G1 enzyme expression on viability and cytotoxicity, HEK-293 cells were transfected with the pCMV-NP868R-(G$_4$S)$_2$-K1ERNAP(R551S) plasmid with or without the pK1E(G)-Luciferase gene reporter plasmid. Empty backbone plasmids were used as controls. Cells were analyzed at selected time points using the MultiTox-Glo Multiplex Cytotoxicity Assay (Promega), a sequential-reagent-addition fluorescent and luminescent assay that measures the relative number of live and dead cells in cell populations.

Cell viability was assayed according to the manufacturer's instructions after addition of 50 μl of the live-cell dipeptidyl

peptidase-1 GF-AFC reagent and analyzed on a fluorescent plate reader using wavelengths of 400 and 505 nm for excitation and emission, respectively. The GF-AFC reagent enters intact cells, where it is cleaved by the dipeptidyl peptidase-1, an enzyme that is restricted to intact viable cells, which release AFC and generate a fluorescent signal that is proportional to the number of viable cells (10). This live-cell protease becomes inactive upon loss of membrane integrity and leakage into the surrounding culture medium.

Cell death, either necrosis or apoptosis, was assayed by addition of 50 μl of a second luminogenic cell-impermeant peptide AAF-aminoluciferin substrate, and then analyzed on a luminescence plate reader. The AAF-aminoluciferin is a peptide substrate for cellular proteases which are released from compromised cells (11). Released proteases liberate aminoluciferin that is measured as luminescence generated by a recombinant Luciferase, whereas uncleaved AAF-aminoluciferin is not a substrate for recombinant luciferase, so viable cells generate a modest luminescence background. The live/dead cell ratio was estimated as the GF-AFC to AAF-aminoluciferin signals.

Confocal laser fluorescence protein microscopy

The subcellular localization of the wild-type T7RNAP was investigated by indirect immunofluorescence. Cells were plated on poly-L-lysine coated coverslips and cotransfected as described above with both pCMV-T7RNAP (encodes the wild-type ORF of T7RNAP under control of the CMV promoter) and pCMVScript-EGFP used as a cytoplasmic control. Cells were also transfected with a plasmid containing the T7RNAP ORF placed downstream to NLS from SV40 large T-antigen (i.e. pCMV-NLS-T7RNAP). Transfected cells were fixed in 4% paraformaldehyde in PBS for 20 min at room temperature. Cells were permeabilized and non-specific binding was blocked by 30 min treatment in PBS with 5% goat serum, 0.1% Triton X-100 and 0.02% sodium azide. The coverslips were incubated overnight at 4°C with the mouse monoclonal IgG$_1$ anti-T7RNAP (1:200; Novagen, Madison, WI), washed with PBS, incubated with secondary Alexa Fluor 555 goat anti–mouse IgG antibody (1:500; Life Technologies) and mounted in the anti-fade Vectashield Mounting Medium (Vector Laboratories, Burlingame, CA, USA).

The tagged C3P3-G1 enzyme was imaged by indirect immunofluorescence. CHO-K1 cells were transfected with the pCMV-FLAGx3-NP868R-(G$_4$S)$_2$-K1ERNAP(R551S), which encodes for the NP868R-(G$_4$S)$_2$-K1ERNAP(R551S) C3P3 enzyme fused in-frame to the C-terminus of the FLAGx3 tag (12). CHO-K1 cells were then stained as above, using a primary polyclonal F7425 rabbit IgG anti-FLAGx3 antibody (1:250; Sigma) (13) followed by secondary Alexa Fluor 568-conjugated goat anti-rabbit (1:800; Life Technologies). Cells transfected with FLAG-RLuc plasmid and cells stained only with secondary antibodies were used as controls. Slides were imaged on a Leica SP8 confocal microscope equipped with the appropriate laser excitation filters under oil immersion (i.e. ×40 lens) with image magnification and processed with Leica LAS-AF and ImageJ analysis software.

To determine whether proteins in different subcellular compartments can be expressed by the C3P3-G1 technology, cells were plated and transfected as described above with plasmids encoding the green EGFP or the red RFP fluorescent proteins fused in-frame with mitochondrial, nuclear or endosomal subcellular signals and imaged by direct fluorescence (Supplementary Table S1). The human transferrin receptor type I (hTFR1) N-terminal signal localizes the fusion protein to the surface of vesicles of the endosomes (14). The mitochondrial presequence of *S. cerevisiae* cytochrome c oxidase subunit 4 (COX4) targets fusion proteins to the inner membrane of the mitochondrial matrix (15). The eNLS is a strong nuclear signal from SV40 for nuclear subcellular localization (16). CHO-K1 cells were co-transfected with pCMV-NP868R-(G$_4$S)$_2$-K1ERNAP(R551S) and the plasmids encoding the different fluorescent protein under control of the K1E promoter. For comparison with standard nuclear expression system, cells were also transfected with plasmids containing the same coding sequences in pCMV-Script backbone. To confirm the localization to the endosomal and mitochondrial compartments of the RFP fluorescent proteins expressed by the C3P3-G1 system, we have investigated their colocalization with EGFP tagged with the human LDLR and rat OCT signal peptides expressed with pCMV-Script plasmids, respectively. Transfected cells were fixed in 4% paraformaldehyde for 20 min at room temperature, then washed with 50 mM glycine solution in PBS for 10 min and permeabilized for 10 min in 0.1% Triton X-100. Cell nuclei were stained with Hoechst 33342 for 15 min, mounted in the anti-fade Vectashield Mounting Medium (Vector Laboratories), and imaged by direct fluorescence as described above.

Production of secreted protein assayed by ELISA

The production yields of secreted or cytoplasmic/secreted proteins were assayed in cell culture medium of CHO-K1 cells. To avoid cross-reactivity in ELISA assay measurements, CHO-K1 cells were cultured and transfected in Panserin PX10 serum-free medium (Pan-Biotech, Aidenbach, Germany), supplemented with 1% penicillin and streptomycin, and 0.25% fungizone. Wild-type ORFs encoding protein precursors were artificially synthesized, i.e. human erythropoietin, human granulocyte colony-stimulating factor and murine alpha-fetoprotein, and subcloned in cassette optimized for the C3P3-G1 enzyme or standard pCMVScript nuclear expression plasmids (Supplementary Table S2).

The concentrations of released hEPO, hG-CSF and mAFP were measured in culture medium collected every single day. Protein concentrations were determined by enzyme-linked immunosorbent assay (ELISA) using commercial kits (Quantikine ELISA, R&D Systems, Minneapolis, MN, USA) according to the manufacturer's instructions. Results were expressed as the optical density at 450 nm, with wavelength correction set to 620 nm. Results were analyzed by 4-parameter nonlinear logistic regression algorithm (17).

Proliferation assay of the erythropoietin-dependent UT7 cell line

To investigate whether the post-translational maturation of proteins produced with the C3P3-G1 expression system in CHO-K1 cells was altered, the functional activity of the human EPO (hEPO) glycoprotein produced with the C3P3-G1 system was assessed. CHO-K1 cells were plated in 24-well plates at a density of 1×10^5 cells per well. To avoid cross-contamination by the cytokines present in the fetal calf serum, CHO-K1 cells were cultured in Panserin X10 serum-free medium. Cells were cotransfected with pCMV-NP868R-$(G_4S)_2$-K1ERNAP(R551S) and pK1E(G)-hEPO plasmids (encoding for the wild-type human erythropoietin precursor gene under control of the K1E phage promoter) using the Lipofectamine 2000 reagent. As a comparator, cells also transfected with the pCMVScript-hEPO plasmid (ORF of the wild-type hEPO precursor under control of the human IE1 CMV promoter/enhancer). Culture medium from CHO-K1 cells was removed on the second day after transfection, and erythropoietin titers were assessed by ELISA as described above.

Human erythropoietin in the culture media was assayed for effects on cell proliferation with the human UT7 leukemic cell line, which has megakaryocytic features and strictly depends on erythropoietin, GM-CSF, or IL3 for its growth [18]. UT7 cells were cultured as described elsewhere with slight modifications [19] in Alpha Modified Minimum Essential Medium Eagle supplemented with 10% fetal calf serum and 2 U/ml hEPO. Before each experiment performed in triplicate, cells were washed, and then serum- and growth factor-deprived by incubation for 24 h in Iscove's modified Dulbecco's medium carrying 0.4% bovine serum albumin. Then, serial two-fold dilutions (from 4 to 0.008 U/ml) of hEPO from the culture medium of transfected CHO-K1 cells or treated with recombinant hEPO (Epoetin β, Roche Laboratories) were added to the culture medium of UT-7cells. Cells were grown in 96-well plates for three days at $37°C$ in $5\% CO_2$ atmosphere at 100% relative humidity. Cell proliferation was then assayed by addition of the Uptiblue reagent (Interchim, Montluçon, France), which is a viable cell counting reagent. Fluorescence intensity was read using a scanner (Typhoon, GE Healthcare), with excitation at 532 nm and emission at 580 nm. Human EPO released in culture medium was compared with the commercial Epoetin β (Roche Pharmaceuticals, Basel, Switzerland). Differences of activity profiles between the commercial Epoetin β and hEPO produced by CHO-K1 cells might be explained by the use of distinct methods for measuring hEPO or other differences related with the processing of Epoetin β [20].

Real-time quantitative Reverse Transcription PCR

The kinetics of mRNA synthesis by the C3P3-G1 expression system was assessed by quantitative RT-PCR. HEK-293 cells were plated in 24-well plates at a density of 1×10^5 cells per well and transfected with either pCMV-NP868R-$(G_4S)_2$-K1ERNAP(R551S) plus the pK1E(G)-Luciferase plasmid, or the pCMVScript-Luciferase plasmid (Luciferase gene under control of the human CMV promoter/enhancer) using the Lipofectamine 2000 reagent.

Total RNA was isolated using the Nucleospin RNA columns (Macherey-Nagel, Düren, Germany) and subjected to TURBO DNase-free treatment (Ambion, Foster City, CA, USA). Total RNA was reverse-transcribed using High Capacity cDNA Reverse Transcription kit with RNase Inhibitor (Life Technologies). The resulting cDNA was amplified by real-time RT-PCR using primer and probe sets for either Luciferase ORF or C3P3-G1 enzyme (Supplementary Table S3) with TaqMan detection. Absolute transcript copy number was reported to calibration curve with known concentrations of each plasmid DNA and normalized to human GAPDH or ACTB expression.

Polysome fractionation

Polysome fractionation was performed as described elsewhere with minor modifications [21]. A single 75 cm^2 tissue culture flask of HEK-293 transfected cells was used for each sucrose gradient. The culture medium was removed 24 h after transfection and replaced with fresh medium. After overnight incubation, the medium was changed again and 2 h later, cycloheximide at 100 μg/ml was added for 10 min. Cells were washed with PBS, collected by trypsinization, and pelleted. The dry cell pellet was resuspended in 500 μl of lysis buffer (50 mM Tris-HCl at pH 7.4, 300 mM KCl, 10 mM Mg-Acetate, 1 mM DTT, 0.05% Nonidet P40) containing 200 units/ml of SUPERaseIn RNAse inhibitor (Invitrogen) and 100 μg/ml of cycloheximide, and lyzed by incubation on ice for 10 min with occasional shaking. Nuclei and cell debris were removed by centrifugation at 1000 × g for 10 min and 400 μl of supernatant was layered directly onto a 12 ml 15–50% (w/v) sucrose gradient in 50 mM Tris-acetate (pH 7.5), 50 mM NH_4Cl, 12 mM $MgCl_2$ and 1 mM DTT. The gradient was centrifuged at 39 000 rpm in a SW41 Beckman rotor for 2.75 h at $4°C$. After centrifugation, optical density (O.D.) at 254 nm was monitored by pumping the gradient through a Retriever 500 (Teledyne Isco) fraction collector.

Statistics

Statistical analyses were performed with paired two-tailed Student's *t*-test. Results are means ($n \geq 4$, except otherwise stated) $\pm$ standard deviation. *P*-value <0.05 was considered statistically significant.

