

Cold-shock induced high-yield protein production in *Escherichia coli*

Guoliang Qing¹, Li-Chung Ma^{2,3}, Ahmad Khorchid⁴, G V T Swapna^{2,3}, Tapas K Mal⁴, Masanori Mitta Takayama⁵, Bing Xia¹, Sangita Phadtare¹, Haiping Ke¹, Thomas Acton^{2,3}, Gaetano T Montelione^{1,2,3}, Mitsuhiro Ikura⁴ & Masayori Inouye^{1,2}

Overexpression of proteins in *Escherichia coli* at low temperature improves their solubility and stability^{1,2}. Here, we apply the unique features of the *cspA* gene to develop a series of expression vectors, termed pCold vectors, that drive the high expression of cloned genes upon induction by cold-shock. Several proteins were produced with very high yields, including *E. coli* EnvZ ATP-binding domain (EnvZ-B) and *Xenopus laevis* calmodulin (CaM). The pCold vector system can also be used to selectively enrich target proteins with isotopes to study their properties in cell lysates using NMR spectroscopy. We have cloned 38 genes from a range of prokaryotic and eukaryotic organisms into both pCold and pET14 (ref. 3) systems, and found that pCold vectors are highly complementary to the widely used pET vectors.

We have developed a series of cold-shock expression vectors, pColdI, II, III and IV, in which protein expression is under the control of the *cspA* promoter (Fig. 1a). These four vectors are identical in that they all have a pUC118 background with the *cspA* promoter, *cspA* 5'-UTR and the *cspA* 3' end transcription terminator site. All the vectors contain the *lac* operator sequence immediately upstream of the *cspA* transcription initiation site. Constitutive expression of the *lacI* gene prevents leaky expression of the cloned genes at 37 °C. Cold-shock induction of gene expression is carried out by simultaneous addition of 1 mM isopropyl-β-thiogalactopyranoside (IPTG) upon temperature downshift. In the pColdI vector, a five-codon sequence, 5'-ATGAATCACAAAGTG-3' (MNHKV), is retained directly after the *cspA* 5'-UTR, which has been shown to enhance translation initiation⁴. Thus it is termed a translation-enhancing element (TEE) (Fig. 1b). The TEE sequence is followed by a hexaHis tag (His₆) and the factor Xa cleavage site (IEGR) (Fig. 1a). The factor Xa site is followed by the multiple cloning sites starting with an *NdeI* site (Fig. 1b). In the pColdII, the factor Xa cleavage site is eliminated, and in the pColdIII, the hexaHis tag is absent (Fig. 1a). In the pColdIV vector, the multiple cloning sites starting with an *NdeI* site are

located directly downstream of the *cspA* 5'-UTR (Fig. 1a). Thus, using pColdIV vector, proteins can be produced as a nonfusion form. As conversion of AAGG to GAGG of the Shine-Dalgarno (SD) sequence results in ~50% increase of the protein expression, the SD sequence of each of the pCold vectors is altered to GAGG (Fig. 1a). Note that further improvement in the complementarity from GAGG to GGAGG resulted in almost complete inhibition of EnvZ-B expression.

To test the versatility of the pCold vectors, 38 genes from four organisms, including *E. coli*, *Caenorhabditis elegans*, *Drosophila melanogaster* and *Homo sapiens*, were inserted into pColdI and the pET14 (ref. 3) vector, respectively. The expression level and solubility of each of the 38 protein targets in these two systems are summarized in Table 1. Although no substantial differences were observed in expression level and solubility for most of the proteins produced in both systems, some interesting differences between these two vector systems were observed. The production levels of ER7, ER130, HR522, WR26 and WR41 were higher in the pET14 system than in the pCold system (Table 1), whereas the production levels of proteins ER6, ER15, FR37, FR48, HR31, HR521, HR529 and WR44 were higher in pCold vectors than in pET14 vectors. The solubilities of ER115, ER130 and FR59 proteins were better in the pET14 system, whereas ER7, ER135, FR2, FR5, FR48, HR31 and WR41 proteins had better solubilities when produced in the pCold system. These data clearly demonstrate that the current generation of pCold vectors provides protein expression and solubility levels complementary to the T7 system.

It has been shown that the *cspA* mRNA has a highly efficient structure for translation initiation⁵. When a nonsense codon is introduced at the 1st, 10th or 30th position in the CspA coding region, induction of *cspA* mRNA at 15 °C using a multiple copy plasmid results in the trapping of most cellular ribosomes and the inhibition of other cellular protein synthesis (LACE)^{6,7}. Based on this unique feature of *cspA* mRNA, it is possible to convert *E. coli* cells into a protein-producing machinery at low temperature by using pCold vectors. To demonstrate the selectivity of the pCold protein production system, we inserted

¹Department of Biochemistry, Robert Wood Johnson Medical School, 675 Hoes Lane, Piscataway, New Jersey 08854, USA. ²Northeast Structural Genomics Consortium, Rutgers University, 679 Hoes Lane, Piscataway, New Jersey 08854, USA. ³Department of Molecular Biology and Biochemistry, Center for Advanced Biotechnology and Medicine (CABM), Rutgers University, 679 Hoes Lane, Piscataway, New Jersey 08854, USA. ⁴Division of Molecular and Structural Biology, Ontario Cancer Institute and Department of Medical Biophysics, University of Toronto, 610 University Avenue, Toronto, Ontario M5G 2M9, Canada. ⁵Takara Bio Inc., Otsu, Shiga, 520-2193, Japan. Correspondence should be addressed to (inouye@umdj.edu).

