Plasmid: **pHsp70A/RbcS2-cEpo**

Synthetic gene encoding the erythropoietin from *Homo sapiens* (GenBank: P01588, aa 28-193) adapted to the nuclear codon usage of *Chlamydomonas reinhardtii* (*cepo*) under the chimeric HSP70A/RBCS2 promoter in plasmid pHsp70A/RbcS2-Chlamy for Cre/*lox*-mediated plasmid fusion (cloned via *Xho*I/*BamH*I, *R6kγ* origin of replication for maintenance in *E.coli* Pir1-cells (available from Invitrogen), kanamycin resistance).The coding sequence is divided by intron sequence 2 of the *Chlamydomonas reinhardtii RBCS2* gene (Genbank: X04472), intron 3 sequence was placed directly downstream of the stop codon. Together with one copy of *RBCS2* intron 1 present in the 5`-untranslated region of the promoter, all three introns occur once and in their physiological order. The extracellular expression of erythropoietin is mediated by the first 21 amino acids of the *ARS2* gene of *Chlamydomonas reinhardtii* (AF333184). A C-terminal His_6-tag allows purification from the culture medium.

cepo-gene: 600 bp
with a recombinant 5`-*Xho*I-restriction site and 3`-*Bam*HI-restriction sites

```
1      AAACTCGAGA TGGGTGCCCT CGCGGTGTTC GCCGTCGCTT GCCTCGCGGC
51     AGTGGCGTCG GTTGCGCATG CGGCCCCGCC TCGCCTGATC TGCGACTCCC
101    GCGTGCTGGA GCGCTACCTG CTGGAGGCGA AGGAGGCCGA GAACATCACG
151    ACCGGCTGCG CCGAGCACTG CTCCCTGAAC GAGAACATTA CCGTGCCCGA
201    TACGAAGGTC AACTTCTACG CGTGGAAGCG CATGGAGGTG GGCCAGCAGG
251    CCGTCGAGGT GTGGCAGGGC CTGGCCCTGC TGAGCGAGGC GGTGCTCCGC
301    GGCCAAGCCC TGCTGGTGAA CAGCAGCCAG CCCTGGGAGC CCCTCCAACT
351    GCACGTGGAC AAGGCGGTGT CCGGCCTGCG CAGCCTCACC ACCCTGCTGC
401    GCGCCCTGGG CGCGCAGAAG GAGGCCATCA GCCCCCCGGA CGCCGCGAGC
451    GCCGCACCCC TGCGCACGAT CACCGCCGAC ACGTTCCGCA AGCTGTTCCG
501    CGTGTACAGC AATTTCCTGC GCGGCAAGCT GAAGCTGTAC ACGGGCGAGG
551    CCTGTCGCAC CGGTGACCGC CACCACCATC ACCACCATTA AGGATCCAAA
```

cepo-ORF: MLE + ARS2 + erythropoietin + His_6 = 196 aa, 21514 Da
aa-sequence same as for human erythropoietin without human signal peptide (GenBank: P01588, aa 28-193)
Note that the ORF encodes for additional aa:

MLE if start-ATG upstream of the *Xho*I-restriction site is used

MGALAVFAVACLAAVASVAHA *ARS2* signal peptide

HHHHHH His_6-tag

Intron 2 was insterted between aa 98/99

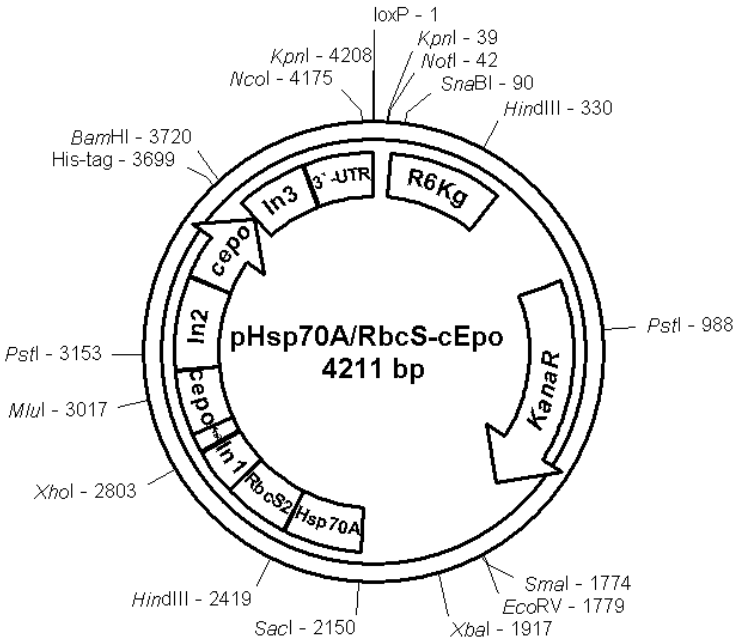
```
1      MLEMGALAVF AVACLAAVAS VAHAAPPRLI CDSRVLERYL LEAKEAENIT
51     TGCAEHCSLN ENITVPDTKV NFYAWKRMEV GQQAVEVWQG LALLSEAVLR
101    GQALLVNSSQ PWEPLQLHVD KAVSGLRSLT TLLRALGAQK EAISPPDAAS
151    AAPLRTITAD TFRKLFRVYS NFLRGKLKLY TGEACRTGDR HHHHHH*
```