RESULTS

Poor translation from T7RNAP-directed transcripts is overcome by co-expression of the Vaccinia Virus capping enzyme

Autographivirinae is a family of bacteriophages encoding their own single-subunit RNAPs showing high specificity for their respective double-stranded promoters. The T7 DNA-dependent RNA polymerase (T7RNAP), their prototype, synthesizes RNA *in vitro* at a rate of 250 nucleotides/second and produces transcripts devoid of 5′-caps [22]. Given the importance of capping for mRNA stability and translation initiation [23], we sought to determine if the translational efficiency of T7-transcripts by mammalian cells could be increased by mRNA capping. To investigate protein expression from T7 promoter-driven transcripts, the wild-type T7RNAP open reading frame (ORF)
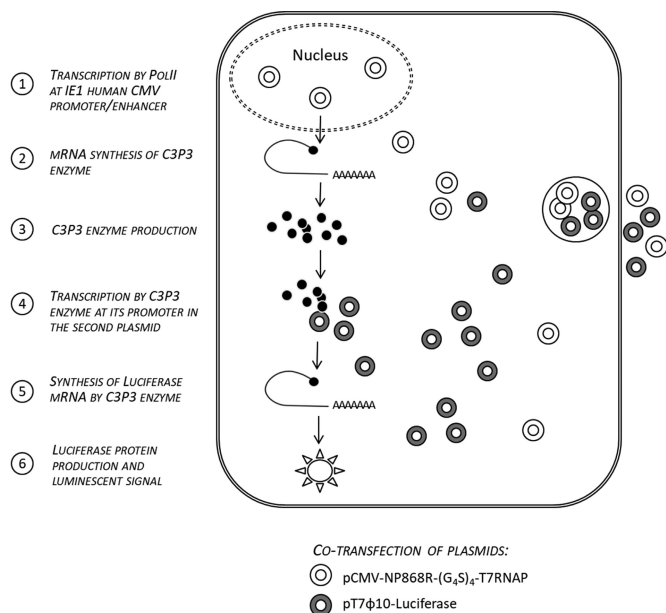
Figure 1. The Firefly Luciferase gene reporter assay used for the optimization of the C3P3-G1 system. A diagram depicts the steps of gene expression from the C3P3-G1 expression system, using Firefly Luciferase as expression gene reporter. C3P3-G1 enzyme and T7 promoter-Luciferase plasmids (pT7φ10-Luciferase) are co-transfected into HEK-293 or CHO-K1 cultured cells. C3P3-G1 mRNA are expressed by the RNA polymerase II-dependent IE1 human CMV promoter/enhancer. The transcripts are subsequently translated into C3P3-G1 protein, which accumulates in the cell cytoplasm where it mediates transcription and capping of luciferase mRNAs from the pT7φ10-Luciferase plasmid, while polyadenylation is encoded in a poly[A]-track from the pT7φ10-Luciferase plasmid. Ultimately, luciferase protein is produced by mRNA translation, which is monitored by luciferin oxidation assay.

was subcloned downstream of the CMV promoter (pCMV-T7RNAP). When expressed in HEK-293 cells, T7RNAP was found in cytoplasm, whereas in-frame fusion of a SV40 nuclear localization signal at the amino-terminus redirected T7RNAP to the nucleus (Supplementary Figure S1) (24). A reporter plasmid was generated by cloning the Firefly luciferase gene under control of the φ10 T7-phage promoter from the major capsid protein (pT7φ10-Luciferase) (25). This reporter plasmid also contains a 40 adenosine track to ensure artificial polyadenylation, followed by the self-cleaving genomic hepatitis D ribozyme (26). Following co-transfection of pCMV-T7RNAP and pT7φ10-Luciferase plasmids in HEK-293 cells, luciferase expression was monitored by conventional luciferin oxidation assays and normalized (Figure 1). As expected, low levels of luciferase expression were seen in transfected HEK-293 cells (Figure 2).

The lack of 5′-RNA cap on T7-transcripts likely contributes to their poor translational efficiency (24,27). In support of this hypothesis, the vaccinia virus-T7RNAP (VV-T7RNAP) expression system was found superior to T7RNAP expression alone (28,29). This elegant expression system relies on infection of mammalian cells by a recombinant vaccinia virus encoding T7RNAP, which transcribes target genes under control of the T7-promoter embedded in a second recombinant vaccinia virus or plasmid. The transcripts produced by this system can receive 5′-cap

and 3′-polyadenylation by the vaccinia virus capping and polyadenylation machineries, respectively (30). Although this expression system is able to generate up to 30% of the total cytoplasmic RNA, only 5–10% of the T7-transcripts are capped and 25% of them are polyadenylated, which suggests insufficient synergy between the transcriptional and post-transcriptional machineries (30,31).

We sought to recapitulate the main components of the vaccinia virus-T7RNAP expression system in a non-viral plasmid-based expression system, therefore devoid of any risk of transmitting infection and virus-mediated cytotoxicity. The vaccinia virus capping enzyme consists of a 97- and a 33-kDa subunit that are encoded by the vaccinia virus-D1R and D12L genes respectively (32). The D1R product has enzymatic activities (33–35), whereas the D12L product acts as its allosteric regulator (36). Both D1R and D12L were cloned into independent plasmids and co-transfected into HEK-293 cells in combination with pCMV-T7RNAP and pT7φ10-Luciferase plasmids. Luciferase expression was increased by 9-fold in cells co-transfected with D1R and D12L plasmids in comparison to T7RNAP plasmid alone, which confirmed the importance of capping for T7-transcripts translational efficiency (Figure 2).

Coupling of T7RNAP and vaccinia virus capping enzyme enhances reporter gene expression

In Nature, eukaryotic RNAPs are commonly coupled with their corresponding capping enzymes, which is likely required for selective capping of transcripts. In contrast, the absence of coupling between the T7RNAP and capping enzyme of the hybrid vaccinia virus-T7RNAP expression system possibly accounts for the poor capping rate of T7-transcripts (30). Hence, we chose to increase the proximity of T7RNAP and vaccinia virus capping enzyme by directly coupling these enzymes together. This approach therefore differs from the strategy used by others (37–39), in which nuclear-directed-T7RNAP was fused with the carboxyl terminal domain from POLR2A (40), which interacts with the host-cell nuclear capping enzyme (41). Although this fusion was sought to bring these enzymes together, it generates neither capped nor spliced transcripts but instead promotes formation of DNA:RNA hybrids (37,38).

T7RNAP and vaccinia virus capping enzyme were coupled using $EE_{1234}L$ and $RR_{1234}L$ leucine-zippers. These amphipathic α-helices form an antiparallel heterodimer with binding affinity of $\sim 10^{-15}$M (42). $RR_{1234}L$ was fused in-frame to the amino-terminal ends of either D1R or D12L vaccinia virus capping enzyme subunits, while $EE_{1234}L$ was fused in-frame to the amino-terminal end of T7RNAP. Two heterotrimeric enzymes were therefore generated by co-transfection: $EE_{1234}L$-T7RNAP that binds to $RR_{1234}L$-D1R and D12L duplex, as well as $EE_{1234}L$-T7RNAP that binds to $RR_{1234}L$-D12L and D1R duplex. The constructs were transfected in HEK-293 cells and assayed for luciferase expression. In comparison to uncoupled enzymes, leucine-zippers on D1R or D12L increased luciferase expression by 7.6- and 5.1-fold respectively (Figure 2).

As an alternative to leucine-zippers, we tested the functional activity of fusing vaccinia virus capping enzyme subunits to T7RNAP. The D1R or D12L ORFs were
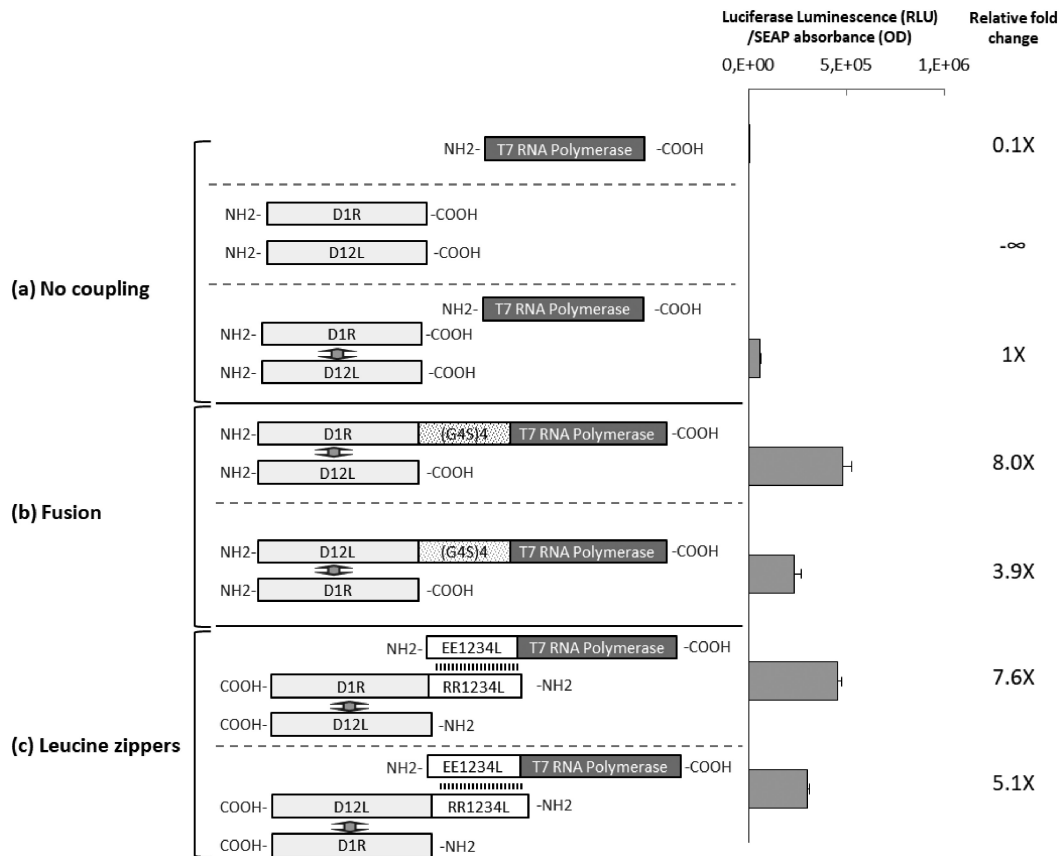
Figure 2. Coupling of T7RNAP with vaccinia virus capping enzyme. Plasmids encoding T7RNAP and/or the D1R and D12L subunits of the heterodimeric vaccinia virus capping enzyme were cotransfected in HEK-293 cells together with the pT7φ10-Luciferase plasmid. The ORFs of corresponding plasmids were designed as follows: (**A**) no coupling domain, (**B**) T7RNAP fused in-frame to the carboxyl-terminal end of either D1R or D12L genes via the $(G_4S)_4$ linker, (**C**) fused in-frame with the leucine-zippers $EE_{1234}L$ and $RR_{1234}L$, which forms a heterodimer in antiparallel orientation. $EE_{1234}L$ was fused to the amino-terminal end of T7RNAP, whereas $RR_{1234}L$ was fused either with D1R or D12L. Luciferase gene expression level is shown, as well as the relative value in comparison to the uncoupled condition. Luminescence was assayed in cell lysates 24 h after transfection and expressed in RLU normalized for hSEAP (secreted embryonic alkaline phosphatase) expression expressed in DO (RLU/DO ratio). Errors bars indicate standard deviation of at least four experiments. Comparisons of cotransfection of T7RNAP, D1R and D12L versus all other constructions: $P < 0.05$, Student's t-test.

fused in-frame to the amino-terminus of T7RNAP, separated by a flexible $(G_4S)_4$ linker, generating D1R-$(G_4S)_4$-T7RNAP and D12L-$(G_4S)_4$-T7RNAP constructs respectively. In comparison to uncoupled enzymes, fusion increased luciferase expression by 8- and 3.9-fold respectively (Figure 2). In both approaches, coupling of D1R to T7RNAP outperformed coupling of D12L to RNAP and confirmed the importance of physically coupling capping enzyme and RNAP enzymes for T7-transcript expression.

Fusion of the single-unit capping enzyme from the African Swine Fever Virus (ASFV) with T7RNAP promotes high cytoplasmic protein expression

In contrast to the two-subunit vaccinia virus capping enzyme, some viruses and episomes encode single-subunit capping enzymes. With the rationale that single-subunit capping enzymes would be simpler to use in routine expression systems, we next tested the activity of distinct single-subunit capping enzymes. Subsequently, artificial enzymes, collectively designated as C3P3-G1 enzymes, were generated by in-frame fusion of candidate capping enzymes to the

amino-terminal end of T7RNAP via $(G_4S)_4$ linkers (Figure 3A).

