



This is a repository copy of *Synthetic G-quadruplex components for predictable, precise two-level control of mammalian recombinant protein expression*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/230140/>

Version: Published Version

---

**Article:**

Pohle, M., Curry, E. [orcid.org/0000-0002-5123-8866](https://orcid.org/0000-0002-5123-8866), Holden, R. et al. (2 more authors) (2025) Synthetic G-quadruplex components for predictable, precise two-level control of mammalian recombinant protein expression. *Nucleic Acids Research*, 53 (14). gkaf732. ISSN 0305-1048

<https://doi.org/10.1093/nar/gkaf732>

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

# Synthetic G-quadruplex components for predictable, precise two-level control of mammalian recombinant protein expression

Melinda Pohle<sup>1,\*</sup>, Edward Curry<sup>1b</sup>, Ryan Holden<sup>1</sup>, Suzanne Gibson<sup>2</sup>, Adam Brown<sup>1,\*</sup>

<sup>1</sup>Department of Chemical and Biological Engineering, University of Sheffield, Sheffield, S1 3JD, United Kingdom

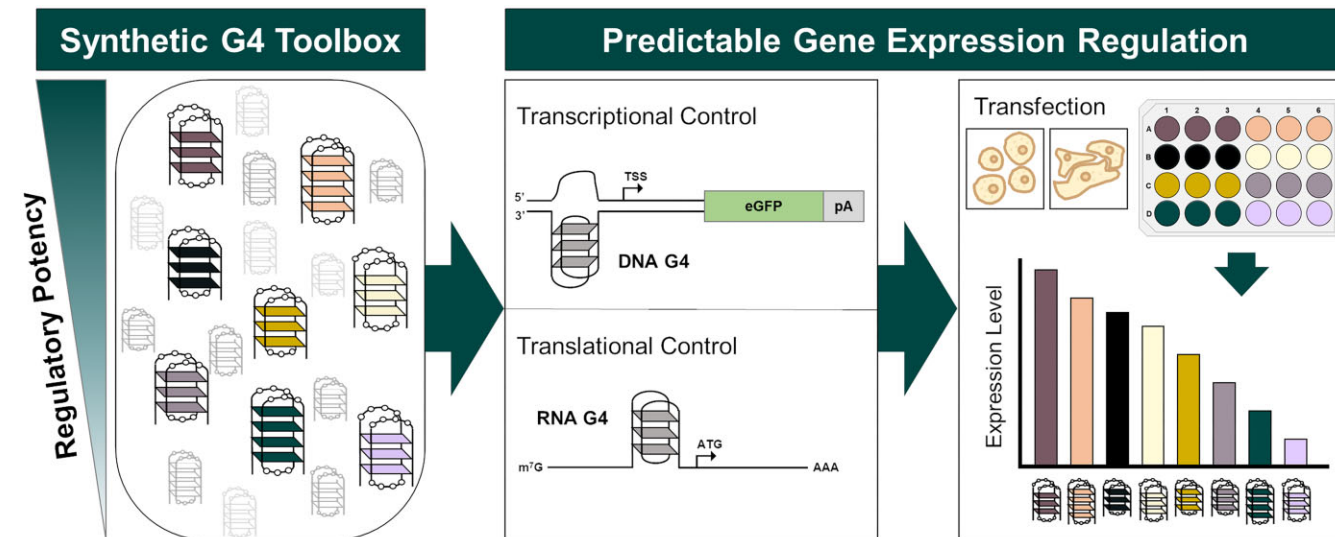
<sup>2</sup>BioPharmaceutical Development, R&D, AstraZeneca, Cambridge, CB2 0AA, United Kingdom

\*To whom correspondence should be addressed. Email: adam.brown@sheffield.ac.uk  
 Correspondence may also be addressed to Melinda Pohle. Email: mgo1@sheffield.ac.uk

## Abstract

Control of mammalian recombinant protein expression underpins the *in vitro* manufacture and *in vivo* performance of all biopharmaceutical products. However, routine optimization of protein expression levels in these applications is hampered by a paucity of genetic elements that function predictably across varying molecular formats and host cell contexts. Herein, we describe synthetic genetic components that are specifically built to simplify bioindustrial expression cassette design processes. Synthetic G-quadruplex elements with varying sequence feature compositions were systematically designed to exhibit a wide range of regulatory activities and inserted into identified optimal positions within a standardized, bioindustry compatible core promoter-5'UTR control unit. The resulting library tuned protein production rates over two orders of magnitude, where DNA and RNA G-quadruplexes could be deployed individually or in combination to achieve synergistic two-level regulatory control. We demonstrate these components can predictably and precisely tailor protein expression levels in (i) varying gene therapy and biomanufacturing cell hosts and (ii) both plasmid DNA and synthetic messenger RNA contexts. As an exemplar use case, a vector design platform was created to facilitate rapid optimization of polypeptide expression ratios for difficult-to-express multichain products. Permitting simple, predictable titration of recombinant protein expression, this technology should prove useful for gene therapy and biopharmaceutical manufacturing applications.

## Graphical abstract



## Introduction

Recombinant protein expression in mammalian cells is underpinned by genetic technologies that control flux through the key steps in product biosynthesis, namely gene transcription, messenger RNA (mRNA) translation, and polypeptide translocation [1]. In most cases, these processes are still con-

trolled by assemblies of viral and endogenous elements that are generally known to encode reasonably high biosynthetic rates, such as the CMV-IE1 promoter, human alpha-globin 5' UTR, and SV40 3'UTR [2-4]. However, increasingly complex protein expression objectives have created a requirement for more sophisticated gene expression control technologies

Received: September 16, 2024. Revised: June 12, 2025. Editorial Decision: July 11, 2025. Accepted: July 18, 2025

© The Author(s) 2025. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

that can precisely tailor transcription and translation rates in molecule- and context-specific manners. These include (i) advanced mammalian cell factory engineering strategies, where multiple genes must be expressed at correct stoichiometric ratios to achieve designed phenotypes [5]; (ii) gene therapy, where target proteins need to be expressed at levels that provide therapeutic efficacy without causing off-target effects [6]; and (iii) protein biopharmaceutical manufacturing, where the expression levels of multiple polypeptide chains must be optimally balanced to facilitate correct protein folding and assembly [7]. Indeed, in the latter example, many promising product molecules are currently considered difficult or impossible to manufacture due to improperly assembled proteins inducing cellular stress responses [8, 9].

Predictable and precise control of recombinant protein expression levels requires design and validation of synthetic genetic component libraries. While synthetic signal peptides can be used to tailor translocation levels, their activity is highly polypeptide-specific, and this lack of orthogonality significantly restricts their routine use in vector design spaces owing to the associated *in vitro* screening burden [10]. Conversely, synthetic promoter and UTR elements have been developed to customize transcription and translation levels, but their performance is typically cell-type and cell-context specific, dependent on relative abundances of cognate nucleic-acid binding proteins [11, 12]. This necessitates design of cell-host specific genetic component libraries, which are only currently available for a handful of mammalian cells [13–15]. Accordingly, for most mammalian protein expression applications, including the vast majority of cell-type-specific gene therapies, it is currently intractable to design genetic construct assemblies that encode optimal biosynthetic rates, owing to a paucity of appropriate validated regulatory components.

Secondary structure-based control elements represent a promising and relatively unexplored means to achieve sophisticated mammalian protein expression control. Nucleic acid strands can fold into a variety of stable structures, including loops, hairpins, helices, and pseudoknots. When present in appropriate positions within promoters and UTRs, their formation can sterically hinder the binding and progression of transcriptional and translational apparatus, thereby facilitating biosynthetic rate tuning by selecting elements with varying inhibitory activities (e.g. varying thermodynamic stabilities) [16]. An elegant example of designing secondary structure components to tailor mammalian protein expression for bioindustrial applications is the study by Eisenhut *et al.*, where 25 distinct RNA hairpins were created by varying GC content, minimum free energy, and positioning relative to the 5' cap [17]. This library was used to vary translation rates over two orders of magnitude in common mammalian bioproduction cell hosts, exemplifying the utility of secondary structure elements to solve complex protein expression challenges. However, the available regulatory component toolbox remains relatively sparsely populated and is particularly lacking technologies that can (i) predictably tune transcription levels and, accordingly, (ii) enact two-level protein expression control by facilitating simultaneous titration of both transcription and translation rates.

G-quadruplexes (G4s) are four-stranded secondary structures that form from guanine-rich nucleic acid sequences, comprising multiple G-tetrads (a square planar arrangement of four guanine bases interconnected through Hoogsteen hydrogen bonding) [18]. Prevalent in both the human genome

and transcriptome, they form in DNA and RNA molecules to regulate transcription and translation rates [19–21]. While some G4s have been shown to upregulate protein expression, e.g. via transcription factor recruitment [22, 23], when appropriately positioned within promoters and 5' UTRs, they can act to reduce protein biosynthesis by disrupting the function of general transcriptional or translational machinery [24–26]. Although hundreds of discrete G4s have been experimentally validated, mainly those that regulate expression of oncogenes such as c-MYC and NRAS [27–29], these endogenous components cannot be easily applied to bioindustrial applications as their functionality has only been profiled within the context of natural human promoters and UTRs (e.g. [30–33]), which typically display unpredictable or undesirable expression dynamics [16, 34, 35]. Moreover, they are usually studied individually, where information on their relative activities is not generally available, preventing their use as amalgamated libraries where elements can be rationally selected and combined to achieve user-defined protein expression rates. Accordingly, harnessing the potential of G4s for cell-agnostic two-level precision protein expression control requires bottom-up creation of built-for-purpose synthetic libraries, comprising well-characterized elements that are designed and validated to function within standardized industrially relevant promoter-UTR contexts.

In this study, we develop the first synthetic G4 element library designed specifically for use in high-value bioindustrial applications. Comprising complementary DNA and RNA G4 components operating at optimized positions within a standardized genetic architecture, this technology enables precise two-level control of mammalian protein expression over two orders of magnitude. Validated to perform predictably in varying cell types and molecular formats (i.e. plasmid DNA, synthetic mRNA, multi-chain products), these synthetic G4 controllers could be utilized to predictably fine-tune recombinant protein expression levels in *in vitro* biomanufacturing and *in vivo* biomedical applications.

## Materials and methods

### Expression cassette and vector construction

The bioindustrial protein expression control unit (BPCU) was constructed by fusing a minimal 41 bp hCMV-IE1 core promoter (GAN M60321.1, nucleotides 1110–1150) to a 57 bp proprietary AstraZeneca 5' UTR [36]. BPCU was inserted upstream of the gene of interest [containing an SV40 polyA 3' UTR (GAN LT727517.1, nucleotides 1449–1676)] and downstream of the hCMV-IE1 proximal promoter (GAN M60321.1, nucleotides 517–1109) in reporter vectors encoding secreted alkaline phosphatase (SEAP), eGFP, mAbT HC, and mAbT LC expression [reporter vector *de novo* synthesized by GeneArt (Regensburg, Germany)]. G4 elements (listed in [Supplementary Table S1](#)) were inserted into BPCU by site-directed mutagenesis using the Q5® Site-Directed Mutagenesis Kit (New England Biolabs, USA). Plasmid templates for *in vitro* transcription (IVT) reactions were generated by cloning appropriate RNA G4-eGFP sequences into pRNA128A [37] using NheI and NotI (New England Biolabs, USA) restriction sites. Vectors for stable transfections were created by inserting appropriate DNA G4-eGFP sequences into pMCS-Gaussia Luc (Thermo Fisher Scientific, USA) using KpnI and NotI restriction sites. All constructed plasmids used

in this study were validated via DNA sequencing and purified to transfection-grade quality using the QIAGEN Plasmid Plus Midi Kit (QIAGEN, USA).

### Cell culture

CHO-K1 derived host cells (AstraZeneca, UK), hereby referred to as Chinese hamster ovary (CHO) cells, and CHO-S (Thermo Fisher Scientific, USA) cells were routinely cultivated in CD-CHO medium (Thermo Fisher Scientific, USA) supplemented with 6 mM L-glutamine and 8 mM L-glutamine, respectively. Expi293F cells (Thermo Fisher Scientific, USA), hereby referred to as HEK cells, were routinely cultivated in Freestyle™ 293 expression medium (Thermo Fisher Scientific, USA). Cultures were maintained at 37°C, 5% CO<sub>2</sub> with 140 rpm orbital shaking. CHO cells were passaged every 3–4 days at a seeding density of  $0.2 \times 10^6$  cells/ml, CHO-S cells were passaged every 2–3 days at a seeding density of  $0.2 \times 10^6$  cells/ml, and HEK cells were passaged every 3–4 days at a seeding density of  $0.3\text{--}0.4 \times 10^6$  cells/ml. HepG2 cells (ATCC HB-8065) were cultured in RPMI 1640 medium (Thermo Fisher Scientific, USA) supplemented with 10% fetal bovine serum and maintained at 37°C under 5% CO<sub>2</sub>. Cell concentration and viability were measured with the VI-CELL BLU automated cell counter and cell viability analyzer (Beckman Coulter, USA).

