Single-Vector mAb Expression Vector - Golden Gate Design

Trastuzumab IgG1 (anti-HER2) | CHO-K1 stable expression | one-pot BsaI assembly

AI Bioinformatician by Labterna | 29 June 2026 | CHO-K1 PICRH-1.0 | de novo design

Vector designed and verified - all design targets met, ready to synthesise
A 6,871 bp single-vector design (~7 kb, within the ~7 kb target) assembles in one BsaI Golden Gate reaction (5 fragments) with **zero internal BsaI/BsmBI/SapI**, CAI 0.94-0.97 (CHO-K1), orthogonal fusion overhangs (no cross-ligation detected), and CHO-genome-specific primers. One design consideration to monitor: the heavy chain is cap-dependent and the light chain IRES-driven, so track the heavy:light ratio during expression (Section 6).

1. Summary of the request

A single-vector monoclonal-antibody (mAb) expression construct was requested for generation of stable CHO-K1 lines: heavy and light chain delivered in one transfection with puromycin selection, assembled in a single BsaI Golden Gate reaction. The brief specified the architecture CMV promoter -> signal peptide -> IgG1 heavy chain -> EMCV IRES -> light chain -> puromycin cassette on an ~7 kb backbone, using the trastuzumab variable domains as a public, well-characterised template. Because two previous attempts had failed at the Golden Gate step on internal BsaI sites in codon-optimised variable regions, **zero internal BsaI was specified as non-negotiable**; internal BsmBI and SapI were also to be removed. Both chains were to be codon-optimised for CHO-K1 (CAI > 0.85), the BsaI fusion-site overhangs taken from the Potapov 2018 high-fidelity set (no palindromes, no pairwise collisions), and a Tm-matched junction colony-PCR primer pair delivered per fragment, BLAST-checked against CHO-K1 PICRH-1.0.

2. Construct design and architecture

The delivered vector is 6,871 bp (circular). A single CMV promoter drives a bicistronic message: the heavy chain is translated cap-dependently, an EMCV internal ribosome entry site (IRES) then drives cap-independent translation of the light chain from the same transcript, and a WPRE plus bGH poly(A) close the unit. Both chains carry a murine Ig-kappa secretion leader (the light chain requires its own leader for secretion; the brief named only one signal peptide, but both are present). A separate SV40-driven puromycin-resistance (pac) cassette provides mammalian selection, and a pUC origin plus AmpR support propagation in E. coli. A chimeric intron in the 5' UTR and the WPRE are standard expression-enhancing elements added to match a production-grade CHO mAb vector; they do not alter the specified element order.