Nine candidate single-unit capping enzymes were selected for analysis (Supplementary Table S4, Supplementary Figure S2), which were inferred by bioinformatic annotation to comprise the three enzymatic domains required for mRNA cap synthesis (Figure 3B): a 5'-triphosphatase (RTPase) that removes the γ-phosphate residue of 5'-triphosphate mRNA ends, a guanylyltransferase (GTase) that transfers GMP from GTP to diphosphate 5'-terminus, and a N7-guanine methyltransferase (N7-MTase) that adds a methyl residue onto nitrogen-7 of guanine to produce ^{m7}GpppN caps [43]. For five of these capping enzymes, at least one of these three enzymatic activities were also previously confirmed biochemically (Bluetongue virus VP4 [44–46], Mimivirus R382 [47], ASFV NP868R [48], ORF3 from the yeast cytoplasmic episome pGKL2 [49,50], Bamboo Mosaic virus ORF1 [51,52]). The remaining four capping enzymes were selected based on having >20% amino-acid identity with NP868R and all three enzymatic capping domains inferred by electronic annotation.
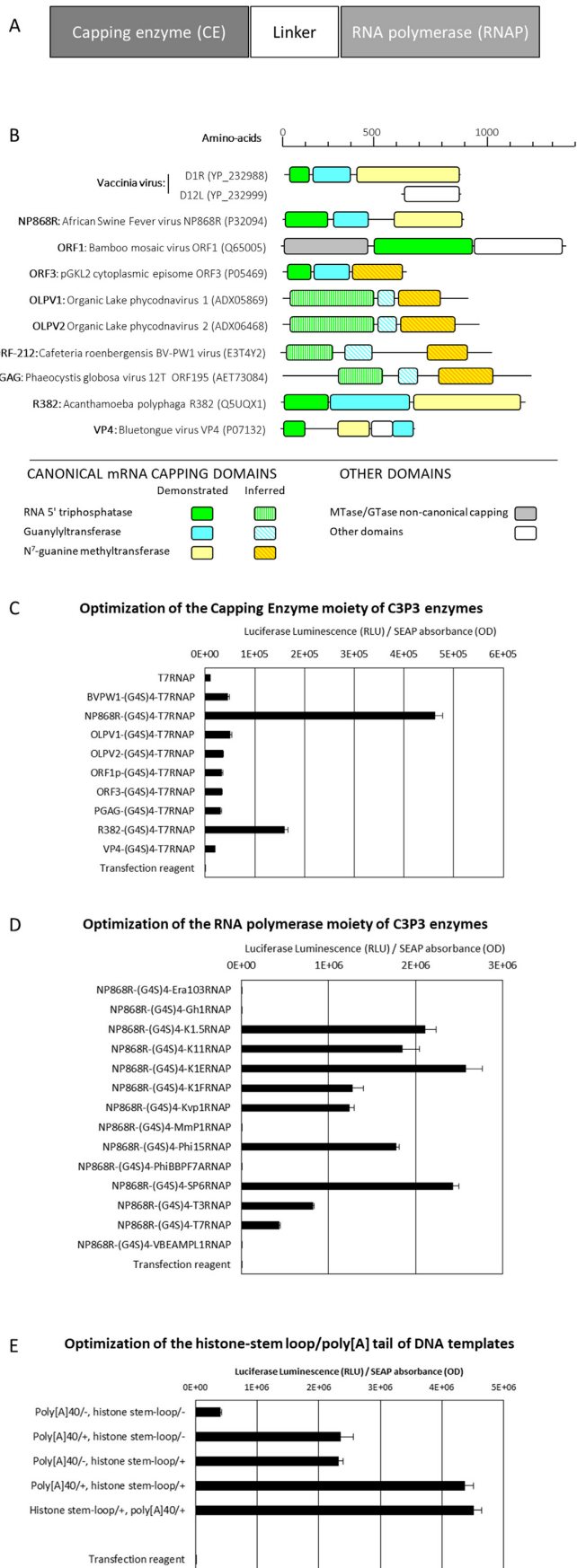
The activity of each single-subunit C3P3-G1 enzyme was tested following co-transfection with pT7φ10-Luciferase reporter plasmid (25). High expression was obtained with NP868R-(G4S)4-T7RNAP and to a lesser extent R382-(G4S)4-T7RNAP, relative to T7RNAP alone (Figure 3C). Coupling of NP868R with T7RNAP by leucine-zippers also increased luciferase expression in comparison to uncoupled enzymes (Supplementary Figure S3). Similar results were obtained when the T7RNAP was substituted by T3 or SP6RNAP (Supplementary Figure S4). Both NP868R-(G4S)4-T7RNAP and R382-(G4S)4-T7RNAP were also active, albeit at reduced efficiency, when the φ10 promoter was substituted by the low-activity φ2.5 promoter that initiates transcription with an adenosine (53) (Supplementary Figure S5). Together, these results show that ASFV NP868R can significantly enhance protein expression when coupled to various phage RNAPs.

To confirm that luciferase expression from NP868R-(G4S)4-T7RNAP depends on T7RNAP activity, transfected cells were exposed to α-amanitin, which inhibits polII but not phage RNAPs (29). As expected, expression by NP868R-(G4S)4-T7RNAP was relatively resistant to α-amanitin, in contrast to nuclear-based expression plasmid, which was effectively inhibited (Supplementary Figure S6).

consists of three main domains (i.e. capping enzyme, linker and DNA-dependent RNA polymerase) which were optimized step-by-step in HEK-293 cells (similar results were obtained in CHO-K1 cells, data not shown). (**B**) Enzymatic domains of the candidate capping enzymes selected to generate C3P3-G1 enzymes. Capping enzymes were selected by bioinformatic sequence analysis showing that the proteins contain the three-enzymatic domains required for mRNA capping (i.e. TPase, GTase and N7-MTase). Blocks stand for different domains and their annotations are listed at the bottom of the figure. The lengths of the blocks are directly proportional to their sequence length. The enzymatic domains of the well-characterized heterodimeric vaccinia virus capping enzyme are also shown. (C) Selection of the mRNA capping moieties of C3P3-G1 enzymes. mRNA capping moieties were fused in-frame to the amino-terminal extremity of the T7RNAP sequence via (G4S)4 linker. Reverse constructions by in-frame fusion to the carboxyl-terminus of T7RNAP were not tested because T7RNAP does not tolerate carboxyl-terminal extensions (85,86). Constructions were tested for firefly luciferase reporter expression assay 24 h after transfection and expressed in RLU normalized for hSEAP expression expressed in DO (RLU/DO ratio). The errors bars indicate the standard deviation of at least four experiments. Comparisons of NP868R-(G4S)4-T7RNAP versus all other C3P3-G1 enzymes: $P < 0.05$, Student's t-test. (**D**) Selection of the phage DNA-dependent RNAPs moieties of C3P3-G1 enzymes. Phage DNA-dependent RNAPs were fused in-frame to the carboxyl-terminal extremity of the NP868R sequence via (G4S)4 flexible linker and tested for firefly luciferase reporter expression as described above. Comparisons of NP868R-(G4S)4-K1ERNAP versus NP868R-(G4S)4-K1.5RNAP and NP868R-(G4S)4-SP6RNAP: NS; NP868R-(G4S)4-K1ERNAP versus all other constructions: $P < 0.05$, Student's t-test. (**E**) Optimization of the polyadenylation/histone stem–loop region of DNA templates. A track of 40 adenosine residues of pK1E-Luciferase, which provides artificial polyadenylation to the transcripts was either removed or associated to the consensus stem-loop from human histone. This RNA element is involved in the regulation of stability and of translation efficiency in the cytoplasm and is functionally similar to a poly(A) tail in that it enhances translational efficiency and is dependent on mRNA capping. The corresponding pK1E-Luciferase with various sequence modifications were co-transfected with the pCMV-NP868R-(G4S)2-K1ERNAP(R551S) plasmid in HEK-293 cells. Comparison of poly[A40] vs. histone tem-loop: NS; poly[A40] versus no poly[A] track or poly[A] track placed before or after consensus stem-loop of human histone: $P < 0.05$, Student's t-test.

Figure 3. Design and optimization of single-unit C3P3-G1 enzymes. (**A**) General design of the C3P3-G1 enzymes. Single-unit C3P3-G1 enzyme

To confirm the requirement of NP868R capping activity, the K282N mutation was introduced into the NP868R moiety of NP868R-$(G_4S)_4$-T7RNAP, which is predicted to suppress GTase activity by blocking NP868R-GMP adduct formation and thereby the transfer of GMP to 5′-diphosphate ends of mRNA (54). As anticipated, the K282N mutation drastically impaired luciferase expression by C3P3-G1 (Supplementary Figure S7).

Finally we devised a semi-quantitative capping enzyme assay to determine the capping efficiency of C3P3-G1 transcripts. Since capping by C3P3 dramatically increased translation of T7p-derived mRNAs, we reasoned that if mRNAs remained uncapped by C3P3, then supplementation with auxiliary NP868R enzyme should further increase translation over C3P3 alone. To direct auxiliary NP868R to luciferase reporter mRNAs, we fused NP868R to the λN peptide of λ-phage which binds boxB RNA hairpins at nanomolar affinity (7), and added four BoxBr hairpins to the 3′UTR of the firefly luciferase gene (Figure 4). As expected, λN-NP868R plus T7RNAP caused higher luciferase expression than T7RNAP alone, demonstrating the functionality λN-NP868R and also the value of capping. λN-NP868R also complemented expression by capping-dead mutant (K282N) C3P3-G1. However, addition of λN-NP868R with C3P3-G1 marginally increased luciferase expression over C3P3 alone, suggesting that capping was already nearly saturated by C3P3. Similar findings we obtained with transcripts produced by standard nuclear expression system using CMV-driven promoter. These findings support that a majority of transcripts are efficiently capped by C3P3.

ASFV NP868R increases protein expression when coupled with various prokaryotic RNAPs, especially the R551S-optimized K1E phage RNAP

Having found that NP868R can boost protein expression when coupled to T7RNAP, we asked if NP868R could function with other bacteriophage RNAPs. We selected 54 complete bacteriophage genome sequences, which were analyzed for candidate phage-promoter sequences using the PHIRE software (55). Twenty-nine RNAPs with well-characterized promoters were then subjected to BLAST+ analysis (56), and clustered into 12 subgroups having <80% amino-acid identity (Supplementary Figure S9). One member of each subgroup was randomly selected and fused to the carboxyl-end of NP868R via $(G_4S)_4$ linker (Supplementary Table S5). For each RNAP, the putative phage promoter from the major capsid protein was inserted upstream of the luciferase gene and co-transfected with the corresponding C3P3-G1 enzyme plasmids (Supplementary Table S5). While highest luciferase expression was achieved with SP6RNAP moiety (Figure 3D), additional phage RNAPs having >80% amino-acid identity with SP6RNAP were examined in a second optimization step. The two closely related K1ERNAP and K1.5RNAP, which share 99% amino-acid identity together and 85% with SP6RNAP (57), were identified and fused in-frame with NP868R with best results given by K1ERNAP (Figure 3D). Overall these studies show that coupling of NP868R to prokaryotic RNAPs generates a novel cytoplasmic expression system.
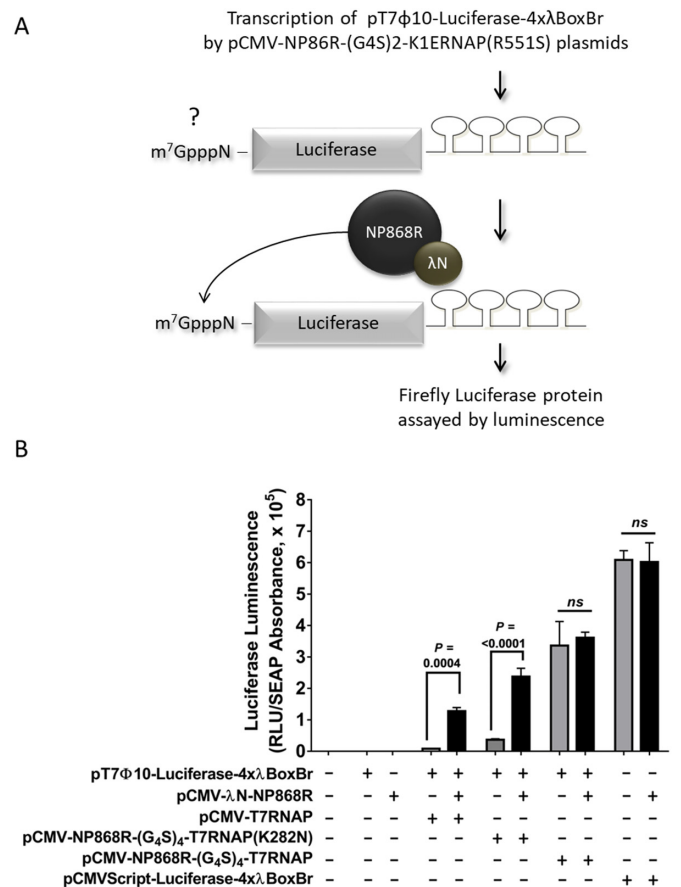

Figure 4. Semi-quantitative capping assay. (**A**) Experimental strategy to investigate the mRNA capping efficiency of the C3P3-G1 system. The African swine fever virus capping enzyme NP868R was tethered to the 3′UTR of the Firefly reporter mRNAs by fusing it to N-terminal peptide of the λN protein and introducing four BoxBr hairpins binding sites in the 3′UTR of the reporter mRNA. (**B**) The tethered NP868R enzyme increases the expression of uncapped reporter mRNA produced by the T7RNAP or by the capping-dead mutant (K282N) C3P3-G1 system. In contrast, the expression of capped reporter mRNA produced by the C3P3-G1 system remains virtually unchanged, which suggests that most of C3P3-G1 transcripts are efficiently capped.