Table 1 Expression and solubility for 38 proteins in pColdI and pET14 systems

<i>E. coli</i> proteins							
Gene (NESG ID ^a)	pET14 ^b		pCold ^b		MW (kDa)	SWISS-PROT/ TrEMBL ID ^a	Putative function
	Expression	Solubility	Expression	Solubility			
ER6	+	NS	+++	NS	38.0	YJHR_ECOLI	Phospholipase D / transphosphatidylase
ER7	+++	NS	++	+	27.2	NAGD_ECOLI	Hydrolase
ER15	+	NS	+++	NS	34.0	YIHR_ECOLI	Unknown
ER19	+++	+++	+++	+++	30.4	RRMA_ECOLI	Ribosomal RNA large subunit methyltransferase A
ER64	NE	NA	NE	NA	29.0	TATD_ECOLI	Deoxyribonuclease tatD
ER85	+	NS	+	NS	27.5	RFAY_ECOLI	Lipopolysaccharide core biosynthesis protein rfaY
ER115	++	+++	++	++	9.5	YRBA_ECOLI	Unknown
ER130	+++	+	++	NS	16.2	FHR_ECOLI	Translational termination
ER135	++	+	++	+++	32.0	YEGS_ECOLI	Unknown
Human proteins^d							
Gene	pET14 ^b		pCold ^b		MW (kDa)	SWISS-PROT/ TrEMBL ID ^a	Putative function
	Expression	Solubility	Expression	Solubility			
HR8	++	+++	++	+++	8.8	15E1_HUMAN	Unknown
HR31	+	NS	++	+++	13.2	C10_HUMAN	Unknown
HR91	++	NS	++	NS	9.0	LEU1_HUMAN	Leukemia-associated protein 1
HR520	NE	NA	NE	NA	19.6	Q9H3L0	Unknown
HR521	++	NS	+++	NS	14.1	Q9H3K8	Unknown
HR522	++	NS	NE	NA	12.9	CHUR_HUMAN	Mediates FGF signaling during neural development
HR524	NE	NA	NE	NA	20.2	Q9H3J9	Unknown
HR529	NE	NA	++	NS	19.0	Q9H3I1	Unknown
HR535	+++	NS	+++	NS	30.4	UT11_HUMAN	Probable U3 small nucleolar RNA-associated protein 11
<i>Drosophila</i> proteins^d							
Gene (NESG ID ^a)	pET14 ^b		pCold ^b		MW (kDa)	SWISS-PROT/ TrEMBL ID ^a	Putative function
	Expression	Solubility	Expression	Solubility			
FR2	++	+	++	++	21.6	Q9VVS0	Mitochondrial substrate carrier
FR4 ^c	NE	NA	NE	NA	14.1	Q9VP56	Insect cuticle protein
FR5 ^c	+++	+	+++	++	10.2	CLP9_DROME	Insect cuticle protein
FR6 ^c	NE	NA	NE	NA	12.0	CLP4_DROME	Larval cuticle protein IV precursor
FR14	++	NS	++	NS	19.9	ESM5_DROME	Enhancer of split m5
FR37	+	++	+++	++	11.9	Q9VIZ0	DNA-directed RNA polymerase activity
FR48	NE	NA	+++	+++	12.5	DYLX_DROME	Cytoplasmic dynein light chain
FR59	++	+++	++	++	21.0	RL1X_DROME	60S ribosomal protein L18a
FR70	+++	++	+++	++	18.5	Q9V9M7	Ribosomal protein L21
FR78	+++	++	+++	++	17.4	Q9VCF9	Ribosomal protein S19e
<i>C. elegans</i> proteins							
Gene	pET14 ^b		pCold ^b		MW (kDa)	SWISS-PROT/ TrEMBL ID ^a	Putative function
	Expression	Solubility	Expression	Solubility			
WR13	+++	NS	+++	NS	12.2	YLK_CAEEL	Unknown
WR24	NE	NA	NE	NA	8.8	yk7717	Unknown
WR26	+++	NS	++	NS	24.8	Q17958	Unknown
WR27	+++	NS	+++	NS	15.1	YHM6_CAEEL	MAP1 LC3 family member
WR33	+++	+++	+++	+++	19.4	YBYK_CAEEL	p25 protein family member
WR35	+++	NS	+++	NS	19.1	P9XWPO	Nucleic acid-binding OB-fold
WR41	+++	++	++	+++	9.8	YOY3_CAEEL	Unknown
WR44	NE	NA	+	NS	14.8	yk598f12	Unknown
WR49	+++	NS	+++	NS	23.3	O01512	Unknown
WR53	+++	+++	+++	+++	17.4	Q9XWK2	Unknown