Plasmid-sequence:	pHsp70A/RbcS2-cEpo	4211 bp
	loxP:	1-34
	HSP70A-promoter:	2150-2419
	RBCS2-promoter:	2420-2649
	intron1 from *RBCS2*:	2650-2803
	cepo:	2804-3716
	leader from *ARS2*	2809-2871
	intron2 from *RBCS2*:	3096-3419
	intron3 from *RBCS2*:	3726-3963
	His_6-tag	3699-3716
	RBCS2-3`-UTR:	3970-4208

```
   1 ATAACTTCGT ATAATGTATG CTATACGAAG TTATGGTACC GCGGCCGCGT
  51 AGAGGATCTG TTGATCAGCA GTTCAACCTG TTGATAGTAC GTACTAAGCT
 101 CTCATGTTTC ACGTACTAAG CTCTCATGTT TAACGTACTA AGCTCTCATG
 151 TTTAACGAAC TAAACCCTCA TGGCTAACGT ACTAAGCTCT CATGGCTAAC
 201 GTACTAAGCT CTCATGTTTC ACGTACTAAG CTCTCATGTT TGAACAATAA
 251 AATTAATATA AATCAGCAAC TTAAATAGCC TCTAAGGTTT TAAGTTTTAT
 301 AAGAAAAAAA AGAATATATA AGGCTTTTAA AGCTTTTAAG GTTTAACGGT
 351 TGTGGACAAC AAGCCAGGGA TGTAACGCAC TGAGAAGCCC TTAGAGCCTC
 401 TCAAAGCAAT TTTGAGTGAC ACAGGAACAC TTAACGGCTG ACATGGGAAT
 451 TAGCTTCACG CTGCCGCAAG CACTCAGGGC GCAAGGGCTG CTAAAGGAAG
 501 CGGAACACGT AGAAAGCCAG TCCGCAGAAA CGGTGCTGAC CCCGGATGAA
 551 TGTCAGCTAC TGGGCTATCT GGACAAGGGA AAACGCAAGC GCAAAGAGAA
 601 AGCAGGTAGC TTGCAGTGGG CTTACATGGC GATAGCTAGA CTGGGCGGTT
 651 TTATGGACAG CAAGCGAACC GGAATTGCCA GCTGGGGCGC CCTCTGGTAA
 701 GGTTGGGAAG CCCTGCAAAG TAAACTGGAT GGCTTTCTTG CCGCCAAGGA
 751 TCTGATGGCG CAGGGGATCA AGATCTGATC AAGAGACAGG ATGAGGATCG
 801 TTTCGCATGA TTGAACAAGA TGGATTGCAC GCAGGTTCTC CGGCCGCTTG
 851 GGTGGAGAGG CTATTCGGCT ATGACTGGGC ACAACAGACA ATCGGCTGCT
 901 CTGATGCCGC CGTGTTCCGG CTGTCAGCGC AGGGGCGCCC GGTTCTTTTT
 951 GTCAAGACCG ACCTGTCCGG TGCCCTGAAT GAACTGCAGG ACGAGGCAGC
1001 GCGGCTATCG TGGCTGGCCA CGACGGGCGT TCCTTGCGCA GCTGTGCTCG
1051 ACGTTGTCAC TGAAGCGGGA AGGGACTGGC TGCTATTGGG CGAAGTGCCG
1101 GGGCAGGATC TCCTGTCATC TCACCTTGCT CCTGCCGAGA AAGTATCCAT
1151 CATGGCTGAT GCAATGCGGC GGCTGCATAC GCTTGATCCG GCTACCTGCC
1201 CATTCGACCA CCAAGCGAAA CATCGCATCG AGCGAGCACG TACTCGGATG
1251 GAAGCCGGTC TTGTCGATCA GGATGATCTG GACGAAGAGC ATCAGGGGCT
1301 CGCGCCAGCC GAACTGTTCG CCAGGCTCAA GGCGCGCATG CCCGACGGCG
1351 AGGATCTCGT CGTGACACAT GGCGATGCCT GCTTGCCGAA TATCATGGTG
1401 GAAAATGGCC GCTTTTCTGG ATTCATCGAC TGTGGCCGGC TGGGTGTGGC
1451 GGACCGCTAT CAGGACATAG CGTTGGCTAC CCGTGATATT GCTGAAGAGC
1501 TTGGCGGCGA ATGGGCTGAC CGCTTCCTCG TGCTTTACGG TATCGCCGCT
1551 CCCGATTCGC AGCGCATCGC CTTCTATCGC CTTCTTGACG AGTTCTTCTG
1601 AGCGGGACTC TGGGGTTCGA AATGACCGAC CAAGCGACGC CCAACCTGCC
1651 ATCACGAGAT TTCGATTCCA CCGCCGCCTT CTATGAAAGG TTGGGCTTCG
1701 GAATCGTTTT CCGGGACGCC GGCTGGATGA TCCTCCAGCG CGGGGATCTC
1751 ATGCTGGAGT TCTTCGCCCA CCCCGGGATA TCCGGATATA GTTCCTCCTT
1801 TCAGCAAAAA ACCCCTCAAG ACCCGTTTAG AGGCCCCAAG GGGTTATGCT
1851 AGTTATTGCT CAGCGGTGGC AGCAGCCAAC TCAGCTTCCT TTCGGGCTTT
1901 GTTAGCAGCC GGATCTTCTA GAATCCCCAG CATGCCTGCT ATTGTCTTCC
1951 CAATCCTCCC CCTTGCTGTC CTGCCCCACC CCACCCCCCA GAATAGAATG
2001 ACACCTACTC AGACAATGCG ATGCAATTTC CTCATTTTAT TAGGAAAGGA