We also tested the influence of point substitutions having predicted effects on K1ERNAP processivity. K1ERNAP amino-acids residues were selected based on conservation among the *Autographivirinae* family members and their known effects on *in vitro* processivity of T7RNAP (58) (Supplementary Figure S10). The R620S and I802S mutations, which were both predicted to reduce RNAP processivity, decreased gene reporter expression. In contrast, the R551S mutation, which was predicted to increase polymerization kinetics, slightly increased reporter gene expression and was selected for further optimization (Supplementary Figure S11).

Finally, various lengths and amino-acid content of linkers between NP868R and K1ERNAP(R551S) moieties were tested, among which two repeats of G_4S was found optimal (Supplementary Figure S11, Supplementary Table S6). The C3P3-G1 enzyme consisting of NP868R- $(G_4S)_2$-K1ERNAP(R551S) was used in further experiments.

NP868R from ASFV is a fully competent capping enzyme

NP868R capping enzyme was previously shown to form a covalent link with GTP, an activity required for GTase activity (48), but RTPase and N7-MTase activities were only inferred from sequence annotation (59). To test if NP868R has all *bona fide* enzymatic activities, we expressed and purified histidine-tagged NP868R by baculovirus expression system and Ni^{2+}-affinity chromatography, respectively (Figure 5A). The NTPase activity of NP868R was evidenced by *in vitro* conversion assay of $[\alpha\text{-}^{32}P]$-GTP or ATP into GDP or ATP. Figure 5B shows that NP868R displayed NTPase activity similar to that of the VVD1R subunit, which hydrolyses $[\alpha\text{-}^{32}P]$-GTP into GDP, or ATP into ADP (not shown). The GTase activity results from two steps reaction process in which the GTase first form a covalent link with GMP which is subsequently transferred on the 5′ end of pRNA. The first step of this process (GTP-covalent binding) was demonstrated through incubation of NP868R with $[\alpha\text{-}^{32}P]$-GTP and $MgCl_2$. A radiolabeled 99 kDa protein was observed under denaturing electrophoresis conditions, suggesting that NP868R forms a covalent adduct with the guanosine nucleotide similar to VVD1R and VVD1R-ΔMTase (Figure 5C). The subsequent transfer of GMP to the 5′-end of RNAs was next monitored by incubating NP868R with synthetic 3′-end radiolabeled RNAs (pppAC4Cp), followed by urea-PAGE RNA product analysis. In presence of $MgCl_2$, NP868R induced a band-shift of RNA products. Migration patterns were similar to those observed when using VVD1R and VVD1R-ΔMTase as positive controls, suggesting similar modification at the 5′-end of RNA blocked at their 3′-end by pCp labeling (Figure 5D). The nature of the cap structure was further analyzed by thin layer chromatography after RNA hydrolysis by nuclease P1. In the presence of the methyl donor S-Adenosyl-Methionine (SAM) and $MgCl_2$, the positive control VVD1R, as well as NP868R could synthesize a $^{m7}GpppA$ cap structure (Figure 5E), indicating that NP868R carries both N7-MTase and GTase activities. NP868R methylated GpppAC4, whereas a blocked N7-atom in the synthetic capped RNA ($^{m7}GpppAC_4$ substrate) abolished ^3H-methyl transfer. NP868R thus carries N7-MTase but not ribose-2′-*O*-methyltransferase activities on these short substrates (Figure 5F). Altogether, these assays demonstrate that NP868R is a fully active capping enzyme.

C3P3-G1 expression system is functional in mammalian cells

Immunofluorescence microcopy showed only cytoplasmic expression of the FLAG-tagged NP868R-(G4S)2-K1ERNAP(R551S) enzyme as expected (Figure 6A-B). In addition, the FLAG-tagged NP868R-(G4S)2-K1ERNAP(R551S) enzyme was expressed in both CHO-K1 and HEK-293 cells as the expected 200-kDa fusion protein (Figure 6C). Cell viability and death, measured by fluorescent dipeptidyl peptidase-1 GF-AFC and luminogenic AAF-aminoluciferin substrates respectively, were similar for empty plasmids and C3P3-expressing plasmids, indicating that the C3P3-G1 enzyme do not exert any toxic effects on HEK-293 cells (Supplementary Figure S12).

Since C3P3-G1 was thus far optimized for expression of luciferase, it was critical to determine whether C3P3-G1 promotes efficient expression, processing, and localization of various proteins. All experiments were performed using DNA templates, which were optimized in both HEK-293 and CHO-K1 cells with similar results (data not shown). As judged by immunofluorescence microcopy, C3P3-G1 drives expression and localization of fluorescent proteins tagged for transport in the expected nuclear, mitochondrial, and endosomal compartments (Figure 7). The localization of these fluorescent proteins to the endosomal and mitochondrial compartments was confirmed by the coexpression of EGFP fused to the human LDLR and rat OCT sorting signals, respectively (Supplementary Figure S13). Noticeably, the morphological patterns of these fluorescent proteins appear similar when expressed using a standard nuclear expression system, therefore suggesting that the C3P3-G1 does not impair protein sorting (Figure 7). To further confirm that the proteins produced by the C3P3-G1 system were properly processed, we assayed in culture medium the functional activity of the released hEPO, a N- and O-glycosylated secreted protein (60). Its activity was similar when produced by C3P3-G1 versus standard nuclear expression systems, showing normal protein folding and disulfide bond formation (Figure 8). Taken together, the findings confirm that the expression of C3P3 transcripts generates normal proteins.

We have optimized components of the DNA templates with no or modest effects of pause site, 5′-UTR, 3′-UTR, and 3′-stop sequences on luciferase expression (Supplementary Figures S14–17, Supplementary Table S7–8). Conversely, the combination of poly[A]40 track and consensus histone stem-loop sequence, which is functionally similar to a poly(A) tail in that it enhances translational efficiency and is dependent on mRNA capping, markedly improved luciferase expression (Figure 3E).

The kinetics of C3P3-G1 gene expression was monitored by quantitative RT-PCR and luciferase assay in HEK-293 cells. Peak of C3P3-G1 enzyme mRNA levels was detected at 24–48 h, whereas C3P3-G1 Firefly Luciferase mRNA and luminescence peaked at 3 days with sustained expression over 4–6 days in HEK-293 cell line (Figure 9A–C). The C3P3-G1 expression system produced at peak 6.7-fold more luciferase mRNA, but 2.2-fold lower luciferase luminescence at peak than standard CMV-driven Firefly Luciferase expression plasmid. This low luminescence:mRNA ratio shows limited translational efficiency of C3P3-G1 Luciferase transcripts, which suggests incomplete maturation of C3P3-driven transcripts in HEK-293 cells.

To investigate if the above results could be caused by global suppression of host-cell translation, the polysome profile of HEK-293 cells expressing the Firefly Luciferase under control of C3P3-G1 or standard CMV-promoter-based nuclear expression plasmid were compared. Polysome profiling separates translated mRNAs on a sucrose gradient according to the number of bound ribosomes. No frank change in ribosome distribution patterns was observed between these two conditions (Figure 10), which supports the notion that expression by C3P3-G1 has no major impact on global translation of human HEK-293 cells.
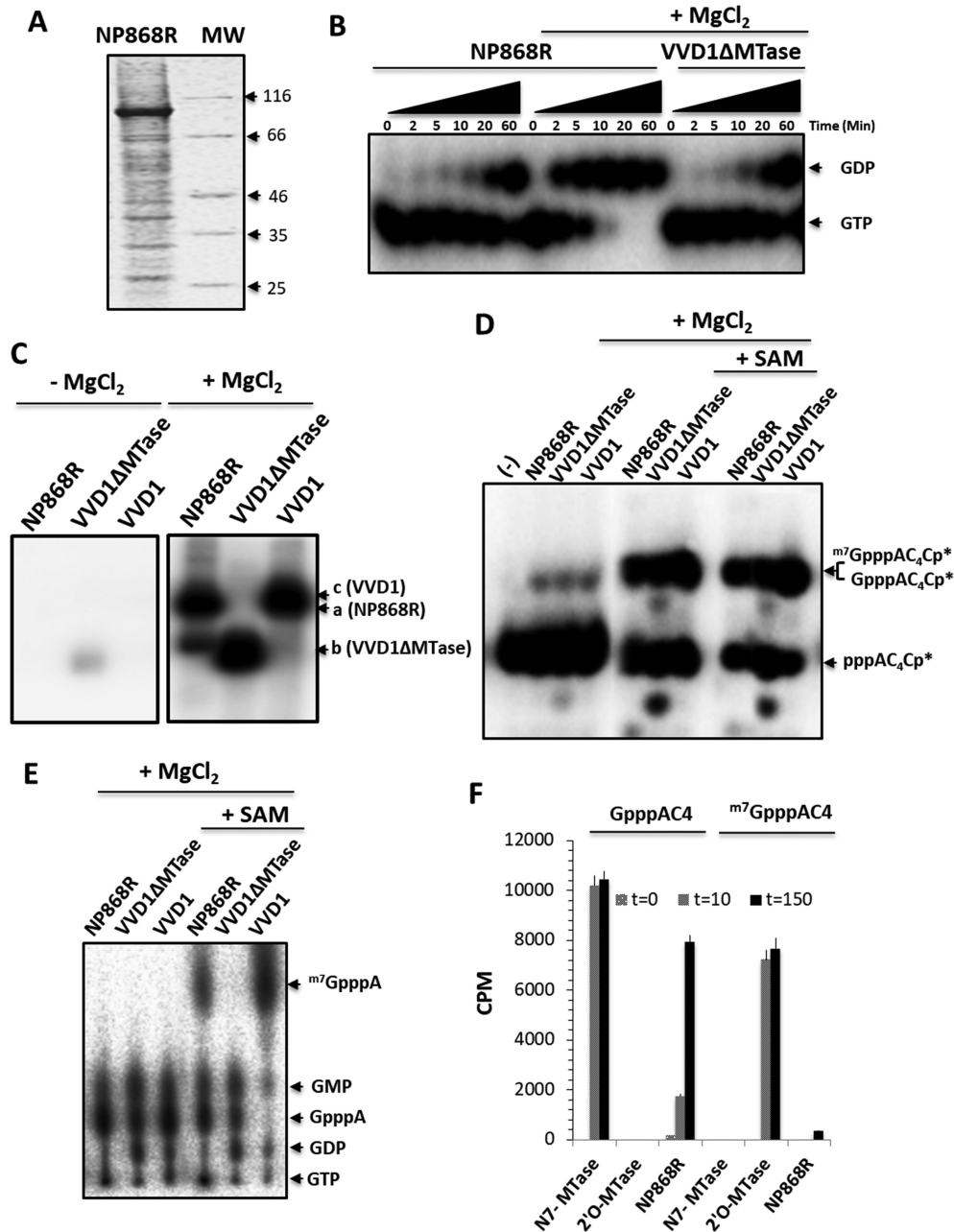
Figure 5. NTPase, GTase and N7-MTase activities of NP868R capping enzyme. (**A**) Coomassie-stained, reducing SDS-PAGE showing the insect-cell derived NP868R protein purified by Ni^{2+}-affinity chromatography migrating at around 100 kDa. Protein ladder is shown on the right lane. (**B**) NTPase activity. Autoradiographs of 20% urea-PAGE gel show that NP868R and the D1R subunit of VV (VVD1ΔMTase) hydrolyze [α-^{32}P] radio-labeled GTP to GDP. The hydrolysis of GTP by NP868R is stimulated by 2.5 mM $MgCl_2$. (**C**) GTP-binding assay. Autoradiographs of 12% SDS-PAGE show that NP868R, VVD1RΔMTase and VVD1R form covalent adducts, after 1 h of incubation with [α-^{32}P] radio-labeled GTP in presence of 2.5 mM $MgCl_2$. The main radiolabeled bands are detected at a molecular weight corresponding to the incubated proteins. (**D**) GTase activity. A synthetic RNA (pppAC$_4$), radiolabeled at its 3′-end using [α-^{32}P]-pCp (pppAC$_4$-Cp*, Ctl) is incubated with NP868R, VVD1RΔMTase and VVD1R in presence of 1 mM GTP. The RNA reaction products, obtained after 1 h of incubation, are separated on a 20% urea–PAGE gel. The VV-capping enzyme (VVD1R and VVD1ΔMTase) used as positive control, convert pppAC$_4$-Cp* into ^{m7}GpppAC$_4$-Cp* or GpppAC$_4$-Cp*, respectively. The autoradiographs show that, in presence of 2.5 mM $MgCl_2$, NP868R convert pppAC$_4$-Cp* RNA to a RNA migrating a molecular weight similar to the capped RNA generated by VVD1R proteins. (**E**) Cap structure identification. pppAC$_4$ RNA was incubated with [α-^{32}P]-GTP in presence of NP868R, VVD1RΔMTase and VVD1R respectively. The reaction products were digested by nuclease P1, and the caps were separated by thin-layer chromatography (TLC). ^{m7}GpppA is the predominant product seen in presence of the methyl-donor (SAM) using the VVD1, whereas VVD1RΔMTase synthesized mainly GpppA, as well as GTP hydrolysis products (GDP and GMP). ^{m7}GpppA cap structure is detected when NP868R is incubated with SAM and 2.5 mM $MgCl_2$, and GpppA in absence of SAM. (**F**) MTase activity. Activity of NP868R was assessed by monitoring the transfer of tritiated methyl (CH$_3$) from SAM to a short synthetic RNA (GpppAC$_4$ or ^{m7}GpppAC$_4$) after 10 and 150 min. The bar chart experiment compares the activities of NP868R with that of human N7-MTase (N7-MTase) and SARS-COV 2′O-MTase known to methylate specifically the N7 position of the RNA cap structure and the 2′O position of capped RNA respectively. NP868R methylate the GpppAC$_4$ but not RNA already methylated on their N7 position (^{m7}GpppAC$_4$) suggesting that the methylation occurs on the N7 position of cap structure. The error bars indicate the standard deviation of three experiments.
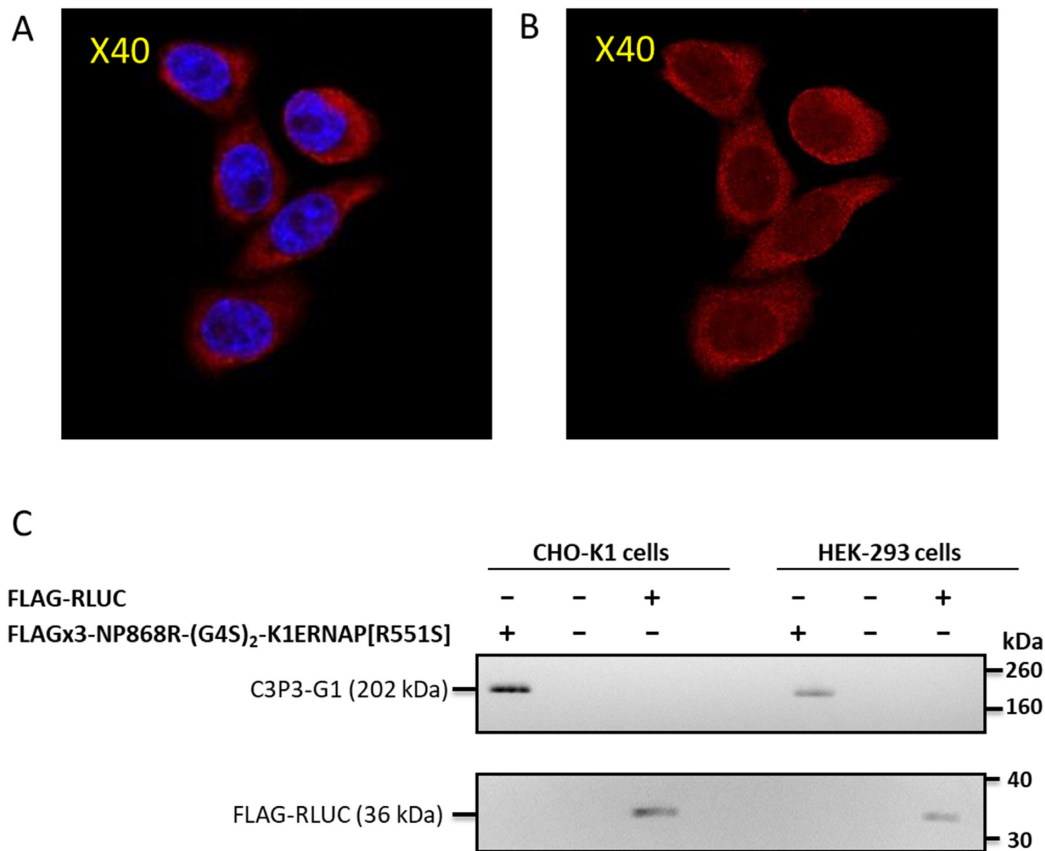
Figure 6. C3P3-G1 enzyme is a 200 kDa cytoplasmic protein. (**A**) Immunofluorescence of FLAGx3-NP868R-(G4S)$_2$-K1ERNAP C3P3-G1 enzyme is found in the cytoplasm of CHO-K1 cells, with or (**B**) without nuclear staining with Hoechst 33342. (**C**) Western blot analysis of HEK-293 and CHO-K1 cells untransfected or transfected with the FLAGx3-NP868R-(G4S)$_2$-K1ERNAP or FLAG-RLUC plasmid used as a positive control. The 200 kDa band corresponds to the expected molecular weight of the full-length flagged C3P3-G1 protein and the 36 kDa to the FLAG-RLUC control.