various genes including *E. coli* EnvZ-B (molecular mass, 20 kDa) and *X. laevis* CaM (molecular mass, 17 kDa) into the pCold vectors. The upper panel in **Figure 2a** shows the time-course expression of EnvZ-B after temperature downshift from 37 to 15 °C. Equal volumes of culture were taken at each time point and used for SDS-PAGE analysis. The production of EnvZ-B increases with time, whereas the background protein remains relatively constant during the cold-shock induction. The lower panel in **Figure 2a** shows the expression level of EnvZ-B from 0 and 36 h samples using western blot analysis. After 36 h induction at 15 °C, there is ~500-fold increase in the amounts of EnvZ-B in the lysate (**Fig. 2a**, lower panel, lanes 2–7). Judging from the band densities of the other cellular proteins at various time points (**Fig. 2a**, upper panel, lanes 1–5), these proteins are produced at a very low level relative to the high expression of EnvZ-B, suggesting that the cells were under the LACE effect upon the induction of EnvZ-B expression. This was further confirmed by pulse-labeling the cold-shocked cells with ³⁵S-methionine and chasing its incorporation into newly synthesized proteins over the same time course (**Fig. 2a**). The LACE effect was observed at every time point as ³⁵S-methionine was mainly incorporated into EnvZ-B, with little incorporation into nontarget proteins (**Fig. 2b**). Note that the synthesis of the EnvZ-B was still maintained at a high level even at 48 h after cold shock (lane 4), indicating that cells in the current system retained the protein-synthesizing capacity for more than 2 d. Removing the *cspA* 5'-UTR from the 5' end down to 15 bases upstream of the SD sequence in the pColdIII vector harboring EnvZ-B, resulting in the loss of LACE, greatly decreases the level of EnvZ-B production (**Fig. 2c**, lanes 3 and 4) and increases the background as compared with the results from the same vector containing intact *cspA* 5'-UTR (**Fig. 2c**, lanes 1 and 2). These data indicate that 5'-UTR of the *cspA* mRNA plays an important role in efficient protein production.

The utility of the cold-shock expression system in isotope-enriched NMR studies of proteins has also been explored in several ways. *E. coli* EnvZ-B was selectively enriched in ¹⁵N isotopes. After 40-h at 15 °C, cells

^aAmino acid sequences for all of these proteins, targets of the Northeast Structural Genomics Consortium (NESG), are available at <http://www-nmr.cabm.rutgers.edu/bioinformatics/ZebaView/index.html>. ^bNS, not soluble; NE, no expression; NA, not available. Expression levels are scored as '+' (5%–10%), '++' (10%–20%) and '+++ (20%–40% of the total cellular proteins). Solubility levels are scored as '+' (10%–20%), '++' (20%–70%) and '+++ (70%–100% of the total expressed protein). ^cProteins FR4, FR5 and FR6 are homologs. ^dSDS-PAGE patterns are in Supplementary Fig. 1.

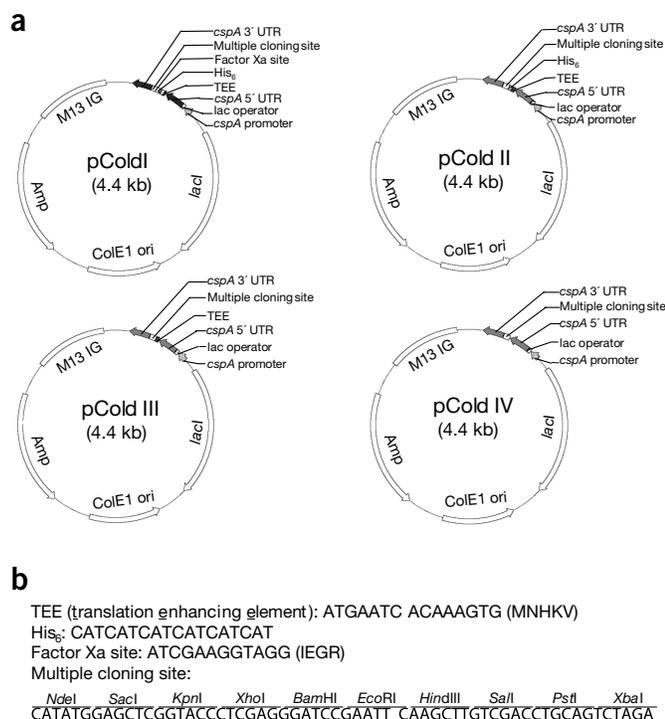


Figure 1 Structures of pCold vectors I, II, III and IV. pCold series were derived from backbone plasmid pUC118. M13 IG is the intergenic region of M13 bacteriophage. **(a)** Schematic maps of pCold vectors. **(b)** DNA sequences for TEE, His₆, Factor Xa site and multiple cloning sites used in the pCold. These vectors are available from Takara Bio, Japan (<http://bio.takara.co.jp/>).

cells using the standard T7 expression system and Ni-NTA column chromatography (Fig. 3b). However, the spectrum obtained for the cold-shock-expressed EnvZ-B was somewhat broader than that of the highly purified counterpart, suggesting there may be some association of the ¹⁵N-labeled protein with other components in the cell lysate (proteins or nucleic acids) or the viscosity of the lysate sample is much higher, giving rise to slow tumbling of the molecule.