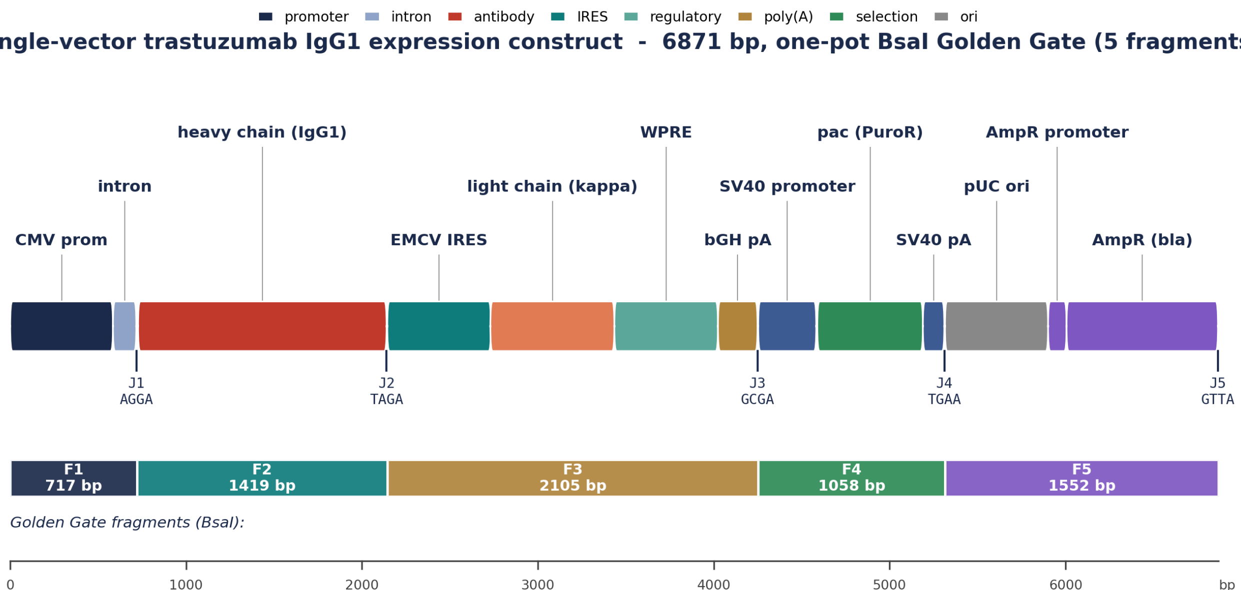

Figure 1. Linear map of the single-vector trastuzumab IgG1 construct. Element boxes are coloured by category; the five BsaI fusion-site overhangs (J1-J5) are marked below the backbone; the lower track shows the five one-pot Golden Gate fragments (F1-F5) and their lengths.

Element	Coordinates (bp)	Length	Function
CMV enhancer/promoter	1-584	584	Strong constitutive promoter (heavy chain, cap-dependent)
Chimeric intron	585-717	133	5'UTR intron; boosts mature-mRNA export/expression

Element	Coordinates (bp)	Length	Function
Heavy chain ORF	728-2,140	1,413	Ig-kappa leader + trastuzumab VH + human IgG1 C (CHO-opt)
EMCV IRES	2,145-2,733	589	Cap-independent re-initiation; 3' ATG = light-chain start
Light chain ORF	2,731-3,435	705	Ig-kappa leader + trastuzumab VL + kappa C (CHO-opt)
WPRE	3,436-4,024	589	Post-transcriptional regulatory element
bGH poly(A)	4,025-4,249	225	Polyadenylation of the antibody transcript
SV40 promoter	4,254-4,583	330	Drives the puromycin-resistance cassette
pac (PuroR) ORF	4,590-5,189	600	Puromycin N-acetyltransferase (CHO-opt); mammalian selection
SV40 poly(A)	5,190-5,311	122	Polyadenylation of the pac cassette
pUC origin	5,316-5,901	586	E. coli origin of replication
AmpR (promoter + bla)	5,902-6,867	966	Ampicillin resistance (E. coli selection)

The five fusion-site scars (J1-J5, 4 bp each) and two Kozak sequences occupy the remaining positions; the light-chain ORF shares its ATG with the 3' end of the EMCV IRES. ORF spans include the stop codon (3 bp more than the optimised CDS lengths in Section 3). Coordinates are for the assembled circular plasmid; the annotated GenBank carries every feature.

3. Codon optimisation for CHO-K1 and Type IIS site removal

The three coding regions were codon-optimised for *Cricetulus griseus* CHO-K1 (NCBI taxid 10029, Kazusa usage) with DNA Chisel, maximising the codon adaptation index (CAI) under hard constraints: no internal BsaI / BsmBI / SapI recognition sites (either strand), GC held within a 40-70% sliding window, and no long homopolymer runs. All three ORFs clear the CAI > 0.85 target with margin.