Assessment of C3P3-G1 expression system for transient mammalian cell expression in CHO-K1 and other cell types

Expression of Firefly Luciferase by the C3P3-G1 expression system was investigated in other cell lines, including the Chinese hamster CHO-K1, the human cervix carcinoma HeLa, the human liver HepG2, and the human fibroblasts WI-38 cells. The expression levels by the C3P3-G1 system were compared to the standard CMV-driven Firefly Luciferase expression plasmid, assayed as previously described. Striking variability was observed across the cell types, with best performances of the C3P3-G1 system observed with the CHO-K1 and HepG2 cell lines (Supplementary Figure S18). The mechanisms responsible for this variability across the different cell types will be investigated in further studies.

To better investigate the performance of the C3P3-G1 system in CHO-K1 cells, which is the predominant mammalian host for therapeutic protein production, the titers of three secreted proteins, human erythropoietin (hEPO), human granulocyte-colony-stimulating factor (hG-CSF), and mouse alpha-fetoprotein (mAFP) were monitored by ELISA. Compared to the standard CMV-promoter-based nuclear expression plasmid, C3P3-G1 produced at peak 1.9- to 3.1-fold higher titers of proteins (Figure 11A–C).

DISCUSSION

We present the C3P3-G1 artificial expression system optimized for its uses in cultured mammalian cells. To the best of our knowledge, this is the first artificial expression system that successfully couples transcription and post-transcription, which are processes of exquisite complexity in eukaryotes. C3P3-G1 relies on a single-unit chimeric enzyme that consists of the fusion of an mRNA capping enzyme with a DNA-dependent RNA polymerase. The RNA polymerase moiety of the C3P3 enzyme transcribes *in cellulo* specific DNA templates under control of the C3P3 promoter and synthesizes RNA chains, which are subsequently capped by the capping enzyme moiety. In contrast, the polyadenylation of the C3P3 transcripts is synthesized by transcription of a polyadenosine track from the DNA templates. Therefore, once primed, the C3P3 expression system autonomously synthesizes RNA chains in the cytoplasm of the host-cell and generates the key modifications requested for their translation.

To ensure mRNA capping of C3P3-G1 transcripts, we pursued an approach that has benefited greatly from the vaccinia virus-T7RNAP expression system. This latter approach makes use of a recombinant vaccinia virus encoding the T7RNAP, while the target gene, under control of a T7 promoter, is either embedded in a plasmid or a sec-
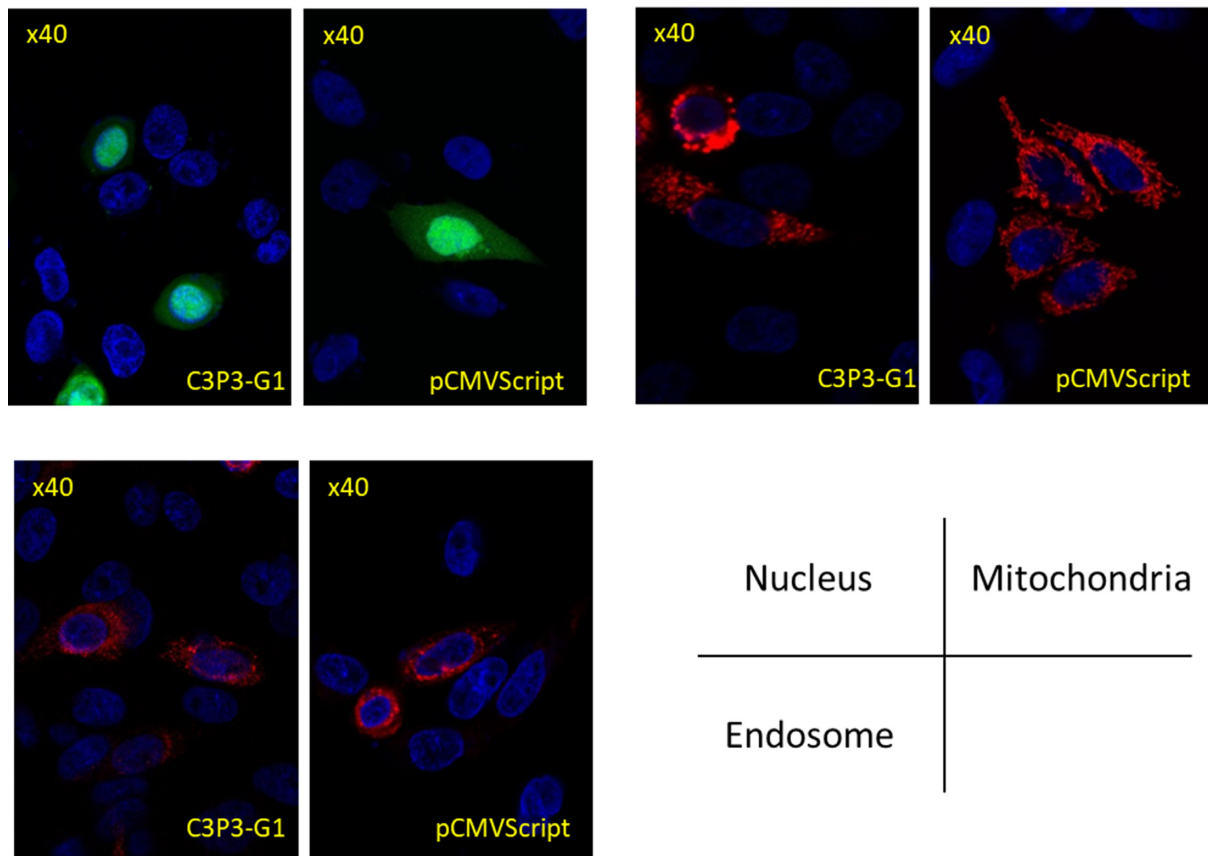
Figure 7. C3P3-G1 expression system generates properly addressed proteins to cell compartments in comparison to standard CMV promoter-driven nuclear expression system. Fluorescent proteins with various trafficking signals were produced by the C3P3-G1 or standard CMV-promoter-driven nuclear expression system (pCMVScript) and were found appropriately addressed to the same corresponding cell compartments (**A**) green GFP fluorescent protein with SV40 eNLS signal featured by homogenous nuclear staining, (**B**) red RFP fused to the mitochondrial COX4 signal, which directs the fusion protein to the inner membrane of the mitochondria, was imaged as small aggregates spread in the cytoplasm, (c) red RFP fused to the transferrin receptor type 1 endosomal trafficking signal was visualized as multiple connected punctuates fluorescent patterns in cell cytoplasm.

ond recombinant vaccinia virus (28,29). As the vaccinia virus is an intracellular parasite that replicates exclusively in the host-cell cytoplasm, the T7-transcripts are 5′-capped and 3′polyadenylated in host-cell cytoplasm by the vaccinia virus enzymes (30). Although the amounts of mRNA made by this system are extremely high, comprising ∼30% of the total steady state RNA in the cytoplasm, only moderate amounts of T7-transcripts are properly 5′-capped (30). A plausible explanation for these findings is provided by the crowded environment of the cytoplasm, which contains a dense network of cytoskeletal filaments, therefore impeding free diffusion of large macromolecules such as proteins, DNA and mRNA (61). Due to the lack of coupling between the T7RNAP and vaccinia virus capping enzyme, the T7-transcripts are poorly 5′-processed by the vaccinia virus capping enzyme. This assumption was central for the development of the C3P3-G1 enzyme, which is preferably the fusion between the African Swine Fever Virus capping enzyme and a mutant bacteriophage K1E RNA polymerase.