We also tested the technology of selective isotope enrichment with the pCold system using *X. laevis* CaM, a eukaryotic protein consisting of 148 residues, which has been extensively used for various NMR studies by many researchers. The HexHis-tagged construct of CaM used in this work includes 160 amino acid residues. SDS-PAGE patterns of the total cellular proteins and the supernatant fraction after ultracentrifugation are shown in Figure 2e, lanes 1 and 2, respectively. The yield of ¹⁵N-labeled *X. laevis* CaM was about 180 mg/l or 35 mg/l unit at OD₆₀₀. In this case, the supernatant fraction after ultracentrifugation was treated with DNase I (10 µg/ml) for 4 h at 4 °C before NMR analysis, as this improved the quality of the resulting NMR spectra. Indeed, selectively-enriched cold-shock-expressed CaM yielded a remarkably clean ¹⁵N-H^N HSQC spectrum (Fig. 4a), which is very similar to the ¹⁵N-H^N HSQC spectrum of purified protein produced with the same vector (Fig. 4b).

We next carried out selective ¹⁵N, ¹³C-enrichment of *X. laevis* CaM using the pCold system. Cell lysate containing ¹⁵N, ¹³C-enriched CaM plus 5% ²H₂O was directly used for triple resonance NMR experiments. The ¹³C^α-H^N projection of the three-dimensional (3D) HNcoCA spectrum, representative of the quality of triple-resonance NMR spectra obtained for the whole cell lysate, is shown in Figure 4c. Analysis of ten 3D NMR spectra used for determining NMR resonance assignments (that is, 3D HNcoCACB, 3D HNCACB, 3D HNCO, 3D HNCA, 3D HNcoCA, 3D HANH, 3D HAcNH, 3D hCCoNH-

expressing *E. coli* EnvZ-B were spun down and lysed. SDS-PAGE patterns of the total cellular proteins and the supernatant fraction after ultracentrifugation are shown in Figure 2d, lanes 1 and 2, respectively. The yield of EnvZ-B protein was about 200 mg/l or 40 mg/l unit at OD₆₀₀. We added 5% D₂O, for NMR spectrometer lock purposes, directly to these cell lysates, and then transferred a sample to a 5-mm NMR sample tube. No further purification was carried out before recording NMR spectra. The ¹⁵N-H^N HSQC spectrum of selectively cold-shock-induced EnvZ-B, providing a peak for each N-H group in the protein, displayed a large number of well-resolved peaks (Fig. 3a), many of which matched exactly those obtained with the EnvZ-B sample that was expressed and purified to 95% homogeneity from *E. coli*

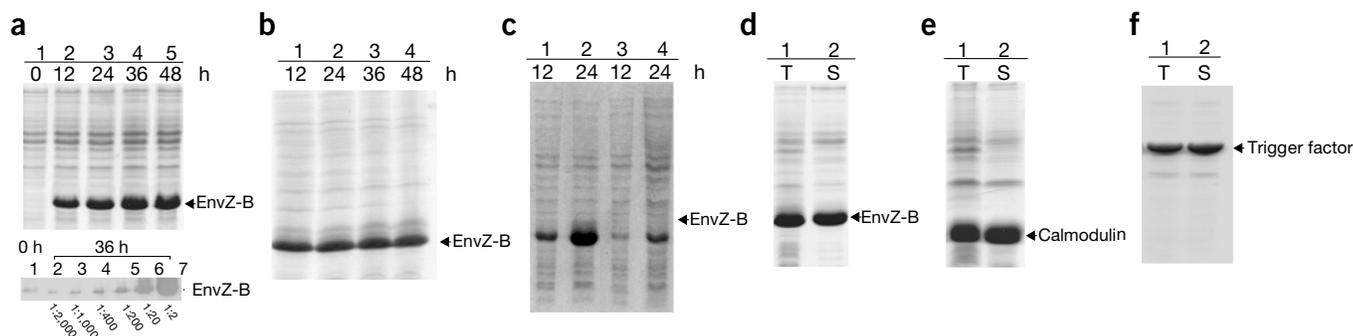


Figure 2 SDS-PAGE patterns of *E. coli* EnvZ ATP-binding domain (EnvZ-B), *X. laevis* calmodulin (CaM) and *E. coli* trigger factor expressed using pCold vectors. CaM, EnvZ-B and trigger factor were inserted into pColdII, III and IV vectors through NdeI and BamHI sites, respectively. **(a)** Upper panel: time-course expression of EnvZ-B in the minimal medium at 15 °C. Lower panel: western blot analysis of EnvZ-B expression. Lane 1, sample from 37 °C (control), the same one as used in the upper panel; lanes 2–7, the sample from 15 °C after 36 h induction, the same one as used in the upper panel but with 2000-, 1000-, 400-, 200-, 20- and twofold dilutions, respectively. **(b)** Pulse-labeling of cold-shocked cells with ³⁵S-methionine at 15 °C. Cells were grown under the same condition as in **a** in the presence of 10 µg/ml cold methionine, and labeled with ³⁵S-methionine for 15 min at the time points indicated. **(c)** Comparison of expression of EnvZ-B in wild-type pColdIII vector (lanes 1 and 2) and *cspA* 5'-UTR deletion mutant vector (lanes 3 and 4). **(d, e)** and SDS-PAGE patterns of (¹⁵NH₄)₂SO₄-labeled EnvZ-B **(d)** and CaM **(e)**, respectively. **(f)** Expression of *E. coli* trigger factor at 15 °C in LB medium. T, total cellular protein; S, supernatant used for the cell lysate NMR.

