Supporting Information

Xia et al. 10.1073/pnas.1003366107

SI Text

SI Materials and Methods. ***Plasmids Construction.*** DNA manipulations were performed according to standard protocols. The constructs containing the cloned fragments were confirmed by automatic DNA sequencing (ABI Prism Model 377, Perkin Elmer Applied Biosystems). *Escherichia coli* TOP10 (Invitrogen), was used for gene cloning studies.

The *glyVXY* genes were amplified from the genomic DNA of *E. coli* W3110 using the primers FglyEco (5′-GCTCGATATCTA ACGACGCAGAAATGCGAAA-3′) and RglyBam (5′-CATTG-GATCCTAAGATTACAGCCTGAGGCTGTG-3′). The amplified DNA, after digestion with *Eco*RV and *Bam*HI, was cloned into plasmid pACYC184 (New England Biolabs) to make pTetglyVXY.

The *tet-glyVXY* cassette was amplified from pTetglyVXY with the primers FglySph (5′-GGCTCGCATGCTCATGTTTGACA GCTTATCATCGA-3′) and RglySaPs (5′-ATTGTCGACTGCTG CAGTAAGATTACAGCCTGAGGCTGTG-3′). The resulting PCR product was cut with *Sph*I and *Sal*I and cloned into pTetglyVXY at the same sites, thus making pTetgly2 harboring two copies of the *glyVXY* genes.

The *glyA* fragment was amplified from −267 bp upstream of the start codon to 209 bp downstream of the stop codon, using primers FglyAn (5′-CCCTGCAATGTAAATGGTTCTTTGG-3′) and RglyASal (5′-AAGTCGACAGATTTGATGGCGCGATA-3′) and W3110 genomic DNA as a template. After digestion with *Sal*I, the amplified DNA fragment was ligated with the 3.2 kb *Sal*I-*Hinc*II fragment of pACYC184 to make p184glyAn.

To overexpress the *glyVXY* and *glyA* genes simultaneously, plasmids pTetgly-glyAn and pTetgly2-glyAn were constructed as follows. The amplified *glyA* fragment, after digestion with *Sal*I as in the construction of p184glyAn, was ligated with the 3.5 kb *Sal*I-*Hinc*II fragment of pTetglyVXY to generate pTetgly-glyAn, or with the 4.0 kb *Sal*I-*Hinc*II fragment of pTetgly2 to generate pTetgly2-glyAn.

Plasmid pSH16a allows the expression of the 16 repeats of a spider dragline silk protein monomer under the control of T7 promoter (1). To express 32 repeats of the protein monomer, plasmid pSH32 was constructed by ligating the 1.7 kb *Spe*I-*Nhe*I fragment of pSH16a to *Spe*I-digested, dephosphorylated pSH16a. Likewise, plasmids pSH48 and pSH64 were created by cloning into the *Spe*I site of pSH32 the 1.7 kb *Spe*I-*Nhe*I fragment of pSH16a, or the 3.4 kb *Spe*I-*Nhe*I fragment of pSH32, respectively. Furthermore, plasmids pSH80 and pSH96 were generated by cloning into the *Spe*I site of pSH64 the *Spe*I-*Nhe*I fragment of pSH16a or pSH32, respectively. Plasmid pSH128 was generated by inserting the 3.4 kb *Spe*I-*Nhe*I fragment of pSH32 into the *Spe*I site of pSH96. The orientation of each insertion was confirmed by double digestion with *Spe*I and *Nhe*I.

Inactivation of Glycine Cleavage System. Glycine cleavage system catalyzes the oxidation of glycine and generates a C1 moiety. The system is composed four subunits: GcvT, GcvH, GvcP (encoded in the *gcvTHP* operon), and lipoamide dehydrogenase (E3), which is shared with pyruvate dehydrogenase and 2-oxoglutarate dehydrogenase. Inactivation of the system was achieved by deleting the chromosomal *gcvTHP* genes. Knocking out the *gcvTHP* genes was conducted by double crossover homologous recombination (2). pECmulox, which contains the lox71-chloramphenicol resistance gene (Cm^R)-lox66 cassette (3), was used as a template for PCR reaction. The primers FgcvKO (5′-TGCCGGAAGGTATTGGCGAAACGGCGATTGTG-CAAATTCGCAACCGTGAAATGCCGACACTATAGAACGC GGCCG-3′) and RgcvKO (5′- AACGGCATTCAGCATTTCT TGCTGTTGCGCGGCGTCCGGTCCGATATGGCGTT-CACCGCATAGGCCACTAGTGGA-3′) were used to amplify the lox71-Cm^R-lox66 cassette. The amplified PCR products were electroporated into *E. coli* BL21(DE3) cells harboring λ-Red recombinase expression plasmid pKD46 (2). Colonies were selected on LB agar plates containing 35 μg/mL of Cm, and successful replacement with the lox71-Cm^R-lox66 cassette was confirmed by PCR with the primers Cupgcv (5′-CAGCAT-GAAGGCATTATCACC-3′) and Cdogcv (5′-CAATAGCCTTG AGTTCTGCCA-3′). The Cm^R marker was subsequently eliminated by a helper plasmid pJW168 (4), which contains a temperature-sensitive replication origin and the IPTG-inducible *cre* recombinase. The loss of Cm^R was further verified by PCR.

