Combinatorial Mutagenesis and Selection of Improved Signal Sequences and Their Application for High-Level Production of Translocated Heterologous Proteins in *Escherichia coli*

Tonje Marita Bjerkan Heggeset,[a] Veronika Kucharova,[b] Ingemar Nærdal,[a] Svein Valla,[b] Håvard Sletta,[a] Trond Erling Ellingsen,[a] Trygve Brautaset[a]

SINTEF Materials and Chemistry, Department of Biotechnology, Trondheim, Norway[a]; Norwegian University of Science and Technology, Department of Biotechnology, Trondheim, Norway[b]

We previously designed the consensus signal peptide (CSP) and demonstrated that it can be used to strongly stimulate heterologous protein production in *Escherichia coli*. A comparative study using CSP and two bacterial signal sequences, pelB and ompA, showed that the effect of signal sequences on both expression level and translocation efficiency can be highly protein specific. We report here the generation of CSP mutant libraries by a combinatorial mutagenesis approach. Degenerated CSP oligonucleotides were cloned in frame with the 5′ end of the *bla* gene, encoding the mature periplasmic β-lactamase released from its native signal sequence. This novel design allows for a direct selection of improved signal sequences that positively affect the expression level and/or translocation efficiency of β-lactamase, based on the ampicillin tolerance level of the *E. coli* host cells. By using this strategy, 61 different CSP mutants with up to 8-fold-increased ampicillin tolerance level and up to 5.5-fold-increased β-lactamase expression level were isolated and characterized genetically. A subset of the CSP mutants was then tested with the alternative reporter gene *phoA*, encoding periplasmic alkaline phosphatase (AP), resulting in an up to 8-fold-increased production level of active AP protein in *E. coli*. Moreover, it was demonstrated that the CSP mutants can improve the production of the medically important human interferon α2b under high-cell-density cultivations. Our results show that there is a clear potential for improving bacterial signal sequences by using combinatorial mutagenesis, and bioinformatics analyses indicated that the beneficial mutations could not be rationally predicted.

Signal sequences target proteins to their site of function, and this has been exploited for the production of a number of recombinant proteins in *Escherichia coli* [1–4]. The periplasmic accumulation of a recombinant protein is known to provide several advantages, including a simplified downstream purification scheme, enhanced biological activity, correct folding, higher product stability and solubility, and the authentic N termini of the mature proteins [5]. Inclusion bodies also form in the periplasm [6], and in general there seem to be no strict correlation between the conformational quality of a protein and its solubility state [7]. Several proteins have been shown to retain biological activity within periplasmic inclusion bodies [7–10], which could be described as dynamic reservoirs consisting of both active and inactive protein with a variety of conformations [11].

There seems to be a rather flexible consensus, both at the DNA level and at the protein level, for signal sequences targeted to the Sec translocation pathway in *E. coli* (for reviews, see references [12], [13], and [14]) and the choice of signal sequence to ensure effective secretion should therefore be individually tested for each new protein [15–17]. Typically, signal sequences are composed of a short, positively charged amino-terminal domain with 1 to 3 basic amino acid residues, followed by a hydrophobic core of 7 to 15 hydrophobic and neutral amino acids [13, 18]. A more polar segment is found immediately preceding the cleavage site, which includes small amino acids at positions −3 and −1, forming a so-called Ala-X-Ala box that is recognized and cleaved by signal peptidase I [13, 18].

Recent studies have shown that signal sequences may play a more intricate role in protein expression than previously anticipated. A whole-genome study of the *E. coli* open reading frames revealed that the highest frequencies of rare codons are found in the signal sequences of secretory proteins [19]. An analysis of the amino acid distribution in the *E. coli* secretory proteins pointed out a bias for the AAA codon (encoding lysine) at the +2 position [20], possibly due to its high rate of translation initiation. The purpose of rare codons in signal sequences is unclear; however, their exchange for more frequent codons has been shown to reduce the mature protein yield of the β-lactamase and the maltose-binding protein [21, 22]. In accordance with these reports is a study of the adaptation between the transcript sequence and the tRNA pool, where it was suggested that the order of frequent and rare codons is important and can be utilized by evolution to design the best possible schedule for the ribosomal flow on the transcripts [23]. We previously designed the consensus signal peptide (CSP), based on sequence alignments of available bacterial signal sequences, and CSP was demonstrated to function efficiently for translocation of two different heterologous proteins produced in *E. coli* [17, 24]. A functional comparison of CSP with the two frequently used bacterial signal sequences pelB and ompA showed that the choice of signal sequence for effective expression and translocation can be highly protein specific and apparently unpredictable in *E. coli*.

Directed evolution using random mutagenesis, followed by screening and selection, was originally invented to improve enzymes (25, 26). Among the many techniques used to create random mutant libraries the most commonly used are error-prone PCR (for a review, see reference 27), DNA shuffling (28), and chemical mutagenesis (29, 30). We have previously applied a combinatorial mutagenesis approach involving the use of degenerated oligonucleotides to improve the bacterial expression system *Pm-xylS* (31–33). By using synthetic oligonucleotide mixtures, the mutagenesis can be focused to specific genetic regions (20 to 100 bp in length), a method that allows for multiple point mutations to be introduced simultaneously. A different approach was applied by Caspers et al. (16) on the amyE signal peptide. By using saturation mutagenesis on the sequence encoding the N domain (amino acids 2 to 7) of amyE, they were able to increase the production yield of their reporter cutinase ~3-fold (16). We report here the application of combinatorial mutagenesis and the selection of CSP to develop new and improved signal sequences useful for high expression and translocation of heterologous proteins in *E. coli*. Libraries were constructed by using degenerated oligonucleotide mixtures encoding CSP sequence variants to substitute for the native signal sequence of the periplasmic β-lactamase. By screening these libraries with respect to increased ampicillin tolerance of the *E. coli* host, a number of improved CSP mutants could be easily selected. The results demonstrate that signal sequences can be improved and optimized for efficient production and translocation of heterologous proteins in *E. coli*, and the genetic tools and technology developed here should be directly useful to improve any bacterial signal sequence.

MATERIALS AND METHODS

Biological materials, DNA manipulations, and growth conditions. The bacterial strains and plasmids used in the present study are listed in Table 1. *E. coli* strains were generally grown at 37°C in lysogeny broth (LB) (10 g of tryptone, 5 g of yeast extract, and 5 g of NaCl liter^{-1}) or on LA (LB with 15 g of agar liter^{-1}) supplemented with ampicillin (0.2 g liter^{-1} [unless stated differently]) or kanamycin (0.05 g liter^{-1}) when appropriate. For expression experiments, recombinant cells were grown at 30°C, and induction of the *Pm-xylS* system was done by adding *m*-toluic acid as indicated. Standard recombinant DNA procedures were performed as described by Sambrook and Russell (34), and transformations were done according to the RbCl method (New England BioLabs). Plasmids were isolated by using a Qiagen midi kit (Qiagen) or a Wizard Plus SV miniprep kit (Promega). DNA was extracted from agarose gel slabs or PCR mixtures by the QIAquick gel extraction or PCR purification kit (Qiagen). PCRs were run using an Expand high-fidelity PCR system kit (Roche). Custom PCR primers were supplied by Eurofins MWG operon, while spiked oligonucleotide mixes were supplied by Medprobe AS. Site-specific mutagenesis was done by using a QuikChange site-directed mutagenesis kit (Stratagene). DNA sequencing was performed by Eurofins MWG operon.