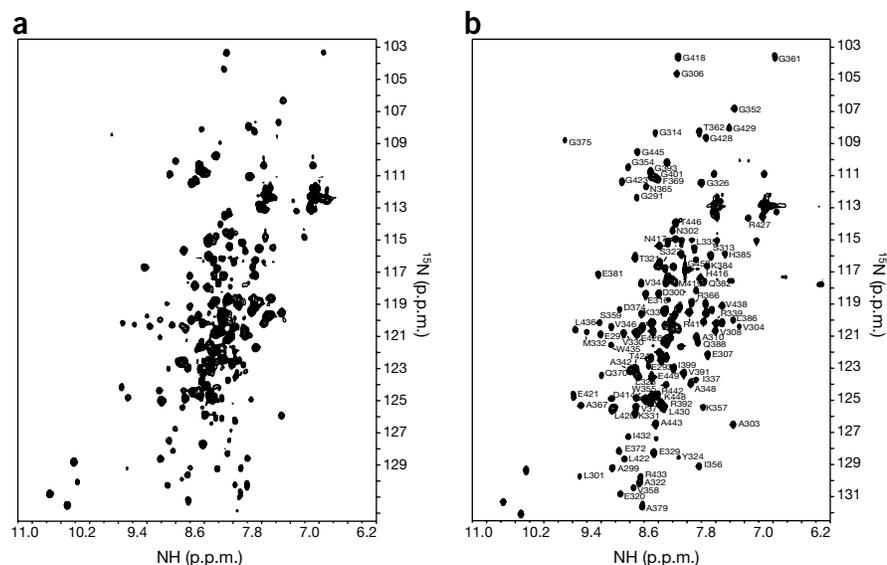


Figure 3 600 MHz ^{15}N -HN HSQC spectra of *E. coli* EnvZ-B. (a,b) Whole cell lysate (a) and purified protein (b) in 20 mM sodium phosphate, pH 7.0, 50 mM KCl, 0.5 mM 4-(2-aminoethyl)-benzenesulphonyl fluoride, 50 mM sodium azide and 5 mM MgCl_2 . Spectra were recorded with a sample temperature of 27 °C, using 576 and 256 complex points in t_1 and t_2 , respectively. Total data acquisition time was about 1 h 40 min for each spectrum. Final data sets comprised 1,024 and 2,048 real points with digital resolution of 1.8 and 4.4 Hz/point in F_1 and F_2 , respectively.

TOCSY, 2D ^{13}C -H, and 2D ^{15}N -HN HSQC spectra, recorded as described in refs. 21 and 22, and summarized in Supplementary Table 1) provided sequence-specific resonance assignments for some 80% of H^α , H^β , ^{15}N , $^{13}\text{C}^\alpha$, $^{13}\text{C}^\beta$ and $^{13}\text{C}'$ nuclei in the 160-residue construct, as well as many side chain resonance assignments. The data establishing these assignments, which match well to the published resonance assignments⁸, are documented in Figure 5 is the analysis of secondary structure indicated by these chemical shift assignments⁹, which is in good agreement with the known 3D structure of CaM. Only small dif-

ferences of chemical shifts are observed between purified CaM and the CaM in the cell lysate. These results demonstrate the feasibility of carrying out resonance assignments and experimental secondary structure determination for a selectively-isotope-enriched protein in a whole cell lysate without a requirement for protein purification.

The present pCold system can also be used for high-level protein production for other biochemical and biophysical studies. For exam-

plified, the pCold technology provides whole cell lysates of ^{15}N , ^{13}C -enriched CaM, which are amenable to extensive resonance assignment analysis. The assignments determined from these data, and the secondary structure indicated by these ^{13}C chemical shifts, are in excellent agreement with assignments and 3D structure that have been reported for purified CaM⁸, suggesting that this approach could be used to determine the secondary structure of an isotope-enriched target protein for which the structure is not previously known.

Thus, the pCold system can in principle provide more selective isotope-enrichment of target proteins than is possible using pET vectors. Consequently, at low temperature most of the translational machinery is dedicated to producing the target protein(s), which persists for a long time without downregulation of the cloned gene. NMR analysis of isotope-enriched proteins in the cell lysate has been previously carried out using T7-based pET expression systems at 37 °C in several reports^{10–12}. This technology is particularly selective in the pCold vectors because the LACE effect of the *cspA* gene inhibits the synthesis of most non-target cellular proteins; consequently, at low temperature most of the translational machinery is dedicated to producing the target protein(s), which persists for a long time without downregulation of the cloned gene.