In addition to ^{m7}GpppN cap at their 5′-ends, virtually all eukaryotic mRNAs contain a poly(A) tail formed by template-independent addition of adenosine monophosphates residues to their 3′-terminal hydroxyl groups. The resulting 3′-poly(A) tail cooperates synergistically with the 5′-capping to stimulate translation with supra-additive potentiation, presumably by circularizing mRNA (62). In addition, poly(A) tails promote stability of the transcripts, while deadenylation is a major rate-limiting step in mRNA decay (63,64). In contrast to the nuclear polyadenylation, the poly(A) tail of the C3P3-G1 is generated by the template-dependent transcription of a short 40 adenosine track from the DNA template, followed by a self-cleaving ribozyme sequence. The length of the poly(A) tail of the C3P3-G1 transcripts encoded by the templates is therefore much shorter than those of mammalian transcripts, which exit the nucleus with a uniform length of ∼250 nucleotides (65–68). As the length of the poly(A) tail is strongly correlated with the level of gene expression (69), the polyadenylation from the C3P3-G1 expression system can be seen as sub-optimal. Further generations of the C3P3 system will improve the polyadenylation process, possibly by tethering a poly(A) polymerase to the C3P3 transcripts.

A current imperfection of the C3P3 system consists in the incomplete translational efficiency of C3P3 transcripts, i.e. the mRNA:luminescence ratio that is lower using the C3P3-G1 system than the standard nuclear expression system. Such incomplete mRNA translational efficiency might be explained by various mechanisms. Firstly, the mRNA
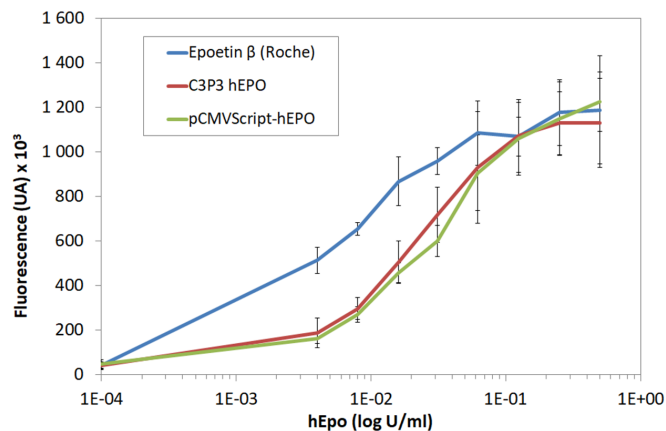
Figure 8. Biological activity of the human erythropoietin released in culture medium by CHO-K1 cells. The human erythropoietin glycoprotein released in culture medium by transfected CHO-K1 cells was assayed by proliferation assay of the erythropoietin-dependent UT7 cell line. Dilutions of hEPO from the culture medium of transfected CHO-K1 cells or recombinant commercial hEPO (Epoetin β, Roche Laboratories) were added to the culture medium of UT-7 cells. The activity of hEPO produced by the C3P3-G1 system was comparable to the nuclear expression plasmid, therefore showing normal protein folding and disulfide found formation. The differences of activity profiles between Epoetin β and hEPO produced in CHO-K1 cells with the standard plasmid or the C3P3-G1 system might be explained by the use of distinct methods for measuring hEPO or other differences in line with the processing of Epoetin β product (20).

modifications performed by the current generation of the C3P3 system might be incomplete. Our semi-quantitative assay (Figure 4) suggests that the majority of mRNAs are capped by C3P3-G1. Nevertheless, even minor amounts of uncapped 5'-triphosphate RNA might induce retinoic acid-inducible protein I (RIG-I)-mediated interferon-α response and host-cell translation shut-down (70,71). Interferon-α response can be also induced by the absence of cap-1 at the 5'-ends of transcripts, which is not accommodated by C3P3-G1 (72). Strikingly, no frank modifications of ribosome distribution patterns were brought into focus by polysome profiling analysis. This finding does not support the hypothesis of global inhibition of host-cell translation as expected in case of strong interferon-response, but closer investigation of content of the polysomal fractions. The length of the polyadenylation tail is also known be critical for mRNA translation and should be increased by further generations of the C3P3 system. Secondly, other post-transcriptional modifications which are not performed by C3P3-G1 might be critical for mRNA translational efficiency. For instance, the reversible N6,2'-*O*-dimethyladenosine methylation at the first encoded nucleotide adjacent to the cap was recently found to influence cellular mRNA fate (73). Alternatively, the absence of other modifications might induce the shut-down of host-cell translation, via interferon pathway or others (70). For instance, internal bases can be subjected to modifications such as pseudouridination or methylation at uridine or cytosine residues, which decrease RNA-dependent protein kinase activation in response to exogenous RNA, then phosphorylates translation initiation factor 2-α (eIF2-α) and inhibits translation (74,75). Thirdly, host-cell proteins, eventually deposited onto pre-
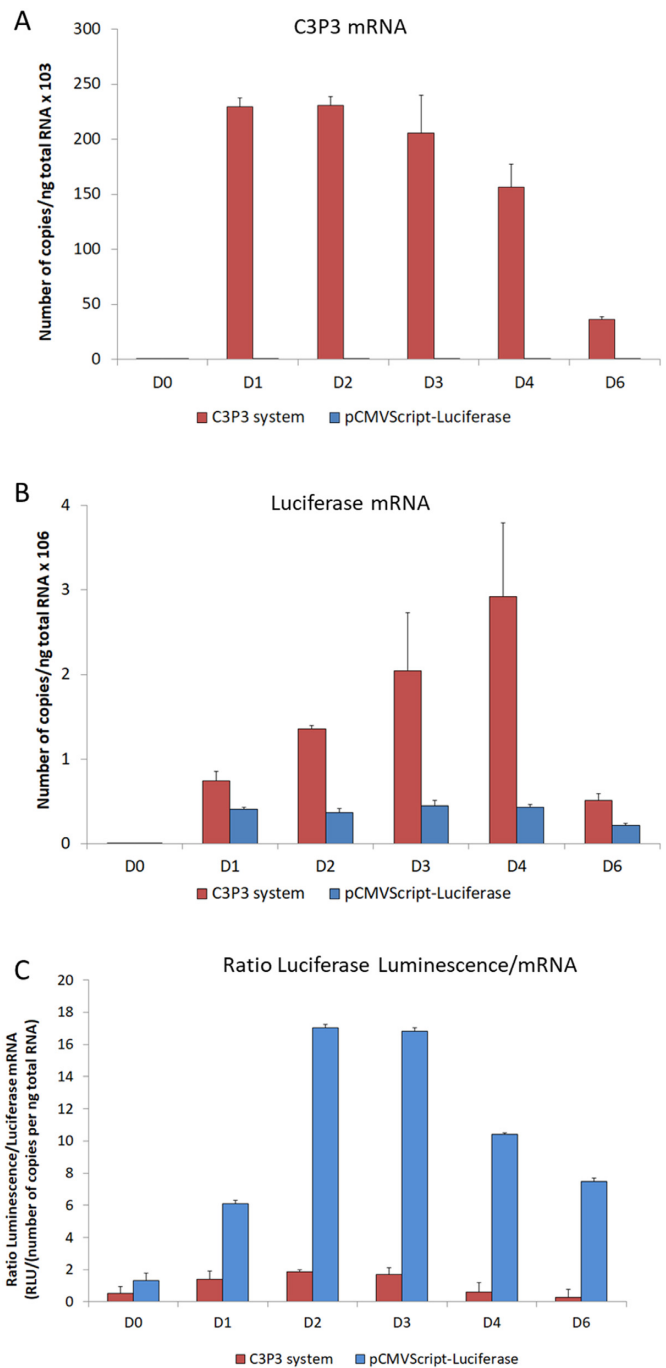

Figure 9. Kinetics of C3P3-G1 gene expression monitored by quantitative RT-PCR and luciferase assay. HEK-293 were either cotransfected with pCMV-NP868R-(G4S)$_2$-K1ERNAP and pK1E-Luciferase (i.e. Firefly Luciferase gene under control of the phage K1E promoter), or pCMVScript-Luciferase plasmids (i.e. Firefly Luciferase gene under control of the RNAP II-dependent IE1 human CMV promoter/enhancer). (**A**) The kinetics of pCMV-NP868R-(G4S)$_2$-K1ERNAP C3P3-G1 mRNA and (**B**) Luciferase mRNA were assayed by quantitative RT-PCR, along with (**C**) Firefly Luciferase expression monitored by luciferin oxidation assay.
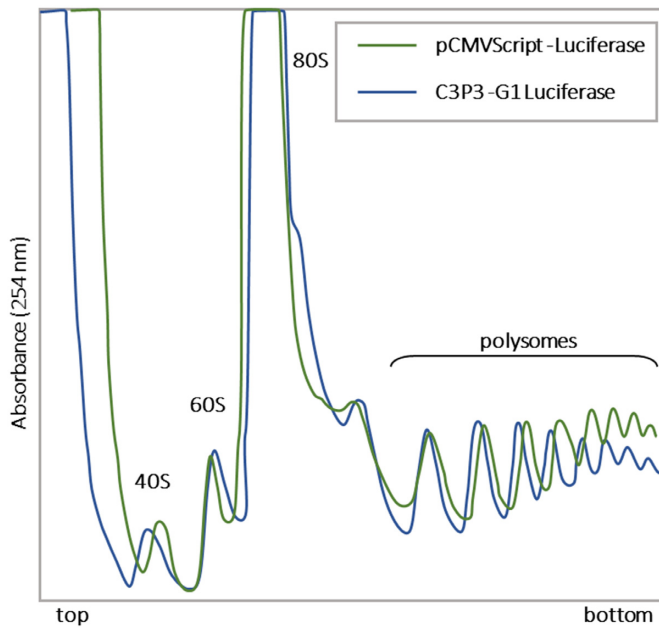
Figure 10. Analysis of translation using polysome profiling. HEK-293 cells were either cotransfected with pCMV-NP868R-(G4S)$_2$-K1ERNAP and pK1E-Luciferase (i.e. C3P3-G1 Luciferase), or pCMVScript-Luciferase plasmids. Cell lysates were fractionated by 15–50% sucrose gradient centrifugation as described in Materials and Methods. The absorbance profile of the gradient is shown (top and bottom of the gradients are indicated). 40S and 60S denote the corresponding ribosomal subunits; 80S, monosome.

messenger RNA or mRNA, might participate in translation. For instance, the exon junction complex, which is assembled onto pre-messenger RNA at the nuclear stage, contributes to both mRNA translational efficiency and decay (76,77). Fourthly, C3P3 transcripts might be inappropriately positioned in the cytoplasmic compartment for their translation by the host-cell machinery. For comparison, large cytoplasmic replicating DNA viruses, such as Poxvirus or Asfarvirus, form viral factories that sequester transcription factors, ribosomes, and chaperonins in circumscribed cytoplasmic sub-compartments. Altogether, these hypotheses will be systemically explored by deeper transcriptional and translational analysis for further improvements of the C3P3 system.