### Transient transfection

CHO cells were transiently transfected by electroporation using the Amaxa Nucleofector system in combination with the SG Cell Line 96-well Nucleofector® Kit (Lonza, Switzerland).  $1.86 \times 10^6$  cells per well were transfected with 750 ng plasmid DNA or 400 ng mRNA following manufacturer's instructions. Transfected cells were transferred to a 24 shallow-well plate and incubated in a total volume of 750 µl for 48 h at 37°C, 5% CO<sub>2</sub> at 240 rpm orbital shaking. IgG1 (mAbT) LC- and HC- encoding plasmids were transfected at ratios maintaining equal HC:LC gene copy numbers, and incubated for 5 days with a 10% (v/v) feed added 48 h post-transfection containing 1:1 (v/v) Efficient Feed A (Thermo Fisher Scientific, USA) and Efficient Feed B (Thermo Fisher Scientific, USA). HEK cells were transiently transfected with PEIpro® (Polyplus, France), whereby  $1.5 \times 10^6$  cells were transfected with 600 ng plasmid DNA mixed with 1.8 µl PEIpro® (1:3 ratio of DNA to PEI) following manufacturer's instructions. Cells were incubated in a 24-shallow-well plate in a total volume of 750 µl for 48 h at 37°C, 5% CO<sub>2</sub> at 240 rpm orbital shaking. HepG2 cells were transfected with Lipofectamine™ 3000 Transfection Reagent (Thermo Fisher Scientific, USA) according to manufacturer's instructions in a 24-well plate scale and incubated for 48 h.

### Stable transfection

CHO-S cells were stably transfected by electroporation using the Amaxa Nucleofector system in combination with the Cell Line Nucleofector® Kit V (Lonza, Switzerland).  $5.0 \times 10^6$  cells per cuvette were transfected with 2 µg plasmid and incubated in T75 flasks at 37°C under 5% CO<sub>2</sub> in 13 ml CD-CHO medium supplemented with 8 mM L-glutamine. Forty-eight hours post-transfection, the medium was changed to CD-CHO selection medium including 8 mM L-glutamine, 10 µl/ml Puromycin, and Anti-Clumping agent (Thermo Fisher Scientific, USA) at a 1:250 (v/v) dilution. Once cultures reached >80%, they were transferred from static to shaking

conditions and passaged every 3 days until they reached a viability of >95%, at which point they were cryopreserved.

### Recombinant protein quantification

eGFP protein expression was quantified 48 h post-transfection using the SpectraMax iD5 Microplate Reader (Molecular Devices, USA) with 485 nm excitation and 535 nm emission. Transfection efficiency was determined using the Countess II FL Automated Cell Counter (Thermo Fisher Scientific, USA). SEAP expression was quantified 48 h post-transfection using the Secreted Alkaline Phosphatase Reporter Gene Assay Kit (Cayman Chemical, USA) according to manufacturer's instructions. For IgG1 mAb protein titer quantification, the ELISA quantification kit for human immunoglobulins G (RD-Biotech, France) was used in accordance with the manufacturer's protocol 5 days post-transfection. All assays were performed using the SpectraMax iD5 Microplate Reader. The synergistic effect of combining two G4 elements within a single construct was predicted using the following equation, where REU is the relative protein expression units (REUs) they encode.

$$\text{REU}_{\text{combined}} = \frac{(\text{REU}_{\text{DNA}} \times \text{REU}_{\text{RNA}})}{100}$$

### Circular dichroism

DNA oligos (Integrated DNA Technologies, USA) were prepared at 5 µM in 50 mM Tris-HCl (pH 7.5) with 100 mM KCl. The samples were denatured at 95°C for 5 min and then cooled down 0.2°C/min to 20°C. Circular dichroism (CD) experiments were performed at 25°C using a JASCO J-810 spectropolarimeter (JASCO, USA) in 1-mm quartz cuvettes. CD scans were taken from 320 to 200 nm at 100 nm/min. Melting curves were obtained by heating the samples from 20 to 100°C at a controlled rate of 2°C/min. Changes in ellipticity were monitored at the wavelength of the highest peak for each oligo every 0.5°C.

### In vitro transcription

Plasmid templates for IVT were linearized with BsaXI and purified by ethanol precipitation. IVT reactions used the HiScribe® T7 High Yield RNA Synthesis Kit (New England Biolabs, USA) following the manufacturer's instructions and 3'-O-Me-m<sup>7</sup>G(5')ppp(5')G RNA cap (New England Biolabs, USA) was used as the cap structure analog. Samples were digested with DNase I (Thermo Fisher Scientific, USA) for 15 min at 37°C and immediately purified using the Monarch® RNA Cleanup Kit (New England Biolabs, USA). RNA samples were diluted to the desired working concentration and stored at –80°C prior to use.

### Chemical ligands

CHO-S cell pools were seeded at  $1 \times 10^6$  cells/ml in 24 shallow-well plates prior to ligand addition. Phen-DC3 (Selleck Chemicals, USA), Pyridostatin (Merck, Germany), TMPyP4 (Selleck Chemicals, USA), and 360A (ApexBio Technology, USA) were added to a final concentration of 25, 5, 25, and 20 µM, respectively (concentrations optimized to maintain cell viabilities >90%), and cells were maintained for 48 h at 37°C, 5% CO<sub>2</sub> with 140 rpm orbital shaking. Phen-DC3 and 360A were dissolved in DMSO and Pyridostatin and TMPyP4 in water.



## Statistical analysis

A one-way ANOVA was performed to evaluate statistical differences with statistical significance being defined as  $P \leq .05$  ( $* = P \leq .05$ ,  $** = P \leq .01$ ,  $*** = P \leq .001$ ,  $**** = P \leq .0001$ ). Tukey's post-hoc test was used to determine statistical differences to other groups and the Dunnett's post-hoc test was used to determine statistical differences to a specific control. Statistical tests were performed using GraphPad Prism 10 software.

## Results and discussion

### Optimization of DNA and RNA G4 element positioning within a standardized genetic chassis

The human CMV-IE1 core promoter is the most widely utilized regulatory element for controlling mammalian recombinant gene transcription in high-value applications, including gene therapy and biopharmaceutical production, and has been shown to function in conjunction with varying proximal promoter partners [38–40]. While a range of human and mouse sequences are available for use as translational controllers, to facilitate deployment in biomedical settings, we combined the CMV-IE1 core with a minimized 5' UTR that has been specifically developed and validated for use in bioproduction processes (AZ 5' UTR). This 98 bp BPCU is an ideal genetic chassis to house regulatory G4s that are designed to control transcription and translation rates by interfering with the activity of general machinery components. Given that G4 function is position-dependent [29, 41], as illustrated in Fig. 1A, we first sought to identify the optimal locations to place DNA and RNA G4s within this sequence context. As shown in Fig. 1B, model DNA (c-MYC) and RNA (NRAS) G4s were each individually inserted into 7 and 6 distinct sites, respectively, within the BPCU in eGFP-reporter plasmids harboring a CMV-IE1 proximal promoter, selected to drive a high protein expression set-point that can be tuned down by G4 activity [40, 42, 43]. Positions were chosen whereby G4 formation should theoretically interfere with binding and movement of biosynthetic machinery, including near key regulatory binding sites such as the TATA box, INR, and Kozak sequence [44, 45]. DNA and RNA G4-encoding sequences were inserted into the template and coding strand, respectively, to facilitate formation of secondary structures at the DNA and RNA level (i.e. RNA G4s form in the 5' UTR of eGFP-mRNA molecules).

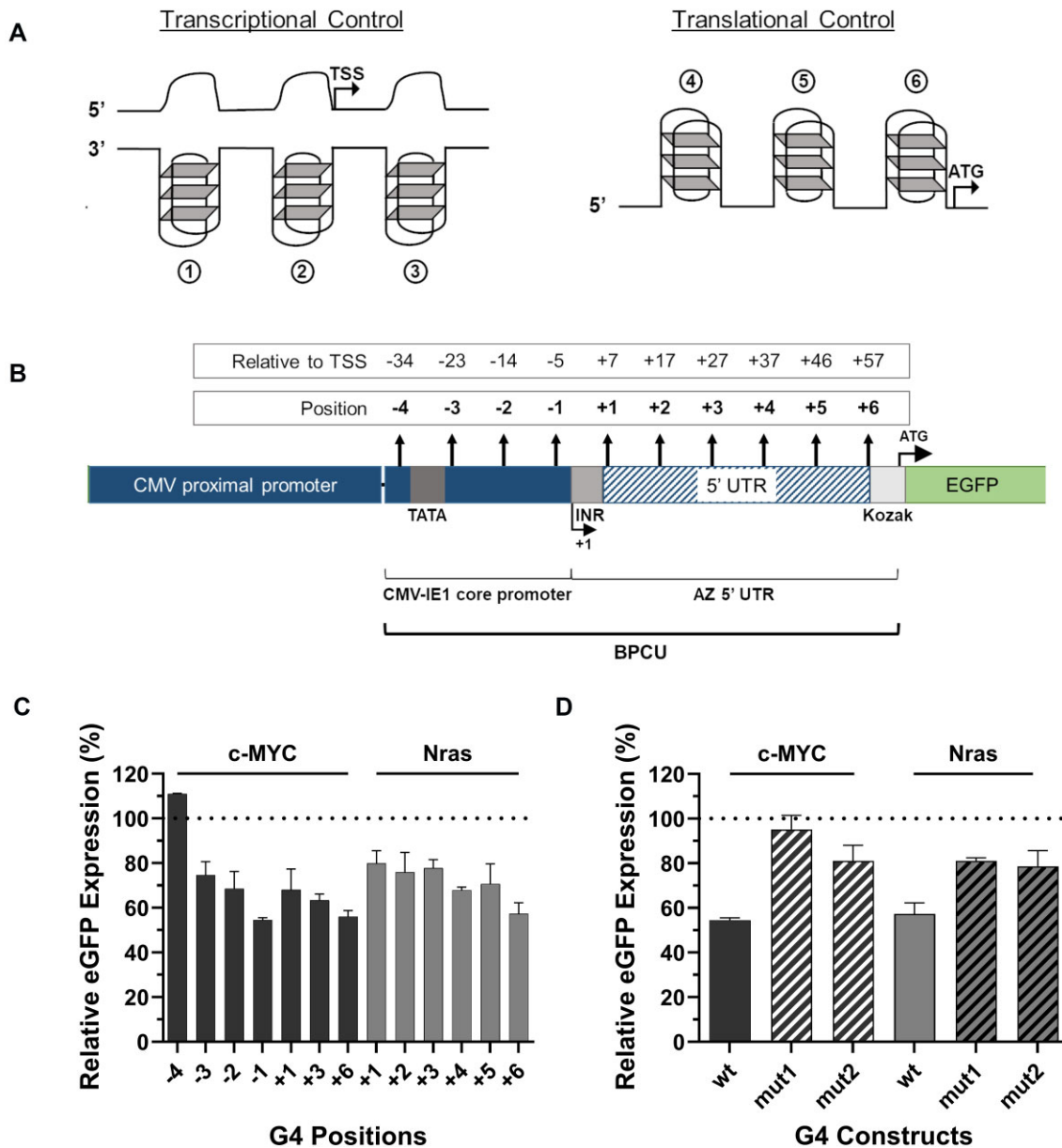
G4-containing reporter plasmids were independently transfected into CHO cells, selected as a model mammalian cell line due to it being the dominant cell factory for biopharmaceutical production, and eGFP expression levels were quantified 48 h later. Transfected cells were analyzed using the Countess II FL Automated Cell Counter to confirm consistently high transfection efficiency (>90%) and normal distribution of fluorescence intensity across the population, enabling quantification of relative G4 activities by measuring average protein expression levels. The data show that G4s are functional within the context of the BPCU, where protein expression was substantially downregulated by G4s acting at both the DNA and RNA level, as compared to the unengineered control construct (i.e. BPCU without G4 insertion, Fig. 1C). Indeed, G4 insertion had a negative regulatory effect on eGFP expression in 12 of 13 tested sequence positions, only being non-functional in the DNA position that was furthest from the transcriptional start site. G4s exerted greatest inhibitory effect when located directly upstream of the transcriptional and transla-

tional start sites, in both cases reducing eGFP expression by ~45% compared to the standard BPCU. These findings are consistent with computational analyses showing that endogenous G4s are preferentially located near transcription start sites [46]. Previous studies have found that positional effects on RNA G4 function are context-specific, depending on the 5' UTR sequence [29, 41]. Ultimately, position-optimization studies should be conducted for any new genetic chassis given that G4 inhibitory activity may be impacted by both the relative location of other regulatory elements, and the nucleotide-composition of flanking sequences either side of the insertion site that can affect G4 formation kinetics [47, 48]. Within the context of the BPCU, we concluded that positions –1 and +6 were optimal locations for DNA and RNA elements, respectively, where inhibitory effects of ~45% facilitate subsequent design of varying strength G4 components that can titrate protein expression levels both above and below this set-point.

