Supplementary Information Appendix

Efficient, ultra-high affinity chromatography in a one-step purification of complex proteins

Marina N. Vassylyeva[a], Sergiy Klyuyev[a], Alexey D. Vassylyev[a], Hunter Wesson[a], Zhuo Zhang[a], Matthew B. Renfrow[a], Hengbin Wang[a], N. Patrick Higgins[a], Louise T. Chow[a,1], Dmitry G. Vassylyev[a,1]

[a] *Department of Biochemistry and Molecular Genetics, University of Alabama at Birmingham, Schools of Medicine and Dentistry, 720 20th Street South, Birmingham, AL 35294, U. S. A*

[1] – to whom correspondence may be addressed:

E-mails: ltchow@uab.edu; dmitry@uab.edu

Content

SI Materials and Methods

Expression

Unless otherwise specified we have used the same procedures for plasmid construction, expression, cell growth and lysis. The commercial pET28a expression vector (Invitrogen; Waltham, MA USA) was used as a template vector, into which the coding nucleotide sequences were inserted at the unique NcoI/XhoI sites. A stop codon was included in the target sequences prior to the XhoI site to exclude the His6-tag (in the commercial construct), which would otherwise tag the C-termini of the target proteins. Using this approach, we derived a variant pET28a plasmid for expressing target protein with the C-terminal CL7-tag. In this new vector, the sequence of the target proteins was cloned using the NcoI/SpeI restriction sites. If N-terminal "expression/affinity" tags were used, we introduced a HindIII restriction site right after the tag sequences to accommodate the cloning of the target protein. The gene sequences were designed through the manual inspection and modification of the native (genomic) sequences to exclude the rare *E. coli* codons and high (G/C) content (where appropriate). Segments of the designed sequences were synthesized commercially (IDT; San Jose, CA, USA) and then merged together either through PCR (Phusion polymerase; NEB; Ipswich, MA, USA) or through ligation. The resulting expression plasmids were transformed into the BL21 Star (DE3) (Invitrogen; Waltham, MA USA) competent cells.

Colonies were grown overnight (37^0C) and plasmids from 2-3 colonies were sequenced to verify the sequences. The cells were cultured in the TB media (http://www.bio-protech.com.tw/databank/DataSheet/Biochemical/DFU-J869.pdf) in 2- or 4-L flasks (for 1- or 2-L cultures, respectively) according to the following protocol. The

bacteria were grown at 37^0C for ~2-2.5 h until the OD_{560} of the cultures reached ~0.7-0.8. The temperature was then reduced to 20^0C and the over-expression was induced overnight (20-24 h) by addition of 0.1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG). The cultures were centrifuged at 4,000 g for ~30 min, and the cell pellets were frozen at -80^0C. For purification, the frozen cell pellet was suspended in the respective lysis buffers (see below) (1 g cells → 10 ml buffer) and then disrupted at 4^0C using the Nano DeBEE high pressure homogenizer (BEE International; Chula Vista, CA, USA) at ~15,000 PSI pressure for ~3 min (for ~3 g cells). The lysates were then centrifuged at 40,000 g for 20 min and filtered through a 45 μm filter (33 mm in diameter). All purifications were carried out using the Acta Prime purification system (GE Healthcare; Marlborough, MA, USA).

Im7 column preparation

The Im7 immobilization unit was expressed as fusion with a PreScission protease (PSC) cleavable thioredoxin (Trx) and His8 tag. One L culture of bacteria expressing the Im7 immobilization unit usually produces ~24 g cells. Purification was carried in the two chromatographic steps (**Fig. 1C**). First, the cell lysate (lysis buffer, i.e., buffer A: 0.5 M NaCl, 20 mM Tris pH8, 5% glycerol, 0.1 mM PMSF) was loaded on the His-Prep FF (20 ml; GE Healthcare) column (flow rate 5 ml/min.) in buffer A with addition of 2% buffer B (1 M imidazole). After loading, the column was washed by the two alternate cycles (2-3 column volume each) of high/low (1 M / 0 M NaCl) salt buffers (with other components identical to those in buffer A) with an addition of 5% buffer B, and then eluted in buffer A with an addition of 25% buffer B. The eluate was dialyzed against buffer A for ~4-5 h in the presence of purified PreScission protease (PSC) to cleave off the Trx and His8 tags

(**Fig. 1C**). Since the Im7-unit resisted ~90^{o}C heat, the dialyzed sample was heated at 70^{0}C for ~45 min to eliminate PSC and then loaded onto the His-Prep column again under the same conditions as at the first step. The flow through (FT) containing the highly purified Im7 (FT2 in **Fig. 1C**) was then concentrated to ~20 - 50 mg/ml and dialyzed against the coupling buffer recommended by Thermo Fisher (Waltham, MA USA) for protein immobilization onto the Sulfo-Link (iodo-acetyl activated) 6B agarose beads. Immobilization was carried out according to the commercial protocol (http://www.funakoshi.co.jp/data/datasheet/PCC/20401.pdf) in a dark room and the reaction was completed in ~15 to 20 min. The typical concentration of the immobilized Im7-unit was ~15 mg/ml beads (or ~0.6 mM). The Im7-coupled beads were then packed into a 20 ml glass, low pressure column for affinity purification (through the respective adaptors) with the Acta Prime system.