In contrast to the standard nuclear expression systems that require access to the host-cell nuclear transcriptional machinery, the C3P3-G1 system has been devised as an autonomous cytoplasmic expression system. Therefore, once expression of the C3P3 enzyme has been primed, it becomes independent of the passive diffusion of the input DNA through the nuclear pores. The diameter of their main channel, which is smaller than the classical diameter of plasmid DNA, explains their limited diffusion to the nucleus and possibly the limited efficacy of transient transfection (78). For example, only 1–10% of the plasmids transfected to cultured cells can effectively reach their nuclei as assessed by quantitative RT-PCR, Southern blot analysis, or electron microscopy (79,80). Hence, the cytoplasmic processivity of the C3P3 system appears appealing for transient non-viral gene therapy, an attractive approach to deliver therapeutic
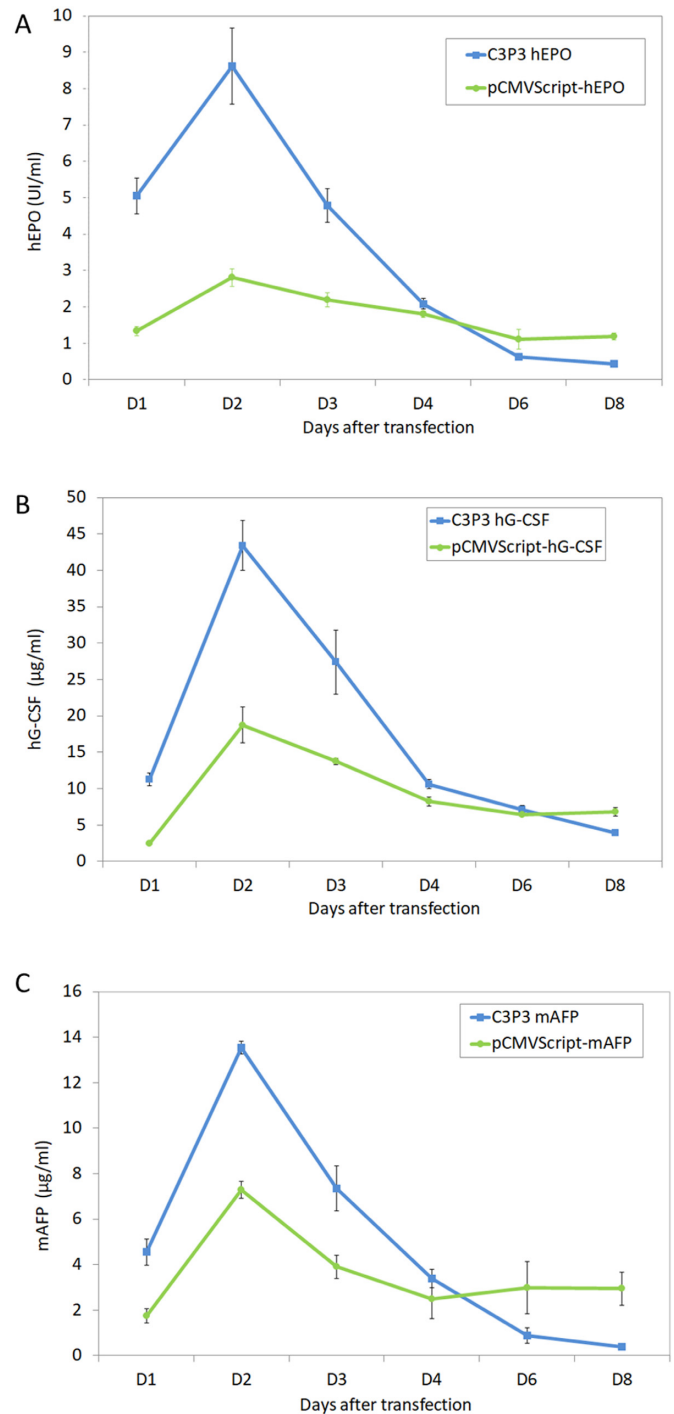

Figure 11. Kinetics of expression of secreted proteins by CHO-K1 cells. Titers of (**A**) the secreted human erythropoietin (hEPO), (**B**) human granulocyte colony-stimulating factor isoform-b (hG-CSF) and, (**C**) secreted/cytoplasmic mouse alpha-fetoprotein (mAFP) released in serum-free culture medium of CHO-K1 cells were monitored by ELISA.

nucleic acids for human therapeutics (81). Non-viral gene therapy appears much safer and less immunogenic than recombinant viruses and has the potential to deliver larger genetic payloads (81). However, its efficacy appears relatively limited, which has greatly hampered its uses (82). We there-

fore anticipate that the C3P3 system can be used for non-viral gene therapy, especially for disorders involving quiescent cells such as hepatocytes, skeletal muscular cells or naive immune B or T cells.

Besides its therapeutics uses, the C3P3 expression system should be also considered for its applications *in cellulo*. For instance, C3P3 expression system can be also used for transient expression studies for phenotypic screening or others. However, the artificial nature of C3P3-G1 system makes it inappropriate for investigation of host-cell transcriptional or translational regulation. In addition, the high performances of the C3P3-G1 system in CHO-K1 cells and the lack of detectable cell toxicity of the C3P3 enzyme suggest that stable C3P3 cell lines could be generated for production of recombinant proteins.

Another interesting application of this system relates to RNA virus production. Recently, Eaton *et al.* (83) have shown that C3P3-G1 system increases reovirus protein expression by 5- to 10-fold and reovirus rescue by 100-fold, in comparison to the current reverse genetics system. Reoviruses are nonpathogenic viruses that demonstrate anti-tumor activity and are being explored for cancer therapy (84). These results are therefore encouraging for using the C3P3-G1 system to produce viruses containing 5′-capped RNAs, including *Retroviridae*, *Rhabdoviridae*, *Paramyxoviridae*, *Filoviridae* or *Reoviridae*.

In conclusion, the C3P3-G1 artificial expression system recapitulates the key mRNA modifications required for mRNA translation, including RNA synthesis, capping, and in part polyadenylation. This system is optimized for mammals, although it can theoretically be adapted to any other eukaryotic species. The C3P3 system should be seen as a versatile system, with multiple *in cellulo* and *in vivo* applications. Importantly, the present C3P3-G1 system should be considered as an advanced prototype, which performances will be strongly improved in further generations without doubt.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Loyter,A., Scangos,G.A. and Ruddle,F.H. (1982) Mechanisms of DNA uptake by mammalian cells: fate of exogenously added DNA monitored by the use of fluorescent dyes. *Proc. Natl. Acad. Sci. U.S.A.*, **79**, 422–426.
2. Zabner,J., Fasbender,A.J., Moninger,T., Poellinger,K.A. and Welsh,M.J. (1995) Cellular and molecular barriers to gene transfer by a cationic lipid. *J. Biol. Chem.*, **270**, 18997–19007.
3. Orphanides,G. and Reinberg,D. (2000) RNA polymerase II elongation through chromatin. *Nature*, **407**, 471–475.
4. Darzacq,X., Shav-Tal,Y., de Turris,V., Brody,Y., Shenoy,S.M., Phair,R.D. and Singer,R.H. (2007) In vivo dynamics of RNA polymerase II transcription. *Nat .Struct. Mol. Biol.*, **14**, 796–806.
5. Raab,D., Graf,M., Notka,F., Schodl,T. and Wagner,R. (2010) The GeneOptimizer Algorithm: using a sliding window approach to cope with the vast sequence space in multiparameter DNA sequence optimization. *Syst. Synth. Biol.*, **4**, 215–225.
6. Das,A. (1993) Control of transcription termination by RNA-binding proteins. *Annu. Rev. Biochem.*, **62**, 893–930.
7. Austin,R.J., Xia,T., Ren,J., Takahashi,T.T. and Roberts,R.W. (2003) Differential modes of recognition in N peptide-boxB complexes. *Biochemistry*, **42**, 14957–14967.
8. Peyrane,F., Selisko,B., Decroly,E., Vasseur,J.J., Benarroch,D., Canard,B. and Alvarez,K. (2007) High-yield production of short GpppA- and 7MeGpppA-capped RNAs and HPLC-monitoring of methyltransfer reactions at the guanine-N7 and adenosine-2'O positions. *Nucleic Acids Res.*, **35**, e26.
9. Bouvet,M., Imbert,I., Subissi,L., Gluais,L., Canard,B. and Decroly,E. (2012) RNA 3'-end mismatch excision by the severe acute respiratory syndrome coronavirus nonstructural protein nsp10/nsp14 exoribonuclease complex. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 9372–9377.
10. Niles,A.L., Moravec,R.A. and Riss,T.L. (2009) In vitro viability and cytotoxicity testing and same-well multi-parametric combinations for high throughput screening. *Curr. Chem. Genomics*, **3**, 33–41.
11. Cho,M.H., Niles,A., Huang,R., Inglese,J., Austin,C.P., Riss,T. and Xia,M. (2008) A bioluminescent cytotoxicity assay for assessment of membrane integrity using a proteolytic biomarker. *Toxicol. In Vitro*, **22**, 1099–1106.
12. Einhauer,A. and Jungbauer,A. (2001) The FLAG peptide, a versatile fusion tag for the purification of recombinant proteins. *J. Biochem. Biophys. Methods*, **49**, 455–465.
13. Chubet,R.G. and Brizzard,B.L. (1996) Vectors for expression and secretion of FLAG epitope-tagged proteins in mammalian cells. *Biotechniques*, **20**, 136–141.
14. Jing,S.Q., Spencer,T., Miller,K., Hopkins,C. and Trowbridge,I.S. (1990) Role of the human transferrin receptor cytoplasmic domain in endocytosis: localization of a specific signal sequence for internalization. *J. Cell Biol.*, **110**, 283–294.
15. Power,S.D., Lochrie,M.A. and Poyton,R.O. (1984) The nuclear-coded subunits of yeast cytochrome c oxidase. III. Identification of homologous subunits in yeast, bovine heart, and Neurospora crassa cytochrome c oxidases. *J. Biol. Chem.*, **259**, 6575–6578.
16. Smale,S.T. and Tjian,R. (1986) T-antigen-DNA polymerase alpha complex implicated in simian virus 40 DNA replication. *Mol. Cell Biol.*, **6**, 4077–4087.

17. Wild,D. (2013) *The Immunoassay Handbook: Theory and Applications of Ligand Binding, ELISA and Related Techniques*. 4th edn.. Elsevier Science, Oxford.

18. Komatsu,N., Nakauchi,H., Miwa,A., Ishihara,T., Eguchi,M., Moroi,M., Okada,M., Sato,Y., Wada,H., Yawata,Y. *et al.* (1991) Establishment and characterization of a human leukemic cell line with megakaryocytic features: dependency on granulocyte-macrophage colony-stimulating factor, interleukin 3, or erythropoietin for growth and survival. *Cancer Res.*, **51**, 341–348.

19. Hermine,O., Dubart,A., Porteux,F., Mayeux,P., Titeux,M., Dumenil,D. and Vainchenker,W. (1996) Inhibition of the erythropoietin-induced erythroid differentiation by granulocyte-macrophage colony-stimulating factor in the human UT-7 cell line is not due to a negative regulation of the erythropoietin receptor. *Blood*, **87**, 1746–1753.

20. Skibeli,V., Nissen-Lie,G. and Torjesen,P. (2001) Sugar profiling proves that human serum erythropoietin differs from recombinant human erythropoietin. *Blood*, **98**, 3626–3634.

21. Verrier,S.B. and Jean-Jean,O. (2000) Complementary between the mRNA 5′ untranslated region and 18S ribosomal RNA can inhibit translation. *RNA*, **6**, 584–597.

22. Yin,Y.W. and Steitz,T.A. (2002) Structural basis for the transition from initiation to elongation transcription in T7 RNA polymerase. *Science*, **298**, 1387–1395.

23. Wilusz,C.J., Wormington,M. and Peltz,S.W. (2001) The cap-to-tail guide to mRNA turnover. *Nat. Rev. Mol. Cell Biol.*, **2**, 237–246.

24. Dower,K. and Rosbash,M. (2002) T7 RNA polymerase-directed transcripts are processed in yeast and link 3′ end formation to mRNA nuclear export. *RNA*, **8**, 686–697.

25. Ikeda,R.A. (1992) The efficiency of promoter clearance distinguishes T7 class II and class III promoters. *J. Biol. Chem.*, **267**, 11322–11328.

26. Been,M.D. and Wickham,G.S. (1997) Self-cleaving ribozymes of hepatitis delta virus RNA. *Eur. J. Biochem.*, **247**, 741–753.

27. Benton,B.M., Eng,W.K., Dunn,J.J., Studier,F.W., Sternglanz,R. and Fisher,P.A. (1990) Signal-mediated import of bacteriophage T7 RNA polymerase into the Saccharomyces cerevisiae nucleus and specific transcription of target genes. *Mol. Cell Biol.*, **10**, 353–360.

28. Fuerst,T.R., Niles,E.G., Studier,F.W. and Moss,B. (1986) Eukaryotic transient-expression system based on recombinant vaccinia virus that synthesizes bacteriophage T7 RNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.*, **83**, 8122–8126.

29. Engleka,K.A., Lewis,E.W. and Howard,B.H. (1998) Mechanisms of replication-deficient vaccinia virus/T7 RNA polymerase hybrid expression: effect of T7 RNA polymerase levels and alpha-amanitin. *Virology*, **243**, 331–339.

30. Fuerst,T.R. and Moss,B. (1989) Structure and stability of mRNA synthesized by vaccinia virus-encoded bacteriophage T7 RNA polymerase in mammalian cells. Importance of the 5′ untranslated leader. *J. Mol. Biol.*, **206**, 333–348.

31. Elroy-Stein,O. and Moss,B. (1990) Cytoplasmic expression system based on constitutive synthesis of bacteriophage T7 RNA polymerase in mammalian cells. *Proc. Natl. Acad. Sci. U.S.A.*, **87**, 6743–6747.

32. Guo,P.X. and Moss,B. (1990) Interaction and mutual stabilization of the two subunits of vaccinia virus mRNA capping enzyme coexpressed in Escherichia coli. *Proc. Natl. Acad. Sci. U.S.A.*, **87**, 4023–4027.