Vector constructions. (i) Vectors with the *bla* reporter gene. The plasmid pCSP1bla, used for construction of mutant libraries (Fig. 1A), is based on the vector pIB11 (32) in which the endogenous secretion signal sequence of the *bla* gene was replaced by the CSP signal sequence from pGM29CSP (17). A series of plasmids were made as intermediates of the final vectors pCSP1bla and pNSP1bla (expressing the *bla* gene without any signal sequence). Plasmid pTMB4 is a pIB11 derivative in which the endogenous NcoI restriction site in *xylS* has been removed. It was constructed by replacing the 0.6-kb BglII-BseRI fragment by the corresponding fragment from pTA13 (31). The *Pm-CSP-bla* expression cassette was then generated in pTMB4 through several intermediates. Plasmid pTMB5 was created by exchanging the 1.1-kb *Pm-bla*-encoding XbaI-KpnI fragment in pTMB4 by the 2.1-kb *Pm-CSP-gm-csf-cmyc-his₆* fusion fragment

from pGM29CSP. The 1.4-kb XbaI-SbfI fragment from pGM29CSP, containing the *Pm-CSP-gm-csf-cmyc-his₆* fusion, was inserted into pLITMUS28 (New England BioLabs), creating pTMB6. Plasmid pTMB7 was generated by subcloning the mature part of the *bla* gene, PCR amplified from pLITMUS28 (primer pair, 5′-AT<u>CCATGG</u>CTCACCCAGAAACGC TG-3′ and 5′-CTC<u>AAGCTT</u>ACCAATGCTTAATCAGTGAGGCAC--3′) into the NcoI-HindIII sites of pTMB6, in frame with the CSP sequence, replacing a 0.5-kb fragment (fusion coding for GM-CSF-cMyc-His₆). Finally, pCSP1bla was made by replacing the 1.4-kb XbaI-SbfI *Pm-CSP-gm-csf-cmyc-his₆* fusion fragment of pTMB5 by the 1.8-kb *Pm-CSP-bla* fragment from pTMB7. pCSP1S1bla to pCSP1S61bla and pCSP1Si1bla to pCSP1Si15bla are pCSP1bla derivatives in which the CSP encoding sequence (NdeI-NcoI) was replaced by annealed, synthetic oligonucleotides corresponding to the various mutant CSP sequences. The pBSP1bla vector was made from pCSP1bla, by replacing the NdeI-NcoI fragment containing the CSP sequence, with annealed oligonucleotides corresponding the native *bla* signal sequence (5′-TATGAGTATTCAACATTTCCGTGT CGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTCG C-3′ and 5′-CATGGCGAAAACAGGAAGGCAAAATGCCGCAAAAAA GGGAATAAGGGCGACACGGAAATGTTGAATACTCA-3′). pTMB14 was constructed by removing the NdeI-NcoI-encoding CSP fragment in pCSP1bla. The DNA ends were filled by Klenow polymerase before the plasmid was circularized. The 3.2-kb BglII-EcoRI fragment from pTMB14 was inserted into pLITMUS28, generating pTMB15, and the 2 bp in the transition between *Pm* and *bla* in pTMB15 were replaced by an NdeI site via site-directed mutagenesis using the primers 5′- CAATAAT AATGGAGTCATGAA<u>CATATG</u>GCTCACCCAGAAAC-3′ and 5′-GTTT CTGGGTGAGC<u>CATATG</u>TTCATGACTCCATTATTATTG-3′, generating pTMB16. Finally, pNSP1bla was constructed by replacing the 3.2-kb BglII-EcoRI fragment in pCSP1bla with the pTMB16 counterpart. All expression vectors constructed were verified by DNA sequencing and established in *E. coli* DH5α.

(ii) Vectors with the *phoA* reporter gene and with the codon-optimized *ifn-α2b$_S$* gene. The NcoI-BstAPI fragment in pCSP1bla, encoding the *bla* gene, was substituted by the *phoA* gene, giving rise to pCSP1phoA. The enhanced *phoA* gene used described by Mandecki et al. (35) was PCR amplified from pHOGscFv141-AP-His (unpublished data, kindly provided by Affitech AS, Oslo) with the primers 5′-TTTCCATGGCACGGG CACCAGAAATGCCTGTTC-3′ and 5′-TTTGCACAATGTGCTTATTT CAGCCCCAGAGCGGCTTTC-3′. In the vectors pCSP1S2phoA, pCSP1S38phoA, pCSP1S48phoA, pCSP1S60phoA, and pCSP1S61phoA, the CSP encoding sequence (NdeI-NcoI) was replaced by annealed, synthetic oligonucleotides corresponding to the various mutant CSP sequences. The pASP1phoA control plasmid was constructed by exchanging the NdeI-NcoI fragment encoding the CSP signal sequence in pCSP1phoA by oligonucleotides corresponding to the endogenous *phoA* signal sequence (36): 5′-TATGAAACAAAGCACTATTGCACTGGCAC TCTTACCGTTACTGTTTACCCCTGTGACAAAAGC-3′ and 5′-CATG GCTTTTGTCACAGGGGTAAACAGTAACGGTAAGAGTGCCAGTGC AATAGTGCTTTGTTTCA-3′. This resulted in a construct with two additional amino acids in the fusion site (due to the NcoI site used for cloning). A vector with the CSP sequence fused in frame with the gene encoding the codon-optimized IFN-α2b was constructed by indirectly replacing the pelB signal sequence in pIFN30SpelB (17) by the CSP signal sequence from pGM29CSP generating pCSP2ifn (Fig. 1B). These plasmids have the same vector backbone and only differ with respect to the choice of signal sequence and target gene. For technical reasons, pCSP2ifn was cloned by replacing the NcoI-NotI fragment containing the *gm-csf* gene in pGM29CSP with the *ifn-α2b$_S$* gene from pIFN30SpelB and not by interchanging the short signal sequence encoding fragments (NdeI-NcoI) directly. Once established, selected CSP mutant signal sequences were introduced as annealed, synthetic oligonucleotides, replacing the corresponding NdeI-NcoI fragment containing the CSP sequence, generating the pCSP2S2ifn and pCSP2S61ifn plasmids. All plasmids were verified by DNA sequencing and established in *E. coli* DH5α. The IFN-α2b-express-

TABLE 1 Bacterial strains and plasmids

| Strain or plasmid(s)	Description[a]	Reference or source
E. coli		
DH5α	λ⁻ φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)*U169 recA1 endA1 hsdR17*(r$_K$⁻ m$_K$⁻) *supE44 thi-1 gyrA relA1*	Bethesda Research Laboratories
RV308	K-12 derivative used for industrial protein production; *lacI*q *su* Δ*lacX74 gal* IS II::OP308	ATCC 31608
Plasmids		
pIB11	RK2-based expression vector containing the *Pm-xylS* promoter system and *bla* with is native signal sequence as a reporter gene for *Pm*; Km^r; 8.1 kb	32
pGM29CSP	RK2-based vector expressing the CSP-GM-CSF-c-*myc*-His$_6$ fusion protein from the *Pm-xylS* promoter system; Ap^r; 8.8 kb	17
pTA13	RK2-based expression vector containing the *Pm-xylS* promoter system and *bla* as a reporter gene for *Pm*; Km^r; 8.1 kb	31
pLITMUS28	General cloning vector; Ap^r; 2.8 kb	New England Biolabs
pHOGscFv141-AP-His	A pBluescript II SK(+) derivative expressing the scFv141-AP-His$_6$ fusion protein from the *P$_{lac}$* promoter; Ap^r; 5.1 kb	Affitech AS (unpublished data)
pIFN30SpelB	RK2-based vector expressing the codon-optimized pelB-IFN-α2b-c-*myc*-His$_6$ fusion protein from the *Pm/xylS* promoter system; Ap^r; 8.9 kb	17
pTMB4	Derivative of pIB11 in which a 0.6-kb BglII-BseRI fragment in the *xylS* gene was replaced by the corresponding fragment from pTA13, removing an endogenic NcoI site; Km^r; 8.1 kb	This study
pTMB5	Derivative of pTMB4 in which the 1.1-kb XbaI-KpnI fragment was replaced by a 2.1-kb fragment from pGM29CSP; Km^r; 9.2 kb	This study
pTMB6	Derivative of pLITMUS28 in which a 1.4-kb XbaI-SbfI fragment from pGM29CSP was inserted; Ap^r; 4.2kb.	This study
pTMB7	Derivative of pTMB6 in which a 0.8kb PCR-fragment containing the *bla* gene generated from pLITMUS28 (see the text) was subcloned into the NcoI-HindIII sites, replacing a 0.5-kb fragment; Ap^r; 4.5 kb	This study
pCSP1bla	RK2-based expression vector containing the *Pm-xylS* promoter system and *bla* with the CSP signal sequence as a reporter gene for *Pm*; Km^r; 9.5 kb	This study
pCSP1S1bla–pCSP1S61bla	Derivatives of pCSP1bla in which the NdeI-NcoI fragment encoding CSP was replaced by oligonucleotides encoding the S1–S61 mutant sequences; Km^r; 9.5 kb	This study
pCSP1Si1bla–pCSP1Si15bla	Derivatives of pCSP1bla in which the NdeI-NcoI fragment encoding CSP was replaced by oligonucleotides encoding the Si1–Si15 mutant sequences; Km^r; 9.5 kb	This study
pBSP1bla	Derivative of pCSP1bla in which the NdeI-NcoI fragment encoding CSP was replaced by oligonucleotides encoding the native β-lactamase signal peptide; Km^r; 9.5 kb	This study
pTMB14	Derivative of pCSP1bla where the NdeI-NcoI CSP fragment was removed, the DNA ends filled by Klenow polymerase and the plasmid circularized; Km^r; 9.4 kb	This study
pTMB15	Derivative of pLITMUS28 in which a 3.2-kb BglII-EcoRI fragment from pTMB14 was inserted; Ap^r; 5.9 kb	This study
pTMB16	Derivative of pTMB15 where 2 bp in the transition between *Pm* and *bla* was removed by site-directed mutagenesis, generating an NdeI site; Ap^r; 5.9 kb	This study
pNSP1bla	Derivative of pCSP1bla encoding the *bla* gene without its native signal sequence; Km^r; 9.4 kb	This study
pCSP1phoA	Derivative of pCSP1bla in which the NcoI-BstAPI fragment encoding the *bla* reporter gene was replaced by a PCR fragment encoding the mature *phoA* gene (see the text); Km^r; 10.0 kb	This study
pCSP1S2phoA–pCSP1S61phoA	Derivatives of pCSP1phoA in which the NdeI-NcoI fragment encoding CSP was replaced by oligonucleotides encoding the S2, S38, S48, S60, and S61mutant sequences; Km^r; 10.0 kb	This study
pASP1phoA	Derivative of pCSP1phoA in which the NdeI-NcoI fragment encoding CSP was replaced by oligonucleotides encoding the native *phoA* signal sequence; Km^r; 10.0 kb	This study
pCSP2ifn	RK2-based vector expressing CSP-IFNα-2b-c-*myc*-His$_6$ fusion protein from the *Pm-xylS* promoter system; trfAcop271; Ap^r; 8.9 kb	This study
pCSP2S2ifn–pCSP2S61ifn	Derivatives of pCSP2ifn in which the NdeI-NcoI fragment encoding CSP was replaced by oligonucleotides encoding the S2 and S61 mutant sequences; Ap^r; 8.9 kb	This study