To validate that decreases in eGFP expression were caused by secondary structure-based biosynthesis inhibition, we inserted mutated versions of the c-MYC and NRAS G4 elements into positions –1 and +6 in the BPCU. Mutants were designed to maintain high sequence similarity and equal length to originator elements while substantially decreasing their propensity to form G4 structures (predicted using QGSR mapper [49]—see [Supplementary Table S2](#)). As shown in Fig. 1D, mutated motifs reduced eGFP expression by 5%–20% compared to the standard BPCU. A generic insertional effect is unsurprising, as previous studies have shown that (i) mutated G4 motifs have a similar impact in other genetic contexts [50], and (ii) small changes to mammalian core promoters and 5' UTRs, including the CMV-IE1 core, typically reduce protein expression levels [51]. However, given that wild-type G4 sequences had a substantially greater inhibitory effect than disabled versions containing a minimal number of G-to-A mutations [ $9.5\times$  ( $P < .0001$ ),  $2.4\times$  ( $P = .0005$ ),  $2.3\times$  ( $P = .001$ ), and  $2.0\times$  ( $P = .003$ ) compared to c-MYC mut1, c-MYC mut2, NRAS mut1, and NRAS mut2, respectively], we inferred that these elements were functioning via secondary structure-based interference of transcription and translation processes. Accordingly, we concluded that this genetic architecture (i.e. the BPCU with optimized DNA and RNA G4 insertion points) could be employed to create a synthetic G4 component library for two-level precision control of recombinant protein expression.

### Synthetic G4 components facilitate predictable, tunable control of recombinant protein expression

G4s have a simple core structure, consisting of alternating runs of guanine tracts (G-tracts) and separating loops, with four possible key input design parameters, namely loop length, G-tract length, loop nucleotide composition, and G-tract number (Fig. 2A). Using a small number of exemplar elements, it has previously been shown that functional synthetic RNA G4s with different regulatory properties can be created by varying one or more of these design features [41]. Accordingly, we rationalized that a component library capable of precisely tailoring protein expression levels from 0% to 100% relative to the standard BPCU could be derived by designing DNA and RNA G4s with varying input parameter combinations. Such G4 variants should exhibit variable propensities to both form and maintain secondary structures within the standardized BPCU genetic chassis, facilitating development of regu-

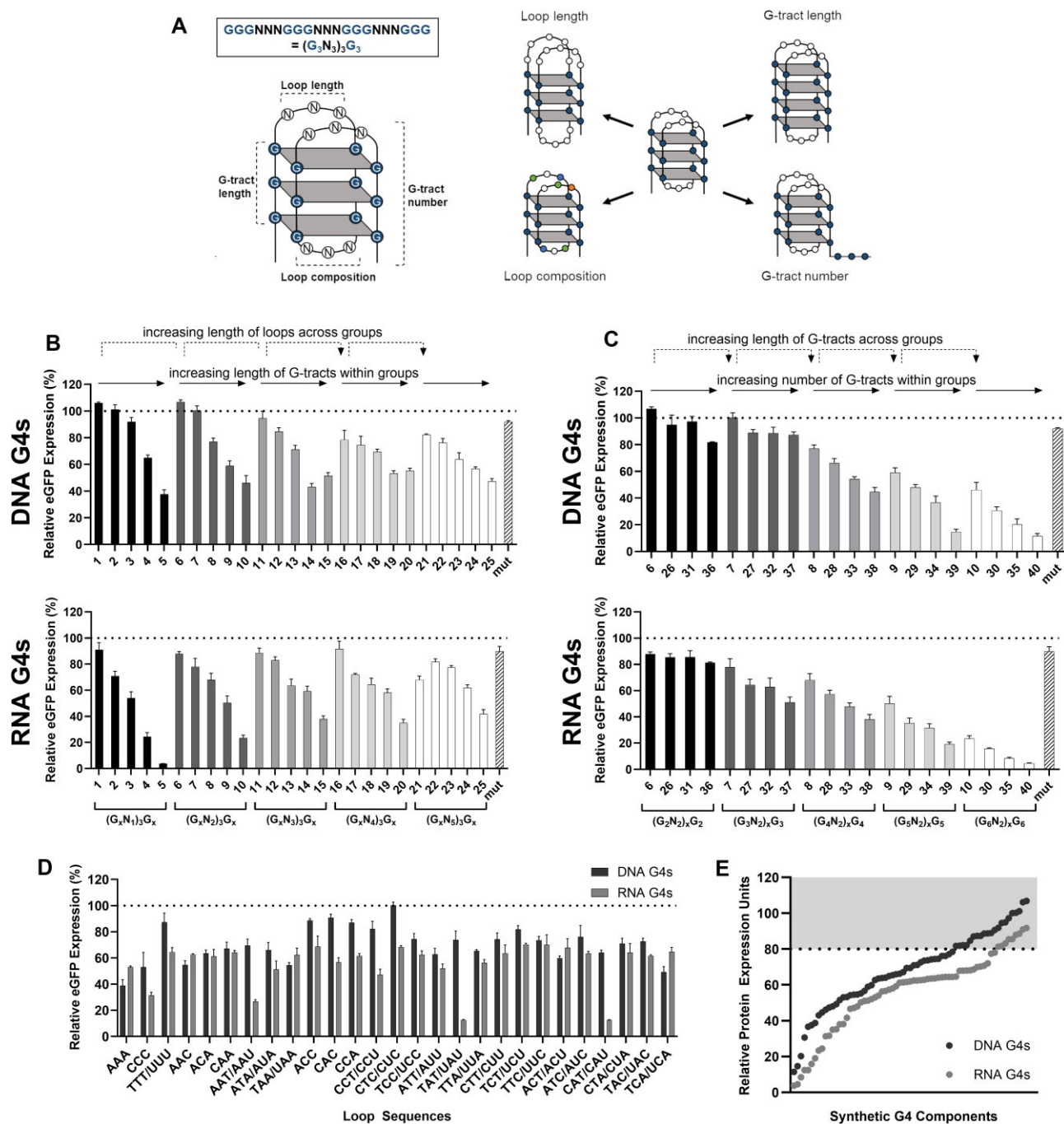


**Figure 1.** Endogenous G4 elements negatively regulate recombinant protein expression rates in a position-dependent manner. **(A)** Schematic presentation of general position-dependent mechanisms by which G4s in core promoter and 5' UTR elements can regulate protein expression rates by disrupting the binding [1, 4], movement [3, 5], and activity [2, 6] of transcription and translation initiation complexes. **(B)** G4 components were inserted into various positions in a BPCU in eGFP-reporter plasmids harboring a CMV-IE1 proximal promoter. CHO cells were transfected with **(C)** reporter vectors containing c-MYC (DNA) and Nras (RNA) G4 elements at different sites within the BPCU and **(D)** eGFP-reporters containing wild-type or mutated versions (see [Supplementary Table S2](#)) of c-MYC and Nras G4s in positions -1 and +6, respectively. eGFP expression was quantified 48 h post-transfection. Data are expressed as a percentage of the production exhibited by the control CMV-IE1 proximal promoter-BPCU ( $\Delta$ G4) construct (dotted line). Values represent the mean  $\pm$  SD of three independent experiments ( $n=3$ , each performed in triplicate). TSS—transcriptional start site.

latory elements with different inhibitory activities (i.e. varying abilities for steric hindrance-mediated downregulation of transcription and translation rates). Based on previous studies showing loop nucleotide composition to have a relatively minimal impact on G4 thermal stability [52], we adopted two distinct design strategies, namely (i) keeping G-tract number constant ( $N = 4$ ) and varying G-tract length ( $N = 2-6$ ) and loop length ( $N = 1-5$ ), and (ii) keeping loop length constant ( $N = 2$ ) and varying G-tract length ( $N = 2-6$ ) and number ( $N = 4-7$ ). In both cases we applied a full-factorial design and tested each element at the DNA and RNA level, resulting in a library of 80 distinct G4 components containing loop se-

quences that maintained equal A:T:C nucleotide ratios within each construct (all synthetic G4 sequences are provided in [Supplementary Table S1](#)).

Synthetic G4s were individually inserted in the identified optimal positions in the BPCU (DNA elements in position -1, RNA elements in position +6, Fig. 1) downstream of a CMV-IE1 proximal promoter in eGFP-reporter vectors and transfected into CHO cells. Quantification of eGFP expression levels 48 h post-transfection showed that engineered G4s displayed a wide range of inhibitory activities, driving expression levels between 11%–107% and 4%–90% at the DNA and RNA levels, respectively, compared to the standard BPCU



**Figure 2.** Synthetic DNA and RNA G4 components enable precision tuning of recombinant protein expression levels. **(A)** Core G4 structure, highlighting four key input design parameters. **(B, C)** Synthetic G4 components (listed in [Supplementary Table S1](#)) with varying sequence feature compositions were inserted into positions  $-1$  (DNA) and  $+6$  (RNA), respectively, in the BPCU (see Fig. 1) downstream of a CMV-IE1 proximal promoter in eGFP-reporter vectors. **(D)** The loop sequence within G4 motif 13 ( $(G_4N_3)_3G_4$ ) was varied at both the DNA and RNA levels. In panels (B–D), CHO cells were transfected with eGFP-reporter plasmids prior to protein quantification after 48 h. Data are expressed as a percentage of the production exhibited by the control CMV-IE1 proximal promoter-BPCU ( $\Delta$ G4) construct (dotted line). Values represent the mean  $\pm$  SD of three independent experiments ( $n=3$ , each performed in triplicate). **(E)** The library of synthetic G4 components encodes a wide range of recombinant protein expression set points; the gray-shaded area represents the generic insertional effects observed when mutated G4s are inserted into the BPCU.



control (Fig. 2B and C). We note that four new mutant variants of varying sizes exhibited the same generic insertional effect as seen previously, reducing eGFP expression by 5%–20% (data not shown). Accordingly, we concluded that only those components that reduced expression by >20%, 57 synthetic G4s in total, were exerting secondary structure-based control of protein biosynthesis. The inhibitory activity of G4 sequences was highly correlated at the DNA and RNA levels (Pearson's correlation coefficient,  $r = 0.85$ ,  $P = < .0001$ , [Supplementary Fig. S1](#)), identifying that discrete designs behave similarly when utilized as transcriptional or translational controllers. As depicted in Fig. 2B and C, G4 performance could be predictably tailored by varying any of the selected input parameters, where inhibitory activities were particularly correlated with increasing G-tract length and number. Indeed, multiple linear regression modeling confirmed that G4 inhibitory activity could be accurately explained as a function of these three features (DNA elements model:  $R^2 = 0.85$   $P = < .0001$ ; RNA elements model:  $R^2 = 0.88$   $P = < .0001$ ). These findings correlate with previous oligonucleotide-based studies that have shown G-tract length, G-tract number, and loop length to be critical determinants of G4 thermal stabilities [41, 52–54]. To our knowledge, this represents the first model linking synthetic G4 sequence feature composition to protein expression regulatory function. We reasoned that the ability to accurately explain the performance of these synthetic components as a function of their sequence parameters will enhance both the confidence with which they can be applied to new applications (i.e. as unknown features are unlikely to cause unpredictable changes to activities) and the ability to (re)engineer future elements with defined characteristics. While the model can accurately predict the regulatory activity of our synthetic G4s as a function of three input design parameters (model summary data shown in [Supplementary Fig. S2](#)), full understanding of their performance in cellular environments would require molecular dynamic simulations. In particular, to deploy them predictably *in vivo*, it would be useful to determine how their ability to form machinery-impeding secondary structures (i.e. the balance of factors including stacking interactions, hydrogen bonding, etc.) is affected by changes in intracellular environment parameters such as ion concentrations [55].