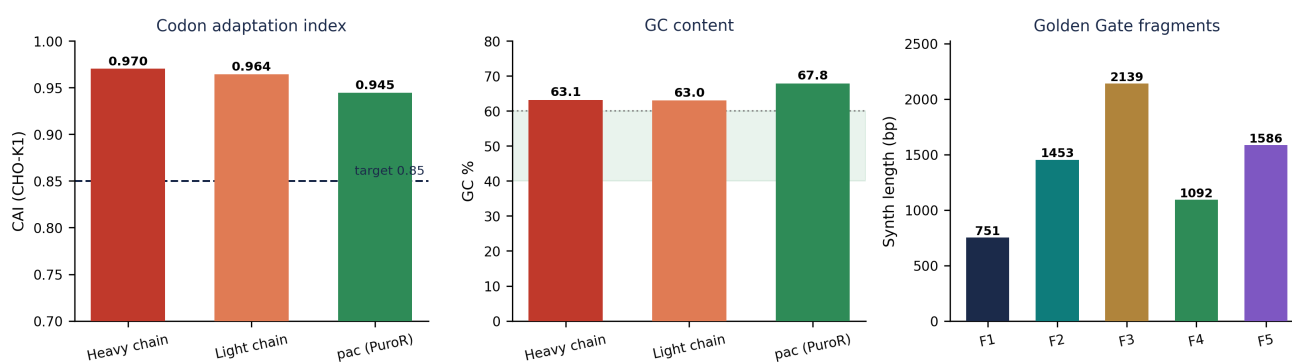

Figure 2. Per-ORF codon adaptation index (left; dashed line = the 0.85 target), GC content (centre; shaded 40-60% band, pac is GC-rich at 67.8%), and the five Golden Gate fragment sizes (right). All ORFs carry zero internal BsaI/BsmBI/SapI sites.

Coding region	aa	bp	CAI	GC %	Internal BsaI/BsmBI/SapI
Heavy chain (leader+VH+IgG1)	470	1,410	0.970	63.1	0 / 0 / 0
Light chain (leader+VL+Ck)	234	702	0.964	63.0	0 / 0 / 0
pac (PuroR)	199	597	0.945	67.8	0 / 0 / 0

3.1 Non-coding elements and domestication

Every regulatory and backbone part was screened on both strands. Seven of nine were already site-free. Two single sites were domesticated without changing function: a BsaI site in the chimeric intron body (one silent substitution, away from the splice donor/acceptor and branch point) and a BsaI site inside the bla (AmpR) coding sequence (a synonymous GGG->GGT, Gly preserved). The two identical 20-aa signal-peptide DNA segments were also diverged at the codon level (15 synonymous swaps) so the construct carries no 60-bp direct repeat.

4. Golden Gate assembly design and fusion-site fidelity

The vector is partitioned into five fragments for a single-tube BsaI (BsaI-HFv2) + T4 ligase reaction. Each fragment is delivered as a synthesisable double-stranded block flanked by inward-facing BsaI sites that excise the fragment with the designed 4-nt overhangs; the recognition sites are carried on the cut-away ends, so **the assembled plasmid contains zero BsaI/BsmBI/SapI sites**. In-silico digestion and re-ligation of the five fragments reproduces the 6,871 bp plasmid exactly.

Frag	Module (5'->3')	Synth (bp)	In-plasmid (bp)	Left OH	Right OH
F1	CMV enhancer/promoter + chimeric intron	751	717	GTTA	AGGA
F2	Kozak + Ig-kappa leader - trastuzumab heavy chain (IgG1) + stop	1,453	1,419	AGGA	TAGA
F3	EMCV IRES + leader - light chain (kappa) + stop + WPRE + bGH poly(A)	2,139	2,105	TAGA	GCGA
F4	SV40 promoter + Kozak + pac (PuroR) + stop + SV40 poly(A)	1,092	1,058	GCGA	TGAA
F5	pUC origin + AmpR (bla)	1,586	1,552	TGAA	GTTA

4.1 Fusion-site overhangs (Potapov / Pryor high-fidelity set)

The five 4-nt fusion overhangs were selected from the Potapov 2018 / Pryor 2020 BsaI-HFv2 ligation-fidelity data. None is palindromic, none is the reverse complement of another, and the minimum pairwise Hamming distance is 2. No cross-ligation is detected among the chosen overhangs in that dataset (model fidelity 100%) - a detection-floor result, not a thermodynamic guarantee. For calibration the same estimator gives ~77% for the common plant/MoClo 11-overhang set (Pryor reports ~81%; the estimator runs a few points conservative), so its relative ranking is reliable.