[a] Ap^r, ampicillin resistance; Km^r, kanamycin resistance.

ing vectors were, in addition to DH5α, also established in the *E. coli* production host strain RV308 to be used for production analysis in high-cell-density cultivations.

Combinatorial mutagenesis for the construction of signal sequence mutant libraries. To introduce mutations in the CSP coding sequence, a strategy involving degenerated, synthetic oligonucleotides, similar to the protocol previously described for the *Pm* promoter and the untranslated region, was used (32, 33). Synthetic oligonucleotides were designed to constitute a double-stranded DNA fragment with the CSP sequence and NdeI and NcoI compatible ends when annealed for subsequent easy cloning into the pCSP1bla vector. Eight different nucleotide mixtures, indicated in the oligonucleotide sequences by the numbers 1 to 8, were used to synthesize degenerated oligonucleotide mixtures (MedProbe AS). The probability of keeping the original base in each position was set to 79%
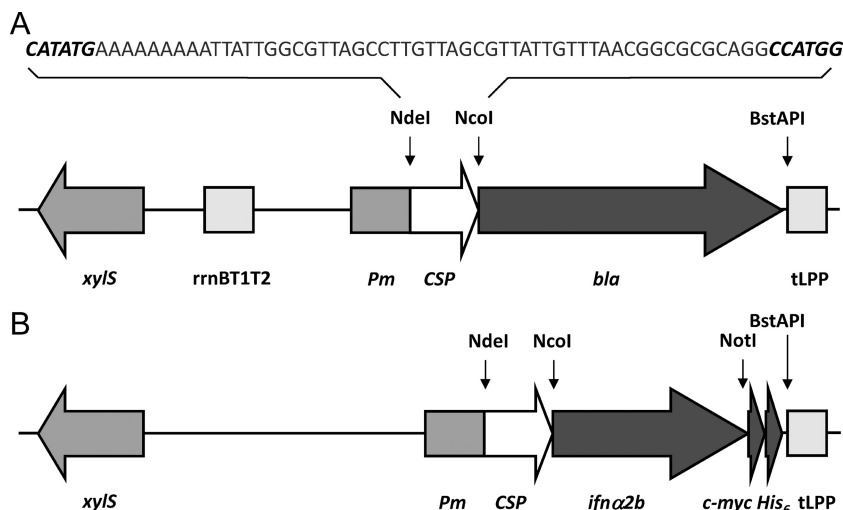
FIG 1 Schematic representation of the expression related features of the pCSP1bla screening vector (A) and the pCSP2ifn protein expression vector (B). In these vectors, the signal peptide and target genes are easily exchanged using the NdeI/NcoI and NcoI/BstAPI (removing the c-myc-His$_6$ tag) or the NcoI/NotI (preserving the c-myc-His$_6$ tag) restriction sites, respectively. *Pm*, positively regulated promoter; *xylS*, gene encoding the *Pm* activator; *bla*, ampicillin resistance gene encoding β-lactamase; tLPP, transcriptional terminator; rrnBT1T2, bidirectional transcriptional terminator; *ifnα2b*, codon optimized gene encoding the human cytokine IFN-α2b. The DNA sequence corresponding to the CSP region is displayed at the top with the unique NdeI and NcoI restriction sites in bold. The vector backbones of pCSP1bla and pCSP2ifn are generally conserved but differ with respect to the selection marker (*kan* in pCSP1bla and *bla* in pCSP2ifn), and pCSP2ifn harbors the cop271C mutant of the *trfA* replication gene, leading to a 3- to 4-fold-increased plasmid copy number (50).

(80% for some bases in the silent library), and the other accepted bases were introduced at equal frequencies. The mixtures used were as follows: 1, 79% A + 7% C + 7% G + 7% T; 2, 7% A + 79% C + 7% G + 7% T; 3, 7% A + 7% C + 79% G + 7% T; 4, 7% A + 7% C + 7% G + 79% T; 5, 80% A + 20% G; 6, 80% C + 20% T; 7, 20% A + 80% G; and 8, 20% C + 80% T. A completely randomized CSP library was generated by replacing the original CSP sequence by an oligonucleotide mixture defined by 5'-TATG11111111144144332344132244344132344144344411233232321 33C-3', while a silent CSP library (Si library) was generated from the oligonucleotide mixture defined by 5'-TATGA5AA5AA58T18T3GC38T1GC28T38T1GC38T18T3TT8AA6GG2GC3CA7GC-3'. The noncoding strands in both libraries were kept complementary to the original CSP sequence (5'-CATGGCCTGCGCGCCGTTAAACAATAACGCTAACAAGGCTAACGCCAATAATTTTTTTTTCA-3'). Wild-type and mutated oligonucleotides were mixed (100 μM each) and phosphorylated using polynucleotide kinase. NaCl was added to a final concentration of 200 mM, and the phosphorylated oligonucleotides were annealed by gradual cooling from 95 to 20°C for 20 min in a thermocycler. Dilutions of the resulting DNA fragments were ligated into the NdeI-NcoI-digested plasmid pCSP1bla, which had been dephosphorylated by calf intestinal phosphatase and purified by using a QIAquick PCR purification kit. The ligated plasmids were transferred into *E. coli* DH5α using kanamycin as selection. Approximately 137,000 transformants containing random CSP mutants were mixed to constitute the CSP library, while the Si library was a mixture of ~150,000 transformants.

Selection of CSP mutant libraries in *E. coli* cells based on increased ampicillin tolerance level. Mutant strains with increased expression and/or translocation of β-lactamase were selected essentially as described for the *Pm* high-level expression mutants by Winther-Larsen et al. (37). In brief, 9 ml of LB with kanamycin (0.05 g liter^{-1}) was inoculated with 90 μl of library cell mixture and grown at 37°C for ~4 h. Then, 10^{-3} to 10^{-8} dilutions were plated onto LA containing *m*-toluic acid (50 to 500 μM) and ampicillin (0.2 to 12 g liter^{-1}) or LA containing kanamycin only. The plates were incubated at 30°C for 2 days. The ampicillin concentration levels tolerated by the candidates identified in the screen were then determined as follow. Individual colonies were inoculated into 100 μl of LB containing kanamycin in 96-well microtiter plates (Nunc), followed by incubation at 30°C overnight. The strain harboring pCSP1bla was in-

cluded as a wild-type control in a few wells. The cells were diluted twice by a 96-pin replicator into microtiter plates with 100 μl of LB in each well and subsequently spotted onto LA with *m*-toluic acid (0 to 100 μM) and various ampicillin concentrations (0 to 10 g liter^{-1}). The plates were incubated at 30°C and inspected for growth after 1 and 2 days. The DNA sequence of the signal sequence region was determined for selected candidates, and all of the mutants were reproduced by annealing synthetic oligonucleotides with the different mutant signal sequences and ligating them into the NdeI-NcoI sites of pCSP1bla. Finally, the mutants were again verified by DNA sequencing.