We predicted that loop nucleotide content would have a negligible impact on G4 performance [52] and accordingly that these “non-functional sections” may provide opportunities to achieve other design objectives such as enhanced sequence efficiency (e.g. by incorporating other regulatory motifs), tailored immunostimulatory effects [56, 57], and optimized GC-content [58]. To evaluate this, we varied loop sequence within a standard G4 architecture (motif 13 - (G<sub>4</sub>N<sub>3</sub>)<sub>3</sub>G<sub>4</sub>), where all 27 possible 3-nucleotide permutations (using A, T, and C nucleotides only) were individually tested (e.g. ATC: GGGATCGGGATCGGGATCGGG). Quantitative evaluation in CHO cells confirmed that loop composition generally has minimal impact on component inhibitory activity, where 17 of 27 DNA and 23 of 27 RNA elements drove eGFP expression levels within one standard deviation of the mean (Fig. 2D). Notably, the three motifs that had the greatest impact on eGFP levels, RNA loops of AAU, UAU, and CAU, all introduced alternative start codons, highlighting the requirement for multi-parameter/objective optimization approaches when designing new synthetic nucleotide parts [59]. Discounting these three sequences (and any components that reduced

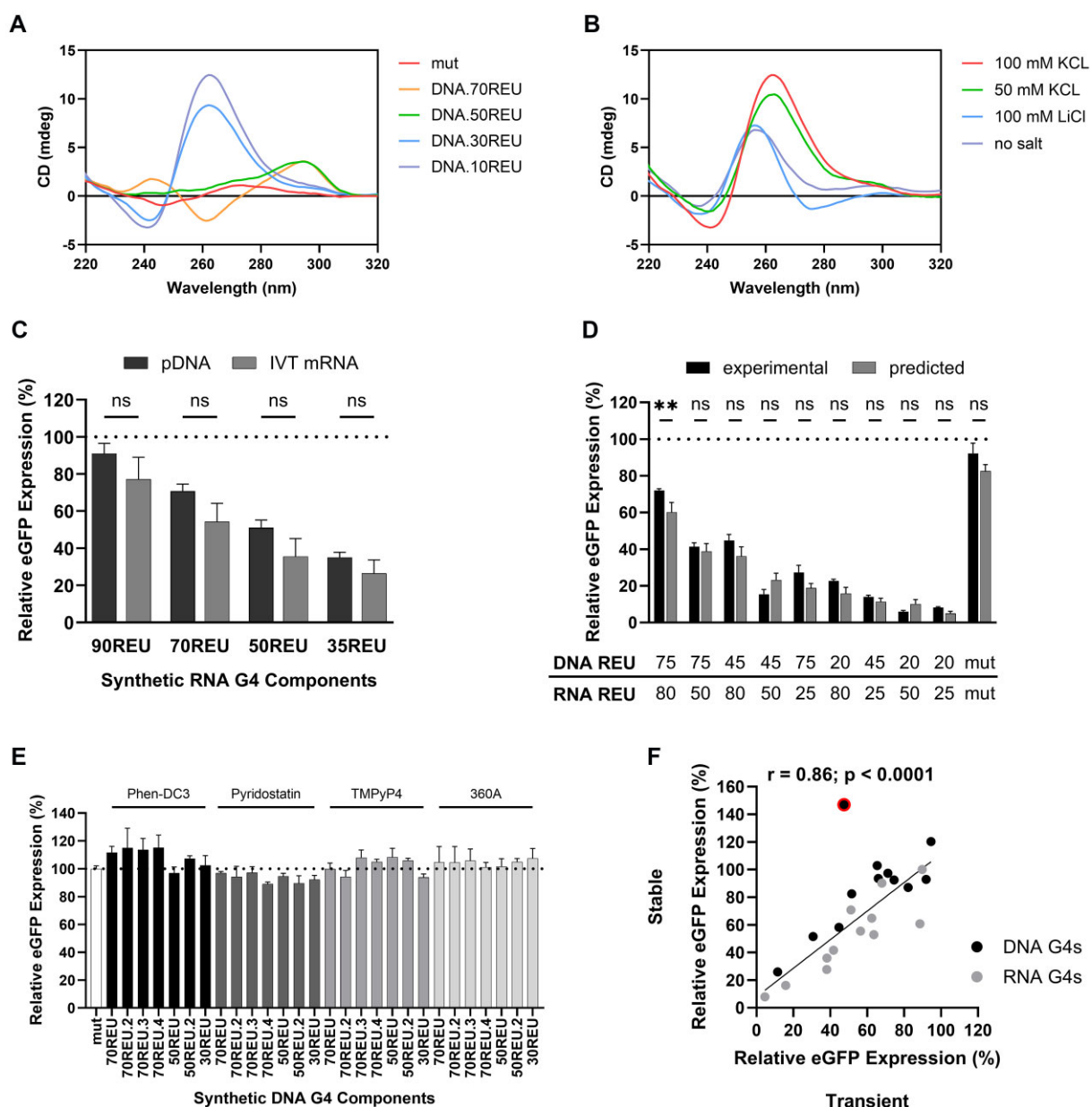
eGFP expression by <20% relative to the standard BPCU) based on the assumption that they encode undesirable variation in translational start sites, our final library comprises 101 distinct regulatory elements (by far the largest library of synthetic G4 controllers described to date) that facilitate predictable two-level precision control of recombinant protein expression over two orders of magnitude within a standardized, industrially relevant genetic chassis (Fig. 2E).

### Mechanistic dissection of synthetic G4 component functionality and performance

To facilitate their adoption in bioindustrial applications, we next sought to obtain detailed mechanistic information on the function and performance of our synthetic G4 components. First, we validated their ability to form secondary structures by CD spectroscopy, a standard technique for G4 structure analysis [60]. DNA G4 components with varying inhibitory activities (see Fig. 2) were analyzed to reveal their structural conformations, namely DNA13, DNA15, DNA30, and DNA40. Henceforth, all synthetic G4 components will be named according to the nucleic acid level at which they operate and the REUs they encode, e.g. DNA.70REU (DNA13), DNA.50REU (DNA15), DNA.30REU (DNA30), and DNA.10REU (DNA40; see [Supplementary Table S1](#) for all G4 nucleic acid sequences and names). As shown in Fig. 3A, all four components produced classical distinctive G4 CD spectra [61], where the mutated control sequence did not show characteristic G4 peaks and troughs. We note that G4s with the highest inhibitory activities (DNA.10REU and DNA.30REU) formed parallel structures (positive peak at 260 nm, negative peak at 240 nm), while medium (DNA.50REU) and low (DNA.70REU) strength components adopted hybrid and anti-parallel (positive peaks at 295 and 240 nm, negative peak at 260 nm) conformations, respectively [62]; the mutated sequence displayed the spectrum typical of unstructured single-stranded DNA [63]. To further assess the stability of the G4 with the strongest inhibitory activity (DNA.10REU), we profiled its conformation in an extended range of salt conditions (Fig. 3B) and performed a melting temperature analysis ([Supplementary Fig. S3](#)), which together confirmed this element forms a highly stable, consistent secondary structure. Based on these data and our previous finding that regulatory activities were correlated with sequence feature compositions, we concluded that synthetic components' relative performances were underpinned by their varying propensities to form and maintain G4 structures. Although we note that some *in vivo* applications, such as gene therapy, may require a more detailed analysis of intracellular secondary structure formation dynamics using G4-specific fluorescent probes [64] or antibodies [65].

Given the dramatic rise in applications utilizing synthetic mRNA, including protein replacement therapy, cancer immunotherapy, and vaccines against infectious diseases [4, 66], we next sought to validate that our RNA components display predictable performance when employed directly in mRNA molecular contexts. eGFP-mRNA molecules harboring RNA elements of varying inhibitory strengths were produced by IVT reactions, where purified transcripts were identical to those created intracellularly from DNA reporter plasmids (i.e. comprising the AZ 5' UTR containing a single G4 element in position +6, upstream of the eGFP CDS; see Fig. 1). Synthetic mRNAs were transfected into CHO cells prior





**Figure 3.** Synthetic G4s function predictably when deployed in combinations, mRNA molecules, and stable expression systems. CD spectra of (A) DNA oligonucleotides containing varying G4 motifs (all G4 motif sequences listed in [Supplementary Table S1](#)) and (B) DNA.10REU in varying salt conditions. CHO cells were transiently transfected with (C) synthetic eGFP-encoding mRNA molecules harboring different RNA G4 components, and (D) eGFP reporter plasmids containing varying combinations of DNA and RNA G4 elements. eGFP levels were quantified 48 h post-transfection and compared to those driven by (C) the same G4s in a DNA plasmid context and (D) predicted values, assuming constituent G4 pairs act synergistically (see Fig. 2). (E) CHO cells were stably transfected with G4-reporter plasmids coexpressing a puromycin selection marker gene. Stable cell pools were created and subsequently exposed to varying chemical ligands for 48 h prior to eGFP quantification. (F) Correlation between the regulatory activity of 24 synthetic G4 components in transient and stable expression systems (stable data represent cell pools measured in exponential phase growth). In panels (C) and (D), data are expressed as a percentage of the production exhibited by the control CMV-IE1 proximal promoter-BPCU ( $\Delta$ G4) construct (dotted line). Values represent the mean  $\pm$  SD of three independent experiments ( $n=3$ , each performed in triplicate). In panel (E), data are expressed as a percentage of the production achieved in control media ( $\Delta$ ligand, containing DMSO or water), normalized to the effect of ligand addition on G4 mutant (white bar). Values represent the mean  $\pm$  SD of three independent experiments ( $n=3$ , each performed in duplicate). In panel (F), data are expressed as a percentage of the production achieved using the control BPCU construct, values represent a single cell pool, and linear regression was conducted excluding the single outlier identified by analysis of standardized residuals (red outline).

to quantification of eGFP protein levels after 48 h. These data show that synthetic element performance in DNA vector and mRNA transcript contexts is highly correlated (Pearson's correlation coefficient,  $r = 0.99$ ,  $P = .008$ ; [Supplementary Fig. S4](#)), where RNA.90REU, RNA.70REU, RNA.50REU, and RNA35.REU drove eGFP expression levels of 75%, 55%, 35%, and 25%, respectively, relative to the unengineered control (Fig. 3C). Accordingly, synthetic RNA G4 components from the library can be utilized predictably to precisely control recombinant protein production in both plasmid DNA (e.g. biopharmaceutical production) and synthetic mRNA (e.g. cancer vaccines) molecular formats. We note that we were unable to synthesize mRNAs containing element RNA.5REU, presumably due to secondary structure-based effects, and therefore use of very strong G4s (i.e. REUs < 20) may require optimization of mRNA manufacturing processes.

Synthetic mammalian transcriptional and translational control elements are typically developed in isolation, where a lack of standardization, such as the use of varying genetic architectures and test systems, complicates efforts to use them in combination [67, 68]. G4-based regulatory controllers operating within a standardized BPCU should permit predictable, simultaneous titration of transcription and translation rates via selection of appropriate DNA and RNA parts. To test this, we created constructs containing all possible combinations of exemplar high, medium, and low strength DNA (DNA.20REU, DNA.45REU, and DNA.75REU) and RNA (RNA.25REU, RNA.50REU, and RNA.80REU) components. As shown in Fig. 3D, dual-controllers drove eGFP expression levels in CHO cells that were highly correlated with predicted values ( $r = 0.96$ ,  $P < .0001$ ; [Supplementary Fig. S5](#)), assuming synergistic effects of constituent elements (see “Materials and methods” section). Indeed, all observed activities were within  $\pm 12$  REU of expected outputs, indicating that composite parts functioned as expected to reduce the rate of both recombinant gene transcription and recombinant mRNA translation. We therefore concluded that DNA and RNA elements from the library can be predictably combined to achieve user-defined, stringent two-level control of recombinant protein expression, which may be particularly advantageous for applications that necessitate product biosynthesis rates to be maintained within defined limits, such as manufacture of toxic molecules.