Purification of a model (CL7M) protein

To test the column performance, we used a model protein (CL7M; **Fig. 1D**) comprising Trx (~12 kDa) tagged at the carboxyl terminus with the CL7 domain (~16 kDa) followed by the SUMO domain (~11 kDa). We tested the Im7-column with this model protein multiple times under the different loading conditions by varying salt (0.3 – 1.2 M NaCl), reducing agent (β-mercaptoethanol up to 15 mM), metal chelating agent (EDTA up to 20 mM), or detergent (DDM up to 1.5%) at varying flow rate (up to 4 ml/min.). We consistently obtained results similar to that shown in **Fig. 1D**. Upon loading, the column was subjected to a few (2-3) alternate cycles of high/low (1 M/ 0 M NaCl) salt buffer washing. The protein was then eluted under the denaturing condition of 6 M Guanidine

hydrochloride (Gdn), following by column cleaning/reactivation using the gradient option of Acta Prime (**Fig. 1B**), i.e., gradually exchanging the Gdn with the physiological buffer (0.5 M NaCl, 20 mM Tris pH8, 5% glycerol) in ~1 h. This cleaning/reactivation step was used at the end of each purification. No significant loss of capacity was detected after over 100 reactivations. The concentrations of the model protein bound to the Im7 column were in the range of ~15 to 20 mg/ml beads. This Im7-column capacity corresponded to ~1:1 molar ratio of CL7M and immobilized Im7-unit, suggesting that ~100% of the immobilized Im7-units retained full binding activity. Notably, we were able to achieve this highly specific and stable immobilization only with the iodo-acetyl (Sulfo-Link, Pierce) beads. Multiple trials with practically all possible alternative amino-coupling resins resulted in ~75-85% loss of the Im7 binding activity.

Purification of ttRNAP and mtRNAP

One L culture of *E. coli* expressing ttRNAP or mtRNAP usually produced ~8 - 10 g cells. The lysis buffer (~30 – 35 ml) contained 0.1 M NaCl, 20 mM Tris pH 8.0, 5% glycerol, 0.5 mM $CaCl_2$, 10 mM $MgCl_2$, 0.1 mM PMSF, ~120-150 μg DNAse I Grade-I, (Roche; Indianapolis, IN, USA), and 1 tablet of inhibitory cocktail (Roche; Indianapolis, IN, USA) for ~3 g cells. The cell lysates were incubated for ~1.5 h at 4^0C in the lysis buffer with addition of 0.05 mM PMSF after each 30 min during incubation. The lysates were then diluted 2 times with the 2-fold loading buffer containing 2.3 M NaCl, 20 mM Tris pH 8.0, 5% glycerol to increase the salt concentration to 1.2 M and loaded onto a 20 ml Im7-column (flow rate of ~1.5-2 ml/min; **Figs. 4C** and **5A**). After loading, the column was washed with 2 or 3 alternate cycles (2 - 3 column volumes each) of high/low (1 M / 0 M NaCl) salt buffer to remove unbound contaminants. The proteins were then eluted using

small amounts (~0.3 mg) of PSC dissolved in ~40 ml of the elution buffer (0.5 M NaCl, 20 mM Tris pH 8.0, 5% glycerol, 0.2 mM EDTA) at ~0.2 ml/min.

For the His-tagged ttRNAP construct (vector MV0, **Fig. 4B**) purification was carried out under essentially the same conditions as that for the Im7 purification. The only differences were the addition of 20 mM, 50 mM and 250 mM imidazole to the loading, washing and elution buffers, respectively. No EDTA was added to the elution buffer.