Preparation of crude extracts for enzyme assays and protein analyses. Recombinant *E. coli* cells were diluted 100-fold from an overnight culture and grown in selective medium at 37°C. At an optical density at 600 nm (OD$_{600}$) of 0.1, the cells were induced by 0.1 mM *m*-toluic acid. The cells were further grown at 30°C for 5 h. Aliquots of cells were harvested by centrifugation (5,000 × g, 10 min, 4°C) and washed once in an assay buffer or 0.9% NaCl, and then the pellets were either kept at −20°C or further processed without storage. Cell extracts for enzyme assays were obtained by resuspending the cell pellets in specific assay buffer, and the cells were lysed by sonication (Branson Sonifier; DSM tip, 3 min on ice, duty cycle, 30%; output control, 2.5). Cell debris was removed by centrifugation at 10,000 × g for 45 min at 4°C.

The β-lactamase assay was performed as previously described (37). Alkaline phosphatase activities were measured by determining the production of *p*-nitrophenol from *p*-nitrophenyl phosphate, which was monitored spectrophotometrically at 405 nm. Total volumes of 160 μl in 96-well plates were used, and the assay mixture consisted of 100 mM Tris (pH 9.5), 5 mM MgCl$_2$, 0.1% Tween 20, and 2.7 mM *p*-nitrophenyl phosphate. Enzymatic assays were performed using enzyme amounts in the linear range. All enzyme assays were carried out in triplicates, and all enzyme activity analyses were repeated at least twice with enzyme extracts obtained from independently grown cultures.

SDS-PAGE and Western blotting was performed essentially as described elsewhere (24), using the primary antibodies Mouse Anti-Tetra-His IgG1 BSA-free (Qiagen, catalog no. 34670) diluted 1:7,500 (for His$_6$-tagged proteins) or mouse monoclonal β-lactamase antibody (Abcam) diluted 1:5,000 (for β-lactamase), and the secondary antibody rabbit anti-mouse IgG conjugated to horseradish peroxidase (Dako, catalog no.

PO260) diluted 1:1,500. Signals were developed by using 3,3′,5,5′-tetramethylbenzidine. For β-lactamase, the same cell-free protein extracts as in the enzymatic activity assays were applied, while total cell extracts were used in high-cell-density cultivation (HCDC) analyses (see below). Before loading, the HCDC cells were resuspended in SDS-PAGE running buffer and lysed by sonication (Branson sonifier, DSM tip, four × 90 s, with 30-s cooling periods, 30% duty control, and output control at 3.0). Protein concentrations were determined in samples diluted 50- to 200-fold in water by using a Bio-Rad DC protein assay (Bio-Rad Laboratories) as described by the manufacturer, using bovine serum albumin as the reference protein.

In silico analysis of mutant signal sequences. A selection of available bioinformatic tools was applied to analyze the subset of 12 CSP mutants in a *bla* context. The 20 amino acids of the signal peptide (CSP or its mutant variant), together with the first 30 amino acids of the β-lactamase protein, were submitted to the SignalP 4.0 server (http://www.cbs.dtu.dk/services/SignalP/ [38]), which predicts the presence and location of signal peptide cleavage sites in amino acid sequences. The net charge of the N domain was found by counting the number of charged amino acids within the first five residues of each signal peptide sequence, with amino acids D and E defined as −1, R and K defined as +1, and any other amino acid defined as 0. The net hydrophobicity was calculated as the sum of hydrophobicities for the amino acids in the hydrophobic core (residues 6 to 17), based on the normalized consensus hydrophobicity scale (47). Translation initiation rates were predicted for the mRNA region −32/+99 (base +1 is the adenine of the translation initiation codon) of the original CSP sequence and the subset of 12 CSP mutant variants using the reverse engineering tool of the RBS calculator website (https://salis.psu.edu/software/doReverseRBS [39]), with input settings selected for the model organism *E. coli* K-12 substrain DH10B. The CSP wild-type mRNA sequence was (the translation initiation codon is underlined): 5′-AACAGAAACAATA ATAATGGAGTCATGAACAT<u>ATG</u>AAAAAAAAAATTATTGGCGTTAG CCTTGTTAGCGTTATTGTTTAACGGCGCGCAGGCCATGGCTCAC CCAGAAACGCTGGTGAAAGTAAAAGATGCT-3′. The direct output of the RBS calculator is a list of start codons in the mRNA transcript, and their predicted translation initiation rates on a relative scale from <1 to 100,000+. In the present study, the values ranged from 522.82 (S46) to 9316.36 (S7), and the best score was generally associated with the position 32 start codon, which correlates to the true translation initiation site. To simplify data comparison, relative translation initiation rates were calculated, and the arbitrary number was set to 1 for the CSP sequence result. For all of the sequences investigated, the calculations were reported to not be at equilibrium, meaning that the mRNA was predicted to contain long-range nucleotide base pairings that are unlikely to form within the relevant time scale. Extending the coding region sequence beyond the length described here did not have any effect on the predictions of the translation initiation rate. The same RNA sequences were submitted to the RNAfold web server of the Vienna RNA package (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi [40]) for calculations of the minimum free energy (kcal mol^{-1}) of the secondary structure of single-stranded RNA.

HCDC. HCDC of *E. coli* RV308 cells harboring pIFN30SpelB, pCSP2ifn, pCSP2S2ifn, or pCSP2S61ifn were performed by automatically controlled fed-batch cultivation in DASGIP_1L fermentors. The defined preculture medium and the main culture medium used were composed as described elsewhere (41), whereas feeding solution 1 (FS1) was slightly modified and contained glucose (500 g kg^{-1} solution) and MgSO$_4$·7H$_2$O (15.8 g kg^{-1} solution). All media were supplemented with ampicillin (0.1 g liter^{-1}). Inducer solution (0.5 M) was prepared by dissolving *m*-toluic acid in ethanol.

The fermentation precultures were prepared as described previously (24). From these cultures, cells were used to inoculate 0.35 liter of prewarmed main culture medium in the fermentor to an initial OD$_{600}$ of 0.05. All fermentations were performed at 30°C, and the pH was maintained at 6.8 by the addition of 12.5% (vol/vol) NH$_3$. Antifoam (Adecanol LG-109; Asahi Denka Kogyo, Japan) was added from the start (50 μl

liter^{-1}) and thereafter when needed. The airflow rate was 0.5 liter per liter of medium per min. The dissolved oxygen was maintained at 20% saturation by automatic adjustments of the stirrer speed (up to 1,200 rpm) and gradual oxygen enrichment at a maximum stirring rate (up to 50%). Inducer (*m*-toluic acid, 1 mM) was added to the fermentors when the cell growth reached an OD$_{600}$ of 100.

The fermentations were divided into three different phases as follows. (i) In the batch phase, glucose (25 g liter^{-1}) was included in the freshly inoculated main culture medium, and cell growth continued until all of the sugar had been consumed. (ii) In the exponential feeding phase, growth was controlled to a specific growth rate (μ) of 0.25 h^{-1} by feeding with FS1. The feed rate of FS1 was initially set to 18.7 g liter^{-1} h^{-1} and thereafter was increased gradually to 45 g liter^{-1} h^{-1}. (iii) In the induction phase, bacterial growth was controlled by automatic adjustments of the airflow oxygen concentration at a constant feed rate of FS1 of 45 g liter^{-1} h^{-1} so that the dissolved oxygen and stirring speed were maintained at 20% saturation and 1,200 rpm, respectively. Throughout the fermentations, pH, dissolved oxygen, feeding rate, airflow, and the molar fraction of CO$_2$ in the exhausted gas were monitored and recorded.

RESULTS AND DISCUSSION

Construction of the pCSP1bla vector, which can be used as an effective selection tool for improved signal sequences. We previously demonstrated that the *bla* gene, encoding periplasmic β-lactamase, can be used as a valuable reporter in *E. coli* enabling direct selection based on the ampicillin tolerance levels of recombinant cells in response to different *bla* expression levels (31–33). In pCSP1bla the native endogenous signal sequence of the *bla* gene is replaced by CSP in such a way that this new sequence can easily be replaced by any CSP variant in a one-step cloning procedure (Fig. 1). Accordingly, pCSP1bla should allow for direct selection of CSP mutants conferring an increased expression level of the *bla* gene and/or increased translocation efficiency of the β-lactamase protein. As controls, we constructed the analogous vectors pBSP1bla (containing the *bla* gene with its native signal sequence) and pNSP1bla (containing the *bla* gene without a signal sequence).