Plasmid Copy Number Analysis. Plasmid copy number (PCN) was estimated as previously described with modifications (5). Briefly, plasmid DNA was extracted from each cell sample ($\sim 6 \times 10^9$ cells), digested with *Eco*RI and *Eco*RV, and separated on 0.8% agarose gels, using the 1 Kb DNA Ladder (Bioneer) as both DNA Mw marker and an internal standard. The 5.4 kb *Eco*RI-*Eco*RV fragment of vector pET-30a(+) was evaluated densitometrically to calculate DNA concentration in each sample. The DNA concentration was divided by the number of cells in the sample and the Mw of the 5.4 kb DNA to calculate PCN.

Two-Dimensional Gel Electrophoresis, Gel Analysis, and Protein Identification. Cells were suspended and mixed with a lysis buffer (8 M urea, 2 M thiourea, 40 mM Tris, 65 mM DTT, 4% (w/v) CHAPS). Soluble proteins (150 μg) were diluted into 340 μL of a rehydration buffer (8 M urea, 2 M thiourea, 20 mM DTT, 2% (w/v) CHAPS, 0.8% (w/v) immobilized pH gradient (IPG) buffer, and 1% (v/v) cocktail protease inhibitor) and then loaded onto Immobiline DryStrip gels (18 cm, pH 3–10 NL; Amersham Biosciences). The loaded IPG strips were rehydrated, focused and equilibrated as described previously (6). The equilibrated strips were transferred to 12% (w/v) SDS-polyacrylamide gels. The 2D image was analyzed using PDQuest 2D Analysis Software (BioRad). The MS/MS analysis of the protein spots and subsequent protein identification were performed as described previously (6).

SDS-PAGE and Protein Quantification. Protein samples were boiled, centrifuged, and loaded onto 10% SDS-polyacrylamide gels with unstained Precision Plus Protein standards (BioRad). The gels were stained with Coomassie brilliant blue R250 (BioRad) and the stained gels were scanned by the UMAX PowerLook 2100XL Scanner. The protein bands were normalized and quantified by BioRad Quantity One® Version 4.2.1 to estimate purity or expression level (% total cellular proteins). Silk production titre was calculated by multiplying expression level with dry cell weight as well as 0.55. Here, 0.55 is the ratio of total cellular proteins to dry cell weight in *E. coli* (7).

Western Blot Analysis. Samples were prepared and run on 10% SDS-polyacrylamide gels as described above. The gels were transferred to PVDF membranes (Roche) using a semidry electrophoretic transfer cell (BioRad). The membrane was blocked with 5% (w/v) nonfat milk in TBS-T (10 mM Tris, 150 mM NaCl, 0.05% Tween 20, pH 8.0), and then incubated with murine mono-

clonal antipolyHistidine, Peroxidase conjugate (Sigma Catalog Number A7058), diluted 1:4,000 for 1 h. Immunoreactive bands were detected using the ECL Western blotting detection reagents (Bionote, Inc. Korea) and medical X- ray film (AGFA).

Proteomic Analysis. The β subunit (GlyS) of glycyl-tRNA synthetase was upregulated 1.95- and 3.65-fold in the cells expressing the 32- and 96-mer silk proteins, respectively. The smaller, α subunit (GlyQ) of glycyl-tRNA synthetase could not be detected in the 2-DE gels; its nondetection has also been described in the *E. coli* protein databases such as SWISS-2DPAGE (http://kr.expasy.org/) and EcoProDB (http://eecoli.kaist.ac.kr). It is known that the α and β subunits of glycyl-tRNA synthetase are transcribed in the same reading frame from α to β (8). Taken together, the data indicated an increased demand of glycyl-tRNA upon the expression of the silk proteins, and this demand became greater in the cells expressing larger silk protein. It should be mentioned that neither of the 32-mer and 96-mer silk proteins was detected in the two-dimensional electrophoresis (2-DE) gels because the silk proteins possess isoelectric points higher than 10.5 that is beyond the separation range of the 2-DE gels (pH 3–10) (Fig. S3).