33. Shuman,S., Surks,M., Furneaux,H. and Hurwitz,J. (1980) Purification and characterization of a GTP-pyrophosphate exchange activity from vaccinia virions. Association of the GTP-pyrophosphate exchange activity with vaccinia mRNA guanylyltransferase. RNA (guanine-7-)methyltransferase complex (capping enzyme). *J. Biol. Chem.*, **255**, 11588–11598.

34. Monroy,G., Spencer,E. and Hurwitz,J. (1978) Characteristics of reactions catalyzed by purified guanylyltransferase from vaccinia virus. *J. Biol. Chem.*, **253**, 4490–4498.

35. Boone,R.F., Ensinger,M.J. and Moss,B. (1977) Synthesis of mRNA guanylyltransferase and mRNA methyltransferases in cells infected with vaccinia virus. *J. Virol.*, **21**, 475–483.

36. Mao,X. and Shuman,S. (1994) Intrinsic RNA (guanine-7) methyltransferase activity of the vaccinia virus capping enzyme D1 subunit is stimulated by the D12 subunit. Identification of amino acid residues in the D1 protein required for subunit association and methyl group transfer. *J. Biol. Chem.*, **269**, 24472–24479.

37. Natalizio,B.J., Robson-Dixon,N.D. and Garcia-Blanco,M.A. (2009) The Carboxyl-terminal Domain of RNA Polymerase II Is Not Sufficient to Enhance the Efficiency of Pre-mRNA Capping or Splicing in the Context of a Different Polymerase. *J Biol Chem*, **284**, 8692–8702.

38. Kaneko,S., Chu,C., Shatkin,A.J. and Manley,J.L. (2007) Human capping enzyme promotes formation of transcriptional R loops in vitro. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 17620–17625.

39. McCracken,S., Fong,N., Rosonina,E., Yankulov,K., Brothers,G., Siderovski,D., Hessel,A., Foster,S., Shuman,S. and Bentley,D.L. (1997) 5′-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Genes Dev.*, **11**, 3306–3318.

40. Meinhart,A. and Cramer,P. (2004) Recognition of RNA polymerase II carboxy-terminal domain by 3′-RNA-processing factors. *Nature*, **430**, 223–226.

41. Ho,C.K. and Shuman,S. (1999) Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme. *Mol. Cell*, **3**, 405–411.

42. Moll,J.R., Ruvinov,S.B., Pastan,I. and Vinson,C. (2001) Designed heterodimerizing leucine zippers with a ranger of pIs and stabilities up to 10(-15) M. *Protein Sci.*, **10**, 649–655.

43. Decroly,E., Ferron,F., Lescar,J. and Canard,B. (2011) Conventional and unconventional mechanisms for capping viral mRNA. *Nat. Rev. Microbiol.*, **10**, 51–65.

44. Martinez-Costas,J., Sutton,G., Ramadevi,N. and Roy,P. (1998) Guanylyltransferase and RNA 5′-triphosphatase activities of the purified expressed VP4 protein of bluetongue virus. *J. Mol. Biol.*, **280**, 859–866.

45. Ramadevi,N., Burroughs,N.J., Mertens,P.P., Jones,I.M. and Roy,P. (1998) Capping and methylation of mRNA by purified recombinant VP4 protein of bluetongue virus. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 13537–13542.

46. Sutton,G., Grimes,J.M., Stuart,D.I. and Roy,P. (2007) Bluetongue virus VP4 is an RNA-capping assembly line. *Nat. Struct. Mol. Biol.*, **14**, 449–451.

47. Benarroch,D., Smith,P. and Shuman,S. (2008) Characterization of a trifunctional mimivirus mRNA capping enzyme and crystal structure of the RNA triphosphatase domain. *Structure*, **16**, 501–512.

48. Pena,L., Yanez,R.J., Revilla,Y., Vinuela,E. and Salas,M.L. (1993) African swine fever virus guanylyltransferase. *Virology*, **193**, 319–328.

49. Larsen,M., Gunge,N. and Meinhardt,F. (1998) Kluyveromyces lactis killer plasmid pGKL2: evidence for a viral-like capping enzyme encoded by ORF3. *Plasmid*, **40**, 243–246.

50. Tiggemann,M., Jeske,S., Larsen,M. and Meinhardt,F. (2001) Kluyveromyces lactis cytoplasmic plasmid pGKL2: heterologous expression of Orf3p and proof of guanylyltransferase and mRNA-triphosphatase activities. *Yeast*, **18**, 815–825.

51. Li,Y.I., Chen,Y.J., Hsu,Y.H. and Meng,M. (2001) Characterization of the AdoMet-dependent guanylyltransferase activity that is associated with the N terminus of bamboo mosaic virus replicase. *J. Virol.*, **75**, 782–788.

52. Huang,Y.L., Hsu,Y.H., Han,Y.T. and Meng,M. (2005) mRNA guanylation catalyzed by the S-adenosylmethionine-dependent guanylyltransferase of bamboo mosaic virus. *J. Biol. Chem.*, **280**, 13153–13162.

53. Coleman,T.M., Wang,G. and Huang,F. (2004) Superior 5′ homogeneity of RNA from ATP-initiated transcription under the T7 phi 2.5 promoter. *Nucleic Acids Res.*, **32**, e14.

54. Cong,P. and Shuman,S. (1995) Mutational analysis of mRNA capping enzyme identifies amino acids involved in GTP binding, enzyme-guanylate formation, and GMP transfer to RNA. *Mol. Cell Biol.*, **15**, 6222–6231.

55. Lavigne,R., Sun,W.D. and Volckaert,G. (2004) PHIRE, a deterministic approach to reveal regulatory elements in bacteriophage genomes. *Bioinformatics*, **20**, 629–635.

56. Camacho,C., Coulouris,G., Avagyan,V., Ma,N., Papadopoulos,J., Bealer,K. and Madden,T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, **10**, 421.

57. Scholl,D., Adhya,S. and Merril,C.R. (2002) Bacteriophage SP6 is closely related to phages K1-5, K5, and K1E but encodes a tail protein very similar to that of the distantly related P22. *J. Bacteriol.*, **184**, 2833–2836.

58. Makarova,O.V., Makarov,E.M., Sousa,R. and Dreyfus,M. (1995) Transcribing of Escherichia coli genes with mutant T7 RNA polymerases: stability of lacZ mRNA inversely correlates with polymerase speed. *Proc. Natl. Acad. Sci. U.S.A.*, **92**, 12250–12254.

59. De la Pena,M., Kyrieleis,O.J. and Cusack,S. (2007) Structural insights into the mechanism and evolution of the vaccinia virus mRNA cap N7 methyl-transferase. *EMBO J.*, **26**, 4913–4925.

60. Dube,S., Fisher,J.W. and Powell,J.S. (1988) Glycosylation at specific sites of erythropoietin is essential for biosynthesis, secretion, and biological function. *J. Biol. Chem.*, **263**, 17516–17521.

61. Dauty,E. and Verkman,A.S. (2005) Actin cytoskeleton as the principal determinant of size-dependent DNA mobility in cytoplasm: a new barrier for non-viral gene delivery. *J. Biol. Chem.*, **280**, 7823–7828.

62. Gallie,D.R. (1991) The cap and poly(A) tail function synergistically to regulate mRNA translational efficiency. *Genes Dev.*, **5**, 2108–2116.

63. Parker,R. and Song,H. (2004) The enzymes and control of eukaryotic mRNA turnover. *Nat. Struct. Mol. Biol.*, **11**, 121–127.

64. Cao,D. and Parker,R. (2001) Computational modeling of eukaryotic mRNA turnover. *RNA*, **7**, 1192–1212.

65. Brawerman,G. (1981) The Role of the poly(A) sequence in mammalian messenger RNA. *CRC Crit. Rev. Biochem.*, **10**, 1–38.

66. Zhao,J., Hyman,L. and Moore,C. (1999) Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol. Mol. Biol. Rev.*, **63**, 405–445.

67. Wahle,E. (1995) Poly(A) tail length control is caused by termination of processive synthesis. *J. Biol. Chem.*, **270**, 2800–2808.

68. Sheets,M.D. and Wickens,M. (1989) Two phases in the addition of a poly(A) tail. *Genes Dev.*, **3**, 1401–1412.

69. Peng,J., Murray,E.L. and Schoenberg,D.R. (2008) In vivo and in vitro analysis of poly(A) length effects on mRNA translation. *Methods Mol. Biol.*, **419**, 215–230.

70. Hornung,V., Ellegast,J., Kim,S., Brzozka,K., Jung,A., Kato,H., Poeck,H., Akira,S., Conzelmann,K.K., Schlee,M. *et al.* (2006) 5'-Triphosphate RNA is the ligand for RIG-I. *Science*, **314**, 994–997.

71. Yoneyama,M., Kikuchi,M., Natsukawa,T., Shinobu,N., Imaizumi,T., Miyagishi,M., Taira,K., Akira,S. and Fujita,T. (2004) The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. *Nat. Immunol.*, **5**, 730–737.

72. Daffis,S., Szretter,K.J., Schriewer,J., Li,J., Youn,S., Errett,J., Lin,T.Y., Schneller,S., Zust,R., Dong,H. *et al.* (2010) 2'-O methylation of the viral mRNA cap evades host restriction by IFIT family members. *Nature*, **468**, 452–456.

73. Mauer,J., Luo,X., Blanjoie,A., Jiao,X., Grozhik,A.V., Patil,D.P., Linder,B., Pickering,B.F., Vasseur,J.J., Chen,Q. *et al.* (2017) Reversible methylation of m(6)Am in the 5' cap controls mRNA stability. *Nature*, **541**, 371–375.

74. Meyer,K.D., Saletore,Y., Zumbo,P., Elemento,O., Mason,C.E. and Jaffrey,S.R. (2012) Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell*, **149**, 1635–1646.

75. Squires,J.E., Patel,H.R., Nousch,M., Sibbritt,T., Humphreys,D.T., Parker,B.J., Suter,C.M. and Preiss,T. (2012) Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. *Nucleic Acids Res.*, **40**, 5023–5033.

76. Le Hir,H., Nott,A. and Moore,M.J. (2003) How introns influence and enhance eukaryotic gene expression. *Trends Biochem. Sci.*, **28**, 215–220.

77. Matsumoto,K., Wassarman,K.M. and Wolffe,A.P. (1998) Nuclear history of a pre-mRNA determines the translational activity of cytoplasmic mRNA. *EMBO J.*, **17**, 2107–2121.

78. Fink,T.L., Klepcyk,P.J., Oette,S.M., Gedeon,C.R., Hyatt,S.L., Kowalczyk,T.H., Moen,R.C. and Cooper,M.J. (2006) Plasmid size up to 20 kbp does not limit effective in vivo lung gene transfer using compacted DNA nanoparticles. *Gene Ther.*, **13**, 1048–1051.

79. Labat-Moleur,F., Steffan,A.M., Brisson,C., Perron,H., Feugeas,O., Furstenberger,P., Oberling,F., Brambilla,E. and Behr,J.P. (1996) An electron microscopy study into the mechanism of gene transfer with lipopolyamines. *Gene Ther.*, **3**, 1010–1017.

80. Tachibana,R., Harashima,H., Shinohara,Y. and Kiwada,H. (2001) Quantitative studies on the nuclear transport of plasmid DNA and gene expression employing nonviral vectors. *Adv. Drug Deliv. Rev.*, **52**, 219–226.

81. Yin,H., Kanasty,R.L., Eltoukhy,A.A., Vegas,A.J., Dorkin,J.R. and Anderson,D.G. (2014) Non-viral vectors for gene-based therapy. *Nat. Rev. Genet.*, **15**, 541–555.

82. Schmidt-Wolf,G.D. and Schmidt-Wolf,I.G. (2003) Non-viral and hybrid vectors in human gene therapy: an update. *Trends Mol. Med.*, **9**, 67–72.

83. Eaton,H.E., Kobayashi,T., Dermody,T.S., Johnston,R.N., Jais,P.H. and Shmulevitz,M. (2017) African swine fever virus NP868R capping enzyme promotes reovirus rescue during reverse genetics by promoting reovirus protein expression, virion assembly, and RNA incorporation into infectious virions. *J. Virol.*, **91**, e02416-16.

84. Duncan,M.R., Stanish,S.M. and Cox,D.C. (1978) Differential sensitivity of normal and transformed human cells to reovirus infection. *J. Virol.*, **28**, 444–449.

85. Mookhtiar,K.A., Peluso,P.S., Muller,D.K., Dunn,J.J. and Coleman,J.E. (1991) Processivity of T7 RNA polymerase requires the C-terminal Phe882-Ala883-COO- or "foot". *Biochemistry*, **30**, 6305–6313.

86. Gardner,L.P., Mookhtiar,K.A. and Coleman,J.E. (1997) Initiation, elongation, and processivity of carboxyl-terminal mutants of T7 RNA polymerase. *Biochemistry*, **36**, 2908–2918.