E. coli DH5α (pCSP1bla) was able to grow on solid media with ampicillin concentrations up to 1.0 g liter^{-1} under induced conditions (Fig. 2), while the corresponding control strains containing pBSP1bla or pNSP1bla were able to grow in the presence of ampicillin concentrations up to 0.2 g liter^{-1} and 5 mg liter^{-1}, respectively. Thus, substitution of the native signal secretion sequence with CSP in itself leads to a 5-fold-higher ampicillin tolerance of the host, and the results also confirmed that the *bla* gene needs a signal sequence to be efficiently produced and translocated to the periplasm, in agreement with previous reports (42–44).

CSP variants selected from a completely randomized CSP mutant library give rise to up to 8-fold-increased ampicillin tolerance level of the host cells. A completely randomized CSP library was constructed in pCSP1bla, using degenerated oligonucleotides, and all nucleotides between the NdeI and NcoI restriction sites were allowed to vary. The resulting library from ~137,000 colonies, growing on LA-Km plates, was screened for mutants able to grow in the presence of elevated ampicillin concentrations (6 to 12 g liter^{-1}) under *Pm*-XylS-induced conditions (20 to 500 μM *m*-toluic acid). Candidates with 6- to 8-increased ampicillin tolerance compared to the pCSP1bla-carrying cells were easily identified and, in order to make sure that the phenotypes were actually caused by mutations in pCSP1bla, plasmid DNA was isolated from promising candidates and retransferred into DH5α cells and functionally confirmed. In this way, a

|	DNA sequence	Amino acid sequence	Ampicillin tolerance (g/L) induced	uninduced
CSP	atgaaaaaaaaaattattggcgttagccttgttagcgttattgtttaacggcgcgcaggccatg	mkkkllalallallfngaqam	1.0	0.1
Si1	...........................t...........c........g...........	....	1.7	0.1
Si2	.......g.............g.........c....................	....	1.9	0.1
Si3	.......g................................g..........	....	1.2	0.1
Si10	......g........t...................................	....	1.4	0.1-
Si13	.....g.........g...........c.......................	....	1.5	0.1
S2	...t.c.g...taa.....t...g.......gc.........a..t...c........	.yrnk.v...c...l..p...	7	0.3
S7	...c..g.........gcca..........a.........a.....t..t...g........	.qe..cq...i........e..	8	0.2
S15	....g..gc.......t....g.g.....tt.a.....gcg..t....g.a.........	.rs...s.gv.li.a..gk..	7+	0.2
S38	......g...	..r................	7-	0.2+
S43	......gc........c..a....aga......t....c............	..s..f..em.v.........	8+	0.2
S46	......g..gt......t...........a.............t.......	..rs..s....e.........	8-	0.2+
S47	......g..c..t...at.....g.........c.....tc..........	..rtf.d....s..s.....	7-	0.2+
S48	......g...........ta....c.....g..c.....c......a...t....	..r...v.....v....eh..	8-	0.3
S49	......g...........ta....g......ag.g...c...gt....t........	..r....i...rvsly.....	8+	0.2+
S54	......g......c..t...g..a....aa....agg...tggg.t..............	..r.ff.vt.i.gfgi.....	7	0.2+
S60	...c..c..........cct.aag..........c..g.......gt.ttc...........	.qq..ssr...pv.lys....	8+	0.2
S61	......g..........g................................	........v...........	8	0.1
NoSP	...--	.-------------------	5·10⁻³	2.5·10⁻³

FIG 2 DNA and amino acid sequences of the original CSP and selected mutant variants. The corresponding ampicillin tolerance levels were obtained from spotting *E. coli* DH5α-strains harboring pCSP1bla and its derivatives with the 5 best silent CSP mutants and the 12 best random CSP mutants onto LA containing either ampicillin and m-toluic acid (20 μM, induced) or ampicillin only (uninduced). Ampicillin tolerance of DH5α (pNSP1bla) is included as a control (no signal sequence [NoSP]). Periods (.) indicate that the base or amino acid is the same as in the unmodified CSP sequence, while "−" indicates gaps in the sequence (in pNSP1bla). "+" behind the ampicillin concentration given indicates that poor growth was observed at even higher concentrations, while a "−" shows that the growth was reduced at the concentration indicated.

total of 61 different CSP mutants causing up to 8-fold-increased ampicillin tolerance levels of the cells were isolated. DNA sequencing revealed that the average number of point mutations was 8, and they were typically randomly distributed throughout the CSP coding region. Among these, the 12 most promising CSP mutants were selected based on their overall characteristics using different combinations of inducer and ampicillin concentrations (data not shown), and the plasmids were then resynthesized to ensure that the observed phenotypes are solely caused by the identified CSP mutations. The DNA and amino acid sequences, as well as the ampicillin tolerance profiles of these mutants, are summarized in Fig. 2.

An alternative codon optimization strategy of CSP resulted in an only 2-fold-increased ampicillin tolerance level of the host cells. Recent studies have highlighted the role of local mRNA folding (45) and a codon ramp (23) for optimizing heterologous gene expression. Consequently, the introduction of alternative codons at the 5′ coding sequence of a recombinant gene may potentially have a great influence on its expression level. Therefore, we next chose to generate an alternative library in pCSP1bla (designated the Si library) that contained only silent mutations in the CSP coding sequence and thus should not affect the function of CSP in translocation. The resulting recombinant cells were tested for ampicillin tolerance on solid medium under *Pm*-XylS-induced conditions (20 μM *m*-toluic acid) such as for the CSP random library. The Si library contained ~150,000 clones, and 14 promising candidates were isolated that could tolerate an up to ~2-fold-higher ampicillin concentration (1.9 g liter⁻¹) compared to those containing the original CSP sequence (see Fig. 2 for the five best Si mutants). Based on these results, it seemed clear that silent mutations in CSP can, to a certain degree, contribute to increased expression levels of the reporter gene. However, these improvements were moderate compared to when completely randomizing CSP (see above), and we therefore chose not to characterize these Si mutants any further.

The production level of active β-lactamase protein is up to 5.5-fold higher in the CSP mutants compared to when using the original CSP. To investigate whether the observed improved am-

picillin tolerance levels were truly accompanied by a concomitant increase in the β-lactamase production levels in the cells, the β-lactamase activities were measured spectrophotometrically in cell extracts from the 12 best CSP mutants. The strains DH5α (pCSP1bla) and DH5α (pNSP1bla) were included as controls. The results showed that the mutants displayed high and different β-lactamase activity profiles (Fig. 3a). For most of them, the β-lactamase activity and host ampicillin tolerance level correlated well, but mutant S61 displayed the same β-lactamase activity as when the original CSP sequence was used. One possible way of explaining this would be to assume that this mutant affects the translocation rate without affecting the total expression level of the β-lactamase, which would also have to imply that such a change affects the host tolerance to ampicillin under the conditions tested. It has previously been reported that increasing the hydrophobicity of the signal peptide can lead to reduced processing and the release of soluble β-lactamase in *E. coli* (46). Although we cannot rule out that this could be the case for S61, it seems unlikely since this particular mutant (like S38) contains only a single point mutation, leading to an alanine-to-valine substitution (Table 2), which only results in an insignificant increase in the net hydrophobicity compared to CSP. The net hydrophobicity of mutant S61 is also considerably lower than for the mutants S54 and S48 (see below), which were among the mutants displaying the highest β-lactamase activities in the assay. We also analyzed the extracts for soluble protein by Western blots to unravel any potential variation between the β-lactamase production level and the corresponding enzymatic activity (Fig. 3b). The results showed that several of the CSP mutants caused an at least 8-fold-increased β-lactamase production level compared to CSP, and there was generally a good correlation between the enzyme amount and the corresponding activity level. The data also confirmed that CSP mutant S61 gives rise to a similar amount of β-lactamase produced as the original CSP sequence, in agreement with the corresponding activity data. For some CSP mutants (e.g., S43 and S46), we noticed the appearance of faint bands presumably representing uncleaved cytoplasmic products (see also below). However, it cannot be ruled out
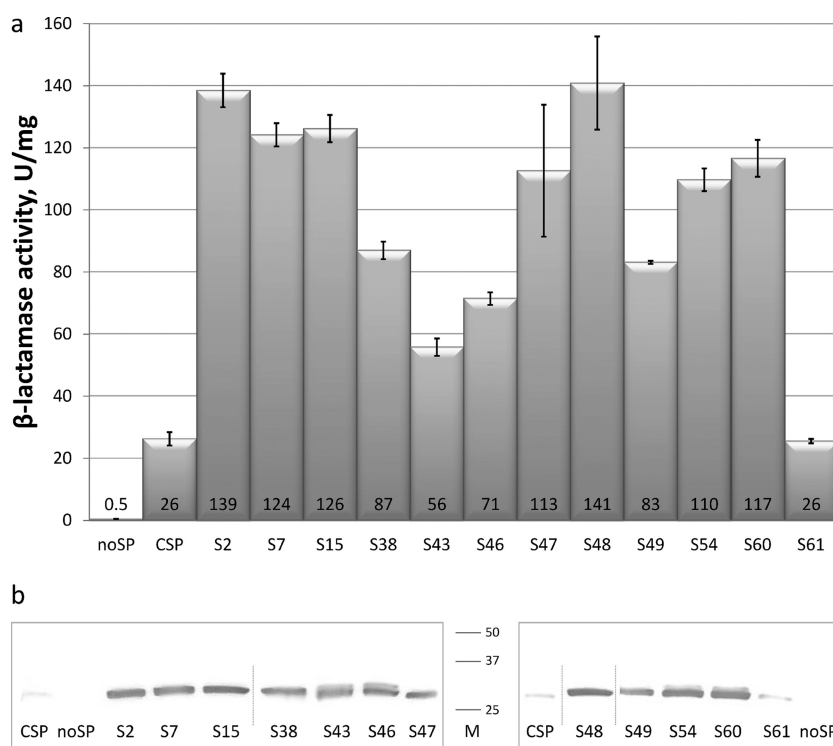
FIG 3 β-Lactamase production in *E. coli* DH5α strains harboring the plasmids pNSP1bla (no signal peptide [noSP]), pCSP1bla (CSP), and the 12 different plasmids pCSP1S2bla-pCSP1S61bla (CSP mutants S2 to S61). Gene expression was induced by 0.1 mM *m*-toluic acid. (a) Specific β-lactamase activities measured from cell extracts. 1 U = the amount of enzyme that catalyzes the transformation of 1 μmol of substrate min^{-1}. The results are the averages of two biological replicas performed in triplicates, and the standard errors are indicated. (b) Western blot detection of soluble β-lactamase in the same extracts as in panel a. The same total protein amount was applied in all wells. The sizes of the mature and the uncleaved β-lactamase proteins are approximately 29 and 32 kDa, respectively. The two boxes show independent blots; vertical lines within the boxes indicate that one or more samples have been removed from the data set. The M column represents the migration of ladder proteins of sizes 25, 37, and 50 kDa, respectively.