We next evaluated the potential to convert our G4-based controllers into inducible expression systems. Many endogenous G4 sequences interact with chemical ligands that can destabilize or stabilize their structure to increase (ON systems) or decrease (OFF systems) protein expression levels [69, 70]. Such genetic switches facilitate sophisticated control strategies, including the ability to enact data- and process-responsive titration of system outputs [71, 72]. We tested four well-known G4 ligands (Phen-DC3, Pyridostatin, TMPyP4, and 360A [70, 73]) in combination with seven distinct DNA elements that were selected to (i) have varying feature compositions, based on findings that ligands may have specific binding preferences [74, 75], and (ii) exhibit medium to low strength (30–70 REU), to permit observation of up- and down-regulatory effects. These 28 expression systems were evaluated in an industrially relevant context, namely stable protein production in CHO cells, where there is a design objective to switch expression “ON” once maximum cell densities are reached in order to decouple growth and production phases during biomanufactur-

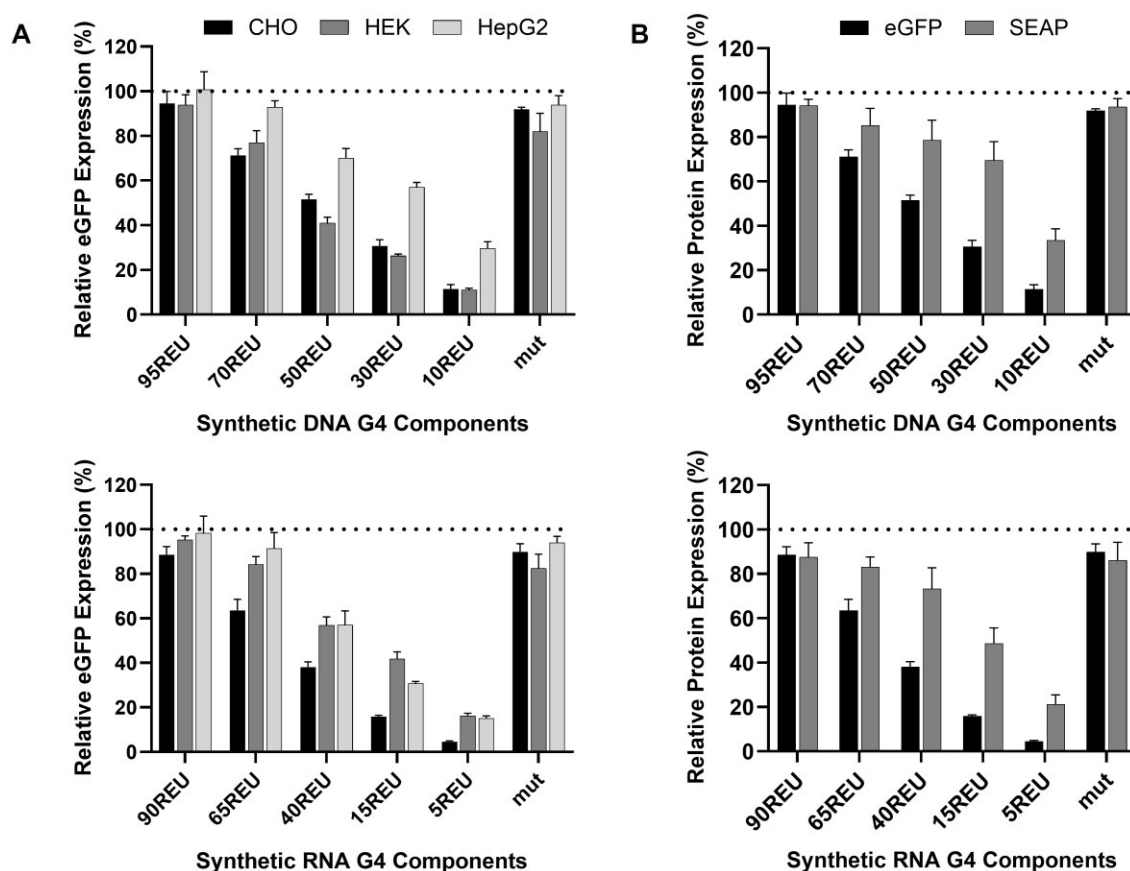
ing processes [76]. As shown in Fig. 3E, eGFP expression levels driven by synthetic G4 components (DNA.30REU, DNA.50REU, DNA.50REU.2, DNA.70REU, DNA.70REU.2, DNA.70REU.3, and DNA.70REU.4) were not significantly affected by any of the tested ligands. Further tests demonstrated that RNA G4 component activity was similarly unchanged by addition of these chemicals (data not shown). Failure to create functional ON and OFF switches may be due to ligand's inability to interact with unnatural synthetic G4 sequences. Alternatively, maximum useable chemical concentrations (not reducing cell viability by >10%) may have been insufficient to significantly regulate the artificially large number of G4 motifs present per cell, where stably transfected CHO cells typically contain a high number of integrated gene copies [77, 78]. These findings, coupled with broad-acting ligands potentially causing off-target effects on host cell function by binding to multiple endogenous structures, indicate that development of inducible systems for bioindustrial contexts will require design of bespoke, highly specific synthetic G4 modulators [75, 79]. Such genetic switches would also have potential use in gene therapy by facilitating dynamic titration of therapeutic protein levels in response to patient data profiles.

Finally, to evaluate the ability to predictably deploy synthetic G4 components in the industrially-relevant context of stable producer cell lines, we compared their relative activities in transient and stable expression systems. Twenty-four elements, including 12 DNA and 12 RNA components of varying strength, were used to create stable eGFP-producing pools by selecting cells with genomically integrated expression cassettes using puromycin. As shown in Fig. 3F, observed activities in a stable expression context were highly correlated with those seen in transient systems (Pearson's correlation coefficient,  $r = 0.72$ ;  $r = 0.86$  with single outlier removed). We concluded that G4 elements behave similarly when positioned in episomal and genomic DNA settings, as previously observed [80], enabling their use in stable biopharmaceutical production systems. However, further work will be required to confirm compatibility with individual industrial platforms, given the wide variety of genome integration and clonal selection procedures employed [81].

### Validating synthetic G4 component performance for bioindustrial applications

As G4 controllers function via interference with the general cellular biosynthetic machinery, rather than site-specific regulators such as transcription factors, they should exhibit modular performances where activities are not affected by the chosen mammalian cell host. Indeed, the ability to predictably deploy these components in new bioindustrial cell contexts is a potential key advantage of this technology, particularly for *in vivo* therapeutic applications, which may target any one of hundreds of distinct human cell types [82, 83]. To test this, we profiled the performance of our G4 elements in two additional bioindustrially-relevant cellular environments, namely (i) human embryonic kidney (HEK) cells, a commonly utilized biomanufacturing host for recombinant protein and viral vector production [84], and (ii) HepG2 cells, a model cell line representing a common target tissue for gene therapy [85, 86].

Reporter plasmids harboring DNA and RNA G4 components of varying activities were individually transfected into each cell type and eGFP expression was quantified 48 h later. As illustrated in Fig. 4A, DNA and RNA components per-

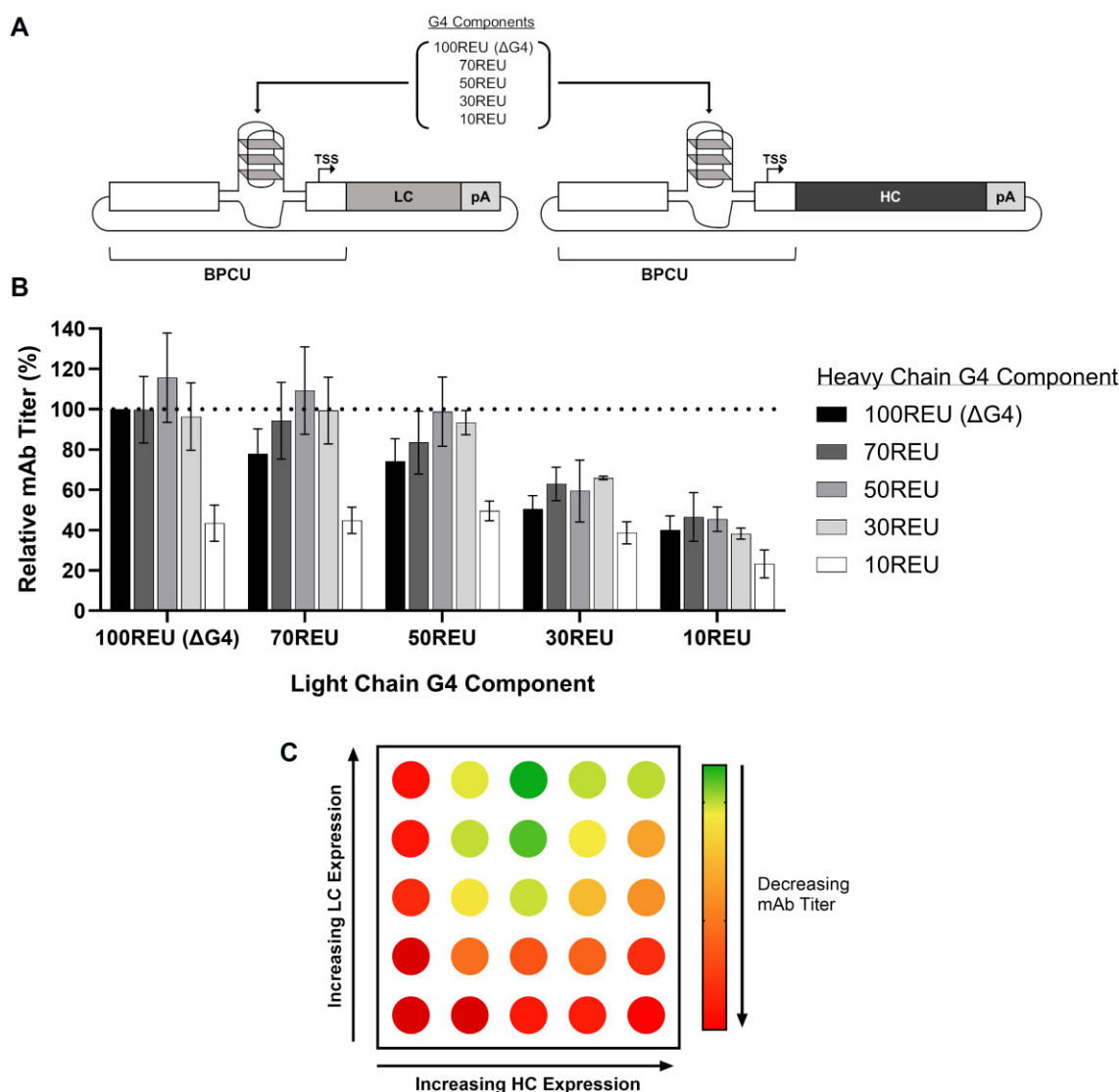


**Figure 4.** Synthetic G4 components predictably titrate recombinant protein expression levels in different cell and product contexts. **(A)** eGFP-reporter plasmids containing DNA or RNA G4-components with varying activities were transfected into CHO, HEK, and HepG2 cells, prior to protein quantification 48 h later. **(B)** CHO cells were transfected with reporter plasmids encoding eGFP or SEAP production under the control of varying-strength DNA or RNA G4 elements. Data are expressed as a percentage of the production exhibited by the control CMV-IE1 proximal promoter-BPCU ( $\Delta$ G4) construct (dotted line). Values represent the mean + SD of three independent experiments ( $n=3$ , each performed in triplicate).

formed as expected in both new cell contexts, each driving five distinct predictable levels of protein production ranging from very low to high. Indeed, previously observed activities in CHO cells were highly correlated with those seen in HEK (Pearson's correlation coefficient,  $r = 0.93$ ,  $P < .0001$ ; [Supplementary Fig. S6A](#)) and HepG2 ( $r = 0.97$ ,  $P < .0001$ ; [Supplementary Fig. S6B](#)). We note that although discrete encoded protein production levels were maintained, construct inhibitory activities were generally slightly weaker in these cell types, as compared to CHO cells, resulting in higher mean expression outputs in both HEK (+7 REU) and HepG2 (+17 REU). For example, RNA components RNA.65REU, RNA.40REU, RNA.15REU, and RNA.5REU drove expression levels of 84%, 57%, 42%, and 16% in HEK and 91%, 57%, 31%, and 15% in HepG2, relative to the unengineered control. This aligns with previous studies showing relative G4 regulatory activities are maintained between cell lines, but overall suppressive effects are somewhat different [41], potentially due to varying ion compositions [87, 88]. While G4s were able to predictably tailor protein expression in both cell types, the maximum inhibitory activity observed was lower in HepG2 (15% relative to control) and HEK (11%) than CHO (5%), potentially due to varying transcriptional/translational machinery compositions [89], indicating that stronger G4s may need to be employed in certain cellular contexts. Nevertheless, our findings confirm that these synthetic G4s can

be predictably utilized as “off-the-shelf” components to optimally titrate recombinant protein expression levels in new cell types of interest.