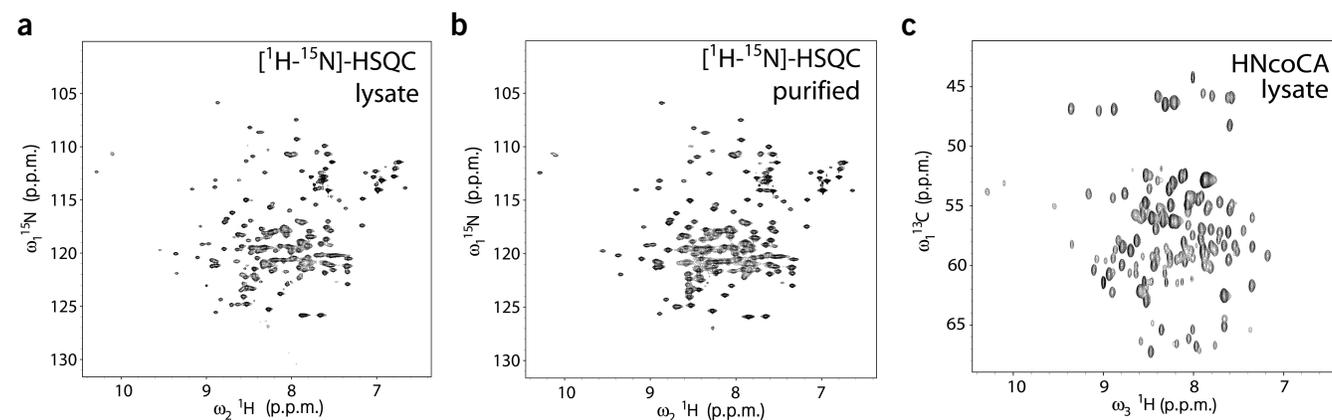
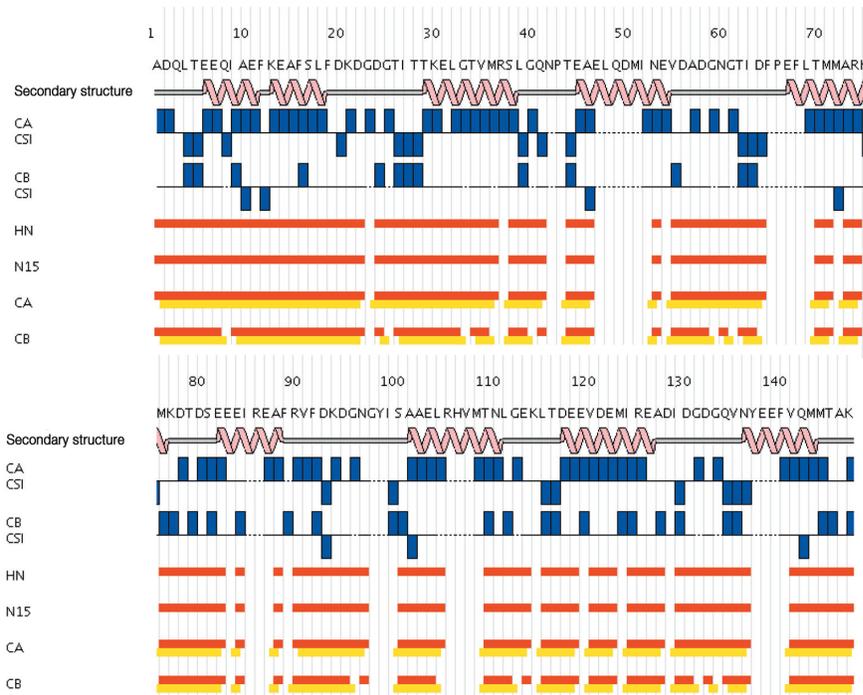


Figure 4 600 MHz NMR spectra of *X. laevis* CaM. (a,b) ^{15}N -HN HSQC for whole cell lysate containing 20 mM Tris HCl, pH 6.8, 100 mM KCl, 5 mM EDTA (a), and ^{15}N -enriched CaM purified from this same sample containing 20 mM Tris HCl, pH 6.8, 100 mM KCl and 5 mM EDTA (b). The total acquisition times for these two-dimensional (2D) ^{15}N -HN HSQC spectra were approximately 30 min each, collecting $200 \times 1,024$ complex points along t_1 and t_2 , respectively. Final data sets comprised 1,024 and 2,048 real points with digital resolution of 1.8 and 4.4 Hz/point in F_1 and F_2 , respectively. (c) 2D ^{13}C -HN projection of 3D HNcoCA spectrum of whole cell lysate containing ^{15}N , ^{13}C -enriched CaM; the $\text{C}^\alpha(i)$ nuclei are frequency-labeled along the ω_1 axis and $\text{H}^\alpha(i+1)$ resonances are on the ω_3 . The 3D HNcoCA was acquired using eight scans per increment and $40 \times 40 \times 1,024$ complex points along t_1 , t_2 and t_3 , respectively, with the total acquisition time of about 20 h. These spectra of CaM were all acquired with sample temperature of 10 °C.