that they could be inner membrane-bound precursors that were released into the soluble fraction due to cell lysis. Whether these are active proteins remained unknown.

Bioinformatics analyses of the improved CSP mutants indicate that they could not be rationally predicted. We next analyzed our 12 selected CSP mutants by using available bioinformatics methods (see Materials and Methods) and the results are summarized in Table 2. With the exception of S7, CSP and the remaining CSP mutants were predicted by the SignalP 4.0 server (38) to be signal sequences that are cleaved off between amino acids 20 and 21 (AQA-MA; CSP, S38, and S61) or between amino acids 22 and 23 (AMA-HP; S2, [S7], S15, S46 to S49, S54, and S60). The net charge of the N domain varied among the signal sequences, with a maximum of +3 (seen for several mutants and the original CSP) and a minimum of 0 (S7). The net hydrophobicity of the hydrophobic core also showed a high variation within the data set and was in particularly low for S60 (4.18) and S49 (5.97) and high for the mutants S54 (9.91), S48 (9.59), S61 (9.13), and S38 (9.11) and the original CSP (9.11). The translational initiation rates for the −32/+99 mRNA region, predicted by the RBS calculator (39), and the minimum free energy scores of the mRNA secondary structures determined by the RNAfold web server (40) emphasize that our collection of signal sequences varies considerably. In general, there seems to be no apparent correlation between the predicted translation rates and the minimum free energy scores of these signal sequences, and these data should imply

that rational methods would not be a feasible strategy to generate improved CSP mutants.

CSP mutants can also be used for efficient expression and translocation of the alternative model protein AP. It was of interest to investigate whether the CSP mutants could also be useful for efficient expression and translocation of alternative heterologous proteins in *E. coli*. As a first model in these studies, we used alkaline phosphatase (AP), encoded by the *phoA* gene, partly because AP must be translocated to the periplasm to be active (48, 49). The pCSP1phoA plasmid was constructed containing the *phoA* gene under the control of *Pm*-XylS, and with the *phoA* native signal sequence replaced by CSP. As a control, we constructed pASP1phoA in which *phoA* is expressed with its native secretion signal sequence. *E. coli* DH5α (pCSP1phoA) and *E. coli* DH5α (pASP1phoA) cells were grown in shake flask cultures, harvested, and assayed for AP activity, and the results clearly showed that CSP could substitute for the native *phoA* signal sequence by leading to a similar production level of active AP protein (Fig. 4). Next, the CSP region of plasmid pCSP1phoA was replaced by a subset of the 12 CSP variants, S2, S38, S48, S60, and S61, yielding plasmids pCSP1S2phoA, pCSP1S38phoA, pCSP1S48phoA, pCSP1S60phoA, and pCSP1S61phoA, respectively. All plasmids were established in *E. coli* DH5α, and the resulting recombinant strains were subjected to AP production analyses as described above. The mutants S2, S38, and S48 displayed 5- to 8-fold-increased AP activities compared to CSP, which were in good agreement with the analo-

TABLE 2 Bioinformatics analysis of selected CSP random mutants: predicted signal sequence properties and translational attributes

Mutant	Amino acid sequence[a]	D score[b]	Cleavage site[b]	Net charge of N domain[c]	Net hydrophobicity of H domain[d]	Relative translation rate prediction (AU)[e]	Minimum free energy (kcal mol^{-1})[f]
CSP	mkkkllalallallfngaqam	0.875	AQA-MA	+3	9.11	1.00	−29.84
S2	.yrnk.v...c...l..p...	0.804	AMA-HP	+2	8.67	0.53	−31.40
S7	.qe..cq...i.......e..	0.520	AMA-HP*	0	7.19	2.57	−28.80
S15	.rs...s.gv.li.a..gk..	0.682	AMA-HP	+2	8.38	1.09	−28.60
S38	..r..................	0.877	AQA-MA	+3	9.11	1.00	−29.84
S43	..s..f..em.v.........	0.678	AMA-HP	+2	7.92	0.99	−24.24
S46	..rs..s.....e........	0.671	AMA-HP	+2	6.95	0.14	−26.50
S47	..rtf.d.....s..s.....	0.689	AMA-HP	+2	6.95	0.58	−27.20
S48	..r...v.....v....eh..	0.725	AMA-HP	+3	9.59	0.33	−27.94
S49	..r....i...rvsly.....	0.707	AMA-HP	+3	5.97	0.33	−28.64
S54	..r.ff.vt.i.gfgi.....	0.615	AMA-HP	+3	9.91	1.64	−25.80
S60	.qq..ssr...pv.lys....	0.666	AMA-HP	+1	4.18	0.87	−26.10
S61	.......v.............	0.884	AQA-MA	+3	9.13	1.00	−29.74

[a] Periods (.) indicate that the amino acid is identical to the one in CSP at this position.

[b] Calculated by SignalP 4.0 (38). S7 did not meet the cutoff criteria. *, the reported cleavage site gave the best score.

[c] Calculated for the first five residues, with amino acids D and E defined as −1, R and K defined as +1 and any other amino acid as 0.

[d] Calculated as the sum of hydrophobicity for the hydrophobic region (residues 6 to 17), based on the normalized consensus hydrophobicity scale (47).

[e] Calculated by the RBS calculator v1.1, reverse engineering mode (39). AU, arbitrary unit(s).

[f] Minimum free energy predicted by RNAfold (40) for the optimal secondary structure of the region surrounding the translation initiation site. The analyzed sequence included 32 bp of the nontranslated region, the signal sequence, and 36 bp of the mature *bla* coding sequence.

gous data obtained when β-lactamase was used as the reporter protein (see Fig. 3). The CSP mutants S60 and S61 generated AP activities in the same range as when the original CSP was used (Fig. 4). For S61, this correlated well with the analogous β-lactamase production results (Fig. 3), while it was not the case for S60, which caused an ~5-fold-increased β-lactamase production level compared to CSP. This may indicate some context dependency between the mutant CSP signal sequence and the recombinant gene with respect to efficient expression and concomitant product translocation, in agreement with previous results (17). Mutant
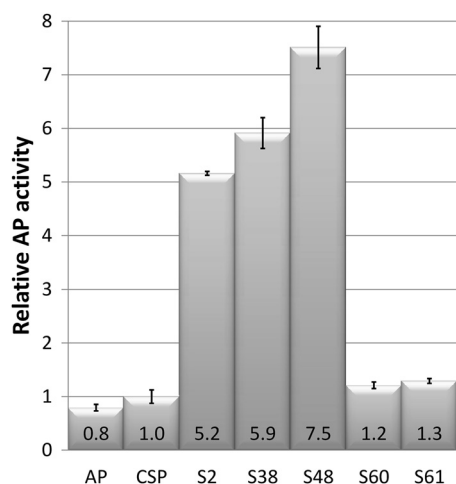

FIG 4 Relative AP activity of *E. coli* DH5α strains harboring relevant expression constructs. All values are relative to the CSP result, which was arbitrarily set to 1. AP, endogenous AP signal sequence (pASP1phoA construct); CSP, original sequence (pCSP1phoA construct), while S2, S38, S48, S60, and S61 indicate the variant CSP sequences in pCSP1S2phoA, pCSP1S38phoA, pCSP1S48phoA, pCSP1S60phoA, and pCSP1S61phoA, respectively. Cultures were induced by 0.1 mM *m*-toluic acid. All results are the averages of two biological replicas with at least three samples, and the standard errors are indicated.