Purification of the YidC protein from the uninduced cells

The cells were grown as described above except that after the cell density reached OD_{560} ~ 0.7-0.8 at 37^0C the temperature was decreased to 20^0C with no IPTG addition. One L culture of uninduced YidC produced ~20 g cells. The 200 ml of filtered lysate in a lysis buffer containing 0.5 M NaCl, 20 mM Tris pH 8.0, 5% glycerol, 0.1 mM PMSF, and 4 inhibitory tablets (Roche; Indianapolis, IN, USA) were ultra-centrifuged at 120,000 g for 1.5 h. The pellet containing the membrane fraction (MF; **Fig. S10B**) was then dissolved in ~100 ml loading buffer (0.9 M NaCl, 20 mM Tris pH 8.0, 5% glycerol, 0.1 mM PMSF, 1.5% DDM) and ultra-centrifuged again at 120,000 g for 30 min. The supernatant was loaded onto a 20 ml Im7-column (flow rate at ~1.2–1.5 ml/min; **Fig. S10B**). The column was then subjected to washing with a few (2-3) alternate cycles (2-3 column volumes each) of high/low (1 M / 0 M NaCl) salt buffers containing 0.1% DDM. The proteins were then eluted using the small amounts (~0.6 mg) of PSC dissolved in ~40 ml of the elution Buffer-E1 (0.5 M NaCl, 20 mM Tris pH 8.0, 5% glycerol, 0.2 mM EDTA, 0.1% DDM) at ~0.2 ml/min.

Purification of the YidC and calnexin proteins from IPTG-induced cells

E. coli cultures expressing the YidC or calnexin (CNX) proteins were grown as described above with a standard over-expression induction (0.1 mM IPTG). One L of the induced cultures produced ~10 g cells. The lysates (~80 ml for YidC and ~120 ml for CNX) in the lysis buffers containing 0.35 M or 0.45M NaCl (for YidC or CNX), 20 mM Tris pH 8.0, 5% glycerol, 0.1 mM PMSF, 1 inhibitory tablet (Roche; Indianapolis, IN, USA) for ~3 g cells were subjected to polyethyleneimine (PEI) precipitation (**Fig. S10A**, right panel, and **Fig. 6C**) as follows. A solution of 10% PEI was added to lysates in three aliquots to a final concentration of 0.06%. At each step, the lysates were gently mixed for ~10 min. After the final step, the suspension was centrifuged at 5,000 g for 15 min. A larger fraction of the CNX protein (SN; **Fig. 6C**) remaining in the soluble fraction after the PEI precipitation than that of YidC (**Fig. S10A**), likely due to a higher salt concentration in the CNX lysate (0.45 M vs 0.35 M NaCl). The PEI pellets were then washed with the solution containing 0.6 M NaCl, 20 mM Tris pH 8.0, 5% glycerol and 1.5% DDM. The soluble fraction was diluted 10-fold with the DDM-free high salt loading buffer to yield 0.9 M NaCl, 20 mM Tris pH 8.0, 5% glycerol and 0.15% DDM. This diluted supernatant was loaded onto a 20 ml Im7-column (flow rate of ~1.5 ml/min; **Figs. 6B** and **6C**). The column was then washed as described in the previous Section. The proteins were eluted using the small amounts (~0.25 – 0.5 mg) of PSC dissolved in ~40 ml of the elution Buffer-E1 at ~0.2 ml/min.

Expression and Purification of bacterial MukBEF condensin complex

A CL7 tag was inserted via recombineering (1) into the chromosomal MukB gene of Salmonella typhimurium at the 3' end using a module that includes a flexible linker and a

PSC cleavage site. To make the module generally useful, a kanamycin resistance gene was included in the module for selection. The CL7 module (1632 bp) was constructed using Gibson assembly (NEB; Ipswich, MA, USA) and then the linear DNA was integrated the *Salmonella* chromosome using plasmid pSIM5 (39), which promotes efficient homologous recombination with linear DNAs having only 40 bp of terminal DNA homology to the chromosome (1). Chromosomal DNAs were sequenced to confirm the strain containing the correct structure of the CL7-modified MukB gene.

Genetically modified cells of were grown (60 L culture) in the UAB fermentation center. Cell paste was frozen in 10 aliquots and MukB-CL7 purification was carried out with one aliquot of 45 g of cells. Cells were thawed and suspended in 200 ml of lysis buffer (10 mM Tris, pH 8; 100 mM NaCl; 5 % glycerol; 0.1 mM PMSF; 1 mM benzamidine; and 1 μg/mL aprotinin) and were then passaged twice through a French press. The lysate was cleared by centrifugation at 40,000 RPM at 4°C for 1 h. NaCl (1 M solution) was added to the supernatant (to yield 0.5 M NaCl concentration in total) followed by addition of polyethelenimine to a final concentration 0.06%. The solution was centrifuged at 30,000 RPM for 30 min and a fraction containing 230 ml of 20 mg/ml protein (4.6 g) was loaded directly onto a 1.5 ml column of Im7 beads over a period of 1 h. The column was washed with 100 ml of high salt buffer (50 mM Tris, pH8.2, 5% glycerol; 800 mM NaCl) followed by 100 ml low salt buffer (100 mM NaCl). The protein was eluted with 6 M Gdn, dialyzed against 8 M urea (to avoid precipitation by SDS) and loaded onto an SDS polyacrylamide gel. Electrophoresis was carried out and proteins were stained with Coomassie Brilliant Blue. Three prominent bands ran at positions expected for MukB (170 kDa), MukF (50 kDa), and MukE (25 kDa). A fourth strong band ran between the 75 and 100 kDa markers

(**Fig. 7A**).