Figure 5 Summary of triple-resonance NMR data obtained on whole-cell lysates establishing sequence-specific resonance assignments for some 80% of ^{15}N , ^1H , ^1H , $^{13}\text{C}^\alpha$, $^{13}\text{C}^\beta$ and $^{13}\text{C}'$ nuclei. Intra (red) and sequential (yellow) connectivity data used to establish resonance assignments at each sequence position are shown. These data were derived from a set of ten 2D and 3D NMR spectra, which are summarized in **Supplementary Table 1**. The derived resonance assignments are summarized in **Supplementary Table 2**. Secondary structure information derived from combined analysis of $^{13}\text{C}^\alpha$, $^{13}\text{C}^\beta$ and $^{13}\text{C}'$ chemical shifts is also plotted along the protein sequence. Bar graphs (blue) represents the Chemical Shift Index (CSI) analysis⁹ of C^α and C^β chemical shifts; for segments of CaM for which backbone resonance assignments were determined, this CSI analysis is in good agreement with known protein structure⁸.



ple, the *E. coli* trigger factor (molecular mass: 45 kDa), was produced at levels as high as 400 mg/l in a soluble form using Luria Bertani (LB) medium with a 24-h induction at 15 °C (Fig. 2f). The pCold system provides a highly effective alternative technology complementing the widely used and very successful pET bacterial production systems.

METHODS

Plasmid construction. Plasmid pCold07, a derivative of pUC118, was used to construct the pCold series. pCold07 was constructed as follows: promoter, 5'-UTR (untranslated region) and transcription terminator regions of the *ospA* gene were separately amplified by PCR, and inserted into plasmid pTV118N (Takara Bio), a derivative of pUC118. The region from -67 to +1 of *ospA*, which was shown to have a full *ospA* promoter activity⁵, was used as a promoter. At the end of the promoter region, a synthetic *lac* operator sequence, 5'-ATTGTGAGCGGATAACAATTGATGTGCTAGCGCATATC-3', was introduced and connected to the 5'-UTR region at +27 in the *ospA* gene. A point mutation from T to C was made at nucleotide +159, just immediately upstream of the initiation codon. Right after the translation initiation site, a translation enhancing element⁴, 5'-ATGAATCACAAAGTG-3' encoding MNHKV, was placed and followed by a histidine hexamer and a Factor Xa cleavage site. This region is then followed by the multiple cloning sites and the *ospA* 3'UTR from +381 to +517, which includes the *ospA* termination codon, TAA. Finally, the designated cold-shock expression unit from the promoter to transcription terminator was cloned into pTV118N between the *AflIII* and *EcoO109I* sites. At the same time, the *E. coli lacI* gene was also inserted immediately upstream of the *ospA* promoter in the opposite direction to the promoter. All the pCold vectors thus constructed were verified by DNA sequencing.

Protein overexpression. pCold vectors harboring different genes of interest were transformed into *E. coli* BL21 cells (containing the rare tRNA expression plasmid pMGK). The transformed cells were grown at 37 °C in LB medium. At mid-log phase, the cultures were shifted to a 15 °C water bath and IPTG (1 mM) was added to induce protein expression. Expression levels were monitored over time (0, 12, 24, 36 and 48 h) after cold-shock induction and analyzed by SDS-PAGE.

The small-scale expression of a set of 38 proteins from four organisms, including *E. coli*, *C. elegans*, *D. melanogaster* and *H. sapiens*, was evaluated in both pCold I and pET 14 (ref. 3) vector systems. All 38 of these proteins are targets of the Northeast Structural Genomics Consortium (NESG)¹³. The first letter of the NESG i.d. refers to the organism from which the proteins have been

cloned: E, *E. coli*; W, *C. elegans*; F, *D. melanogaster*; and H, *H. sapiens*. The level of expression and the degree of solubility for each of these proteins was estimated by the total lysate and the supernatant from the harvested cell pellet of each protein on SDS-PAGE gels in both systems. For the expression of pET14 system, a single colony was cultured in MJ9 (ref. 14) minimal medium. Initial growth was carried out at 37 °C until the OD₆₀₀ of the culture reached 0.8–1.0 units. The incubation temperature was then decreased to 17 °C and protein expression was induced by the addition of IPTG at a final concentration of 1 mM. After overnight incubation at 17 °C, the cells were harvested by centrifugation. Protein expression in pColdI vector was done the same way as that in pET14 vector at 15 °C.

For NMR studies, uniformly ^{15}N or both ^{15}N - and ^{13}C -enriched proteins were produced in pCold vectors after a change to isotope-enriched minimal medium coupled with cold shock. Transformed cells were first grown in 25 ml LB medium at 37 °C to OD₆₀₀ = 1.0, and then transferred to prewarmed MJ9 medium¹⁴ to a final volume of 250 ml. At mid-log phase (OD₆₀₀ = 0.6), the cultures were shifted to a 15 °C water bath for 30 min, and then centrifuged. The pelleted cells were resuspended in 125-ml isotope-enriched MJ9 medium, containing 1 mM IPTG and 0.1% ($^{15}\text{NH}_4$)₂SO₄ for ^{15}N -labeling experiment or 1 mM IPTG and 0.1% ($^{15}\text{NH}_4$)₂SO₄ plus 0.4% ^{13}C -glucose for ^{15}N , ^{13}C -enrichment. The induced cells were harvested after 40 h at 15 °C.