Effects of Elevating Glycine Pool. Addition of 1 to 10 mM glycine to the culture medium did not increase the cell growth or expression levels. A similar result was reported in the expression of an elastin-like polypeptide fusion protein having abundant glycine (9). This might be due to metabolic imbalance because exogenous glycine induced the glycine cleavage system (10), and inhibited *serA*-encoded 3-phosphoglycerate dehydrogenase, which catalyzed the first step of serine biosynthesis (11). The second option was not desirable because glycine cleavage system mutation caused a slight but significant decrease in the host growth. Nonetheless, inactivation of the system exhibited effects on silk expression similar with *glyA* overexpression (Fig. S5). Thus, *glyA* overexpression was chosen.

Effects of Dope Concentration on the Properties of Spun Fibers. To test the possible effects of dope protein concentration on the properties of spun fibers, the 16-, 32-, and 64-mer proteins were also spun at their respective maximum operational concentrations. When the 16-mer protein concentration was increased from 20% to its maximum (32%) for spinning, improvements in fiber properties were significant (Fig. S8). For the 32-mer protein, the tenacity and the breaking strain increased as the protein concentration increased from 20% to its maximum (27%), but with less statistical significance. For the 64-mer protein, the maximum operational concentration was 23%, which is not much different from the already tested 20%. Material properties did not change much (Fig. S8).

1. Wong Po Foo C, Bini E, Huang J, Lee SY, Kaplan DL (2006) Solution behavior of synthetic silk peptides and modified recombinant silk proteins. *Appl Phys A* 82:193–203.
2. Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci USA* 97:6640–6645.
3. Kim JM, Lee KH, Lee SY (2008) Development of a markerless gene knock-out system for *Mannheimia succiniciproducens* using a temperature-sensitive plasmid. *FEMS Microbiol Lett* 278:78–85.
4. Palmeros B, et al. (2000) A family of removable cassettes designed to obtain antibiotic-resistance-free genomic modifications of *Escherichia coli* and other bacteria. *Gene* 247:255–264.
5. O'Connell HA, Niu C, Gilbert ES (2007) Enhanced high copy number plasmid maintenance and heterologous protein production in an *Escherichia coli* biofilm. *Biotechnol Bioeng* 97:439–446.
6. Xia XX, Han MJ, Lee SY, Yoo JS (2008) Comparison of the extracellular proteomes of *Escherichia coli* B and K-12 strains during high cell density cultivation. *Proteomics* 8:2089–2103.
7. Neidhardt FC (1987) *Escherichia coli* and *Salmonella typhimurium*: Cellular and molecular biology, eds Neidhardt FC, et al. (American Society for Microbiology, Washington, DC), pp 3–6.
8. Keng T, Webster TA, Sauer RT, Schimmel P (1982) Gene for *Escherichia coli* glycyl-tRNA synthetase has tandem subunit coding regions in the same reading frame. *J Biol Chem* 257:12503–12508.
9. Chow DC, Dreher MR, Trabbic-Carlson K, Chilkoti A (2006) Ultra-high expression of a thermally responsive recombinant fusion protein in *E. coli*. *Biotechnol Prog* 22:638–646.
10. Meedel TH, Pizer LI (1974) Regulation of one-carbon biosynthesis and utilization in *Escherichia coli*.*J Bacteriol* 118:905–910.
11. Zhao G, Winkler ME (1996) A novel alpha-ketoglutarate reductase activity of the *serA*-encoded 3-phosphoglycerate dehydrogenase of *Escherichia coli* K-12 and its possible implications for human 2-hydroxyglutaric aciduria. *J Bacteriol* 178:232–239.

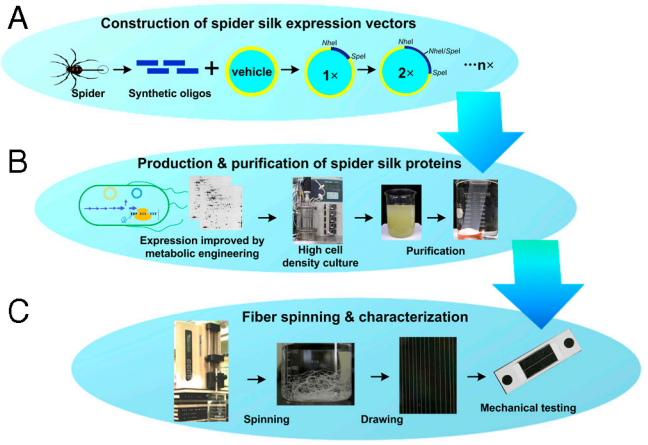

Fig. S1. A schematic of making artificial fibers. (*A*) Construction of expression plasmid for recombinant spider dragline silk proteins. According to partial cDNA sequences of the spider *N. clavipes* spidroin I, a silk monomer gene was designed and assembled from chemically synthesized oligonucleotides. Multimerization of the silk monomer is achieved through the "head-to-tail" strategy relying on two compatible but nonregenerable restriction enzyme sites, *Nhe*I and *Spe*I. (*B*) Metabolic engineering of the bacterial expression host, *Escherichia coli* allows high expression of the silk proteins. The resulting engineered strains are subjected to high cell density cultivation for high-level production of silk proteins, followed by purification. (*C*) Making fibers by spinning of recombinant silk proteins and mechanical testing.