We next evaluated the ability to utilize synthetic G4 constructs to routinely tailor expression of new target recombinant proteins, using SEAP as an exemplar secreted, glycosylated molecule. Reporter vectors encoding SEAP expression driven by G4 controllers of varying strength were transfected into CHO cells prior to product quantification after 48 h. SEAP outputs from G4 elements were highly correlated with expected values ( $r = 0.91$ ,  $P = .0002$ ; [Supplementary Fig. S7](#)), although product expression levels relative to the standard BPCU were higher than those observed for eGFP (Fig. 4B). The latter observation is likely due to product-specific biosynthetic processing kinetics (i.e. impact of post-translational processing steps), where relative mature protein levels do not directly correlate with relative mRNA transcript and nascent polypeptide quantities [90–92]. Accordingly, where very low protein levels are required (i.e. <10% of that driven by the standard expression cassette lacking a G4 element), we recommend creating vector-optimization spaces that include dual-controllers encoding minimal expression outputs (e.g. DNA.10REU and RNA.5REU utilized in combination, see Fig. 3D). However, given the observed overall increase in SEAP expression levels relative to those achieved with eGFP, data using a wider range of sequence-diverse proteins will likely be required to finalize



**Figure 5.** Synthetic G4 toolboxes facilitate simplified optimization of polypeptide expression ratios for multichain protein products. **(A)** Vector design platform to optimize polypeptide expression ratios for multichain biopharmaceutical products. **(B)** CHO cells were co-transfected with separate plasmids (at an equimolar ratio) encoding mAbT HC and LC expression under the control of varying-strength G4 components (sequences listed in [Supplementary Table S1](#)). Cells were incubated for 5 days, including addition of feed 48 h post-transfection, prior to quantification of mAb titer. Data are shown as a percentage of the production achieved using the unengineered BPCU ( $\Delta$ G4) construct to drive expression of both HC and LC (dotted line). Values represent the mean + SD of three independent experiments ( $n=3$ , each performed in triplicate). **(C)** Vector solution space identifying mAbT product-specific design rules.

panels of G4-containing vectors that can predictably optimize product transcription/translation rates. Nevertheless, the selected G4 component mini library performed as designed by facilitating six discrete expression set points for a new protein of interest (i.e. 20, 30, 50, 70, 85, and 100 REUs), validating their product-agnostic functionality.

Finally, we incorporated our G4 controller technology into a current high-value bioindustrial process, namely cell line development for protein biopharmaceutical production, which involves derivation of optimized CHO cell host DNA expression vector combinations. The dominant class of biopharmaceuticals, monoclonal antibodies (mAb), consists of separate heavy chain (HC) and light chain (LC) polypeptides, where it is well known that the optimal HC:LC expression ratio is highly product specific, ranging from 1:1 to 1:10 depending on antibody folding and assembly dynamics [93–95]. We utilized

four DNA G4 components of varying strength, DNA.REU70, DNA.REU50, DNA.REU30, and DNA.REU10, alongside the unengineered BPCU construct (i.e. 100 REU) to independently control LC and HC expression for an exemplar AZ mAb (mAbT; Fig. 5A). All possible component combinations were tested in transient fed-batch production processes, where the 25 unique encoded HC:LC expression ratios facilitated a wide range of mAbT titers (Fig. 5B). As illustrated in Fig. 5C, rapidly testing a standardized vector solution space revealed that (i) an encoded HC:LC expression ratio of ~1:2 enabled highest product titer, (ii) mAbT yields were positively correlated with increasing LC quantities, and (iii) HC levels needed to be precisely controlled to achieve an optimal set-point, presumably due to excess intracellular HC causing adverse effects on cell factory performance [93, 96]. While polypeptide-specific processing dynamics will likely result in variable ab-



solute expression set points for each G4 component pair when combined with new test mAbs, our previous data (see Fig. 4) indicate that this toolbox approach will routinely encode a wide range of predictable, relevant HC:LC ratios to facilitate expression cassette optimization. We therefore concluded that a G4 element-based screening platform could derive improved expression vector designs for new proteins and could also be utilized as a diagnostic tool to interrogate the cell's relative ability to synthesize a given polypeptide, guiding product-specific cell engineering solutions [97–99]. We anticipate that this technology will be particularly useful for the development of difficult-to-express proteins, where suboptimal polypeptide biosynthesis kinetics dramatically restricts product yield and quality, and next-generation molecules that require expression of  $\geq 3$  composite chains to be precisely stoichiometrically balanced [100–102]. We note that titer optimization is not the only objective for biopharmaceutical expression vector design, and future work will focus on evaluating the ability of this platform to enhance product quality characteristics, such as aggregation and glycosylation profiles, in industrially relevant scaled-up manufacturing processes [103].

## Conclusion

The synthetic G4 components presented in this study can be used predictably to precisely tune mammalian recombinant protein expression levels. Specifically created to function within a bioindustry-compatible genetic chassis, these elements can provide simple off-the-shelf solutions to current protein production challenges. Shown to reliably encode a wide range of discrete protein biosynthesis set points in both CHO and HEK cells, they can be harnessed to (i) optimize polypeptide expression ratios for difficult-to-express multi-chain molecules, such as trispecific antibodies, and adeno-associated viral vectors, and (ii) enact complex multigene cell factory engineering strategies [5, 102, 104]. We have validated a toolbox approach to achieve this, employing G4s with varying regulatory activities; however, given polypeptide-specific processing dynamics, the element composition of vector design platforms may need to be refined dependent on individual applications or product portfolios. Having demonstrated the technology's performance predominantly for transient production contexts [105], we note that implementation in stable expression systems will require more comprehensive validation in relevant industrial manufacturing platforms (e.g. using standard genome integration methodologies, clonal selection procedures, etc [81]).

Demonstrated to titrate protein expression rates predictably in varying molecular formats (i.e. mRNA and DNA) and mammalian cell types, synthetic G4s could be deployed in gene therapy applications to coordinate therapeutic efficacy with product safety by driving optimized expression levels that achieve intended medical outcomes without causing off-target effects on cell behavior [106]. Moreover, for the growing number of products that target multigenic disorders, they could be used to precisely balance the intracellular levels of multiple therapeutic proteins in order to achieve required, physiologically relevant expression ratios [107]. Given that in *in vivo* applications, genetic regulatory elements become “part of the product” introduced into patients, their successful incorporation into such therapeutics will require extensive testing in animal models. It will also be beneficial to characterize the long-term interaction between G4 el-

ements and target host cells to assess potential off-target effects. While the components were designed to function via steric hindrance-based mechanisms, ChIP-seq studies would be useful to quantitatively profile their ability to bind to and potentially sequester host cell proteins [108]. Transcriptomic and proteomic analyses would confirm if impeding movement of key transcriptional and translational machinery leads to unintended changes in global gene expression patterns, and comet assays could be employed to assess whether secondary structures induce local DNA damage, e.g. by causing breakage of DNA strands [109, 110]. Depending on the intended application, G4-host cell relationships may need to be studied in extended animal studies or scaled-up bioreactor processes. Finally, to deploy this technology predictably across the wide range of different potential host cells, it would be beneficial to fully characterize how synthetic G4 structures are affected by varying cellular parameters, e.g. by conducting CD analysis in a wide range of salt conditions [111] and performing molecular dynamic simulations [112].

As they enable simultaneous precision control of both transcription and translation rates, G4 components can also be utilized in combination with other gene expression modulation technologies, including cell-type-specific promoters [11] and exogenous additives that act as ON or OFF switches [113]. Indeed, the ability to enact two-level protein expression control with a single motif (i.e. using the same G4 sequence in the core promoter and 5' UTR) provides an ideal basis for constructing non-leaky inducible expression systems that can coordinate product biosynthesis kinetics with bioreactor processes, cell line development workflows, and patient requirements [76, 114, 115]. However, this will require the design of cognate regulatory-compliant chemical ligands that exhibit highly G4-specific stabilization or destabilization activities without inducing off-target effects that disrupt other required host cell functionalities [116]. Indeed, the absence of such ligands is a primary obstacle currently limiting the full potential of this technology for *in vivo* applications. Identifying optimal chemicals is challenging given the complex design criteria required, including the need to (i) develop/identify manufacturing processes that can produce the ligand at requisite quality, and (ii) comprehensively determine their activity and safety profiles in animal models. However, the initial obstacle of designing bespoke highly-specific ligands will be aided by recent progress in this area, including CRISPR-guided techniques [117] and methods that conjugate chemicals with oligonucleotides recognizing sequences adjacent to target G4s [79]. In conclusion, the synthetic genetic elements introduced in this study provide a new toolkit to precisely tailor mammalian recombinant protein expression in high-value bioindustrial applications, where future work will focus on the development of next-generation small molecule-inducible ON and OFF switches.

## Acknowledgements

**Author contributions:** Melinda Pohle: Conceptualization, methodology, analysis, investigation, and manuscript preparation. Edward Curry: Investigation, methodology, analysis, and manuscript preparation. Ryan Holden: Investigation, methodology, and validation. Suzanne Gibson: Conceptualization and analysis. Adam Brown: Conceptualization, methodology, analysis, manuscript preparation, and funding acquisition.

## Supplementary data

Supplementary data is available at NAR online.

## Conflict of interest

None declared.

## Funding

This study was supported by AstraZeneca and the Engineering and Physical Sciences Research Council (Grant No. EP/T517835/1). Funding to pay the Open Access publication charges for this article was provided by University of Sheffield.

## Data availability

The data underlying this article are available in the article and in its online supplementary material.