2051 CAGTGGGAGT GGCACCTTCC AGGGTCAAGG AAGGCACGGG GGAGGGGCAA
2101 ACAACAGATG GCTGGCAACT AGAAGGCACA GTCGAGGCTG ATAGCGAGCT
2151 CGCTGAGGCT TGACATGATT GGTGCGTATG TTTGTATGAA GCTACAGGAC
```

```
2201 TGATTTGGCG GGCTATGAGG GCGGGGGAAG CTCTGGAAGG GCCGCGATGG
2251 GGCGCGCGGC GTCCAGAAGG CGCCATACGG CCCGCTGGCG GCACCCATCC
2301 GGTATAAAAG CCCGCGACCC CGAACGGTGA CCTCCACTTT CAGCGACAAA
2351 CGAGCACTTA TACATACGCG ACTATTCTGC CGCTATACAT AACCACTCAG
2401 CTAGCTTAAG ATCCCATCAA GCTTGCATGC CGGGCGCGCC AGAAGGAGCG
2451 CAGCCAAACC AGGATGATGT TTGATGGGGT ATTTGAGCAC TTGCAACCCT
2501 TATCCGGAAG CCCCCTGGCC CACAAAGGCT AGGCGCCAAT GCAAGCAGTT
2551 CGCATGCAGC CCCTGGAGCG GTGCCCTCCT GATAAACCGG CCAGGGGGCC
2601 TATGTTCTTT ACTTTTTTAC AAGAGAAGTC ACTCAACATC TTAAAATGGC
2651 CAGGTGAGTC GACGAGCAAG CCCGGCGGAT CAGGCAGCGT GCTTGCAGAT
2701 TTGACTTGCA ACGCCCGCAT TGTGTCGACG AAGGCTTTTG GCTCCTCTGT
2751 CGCTGTCTCA AGCAGCATCT AACCCTGCGT CGCCGTTTCC ATTTGCAGGA
2801 TGCTCGAGAT GGGTGCCCTC GCGGTGTTCG CCGTCGCTTG CCTCGCGGCA
2851 GTGGCATCGG TTGCGCATGC GGCCCCGCCT CGCCTGATCT GCGACTCCCG
2901 CGTGCTGGAG CGCTACCTGC TGGAGGCGAA GGAGGCCGAG AACATCACGA
2951 CCGGCTGCGC CGAGCACTGC TCCCTGAACG AGAACATTAC CGTGCCCGAT
3001 ACGAAGGTCA ACTTCTACGC GTGGAAGCGC ATGGAGGTGG GCCAGCAGGC
3051 CGTCGAGGTG TGGCAGGGCC TGGCCCTGCT GAGCGAGGCG GTGAGCTTGC
3101 GGGGTTGCGA GCAACACTCC AGCAACGAAC AGTGCCCAAG TCAGGAATCT
3151 GCAGTCAGCC TGGGCTTTCG GCGGCTTTTT CTTGGGCAAA CAGCTTGCAC
3201 TCATGCCAGC GCGGCTTGTC CAGCCTCACT TGAGCTTTCC AGCTGCTACC
3251 AGCCGGGCTA TACGACAGCG ACAGAGCCAT AGCGTGGAAT CACTTATTTG
3301 GGTTGCCGAA GTAGCGGTCG GAGCGTGAGT TCTTGGTCAA GCCGCCCCTT
3351 ATCCGGTTCC TGTCCGTGTC TTTGTCCCTC GTTCACCCTT CGCGGCACCC
3401 TTCATCCCCT TGCTTGCAGG TGCTCCGCGG CCAAGCCCTG CTGGTGAACA
3451 GCAGCCAGCC CTGGGAGCCC CTCCAACTGC ACGTGGACAA GGCGGTGTCC
3501 GGCCTGCGCA GCCTCACCAC CCTGCTGCGC GCCCTGGGCG CGCAGAAGGA
3551 GGCCATCAGC CCCCCGGACG CCGCGAGCGC CGCACCCCTG CGCACGATCA
3601 CCGCCGACAC GTTCCGCAAG CTGTTCCGCG TGTACAGCAA TTTCCTGCGC
3651 GGCAAGCTGA AGCTGTACAC GGGCGAGGCC TGTCGCACCG GTGACCGCCA
3701 CCACCATCAC CACCATTAAG GATCCGTAAG TCTGGCGAGA GCCCGACGGG
3751 TCCACTGTGG CACTGGGTTA GCTTTTGGCA CACGGGTCCA CTGTGGCACT
3801 GGTTAGCTTG GCACCGGGAC AGCGCCTATC TCACCGCGGG GAACTGACGC
3851 ATACCCCTGC TCGTGCTTCA GCACGGAAAA GCAAGGGGCC CAATTCCATC
3901 TTTGGTGGTT CTGTGCGCTG GTGACTGAAC CTCTTCTCCC TCCCATTTCC
3951 CGTGCGCCCG CAGAGATCCC CGCTCCGTGT AAATGGAGGC GCTCGTTGAT
4001 CTGAGCCTTG CCCCCTGACG AACGGCGGTG GATGGAAGAT ACTGCTCTCA
4051 AGTGCTGAAG CGGTAGCTTA GCTCCCCGTT TCGTGCTGAT CAGTCTTTTT
4101 CAACACGTAA AAAGCGGAGG AGTTTTGCAA TTTTGTTGGT TGTAACGATC
4151 CTCCGTTGAT TTTGGCCTCT TTCTCCATGG GCGGGCTGGG CGTATTTGAA
4201 GCGGGTACCC C
```

Plasmid-map:



Reference: Heitzer, M. and Zschoernig, B (2007) Construction of modular tandem expression vectors for the green alga *Chlamydomonas reinhardtii* using the Cre/*lox*-system. Biotechniques 43(3), 324-32.