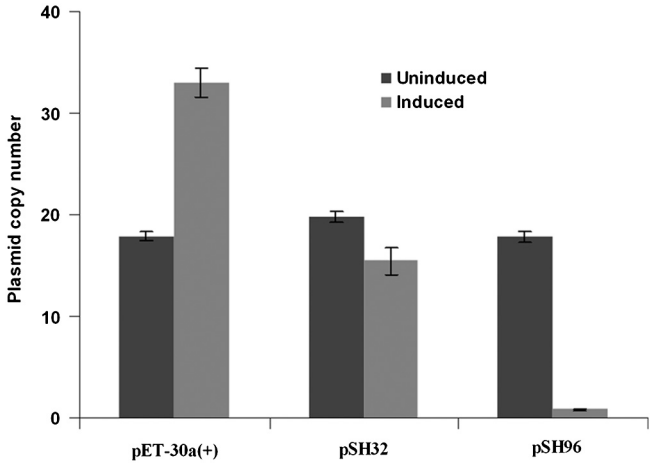

Fig. S2. PCN analysis. Uninduced and induced BL21(DE3) cells harboring pET-30a(+) (5.4 kb, empty vector), pSH32 (8.8 kb, 32-mer silk protein) or pSH96 (15.5 kb, 96-mer silk protein) were subjected to PCN analysis as described in online methods. The PCN becomes lower for the plasmid expressing a larger silk protein upon induction, probably due to the burden to the cell. However, the difference of PCN was not due to the plasmid size as the three plasmids were maintained at comparable copies under uninduced condition. All error bars represent s.d. ($n = 3$).

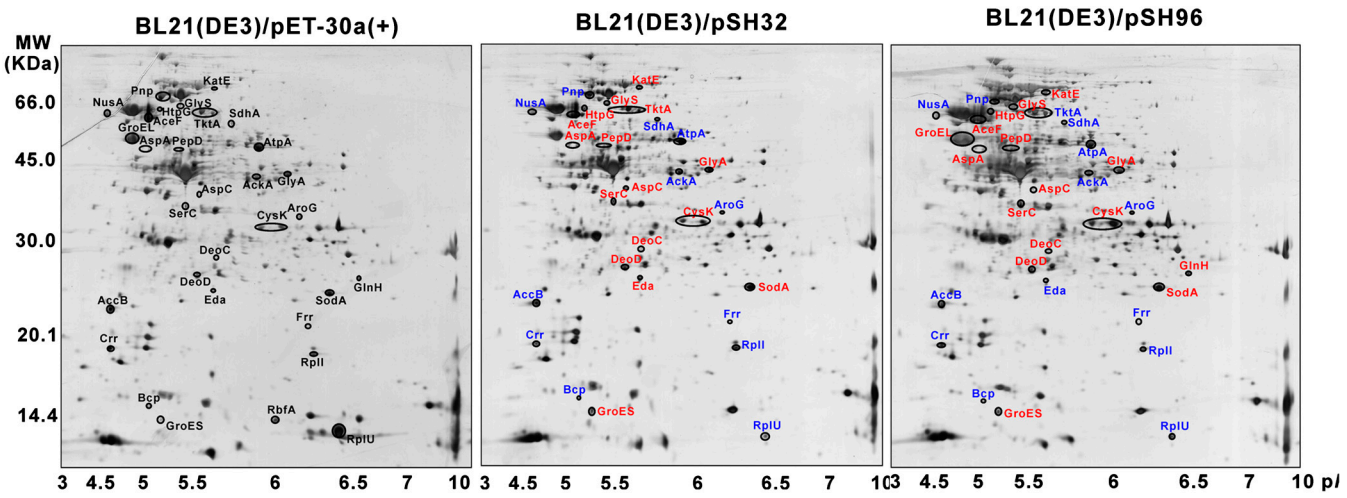

Fig. S3. The proteomes of *E. coli* BL21(DE3) cells harboring plasmid pET-30a(+) (control), pSH32 (encoding 32-mer silk protein) or pSH96 (encoding 96-mer silk protein). Proteins were prepared and separated by 2-DE in a pH range of 3–10. The upregulated proteins (labeled in red) and downregulated proteins (labeled in blue) are indicated. The detailed comparison results are shown in Table S1.