S38 contains only one amino acid substitution (Lys to Arg) compared to the original CSP (Fig. 2), and this particular substitution is also present in CSP mutants S2 and S48. CSP mutants S60 and S61 do not harbor this substitution, which presumably has a positive effect on the production of active β-lactamase and AP proteins under the conditions tested. Summarized, these results demonstrated that CSP mutants can also be highly useful for the effective production of alternative translocated proteins *in E. coli*.

CSP mutants S2 and S61 can be used for efficient recombinant production of the human cytokine IFN-α2b under high-cell-density cultivation conditions. We previously demonstrated that recombinant production under HCDC of the human cytokine α2b (IFN-α2b), encoded by a synthetic and codon-optimized gene (*ifn-α2b$_S$*), was dramatically stimulated when fused to the pelB signal sequence, yielding ~0.6 g liter^{-1} (17). Most of the product was insoluble protein, and in the absence of a pelB signal, no detectable IFN-α2b protein was produced. It was here of interest to test whether CSP mutants could be useful for the efficient production of IFN-α2b under HCDC, and both the original CSP and its mutants S2 and S61 were used to substitute for pelB in plasmid pIFN30SpelB (17), yielding plasmids pCSP2ifn (Fig. 1B), pCSP2S2ifn, and pCSP2S61ifn, respectively. S2 was chosen based on the general good production results obtained with both Bla and PhoA (see above), while S61 was chosen based on the assumption that it might specifically confer efficient protein translocation (see above). The plasmids were transferred into the production host *E. coli* RV308, and the resulting recombinant strains were analyzed for IFN-α2b production under HCDC by Western blotting. Initial analyses of crude extracts showed that the IFN-α2b protein was mainly detected in the insoluble fraction (data not shown) in agreement with previous observations (34), and total cell extracts were therefore used here to compare the total production levels in the different strains. Based on visual inspection of the Western blot (Fig. 5), it appeared that CSP gave production levels approximately similar to those with pelB, while CSP mutants S2 and S61 caused at least 2-fold-higher IFN-α2b production levels. We no-
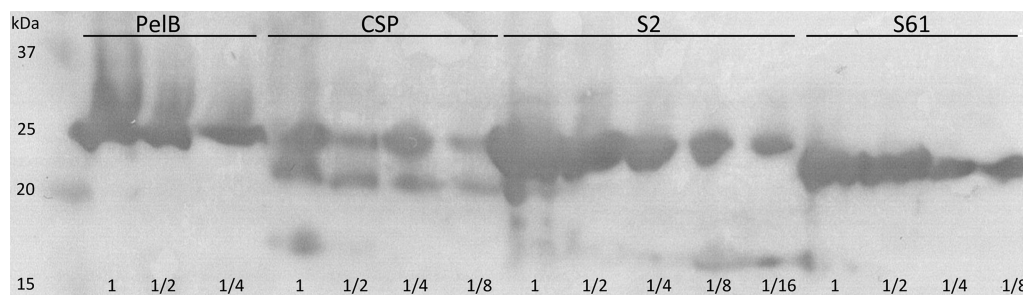
FIG 5 Western blot analysis of *E. coli* RV308 strains (pIFN30SpelB, pCSP2ifn, pCSP2S2ifn, or pCSP2S61ifn) producing IFN-α2b fusion proteins with different secretion signal peptides in high-cell-density cultivations. Cell samples were withdrawn from each culture 6.3 h after induction. The cells were lysed by sonication, and the total crude cell extracts were subjected to analysis. Dilution series of decreasing total protein amount were loaded for each sample and are shown relative to the highest total protein concentration loaded (set arbitrarily to 1). Signal peptide–IFN-α2b–cMyc–His$_6$ is 24.4 kDa, while the protein complex in which a signal peptide has been cleaved off is 22.4 kDa.

ticed that the protein products appeared in two slightly different sizes among the strains, presumably reflecting the cleavage status of the different signal peptides. The proteins expressed with pelB and mutant S2 seemed dominantly to yield an uncleaved version of the protein, while the original CSP gave a mixture of cleaved and uncleaved product. When mutant S61, on the other hand, was used, the protein product was predominantly cleaved off, and the biological reasons for these discrepancies remained unknown. These results indicated that IFN-α2b formed inclusion bodies also when being translocated to the periplasm. In any case, since all IFN-α2b protein produced by any recombinant strain tested here remained as insoluble protein, we chose to not investigate this any further. Summarized, these results show that CSP mutants have a potential for increasing the expression of difficult-to-express proteins, such as IFN-α2b, which even in a codon-optimized variant seems to be highly dependent on a signal sequence to be recombinantly expressed in *E. coli* (17).

Concluding remarks. We presented here an approach enabling the effective selection of improved signal sequences from large randomized mutant libraries by directly scoring for increased antibiotic tolerance levels of the *E. coli* host cells. Our previously designed signal sequence CSP was used as a model, and we show that this method enables selection of CSP mutants causing increased expression level and improved translocation of presumably any recombinant protein in *E. coli*. Our results support and extend previous reports demonstrating that signal sequences can be highly useful in promoting high-level recombinant production of heterologous proteins that are generally difficult to express in *E. coli*. Bioinformatics analyses of the CSP mutants indicated that they could not have been predicted by rational methods, further emphasizing the advantage of this combinatorial mutagenesis and selection approach for the generation of superior signal sequences useful for recombinant expression in bacteria.

REFERENCES

1. **Choi JH, Lee SY.** 2004. Secretory and extracellular production of recombinant proteins using *Escherichia coli*. Appl. Microbiol. Biotechnol. **64:** 625–635.
2. **Esposito D, Chatterjee DK.** 2006. Enhancement of soluble protein expression through the use of fusion tags. Curr. Opin. Biotechnol. **17:**353–358.
3. **Jonet MA, Mahadi NM, Murad AMA, Rabu A, Bakar FDA, Rahim RA, Low KO, Illias RM.** 2012. Optimization of a heterologous signal peptide by site-directed mutagenesis for improved secretion of recombinant proteins in *Escherichia coli*. J. Mol. Microbiol. Biotechnol. **22:**48–58.
4. **Mergulhão FJM, Summers DK, Monteiro GA.** 2005. Recombinant protein secretion in *Escherichia coli*. Biotechnol. Adv. **23:**177–202.
5. **Makrides SC.** 1996. Strategies for achieving high-level expression of genes in *Escherichia coli*. Microbiol. Rev. **60:**512–538.
6. **Baneyx F, Mujacic M.** 2004. Recombinant protein folding and misfolding in *Escherichia coli*. Nat. Biotechnol. **22:**1399–1408.
7. **González-Montalbán N, García-Fruitós E, Villaverde A.** 2007. Recombinant protein solubility: does more mean better? Nat. Biotechnol. **25:** 718–720.
8. **Arié JP, Miot M, Sassoon N, Betton JM.** 2006. Formation of active inclusion bodies in the periplasm of *Escherichia coli*. Mol. Microbiol. **62:** 427–437.
9. **García-Fruitós E, González-Montalbán N, Morell M, Vera A, Ferraz RM, Arís A, Ventura S, Villaverde A.** 2005. Aggregation as bacterial inclusion bodies does not imply inactivation of enzymes and fluorescent proteins. Microb. Cell Fact. **4:**27.
10. **Stampolidis P, Kaderbhai NN, Kaderbhai MA.** 2009. Periplasmically exported lupanine hydroxylase undergoes transition from soluble to functional inclusion bodies in *Escherichia coli*. Arch. Biochem. Biophys. **484:** 8–15.
11. **Ventura S, Villaverde A.** 2006. Protein quality in bacterial inclusion bodies. Trends Biotechnol. **24:**179–185.
12. **Dalbey RE, Wang P, van Dijl JM.** 2012. Membrane proteases in the bacterial protein secretion and quality control pathway. Microbiol. Mol. Biol. Rev. **76:**311–330.
13. **Natale P, Brüser T, Driessen AJ.** 2008. Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane: distinct translocases and mechanisms. Biochim. Biophys. Acta **1778:**1735–1756.
14. **Rusch SL, Kendall DA.** 2007. Interactions that drive Sec-dependent bacterial protein transport. Biochemistry **46:**9665–9673.
15. **Brockmeier U, Caspers M, Freudl R, Jockwer A, Noll T, Eggert T.** 2006. Systematic screening of all signal peptides from *Bacillus subtilis*: a powerful strategy in optimizing heterologous protein secretion in Gram-positive bacteria. J. Mol. Biol. **362:**393–402.
16. **Caspers M, Brockmeier U, Degering C, Eggert T, Freudl R.** 2010. Improvement of Sec-dependent secretion of a heterologous model protein in *Bacillus subtilis* by saturation mutagenesis of the N-domain of the AmyE signal peptide. Appl. Microbiol. Biotechnol. **86:**1877–1885.
17. **Sletta H, Tøndervik A, Hakvåg S, Aune TEV, Nedal A, Aune R, Evensen G, Valla S, Ellingsen TE, Brautaset T.** 2007. The presence of N-terminal secretion signal sequences leads to strong stimulation of the total expression levels of three tested medically important proteins during high-cell-density cultivations of *Escherichia coli*. Appl. Environ. Microbiol. **73:**906–912.
18. **Auclair SM, Bhanu MK, Kendall DA.** 2012. Signal peptidase I: cleaving the way to mature proteins. Protein Sci. **21:**13–25.