Junction	Overhang (5'->3')	GC	Palindrome?
J1	AGGA	2/4	no
J2	TAGA	1/4	no
J3	GCGA	3/4	no
J4	TGAA	1/4	no
J5	GTTA	1/4	no

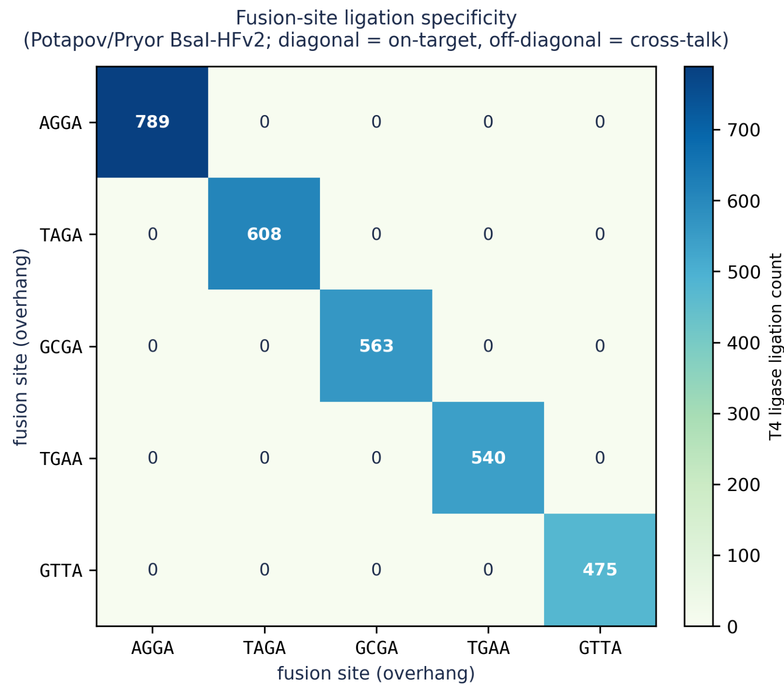

Figure 3. Ligation-specificity matrix for the five chosen overhangs (Potapov/Pryor BsaI-HFv2 data). Diagonal = on-target ligation; every off-diagonal (cross-talk) cell is 0.

5. Junction colony-PCR primers and specificity

One junction-spanning colony-PCR pair was designed per fragment seam (five pairs). Each pair flanks a Golden Gate junction, so a band of the listed size forms only when the two adjacent fragments are correctly joined; the five reactions together confirm the entire assembly. All pairs are Tm-matched (within 0.5 degC), carry a 3' G/C clamp, and give distinguishable 263-412 bp products.

Primer	Sequence (5'->3')	nt	Tm	GC%	Amplicon
J1-F	AACTGGGCTTGTCGAGACAG	20	60.0	55	340 bp
J1-R	GTACCTGGTGTAGCCGTTGG	20	60.4	60	340 bp
J2-F	AAGCTGACCGTGGACAAGAG	20	60.0	55	302 bp
J2-R	TTTGGCGAGAGGGGAAAGAC	20	60.0	55	302 bp
J3-F	CCACTCCCACTGTCCTTTCC	20	60.0	60	263 bp
J3-R	CTGGGGACTTTCCACACCTG	20	60.2	60	263 bp
J4-F	CTTCCTGGAGACATCTGCCC	20	59.8	60	412 bp
J4-R	GACAGGTATCCGGTAAGCGG	20	60.0	60	412 bp
J5-F	GCGGTATCATTGCAGCACTG	20	60.0	55	290 bp
J5-R	TATTGACGTCAATGGGCGGG	20	60.5	55	290 bp

Specificity vs CHO-K1 PICRH-1.0 (GCF_003668045.3): PASS - construct-specific (no CHO-K1 host amplicon). In-silico PCR (isPcr, sensitive minPerfect=8, maxSize 4 kb) of all five designed pairs plus all ten single-primer self-pairs produced **0 host amplicons**. The longest single-primer perfect 3' match in the 2.4 Gb host genome was 17 nt - within chance expectation and, with no convergent partner, not priming-competent. The primers are construct-specific and safe for genotyping stable CHO lines.