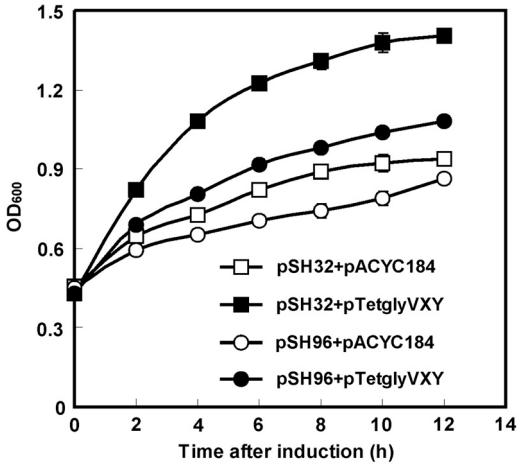

Fig. S4. Elevating $tRNA^{Gly}$ pool (in the presence of plasmid pTetglyVXY) increases the cell yields. Used as controls were the BL21(DE3) cells cotransformed with pACYC184 and each of the silk protein expression plasmids. Again, the 32- and 96-mer silk proteins were shown as typical examples. In both instances, a higher cell yield was observed compared to the respective control. A comparable cell growth-enhancing effect was observed for the remaining silk proteins. It should be mentioned that higher $tRNA^{Gly}$ pool (in the presence of pTetgly2) did not further increase the cell growth following induction, and that elevating the $tRNA^{Gly}$ pool did not affect the cell growth under uninduced condition. All error bars represent s.d. ($n = 3$).

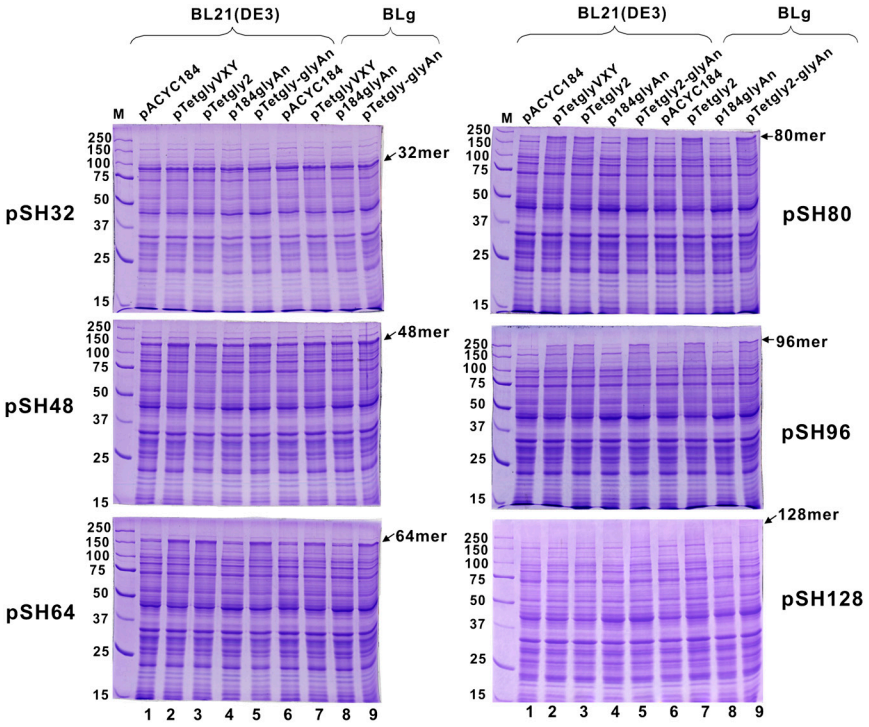

Fig. S5. Expression of the silk proteins in *E. coli* BL21(DE3) and its glycine cleavage system-deficient mutant (BLg) upon overexpression of $tRNA^{Gly}$ *glyVXY* genes, serine hydroxymethyltransferase (*glyA*) or in combination. Cells were cotransformed with a silk expression plasmid and one of the following plasmids: pACYC184, control; pTetglyVXY, *glyVXY* overexpression; pTetgly2, higher *glyVXY* overexpression; p184glyAn, *glyA* overexpression; pTetgly-glyAn, *glyA* and *glyVXY* overexpression; pTetgly2-glyAn, *glyA* and higher *glyVXY* overexpression. Proteins from cell samples (0.024 mg) were separated on 10% SDS-PAGE gels. Note that the first six lanes are the same as those in Fig. 2.