## References

- Brown AJ, Gibson SJ, Hatton D *et al.* Whole synthetic pathway engineering of recombinant protein production. *Nat Cell Biol* 2019;116:1212–24. <https://doi.org/10.1002/bit.26855>
- Wang TY, Guo X. Expression vector cassette engineering for recombinant therapeutic production in mammalian cell systems. *Appl Microbiol Biotechnol* 2020;104:5673–88. <https://doi.org/10.1007/s00253-020-10640-w>
- Poletti V, Mavilio F. Designing lentiviral vectors for gene therapy of genetic diseases. *Viruses* 2021;13:1526. <https://doi.org/10.3390/v13081526>
- Qin S, Tang X, Chen Y *et al.* mRNA-based therapeutics: powerful and versatile tools to combat diseases. *Signal Transduct Target Ther* 2022;7:166.
- Eisenhut P, Marx N, Borsi G *et al.* Manipulating gene expression levels in mammalian cell factories: an outline of synthetic molecular toolboxes to achieve multiplexed control. *N Biotechnol* 2024;79:1–19. <https://doi.org/10.1016/j.nbt.2023.11.003>
- Cring MR, Sheffield VC. Gene therapy and gene correction: targets, progress, and challenges for treating human diseases. *Gene Ther* 2022;29:3–12. <https://doi.org/10.1038/s41434-020-00197-8>
- Sou SN, Harris CL, Williams R *et al.* CHO synthetic promoters improve expression and product quality of biotherapeutic proteins. *Biotechnol Prog* 2023;39:e3348. <https://doi.org/10.1002/btpr.3348>
- Li ZM, Fan ZL, Wang XY *et al.* Factors affecting the expression of recombinant protein and improvement strategies in Chinese hamster ovary cells. *Front Bioeng Biotechnol* 2022;10:880155.
- Samoudi M, Masson HO, Kuo CC *et al.* From omics to cellular mechanisms in mammalian cell factory development. *Curr Opin Chem Eng* 2021;32:100688. <https://doi.org/10.1016/j.coche.2021.100688>
- O'Neill P, Mistry RK, Brown AJ *et al.* Protein-specific signal peptides for mammalian vector engineering. *ACS Synth Biol* 2023;12:2339–52.
- Johnson AO, Fowler SB, Webster CI *et al.* Bioinformatic design of dendritic cell-specific synthetic promoters. *ACS Synth Biol* 2022;11:1613–26. <https://doi.org/10.1021/acssynbio.2c00027>
- Cao J, Novoa EM, Zhang Z *et al.* High-throughput 5' UTR engineering for enhanced protein production in non-viral gene therapies. *Nat Commun* 2021;12:4138. <https://doi.org/10.1038/s41467-021-24436-7>
- Johari YB, Mercer AC, Liu Y *et al.* Design of synthetic promoters for controlled expression of therapeutic genes in retinal pigment epithelial cells. *Biotechnol Bioeng* 2021;118:2001–15. <https://doi.org/10.1002/bit.27713>
- Skopenkova VV, Egorova TV, Bardina MV. Muscle-specific promoters for gene therapy. *Acta Naturae* 2021;13:47–58. <https://doi.org/10.32607/actanaturae.11063>
- McGraw CE, Peng D, Sandoval NR. Synthetic biology approaches: the next tools for improved protein production from CHO cells. *Curr Opin Chem Eng* 2020;30:26–33. <https://doi.org/10.1016/j.coche.2020.06.003>
- Leppek K, Das R, Barna M. Functional 5' UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nat Rev Mol Cell Biol* 2018;19:158–74. <https://doi.org/10.1038/nrm.2017.103>
- Eisenhut P, Mebrahtu A, Moradi Barzadd M *et al.* Systematic use of synthetic 5'-UTR RNA structures to tune protein translation improves yield and quality of complex proteins in mammalian cell factories. *Nucleic Acids Res* 2020;48:e119. <https://doi.org/10.1093/nar/gkaa847>
- Sen D, Gilbert W. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature* 1988;334:364–6. <https://doi.org/10.1038/334364a0>
- Chambers VS, Marsico G, Boutell JM *et al.* High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat Biotechnol* 2015;33:877–81. <https://doi.org/10.1038/nbt.3295>
- Kwok CK, Marsico G, Sahakyan AB *et al.* RG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. *Nat Methods* 2016;13:841–4. <https://doi.org/10.1038/nmeth.3965>
- Masai H, Tanaka T. G-quadruplex DNA and RNA: their roles in regulation of DNA replication and other biological functions. *Biochem Biophys Res Commun* 2020;531:25–38. <https://doi.org/10.1016/j.bbrc.2020.05.132>
- Lago S, Nadai M, Cernilogar FM *et al.* Promoter G-quadruplexes and transcription factors cooperate to shape the cell type-specific transcriptome. *Nat Commun* 2021;12:3885. <https://doi.org/10.1038/s41467-021-24198-2>
- Robinson J, Raguseo F, Nuccio SP *et al.* DNA G-quadruplex structures: more than simple roadblocks to transcription? *Nucleic Acids Res* 2021;49:8419–31. <https://doi.org/10.1093/nar/gkab609>
- Varshney D, Spiegel J, Zyner K *et al.* The regulation and functions of DNA and RNA G-quadruplexes. *Nat Rev Mol Cell Biol* 2020;21:459–74. <https://doi.org/10.1038/s41580-020-0236-x>
- Song J, Perreault J-P, Topisirovic I *et al.* RNA G-quadruplexes and their potential regulatory roles in translation. *Translation* 2016;4:e1244031. <https://doi.org/10.1080/21690731.2016.1244031>
- Bochman ML, Paeschke K, Zakian VA. DNA secondary structures: stability and function of G-quadruplex structures. *Nat Rev Genet* 2012;13:770–80. <https://doi.org/10.1038/nrg3296>
- Wang W, Hu S, Gu Y *et al.* Human MYC G-quadruplex: from discovery to a cancer therapeutic target. *Biochim Biophys Acta Rev Cancer* 2020;1874:188410. <https://doi.org/10.1016/j.bbcan.2020.188410>
- Morgan RK, Batra H, Gaerig VC *et al.* Identification and characterization of a new G-quadruplex forming region within the KRAS promoter as a transcriptional regulator. *Biochim Biophys Acta* 2016;1859:235–45.
- Kumari S, Bugaut A, Balasubramanian S. Position and stability are determining factors for translation repression by an RNA G-quadruplex-forming sequence within the 5' UTR of the NRAS proto-oncogene. *Biochemistry* 2008;47:12664–9. <https://doi.org/10.1021/bi8010797>
- Yan J, Zhao D, Dong L *et al.* A novel G-quadruplex motif in the human MET promoter region. *Biosci Rep* 2017;37:BSR20171128. <https://doi.org/10.1042/BSR20171128>
- Onel B, Carver M, Wu G *et al.* A new G-quadruplex with hairpin loop immediately upstream of the human BCL2 P1 promoter

- modulates transcription. *J Am Chem Soc* 2016;138:2563–70. <https://doi.org/10.1021/jacs.5b08596>
32. Shahid R, Bugaut A, Balasubramanian S. The BCL-2 5' untranslated region contains an RNA G-quadruplex-forming motif that modulates protein expression. *Biochemistry* 2010;49:8300–6. <https://doi.org/10.1021/bi100957h>
  33. Huang W, Smaldino PJ, Zhang Q *et al.* Yin Yang 1 contains G-quadruplex structures in its promoter and 5'-UTR and its expression is modulated by G4 resolvase 1. *Nucleic Acids Res* 2012;40:1033–49. <https://doi.org/10.1093/nar/gkr849>
  34. Romanova N, Noll T. Engineered and natural promoters and chromatin-modifying elements for recombinant protein expression in CHO cells. *Biotechnol J* 2018;13:1700232. <https://doi.org/10.1002/biot.201700232>
  35. Brown AJ, James DC. Precision control of recombinant gene transcription for CHO cell synthetic biology. *Biotechnol Adv* 2016;34:492–503. <https://doi.org/10.1016/j.biotechadv.2015.12.012>
  36. Patel YD, Brown AJ, Zhu J *et al.* Control of multigene expression stoichiometry in mammalian cells using synthetic promoters. *ACS Synth Biol* 2021;10:1155–65. <https://doi.org/10.1021/acssynbio.0c00643>
  37. Williams DJ, Puhl HL, Ikeda SR. A simple, highly efficient method for heterologous expression in mammalian primary neurons using cationic lipid-mediated mRNA transfection. *Front Neurosci* 2010;4:1761. <https://doi.org/10.3389/fnins.2010.00181>
  38. Brown AJ, Sweeney B, Mainwaring DO *et al.* Synthetic promoters for CHO cell engineering. *Biotechnol Bioeng* 2014;111:1638–47. <https://doi.org/10.1002/bit.25227>
  39. Johari YB, Scarrott JM, Pohle TH *et al.* Engineering of the CMV promoter for controlled expression of recombinant genes in HEK293 cells. *Biotechnol J* 2022;17:2200062. <https://doi.org/10.1002/biot.202200062>
  40. Au HKE, Isalan M, Mielcarek M. Gene therapy advances: a meta-analysis of AAV usage in clinical settings. *Front Med* 2022;8:809118. <https://doi.org/10.3389/fmed.2021.809118>
  41. Halder K, Wieland M, Hartig JS. Predictable suppression of gene expression by 5'-UTR-based RNA quadruplexes. *Nucleic Acids Res* 2009;37:6811–7. <https://doi.org/10.1093/nar/gkp696>
  42. Xia W, Bringmann P, McClary J *et al.* High levels of protein expression using different mammalian CMV promoters in several cell lines. *Protein Expr Purif* 2006;45:115–24. <https://doi.org/10.1016/j.pep.2005.07.008>
  43. Qin JY, Zhang L, Clift KL *et al.* Systematic comparison of constitutive promoters and the doxycycline-inducible promoter. *PLoS One* 2010;5:10611. <https://doi.org/10.1371/journal.pone.0010611>
  44. Juven-Gershon T, Kadonaga JT. Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Dev Biol* 2010;339:225–9. <https://doi.org/10.1016/j.ydbio.2009.08.009>
  45. Kozak M. Circumstances and mechanisms of inhibition of translation by secondary structure in eucaryotic mRNAs. *Mol Cell Biol* 1989;9:5134–42.
  46. Huppert JL, Balasubramanian S. G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res* 2007;35:406–13. <https://doi.org/10.1093/nar/gkl1057>
  47. Beaudoin JD, Perreault JP. 5'-UTR G-quadruplex structures acting as translational repressors. *Nucleic Acids Res* 2010;38:7022–36. <https://doi.org/10.1093/nar/gkq557>
  48. Beaudoin JD, Jodoin R, Perreault JP. New scoring system to identify RNA G-quadruplex folding. *Nucleic Acids Res* 2014;42:1209–23. <https://doi.org/10.1093/nar/gkt904>
  49. Kikin O, D'Antonio L, Bagga PS. QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res* 2006;34:W676–82. <https://doi.org/10.1093/nar/gkl253>
  50. Agarwal T, Roy S, Kumar S *et al.* In the sense of transcription regulation by G-quadruplexes: asymmetric effects in sense and antisense strands. *Biochemistry* 2014;53:3711–8. <https://doi.org/10.1021/bi401451q>
  51. Patwardhan RP, Lee C, Litvin O *et al.* High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. *Nat Biotechnol* 2009;27:1173–5. <https://doi.org/10.1038/nbt.1589>
  52. Pandey S, Agarwala P, Maiti S. Effect of loops and G-quartets on the stability of RNA G-quadruplexes. *J Phys Chem B* 2013;117:6896–905. <https://doi.org/10.1021/jp401739m>
  53. Zhang AYQ, Bugaut A, Balasubramanian S. A sequence-independent analysis of the loop length dependence of intramolecular RNA G-quadruplex stability and topology. *Biochemistry* 2011;50:7251–8. <https://doi.org/10.1021/bi200805j>
  54. Petraccone L, Erra E, Duro I *et al.* Relative stability of quadruplexes containing different number of G-tetrads. *Nucleosides Nucleotides Nucleic Acids* 2005;24:757–60. <https://doi.org/10.1081/NCN-200060080>
  55. Castelli M, Doria F, Freccero M *et al.* Studying the dynamics of a complex G-quadruplex system: insights into the comparison of MD and NMR data. *J Chem Theory Comput* 2022;18:4515–28. <https://doi.org/10.1021/acs.jctc.2c00291>
  56. Yu YZ, Ma Y, Xu WH *et al.* Combinations of various CpG motifs cloned into plasmid backbone modulate and enhance protective immunity of viral replicon DNA anthrax vaccines. *Med Microbiol Immunol* 2015;204:481–91. <https://doi.org/10.1007/s00430-014-0359-9>
  57. Chan YK, Wang SK, Chu CJ *et al.* Engineering adeno-associated viral vectors to evade innate immune and inflammatory responses. *Sci Transl Med* 2021;13:3438. <https://doi.org/10.1126/scitranslmed.abd3438>
  58. Hoose A, Vellacott R, Storch M *et al.* DNA synthesis technologies to close the gene writing gap. *Nat Rev Chem* 2023;7:144–61. <https://doi.org/10.1038/s41570-022-00456-9>
  59. Fath S, Bauer AP, Liss M *et al.* Multiparameter RNA and codon optimization: a standardized tool to assess and enhance autologous mammalian gene expression. *PLoS One* 2011;6:e17596. <https://doi.org/10.1371/journal.pone.0017596>
  60. Grün JT, Schwalbe H. Folding dynamics of polymorphic G-quadruplex structures. *Biopolymers* 2022;113:e23477.
  61. del Villar-Guerra R, Trent JO, Chaires JB. G-quadruplex secondary structure obtained from circular dichroism spectroscopy. *Angew Chem Int Ed Engl* 2018;57:7171–5. <https://doi.org/10.1002/anie.201709184>
  62. Kejnovská I, Renčiuk D, Palacký J *et al.* CD study of the G-quadruplex conformation. In: Yang D, Lin C (eds.), *Methods in Molecular Biology*. Vol. 2035. Humana Press Inc., 2019, 25–44.
  63. Johnson WC. Determination of the conformation of nucleic acids by electronic CD. In: Fasman GD (ed.), *Circular Dichroism and the Conformational Analysis of Biomolecules*. Boston, MA: Springer, 1996, 433–68. <https://doi.org/10.1007/978-1-4757-2508-7>
  64. Di Antonio M, Ponjavic A, Radzevičius A *et al.* Single-molecule visualization of DNA G-quadruplex formation in live cells. *Nat Chem* 2020;12:832–7. <https://doi.org/10.1038/s41557-020-0506-4>
  65. Biffi G, Tannahill D, McCafferty J *et al.* Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat Chem* 2013;5:182–6. <https://doi.org/10.1038/nchem.1548>
  66. Vavilis T, Stamoula E, Ainatoglou A *et al.* mRNA in the context of protein replacement therapy. *Pharmaceutics* 2023;15:166. <https://doi.org/10.3390/pharmaceutics15010166>
  67. Sample PJ, Wang B, Reid DW *et al.* Human 5' UTR design and variant effect prediction from a massively parallel translation assay. *Nat Biotechnol* 2019;37:803–9. <https://doi.org/10.1038/s41587-019-0164-5>
  68. Cazier AP, Blazeck J. Advances in promoter engineering: novel applications and predefined transcriptional control. *Biotechnol J* 2021;16:2100239. <https://doi.org/10.1002/biot.202100239>