6. Target biology and design rationale

Trastuzumab is a humanised IgG1/kappa monoclonal antibody against the HER2 (ERBB2) receptor; its variable domains are public and structurally defined (PDB 1N8Z), which is why the brief chose them as a template. The VH and VL were taken from that structure and verified against it; because 1N8Z is a Fab, the CH2-CH3 was completed from the human IgG1 heavy constant (UniProt P01857, secreted CH1-hinge-CH2-CH3), and the light chain uses the human kappa constant (UniProt P01834). The IgG1 isotype provides the standard effector-competent format used for most therapeutic and research mAbs.

Bicistronic expression and the heavy:light balance. The heavy chain is translated cap-dependently from the CMV transcript while the light chain is translated from the EMCV IRES, which gives lower and less tunable second-cistron output. The heavy chain is therefore expected in molar excess of the light chain - yet for many IgGs an *excess of light chain* aids folding and secretion of assembled antibody and limits heavy-chain aggregation. The brief-specified heavy-chain-first order is thus the empirically less-favoured arrangement for single-vector IRES mAbs, and a heavy-over-light ratio can cap assembled-mAb titre. Where titre matters, an LC-first swap or a dual-promoter / dual-vector format is worth considering; this is the main design consideration to watch (Section 7).

7. Caveats and recommendations

- **Heavy:light expression balance.** If assembled-mAb titre is low and heavy-chain aggregation is seen, swap the chain order (light chain cap-dependent, heavy chain via the IRES) - a one-step rearrangement of the F2/F3 coding fragments and their overhangs - or move to two promoters.
- **pac is GC-rich (67.8%).** CAI is high (0.945) and there are no internal Type IIS sites, but specify a GC-rich synthesis protocol for the F4 fragment.
- **De novo CDS screened for cryptic signals.** The optimised heavy, light and pac ORFs carry no canonical polyadenylation signal (AATAAA) and no strong 5' splice-donor motif (GT[AG]AGT), lowering the risk of truncated transcripts; confirm full-length expression at the bench.
- **IRES initiation context.** The light chain is fused to the wild-type EMCV (M81861) 3' initiator ATG rather than an engineered pIRES initiator; second-cistron initiation efficiency should be confirmed empirically (Bochkov & Palmenberg 2006).
- **In-silico only.** Assembly correctness, zero-site status, CAI, fidelity and primer specificity are all computational. Splice activity of the chimeric intron, EMCV IRES activity in CHO, expression level, and secreted-mAb titre must be confirmed at the bench.
- **Order integrity.** Confirm each synthesised fragment carries exactly its two terminal BsaI sites and no internal ones (vendor QC), and sequence-verify the assembled plasmid across all five junctions before scale-up.

8. Recommended protocol (starting point)

- **Assembly:** equimolar fragments (~75 fmol each) + 50 ng destination context if used; NEBridge Golden Gate Assembly Kit (BsaI-HFv2) + T4 ligase. Cycle 30x (37 degC 5 min / 16 degC 5 min), then 60 degC 5 min, 80 degC 5 min. Transform into a recA- E. coli (e.g. DH5-alpha); select on ampicillin.
- **Screen:** colony PCR with the five junction pairs (Section 5); a correctly assembled clone gives all five expected bands. Sanger-sequence the junctions of positive clones.
- **Express:** transfect CHO-K1, select with puromycin (start 5-10 ug/mL; titrate per line), expand stable pools/clones, and quantify secreted IgG (e.g. Protein A / ELISA) with an SDS-PAGE check of the heavy:light ratio.