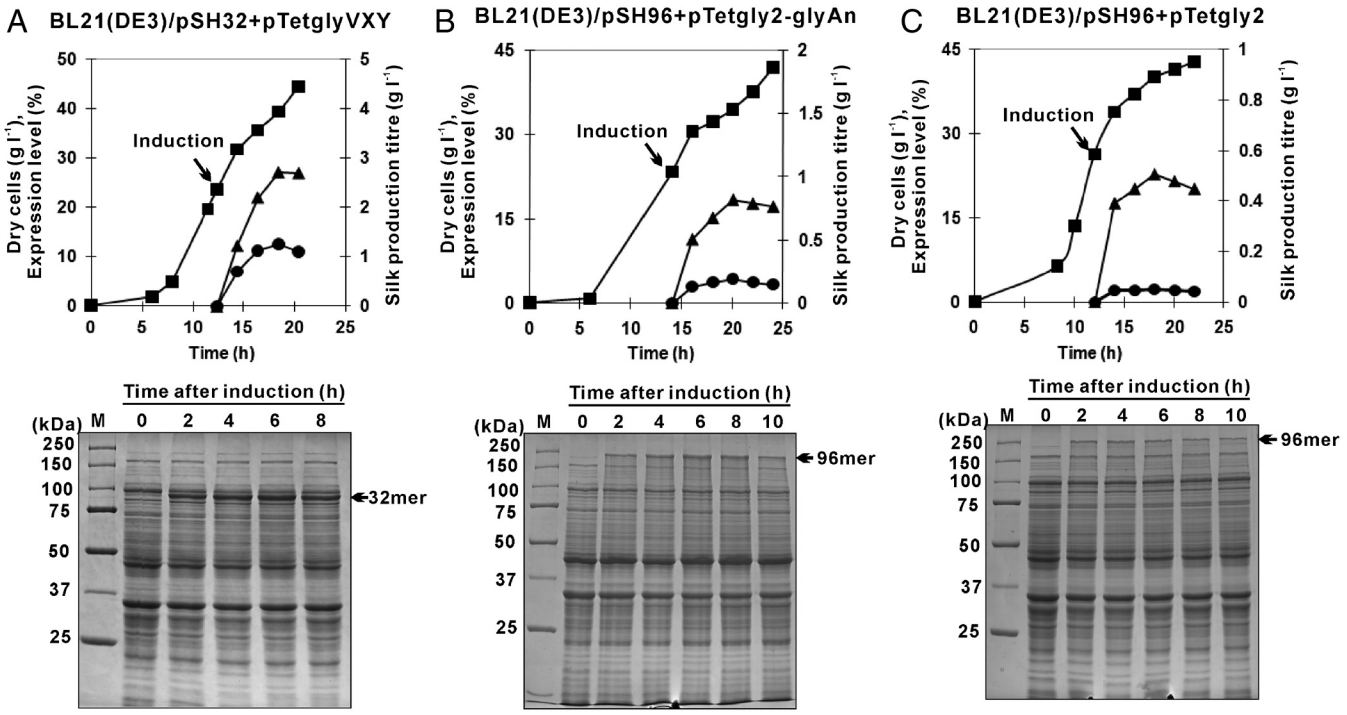

Fig. S6. Production of recombinant dragline silk proteins by high cell density cultivation (HCDC). (*A*) BL21(DE3) carrying pSH32 and pTetglyVXY. (*B*) BL21(DE3) carrying pSH96 and pTetgly2-glyAn. (*C*) BL21(DE3) carrying pSH96 and pTetgly2. Cells were cultivated in a 6.6 L jar fermentor and induced with 1 mM IPTG at indicated time. Time profiles of dry cells (■), expression level (•), and silk protein titre (▲) are shown (*Upper*). Proteins from cell samples (0.024 mg) were separated on 10% SDS-PAGE gels (*Lower*). It should be mentioned that Coomassie stains much less for the glycine-rich proteins (e.g., spider silk proteins), and thus the concentrations reported here are the minimum ones.