19. **Power PM, Jones RA, Beacham IR, Bucholtz C, Jennings MP.** 2004. Whole genome analysis reveals a high incidence of non-optimal codons in secretory signal sequences of *Escherichia coli.* Biochem. Biophys. Res. Commun. **322**:7.

20. **Zalucki YM, Power PM, Jennings MP.** 2007. Selection for efficient translation initiation biases codon usage at second amino acid position in secretory proteins. Nucleic Acids Res. **35**:5748–5754.

21. **Zalucki YM, Gittins KL, Jennings MP.** 2008. Secretory signal sequence non-optimal codons are required for expression and export of β-lactamase. Biochem. Biophys. Res. Commun. **366**:135–141.

22. **Zalucki YM, Jones CE, Ng PSK, Schulz BL, Jennings MP.** 2010. Signal sequence non-optimal codons are required for the correct folding of mature maltose binding protein. Biochim. Biophys. Acta **1798**:1244–1249.

23. **Tuller T, Carmi A, Vestsigian K, Navon S, Dorfan Y, Zaborske J, Pan T, Dahan O, Furman I, Pilpel Y.** 2010. An evolutionarily conserved mechanism for controlling the efficiency of protein translation. Cell **141**:344–354.

24. **Sletta H, Nedal A, Aune TEV, Hellebust H, Hakvåg S, Aune R, Ellingsen TE, Valla S, Brautaset T.** 2004. Broad-host-range plasmid pJB658 can be used for industrial-level production of a secreted host-toxic single-chain antibody fragment in *Escherichia coli.* Appl. Environ. Microbiol. **70**:7033–7039.

25. **Hibbert E, Dalby P.** 2005. Directed evolution strategies for improved enzymatic performance. Microb. Cell Fact. **4**:29.

26. **Yuan L, Kurek I, English J, Keenan R.** 2005. Laboratory-directed protein evolution. Microbiol. Mol. Biol. Rev. **69**:373–392.

27. **Pritchard L, Corne D, Kell D, Rowland J, Winson M.** 2005. A general model of error-prone PCR. J. Theor. Biol. **234**:497–509.

28. **Stemmer WPC.** 1994. Rapid evolution of a protein in vitro by DNA shuffling. Nature **370**:389–391.

29. **Encell LP, Coates MM, Loeb LA.** 1998. Engineering human DNA alkyltransferases for gene therapy using random sequence mutagenesis. Cancer Res. **58**:1013–1020.

30. **Lai Y-P, Huang J, Wang L-F, Li J, Wu Z-R.** 2004. A new approach to random mutagenesis in vitro. Biotechnol. Bioeng. **86**:622–627.

31. **Aune TEV, Bakke I, Drabløs F, Lale R, Brautaset T, Valla S.** 2010. Directed evolution of the transcription factor XylS for development of improved expression systems. Microb. Biotechnol. **3**:38–47.

32. **Bakke I, Berg L, Aune TE, Brautaset T, Sletta H, Tøndervik A, Valla S.** 2009. Random mutagenesis of the PM promoter as a powerful strategy for improvement of recombinant-gene expression. Appl. Environ. Microbiol. **75**:2002–2011.

33. **Berg L, Lale R, Bakke I, Burroughs N, Valla S.** 2009. The expression of recombinant genes in *Escherichia coli* can be strongly stimulated at the transcript production level by mutating the DNA-region corresponding to the 5′-untranslated part of mRNA. Microb. Biotechnol. **2**:379–389.

34. **Sambrook J, Russell DW.** 2001. Molecular cloning: a laboratory manual, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

35. **Mandecki W, Shallcross MA, Sowadski J, Tomazic-Allen S.** 1991. Mu-tagenesis of conserved residues within the active site of *Escherichia coli* alkaline phosphatase yields enzymes with increased kcat. Protein Eng. **4**:801–804.

36. **Kikuchi Y, Yoda K, Yamasaki M, Tamura G.** 1981. The nucleotide sequence of the promoter and the amino-terminal region of alkaline phosphatase structural gene (phoA) of *Escherichia coli.* Nucleic Acids Res. **9**:8.

37. **Winther-Larsen HC, Blatny JM, Valand B, Brautaset T, Valla S.** 2000. Pm promoter expression mutants and their use in broad-host-range RK2 plasmid vectors. Metab. Eng. **2**:92–103.

38. **Petersen TN, Brunak S, von Heijne G, Nielsen H.** 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods **8**:785–786.

39. **Salis HM, Mirsky EA, Voigt CA.** 2009. Automated design of synthetic ribosome binding sites to control protein expression. Nat. Biotechnol. **27**:946–950.

40. **Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL.** 2008. The Vienna RNA Websuite. Nucleic Acids Res. **36**:W70–W74.

41. **Horn U, Strittmatter W, Krebber A, Knüpfer U, Kujau M, Wenderoth R, Müller K, Matzku S, Plückthun A, Riesenberg D.** 1996. High volumetric yields of functional dimeric miniantibodies in *Escherichia coli,* using an optimized expression vector and high-cell-density fermentation under non-limited growth conditions. Appl. Microbiol. Biotechnol. **46**:524–532.

42. **Kadonaga JT, Gautier AE, Straus DR, Charles AD, Edge MD, Knowles JR.** 1984. The role of the beta-lactamase signal sequence in the secretion of proteins by *Escherichia coli.* J. Biol. Chem. **259**:2149–2154.

43. **Smith H, Bron S, van Ee J, Venema G.** 1987. Construction and use of signal sequence selection vectors in *Escherichia coli* and *Bacillus subtilis.* J. Bacteriol. **169**:3321–3328.

44. **Smith H, de Jong A, Bron S, Venema G.** 1988. Characterization of signal-sequence-coding regions selected from the *Bacillus subtilis* chromosome. Gene **70**:351–361.

45. **Kudla G, Murray AW, Tollervey D, Plotkin JB.** 2009. Coding-sequence determinants of gene expression in *Escherichia coli.* Science **324**:255–258.

46. **van Dijl JM, de Jong A, Vehmaanpera J, Venema G, Bron S.** 1992. Signal peptidase I of *Bacillus subtilis*: patterns of conserved amino acids in prokaryotic and eukaryotic type I signal peptidases. EMBO J. **11**:2819–2828.

47. **Eisenberg D, Schwarz E, Komaromy M, Wall R.** 1984. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. J. Mol. Biol. **179**:125–142.

48. **Hoffman CS, Wright A.** 1985. Fusions of secreted proteins to alkaline phosphatase: an approach for studying protein secretion. Proc. Natl. Acad. Sci. U. S. A. **82**:5107–5111.

49. **Manoil C, Beckwith J.** 1985. TnphoA: a transposon probe for protein export signals. Proc. Natl. Acad. Sci. U. S. A. **82**:8129–8133.

50. **Blatny JM, Brautaset T, Winther-Larsen HC, Karunakaran P, Valla S.** 1997. Improved broad-host-range RK2 vectors useful for high and low regulated gene expression levels in gram-negative bacteria. Plasmid **38**:35–51.