69. Chaudhary S, Kumar M, Kaushik M. Interface of G-quadruplex with both stabilizing and destabilizing ligands for targeting various diseases. *Int J Biol Macromol* 2022;219:414–27. <https://doi.org/10.1016/j.ijbiomac.2022.07.248>
70. Yan MP, Wee CE, Yen KP *et al.* G-quadruplex ligands as therapeutic agents against cancer, neurological disorders and viral infections. *Future Med Chem* 2023;15:1987–2009. <https://doi.org/10.4155/fmc-2023-0202>
71. Pardi ML, Wu J, Kawasaki S *et al.* Synthetic RNA-based post-transcriptional expression control methods and genetic circuits. *Adv Drug Deliv Rev* 2022;184:114196. <https://doi.org/10.1016/j.addr.2022.114196>
72. Verbič A, Praznik A, Jerala R. A guide to the design of synthetic gene networks in mammalian cells. *FEBS J* 2021;288:5265–88.
73. Figueiredo J, Mergny JL, Cruz C. G-quadruplex ligands in cancer therapy: progress, challenges, and clinical perspectives. *Life Sci* 2024;340:122481. <https://doi.org/10.1016/j.lfs.2024.122481>
74. Wu G, Tillo D, Ray S *et al.* Custom G4 microarrays reveal selective G-quadruplex recognition of small molecule BMVC: a large-scale assessment of ligand binding selectivity. *Molecules* 2020;25:3465. <https://doi.org/10.3390/molecules25153465>
75. Duarte AR, Cadoni E, Ressurreição AS *et al.* Design of modular G-quadruplex ligands. *ChemMedChem* 2018;13:869–93. <https://doi.org/10.1002/cmdc.201700747>
76. Donaldson JS, Dale MP, Rosser SJ. Decoupling growth and protein production in CHO cells: a targeted approach. *Front Bioeng Biotechnol* 2021;9:349. <https://doi.org/10.3389/fbioe.2021.658325>
77. Lee Z, Wan J, Shen A *et al.* Gene copy number, gene configuration and LC/HC mRNA ratio impact on antibody productivity and product quality in targeted integration CHO cell lines. *Biotechnol Prog* 2024;40:e3433. <https://doi.org/10.1002/btpr.3433>
78. Joubert S, Stuible M, Lord-Dufour S *et al.* A CHO stable pool production platform for rapid clinical development of trimeric SARS-CoV-2 spike subunit vaccine antigens. *Biotechnol Bioeng* 2023;120:1746–61. <https://doi.org/10.1002/bit.28387>
79. Berner A, Das RN, Bhuma N *et al.* G4-Ligand-conjugated oligonucleotides mediate selective binding and stabilization of individual G4 DNA structures. *J Am Chem Soc* 2024;146:6926–35. <https://doi.org/10.1021/jacs.3c14408>
80. Shin YJ, Kumarasamy V, Camacho D *et al.* Involvement of G-quadruplex structures in regulation of human RET gene expression by small molecules in human medullary thyroid carcinoma TT cells. *Oncogene* 2015;34:1292–9. <https://doi.org/10.1038/onc.2014.65>
81. Tihanyi B, Nyitray L. Recent advances in CHO cell line development for recombinant protein production. *Drug Discov Today Technol* 2020;38:25–34. <https://doi.org/10.1016/j.ddtec.2021.02.003>
82. Regev A, Teichmann SA, Lander ES *et al.* The human cell atlas. *eLife* 2017;6:e27041. <https://doi.org/10.7554/eLife.27041>
83. Shchaslyvyi AY, Antonenko SV, Tesliuk MG *et al.* Current state of human gene therapy: approved products and vectors. *Pharmaceuticals* 2023;16:1416. <https://doi.org/10.3390/ph16101416>
84. Tan E, Chin CSH, Lim ZFS *et al.* HEK293 cell line as a platform to produce recombinant proteins and viral vectors. *Front Bioeng Biotechnol* 2021;9:796991. <https://doi.org/10.3389/fbioe.2021.796991>
85. Arzumaniyan VA, Kiseleva OI, Poverennaya EV. The curious case of the HepG2 cell line: 40 years of expertise. *Int J Mol Sci* 2021;22:13135. <https://doi.org/10.3390/ijms222313135>
86. Zabaleta N, Unzu C, Weber ND *et al.* Gene therapy for liver diseases—progress and challenges. *Nat Rev Gastroenterol Hepatol* 2023;20:288–305. <https://doi.org/10.1038/s41575-022-00729-0>
87. Bhattacharyya D, Arachchilage GM, Basu S. Metal cations in G-quadruplex folding and stability. *Front Chem* 2016;4:38. <https://doi.org/10.3389/fchem.2016.00038>
88. Zhang ZH, Qian SH, Wei D *et al.* *In vivo* dynamics and regulation of DNA G-quadruplex structures in mammals. *Cell Biosci* 2023;13:117. <https://doi.org/10.1186/s13578-023-01074-8>
89. Mendoza O, Bourdoncle A, Boulé JB *et al.* G-quadruplexes and helicases. *Nucleic Acids Res* 2016;44:1989–2006. <https://doi.org/10.1093/nar/gkw079>
90. Mathias S, Wippermann A, Raab N *et al.* Unraveling what makes a monoclonal antibody difficult-to-express: from intracellular accumulation to incomplete folding and degradation via ERAD. *Biotechnol Bioeng* 2020;117:5–16. <https://doi.org/10.1002/bit.27196>
91. Reinhardt D, Sommeregger W, Debreczeny M *et al.* In search of expression bottlenecks in recombinant CHO cell lines—a case study. *Appl Microbiol Biotechnol* 2014;98:5959–65. <https://doi.org/10.1007/s00253-014-5584-z>
92. Reisinger H, Steinfellner W, Stern B *et al.* The absence of effect of gene copy number and mRNA level on the amount of mAb secretion from mammalian cells. *Appl Microbiol Biotechnol* 2008;81:701–10. <https://doi.org/10.1007/s00253-008-1701-1>
93. Schlatter S, Stansfield SH, Dinnis DM *et al.* On the optimal ratio of heavy to light chain genes for efficient recombinant antibody production by CHO cells. *Biotechnol Prog* 2005;21:122–33. <https://doi.org/10.1021/bp049780w>
94. Wijesuriya SD, Pongo E, Tomic M *et al.* Antibody engineering to improve manufacturability. *Protein Expr Purif* 2018;149:75–83. <https://doi.org/10.1016/j.pep.2018.04.003>
95. Pybus LP, James DC, Dean G *et al.* Predicting the expression of recombinant monoclonal antibodies in Chinese hamster ovary cells based on sequence features of the CDR3 domain. *Biotechnol Prog* 2014;30:188–97. <https://doi.org/10.1002/btpr.1839>
96. Vanhove M, Usherwood YK, Hendershot LM. Unassembled Ig heavy chains do not cycle from BiP *in vivo* but require light chains to trigger their release. *Immunity* 2001;15:105–14. [https://doi.org/10.1016/S1074-7613\(01\)00163-7](https://doi.org/10.1016/S1074-7613(01)00163-7)
97. Cartwright JF, Arnall CL, Patel YD *et al.* A platform for context-specific genetic engineering of recombinant protein production by CHO cells. *J Biotechnol* 2020;312:11–22. <https://doi.org/10.1016/j.jbiotec.2020.02.012>
98. Donaldson J, Kleinjan DJ, Rosser S. Synthetic biology approaches for dynamic CHO cell engineering. *Curr Opin Biotechnol* 2022;78:102806. <https://doi.org/10.1016/j.copbio.2022.102806>
99. Fischer S, Otte K. CHO cell engineering for improved process performance and product quality. In: Min Lee G, Kildegaard HF (eds.), *Cell Culture Engineering*. John Wiley & Sons, Ltd, 2019, 207–50. <https://doi.org/10.1002/9783527811410>
100. Tadauchi T, Lam C, Liu L *et al.* Utilizing a regulated targeted integration cell line development approach to systematically investigate what makes an antibody difficult to express. *Biotechnol Prog* 2019;35:e2772. <https://doi.org/10.1002/btpr.2772>
101. Kaneyoshi K, Kuroda K, Uchiyama K *et al.* Secretion analysis of intracellular “difficult-to-express” immunoglobulin G (IgG) in Chinese hamster ovary (CHO) cells. *Cytotechnology* 2019;71:305–16. <https://doi.org/10.1007/s10616-018-0286-5>
102. Sebastião MJ, Hoffman M, Escandell J *et al.* Identification of mispairing omic signatures in Chinese hamster ovary (CHO) cells producing a tri-specific antibody. *Biomedicine* 2023;11:2890. <https://doi.org/10.3390/biomedicine11112890>
103. Zhang JH, Shan LL, Liang F *et al.* Strategies and considerations for improving recombinant antibody production and quality in Chinese hamster ovary cells. *Front Bioeng Biotechnol* 2022;10:856049. <https://doi.org/10.3389/fbioe.2022.856049>
104. Fu Q, Polanco A, Lee YS *et al.* Critical challenges and advances in recombinant adeno-associated virus (rAAV) biomanufacturing. *Biotechnol Bioeng* 2023;120:2601–21. <https://doi.org/10.1002/bit.28412>
105. Gutiérrez-Granados S, Cervera L, Kamen AA *et al.* Advancements in mammalian cell transient gene expression



- (TGE) technology for accelerated production of biologics. *Crit Rev Biotechnol* 2018;38:918–40. <https://doi.org/10.1080/07388551.2017.1419459>
106. Ingusci S, Verlengia G, Soukupova M *et al.* Gene therapy tools for brain diseases. *Front Pharmacol* 2019;10:724. <https://doi.org/10.3389/fphar.2019.00724>
  107. Castro B, Steel JC, Layton CJ. AAV-mediated gene therapies for glaucoma and uveitis: are we there yet? *Expert Rev Mol Med* 2024;26:e9. <https://doi.org/10.1017/erm.2024.4>
  108. Valouev A, Johnson DS, Sundquist A *et al.* Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. *Nat Methods* 2008;5:829–34. <https://doi.org/10.1038/nmeth.1246>
  109. Xu H, Di Antonio M., McKinney S. *et al.* CX-5461 is a DNA G-quadruplex stabilizer with selective lethality in BRCA1/2 deficient tumours. *Nat Commun* 2017;8:14432. <https://doi.org/10.1038/ncomms14432>
  110. van Kregten M, Tijsterman M. The repair of G-quadruplex-induced DNA damage. *Exp Cell Res* 2014;329:178–83. <https://doi.org/10.1016/j.yexcr.2014.08.038>
  111. Olejko L, Dutta A, Shahsavari K *et al.* Influence of different salts on the G-quadruplex structure formed from the reversed human telomeric DNA sequence. *Int J Mol Sci* 2022;23:12206. <https://doi.org/10.3390/ijms232012206>
  112. Verdian Doghaei A, Housaindokht MR, Bozorgmehr MR. Molecular crowding effects on conformation and stability of G-quadruplex DNA structure: insights from molecular dynamics simulation. *J Theor Biol* 2015;364:103–12. <https://doi.org/10.1016/j.jtbi.2014.09.015>
  113. Hussein MK, Papež M, Dhiman H *et al.* *In silico* design of CMV promoter binding oligonucleotides and their impact on inhibition of gene expression in Chinese hamster ovary cells. *J Biotechnol* 2022;359:185–93. <https://doi.org/10.1016/j.jbiotec.2022.10.003>
  114. Jalšić L, Lytvyn V, Elahi SM *et al.* Inducible HEK293 AAV packaging cell lines expressing Rep proteins. *Mol Ther Methods Clin Dev* 2023;30:259–75.
  115. Maltais JS, Lord-Dufour S, Morasse A *et al.* Repressing expression of difficult-to-express recombinant proteins during the selection process increases productivity of CHO stable pools. *Biotechnol Bioeng* 2023;120:2840–52. <https://doi.org/10.1002/bit.28435>
  116. Chen L, Dickerhoff J, Sakai S *et al.* DNA G-quadruplex in human telomeres and oncogene promoters: structures, functions, and small molecule targeting. *Acc Chem Res* 2022;55:2628–46. <https://doi.org/10.1021/acs.accounts.2c00337>
  117. Qin G, Liu Z, Yang J *et al.* Targeting specific DNA G-quadruplexes with CRISPR-guided G-quadruplex-binding proteins and ligands. *Nat Cell Biol* 2024;26:1212–24. <https://doi.org/10.1038/s41556-024-01448-1>