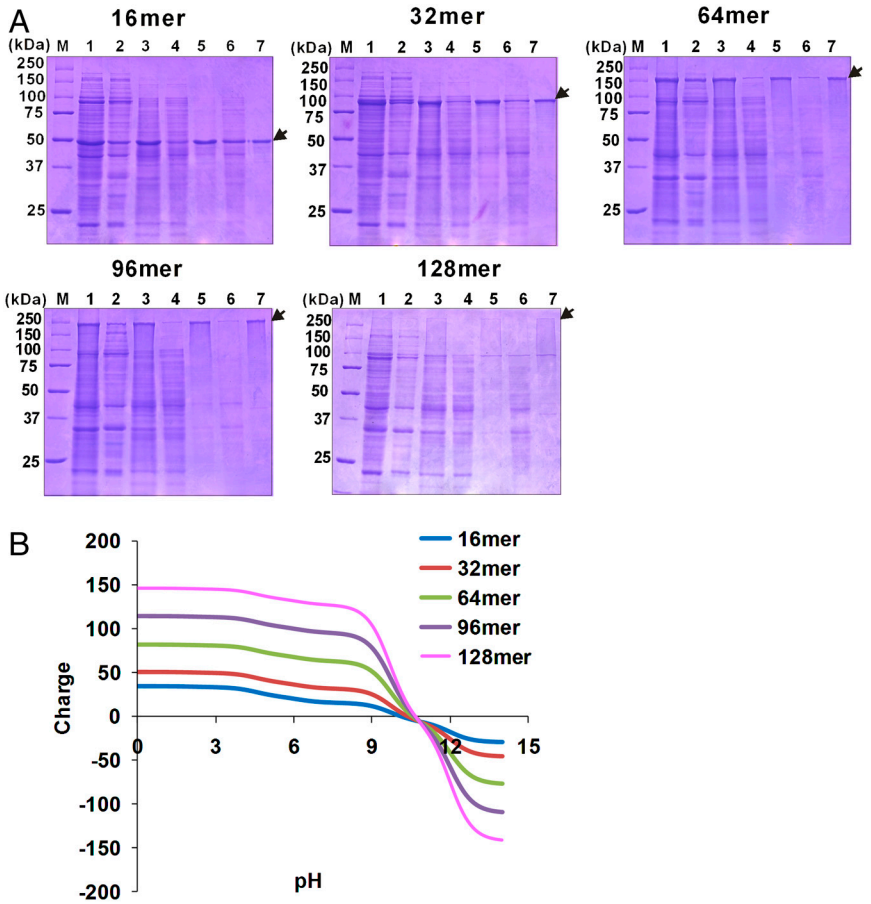

Fig. S7. Purification of silk proteins. (*A*) Coomassie-stained 10% SDS-PAGE gel analysis of protein samples in the purification process as shown in Fig. 3*A*. Lane 1, cell lysates; Lane 2, pellets from lysed cells at pH 4.0; Lane 3, cleared supernatant from lysed cells at pH 4.0; Lane 4, precipitates after 1.32 M $(NH_4)_2SO_4$ treatment; Lane 5, precipitates after 2.8 M $(NH_4)_2SO_4$ treatment; Lane 6, precipitates after dialysis; Lane 7, cleared supernatant after dialysis. (*B*) Protein charge analysis of the recombinant silk proteins using CLC Main Workbench 5.0 (CLC bio).

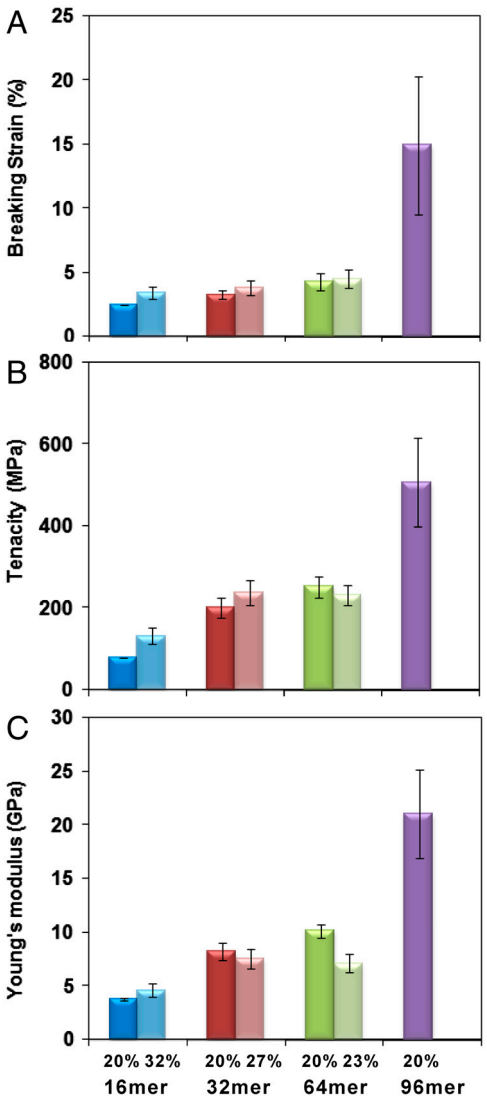

Fig. S8. Effects of dope protein concentration on the mechanical properties of the spun fibers. (*A*) Breaking strain. (*B*) Tenacity. (*C*) Young's modulus. The silk proteins were spun at 20% or the respective maximum concentrations as indicated. Notably, the maximum achievable level for the native-sized 96-mer protein was 20%. All error bars represent s.d. ($n = 10$).