Lysate preparation. The cell pellet was washed and suspended in 5 ml NMR buffer (20 mM sodium phosphate, pH 7.0, 50 mM KCl, 5 mM MgCl₂, 5 mM β-ME and 0.5 mM NaN₃ for EnvZ-B; 20 mM Tris-HCl, pH 6.8, 100 mM KCl and 5 mM EDTA for CaM). After sonication, the cell debris was pelleted by centrifugation at 35,000g for 30 min, and the supernatant was further centrifuged at 100,000g for 1 h. The supernatant thus obtained was subjected to further NMR analysis.

NMR spectroscopy. The NMR data were obtained on Varian Unity Inova 600 MHz spectrometers, each equipped with 5-mm triple resonance probes and operating proton frequency of 600.256 MHz. For EnvZ-B, these data were processed and analyzed with NMRpipe and NMRDraw¹⁵ software. Chemical shifts were referenced to external DSS. Backbone resonance assignments for EnvZ-B have been described elsewhere¹⁶. Backbone resonance assignments for CaM were made on a whole lysate sample with selective ^{13}C , ^{15}N -enrichment of CaM, using standard triple resonance NMR experiments^{17,18}. Spectra of CaM

were processed using the NMRpipe 2.1 software¹⁵ and analyzed using SPARKY (<http://www.cgl.ucsf.edu/home/sparky>).

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS

We thank Y. Chiang for helpful discussions. This work was supported by NIH grants R21-GM067061 (to M.I.) and P50-GM62413 (to G.T.M.), a grant from Takara-Bio Inc., Japan (to M.I.), and a Canadian Institutes of Health Research (CIHR) grant (to M.I.). M.I. is a CIHR senior investigator.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 20 January; accepted 27 April 2004

Published online at <http://www.nature.com/naturebiotechnology/>

- Schein, C.H. Solubility as a function of protein structure and solvent components. *Biotechnology* **8**, 308–317 (1989).
- Schirano, Y. & Shibata, D. Low temperature cultivation of *Escherichia coli* carrying a rice lipoxygenase L-2 cDNA produces a soluble and active enzyme at a high level. *FEBS Lett.* **271**, 128–130 (1990).
- Studier, F.W. & Moffatt, B.A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189**, 113–130 (1986).
- Etchegaray, J.P. & Inouye, M. Translational enhancement by an element downstream of the initiation codon in *Escherichia coli*. *J. Biol. Chem.* **274**, 10079–10085 (1999).
- Mitta, M., Fang, L. & Inouye, M. Deletion analysis of cspA of *Escherichia coli*: requirement of the AT-rich UP element for cspA transcription and the downstream box in the coding region for its cold-shock induction. *Mol. Microbiol.* **26**, 321–335 (1997).
- Xia, B., Etchegaray, J.P. & Inouye, M. Nonsense Mutations in cspA Cause Ribosome Trapping Leading to Complete Growth Inhibition and Cell Death at Low Temperature in *Escherichia coli*. *J. Biol. Chem.* **276**, 35581–35588 (2001).
- Bae, W., Jones, P.G. & Inouye, M. CspA, the major cold shock protein of *Escherichia coli*, negatively regulates its own gene expression. *J. Bacteriol.* **179**, 7081–7088 (1997).
- Zhang, M., Tanaka, T. & Ikura, M. Calcium-induced conformational transition revealed by the solution structure of apo calmodulin. *Nat. Struct. Biol.* **2**, 758–767 (1995).
- Wishart, D.S. & Sykes, B.D. The ¹³C chemical-shift index: a simple method for the identification of protein secondary structure using ¹³C chemical-shift data. *J. Biomol. NMR* **4**, 171–180 (1994).
- Gronenborn, A.M. & Clore, G.M. Rapid screening for structural integrity of expressed proteins by heteronuclear NMR spectroscopy. *Protein Sci.* **5**, 174–177 (1996).
- Erbel, P.J. *et al.* Identification and biosynthesis of cyclic enterobacterial common antigen in *Escherichia coli*. *J. Bacteriol.* **185**, 1995–2004 (2003).
- Almeida, F.C. *et al.* Selectively labeling the heterologous protein in *Escherichia coli* for NMR studies: a strategy to speed up NMR spectroscopy. *J. Magn. Reson.* **148**, 142–146 (2001).
- Wunderlich, Z. *et al.* The protein target list of the Northeast Structural Genomics Consortium. (in the press) (2004). Published online May 2004.
- Jansson, M. *et al.* High level production of uniformly ¹⁵N- and ¹³C-enriched fusion proteins in *Escherichia coli*. *Biomol. NMR* **7**, 131–141 (1996).
- Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
- Tanaka, T. *et al.* NMR structure of the histidine kinase domain of the *E. coli* osmosensor EnvZ. *Nature* **396**, 88–92 (1998).
- Cavanaugh, J., Fairbrother, W.J., Palmer, A.G., III & Skelton, N.J. Protein NMR Spectroscopy (Academic Press, San Diego, 1996).
- Montelione, G.T., Rios, C.B., Swapna, G.V.T. & Zimmerman, D.E. NMR pulse sequences and computational approaches for analysis of sequence-specific backbone resonance assignments of proteins. in *Biological Magnetic Resonance* (Kluwer Academic/Plenum, New York, 1999) **17**, 81–130.