9. Output files

- **Trastuzumab_IgG1_GoldenGate_CHO_Report.pdf** - this report.
- **Trastuzumab_IgG1_GoldenGate_CHO_construct.gb** - annotated, circular full-construct GenBank.
- **Trastuzumab_IgG1_GoldenGate_CHO_Data.xlsx** - synthesis fragments, colony-PCR primers, fusion overhangs, codon-optimisation QC.
- **sequences.fasta** - assembled plasmid, five BsaI-flanked synthesis fragments, ten primers.
- **figures/** - Figures 1-3 (PNG 300 dpi + vector PDF).

scripts/ contains the reproducible numbered pipeline (sequence sourcing -> codon optimisation -> Type IIS screening -> fusion-site selection -> assembly -> primers -> specificity -> figures -> report).

Provenance & tools. Trastuzumab V-domains PDB 1N8Z; constants UniProt P01857 (IGHG1) / P01834 (IGKC); puromycin resistance UniProt P13249; EMCV IRES GenBank M81861; regulatory parts from pcDNA3.1(+) / pCI-neo / pLVX-Puro; host genome CHO-K1 PICRH-1.0 (GCF_003668045.3); CHO-K1 codon usage from Kazusa (taxid 10029). Tools: DNA Chisel 3.2.16, Primer3-py 2.3.0, Biopython 1.87, BLAST+ 2.12.0, UCSC isPcr v39. Designed 29 June 2026.

10. References

[1] Cho HS, Mason K, Ramyar KX, et al. (2003) Structure of the extracellular region of HER2 alone and in complex with the Herceptin Fab. *Nature* 421:756-760. (PDB 1N8Z)

[2] Carter P, Presta L, Gorman CM, et al. (1992) Humanization of an anti-p185HER2 antibody for human cancer therapy. *Proc Natl Acad Sci USA* 89:4285-4289.

[3] Potapov V, Ong JL, Kucera RB, et al. (2018) Comprehensive profiling of four-base overhang ligation fidelity by T4 DNA ligase and application to DNA assembly. *ACS Synth Biol* 7:2665-2674.

[4] Pryor JM, Potapov V, Kucera RB, et al. (2020) Enabling one-pot Golden Gate assemblies of unprecedented complexity using data-optimized assembly design. *PLoS ONE* 15:e0238592.

[5] Engler C, Kandzia R, Marillonnet S (2008) A one pot, one step, precision cloning method with high throughput capability. *PLoS ONE* 3:e3647.

[6] Jang SK, Krausslich HG, Nicklin MJ, et al. (1988) A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes. *J Virol* 62:2636-2643. (EMCV IRES; GenBank M81861)

[7] Bochkov YA, Palmenberg AC (2006) Translational efficiency of EMCV IRES in bicistronic vectors is dependent upon IRES sequence and gene location. *BioTechniques* 41:283-292.

[8] Zufferey R, Donello JE, Trono D, Hope TJ (1999) Woodchuck hepatitis virus posttranscriptional regulatory element enhances expression of transgenes delivered by retroviral vectors. *J Virol* 73:2886-2892.

[9] Kozak M (1987) An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res* 15:8125-8148.

[10] Sharp PM, Li WH (1987) The codon adaptation index. *Nucleic Acids Res* 15:1281-1295.

[11] Rupp O, MacDonald ML, Li S, et al. (2018) A reference genome of the Chinese hamster based on a hybrid assembly strategy. *Biotechnol Bioeng* 115:2087-2100. (CHO-K1 PICRH-1.0, GCF_003668045.3)

[12] Zulkower V, Rosser S (2020) DNA Chisel, a versatile sequence optimizer. *Bioinformatics* 36:4508-4509.

[13] Untergasser A, Cutcutache I, Koressaar T, et al. (2012) Primer3 - new capabilities and interfaces. *Nucleic Acids Res* 40:e115.

[14] Kent WJ (2002) BLAT - the BLAST-like alignment tool. *Genome Res* 12:656-664. (isPcr in-silico PCR)