Table S1. Differentially expressed proteins revealed by two-dimentional electrophoresis gel analysis of *E. coli* BL21(DE3) cells carrying empty vector (S1) and those expressing the 32- (S2) or 96-mer silk protein (S3)

Protein name	Description	p*I*	*Mw* (kDa)	Accession no.†	Fold change*		Type of analysis‡
					S2/S1	S3/S1	
Central carbon metabolism							
Tricarboxylic acid cycle							
SdhA	Succinate dehydrogenase flavoprotein subunit	5.85	64.4	P0AC41	0.70	0.52	GM
Pyruvate dehydrogenase							
AceF	Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex	5.09	66.1	P06959	1.31	1.75	MS/MS
Pentose phosphate pathway							
TktA	Transketolase 1	5.43	72.1	P27302	4.14	5.04	GM
Entner-Doudoroff pathway							
Eda	KHG/KDPG aldolase	5.57	22.3	P0A955	2.10	2.32	MS/MS
Respiration							
AckA	Acetate kinase	5.85	43.3	P0A6A3	0.86	0.65	MS/MS
ATP-proton motive force interconversion							
AtpA	ATP synthase subunit alpha	5.80	55.3	P0ABB0	0.65	0.50	GM
Building block biosynthesis and degradation							
Amino acid biosynthesis							
SerC	Phosphoserine aminotransferase	5.37	39.7	P23721	1.28	2.00	MS/MS
CysK	Cysteine synthase A	5.83	34.4	P0ABK5	1.20	1.76	MS/MS
GlyA	Serine hydroxymethyltransferase	6.03	45.3	P0A825	1.27	1.30	MS/MS
AroG	Phospho-2-dehydro-3-deoxyheptonate aldolase, Phe-sensitive	6.14	38.0	P0AB91	0.62	0.41	GM
Fatty acid biosynthesis							
AccB	Enoyl-[acyl-carrier-protein] reductase [NADH]	5.58	27.8	P0AEK4	0.88	0.73	MS/MS
Amino acid degradation							
AspA	Aspartate ammonia-lyase	5.19	52.4	P0AC38	2.26	2.49	GM
AspC	Aspartate Aminotransferase	5.54	43.5	P00509	2.30	2.87	GM
Nucleotide degradation							
DeoC	2-deoxyribose-5-phosphate aldolase	5.50	27.7	P0A6L0	1.52	1.67	GM
DeoD	Purine-nucleoside phosphorylase	5.42	25.9	P0ABP8	1.30	1.64	GM
RNA degradaton							
Pnp	Polyribonucleotide nucleotidyltransferase	5.39	80.5	Q8FD87	0.74	0.51	MS/MS
Protein degradation							
PepD	Aminoacyl-histidine dipeptidase	5.20	52.9	P15288	2.48	3.74	MS/MS
Protein translation							
GlyS	Glycyl-tRNA synthetase beta subunit	5.29	76.7	P00961	1.95	3.65	GM
RplI	50S ribosomal protein L9	6.17	15.7	P0A7R1	0.82	0.45	MS/MS
RbfA	Ribosome-binding factor A	5.93	15.2	Q1R6H1	0	0	GM
RplU	50S ribosomal protein L21	6.71	11.0	P0AG48	0.10	0.09	GM
Frr	Ribosome-recycling factor	6.44	20.6	P0A805	0.65	0.46	GM
Cell process and regulation							
Chaperones							
HtpG	Chaperone protein htpG	5.09	71.4	P0A6Z3	3.98	5.24	MS/MS
GroEL	60 kDa chaperonin	4.85	57.2	P0A6F5	0.85	1.47	GM
GroES	10 kDa chaperonin	5.15	10.4	P0A6F9	1.85	2.15	GM
Protection							
SodA	Superoxide dismutase [Mn]	6.44	23.0	P00448	1.64	1.99	GM
Bcp	Putative peroxiredoxin bcp	5.03	17.6	P0AE52	0.60	0.40	MS/MS
KatE	Catalase HPII	5.54	84.2	P21179	1.57	4.15	GM
Transport and binding							
GlnH	Glutamine-binding periplasmic protein [Precursor]	6.87	25.0	P0AEQ3	1.18	1.53	GM
Crr	Glucose-specific phosphotransferase enzyme IIA component	4.73	18.1	P69783	0.58	0.45	GM
Transcriptional regulation							
NusA	Transcription elongation protein nusA	4.53	54.9	P0AFF6	0.40	0.39	GM

*Fold change indicates spot density of the protein in the cells with silk expression relative to the control.
†Accession number refers to the SWISSPROT accession code.
‡MS, mass spectrometry; GM, gel matching.