To identify every protein, stained bands were excised from the gel and subjected to liquid chromatography high resolution Mass Spectrometry. The mass spectrometry raw files were searched against the protein sequence database and identified all three components with multiple peptides covering >50% of the amino acid sequence for each MukB protein at the >95% confidence level. Bands for MukB, MukE, and MukF were present at the expected molar ratios. A fourth, additional band (above) was present at equimolar as MukB. Mass Spec revealed this protein to be DnaK.

Purification of RSF1 recombinant protein and its truncated variants

The expression of recombinant double-tagged (His8/CL7) RSF1 protein and its truncated variants was induced in *E. coli* (BL21 DE3) by using 0.5 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) (Cat. # R1171, Fisher, Waltham, MA, USA) overnight. The cells were pelleted by centrifugation at 4000 rpm, 4ºC, 20 min and lysed in buffer A (20 mM Tris-HCL, pH7.6, 150 mM NaCl, 10% glycerol, 1 mM PMSF, 1 mM DTT) plus 1% protease inhibitor cocktail (Cat #78430, Fisher, Waltham, MA, USA) with sonication. After centrifugation (13,000 rpm, 4ºC, 10 min), the supernatant (supplemented with 15 mM imidazole) was loaded onto Buffer A pre-washed His-column (HisTrap HP, Cat. #17-5247-01, GE, Pittsburgh, PA, USA). The loaded column was washed with PBS supplemented with 15 mM imidazole (Cat. # 10284730, Fisher, Waltham, MA, USA) until no protein was recovered. The His-tagged RSF1 protein was then eluted with PBS supplemented with 300 mM imidazole. To bind the recombinant His-tagged RSF1 to the Im7 beads, the elution from His-column was incubated for 2 hours at 4ºC with 100 µL Im7

beads, which were prewashed with BC50 (20 mM Tris-HCl, pH 7.9, 50 mM KCl, 10% glycerol) 3 times. The Im7 beads were then washed 3 times with BC500 (20 mM Tris-HCl, pH 7.9, 500 mM KCl, 10% glycerol), twice with BC50 (20 mM Tris-HCl, pH 7.9, 50 mM KCl, 10% glycerol) supplemented with 2 M NaCl, and twice with BC50. A one-step purification of the full-length RSF1 and its F10 fragment was carried out in essentially the same manner as that of RNAPs (see above), beginning with loading of the lysate on the Im7 beads in high (1 M NaCl) salt, except that DNAse treatment was not used.

References

1. Datta S, Costantino N, & Court DL (2006) A set of recombineering plasmids for gram-negative bacteria. *Gene* 379:109-115.
2. Elrod-Erickson M, Benson TE, & Pabo CO (1998) High-resolution structures of variant Zif268-DNA complexes: implications for understanding zinc finger-DNA recognition. *Structure* 6(4):451-464.
3. Symersky J*, et al.* (2006) Regulation through the RNA polymerase secondary channel. Structural and functional variability of the coiled-coil transcription factors. *J Biol Chem* 281(3):1309-1312.
4. Perederina AA*, et al.* (2006) Cloning, expression, purification, crystallization and initial crystallographic analysis of transcription elongation factors GreB from Escherichia coli and Gfh1 from Thermus thermophilus. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 62(Pt 1):44-46.
5. Vassylyeva MN*, et al.* (2007) The carboxy-terminal coiled-coil of the RNA polymerase beta'-subunit is the main binding site for Gre factors. *EMBO Rep* 8(11):1038-1043.
6. Artsimovitch I, Svetlov V, Murakami KS, & Landick R (2003) Co-overexpression of Escherichia coli RNA polymerase subunits allows isolation and analysis of mutant enzymes lacking lineage-specific sequence insertions. *J Biol Chem* 278(14):12344-12355.
7. Svetlov V & Artsimovitch I (2015) Purification of bacterial RNA polymerase: tools and protocols. *Methods Mol Biol* 1276:13-29.
8. Ko TP, Liao CC, Ku WY, Chak KF, & Yuan HS (1999) The crystal structure of the DNase domain of colicin E7 in complex with its inhibitor Im7 protein. *Structure* 7(1):91-102.
9. Wang YT, Yang WJ, Li CL, Doudeva LG, & Yuan HS (2007) Structural basis for sequence-dependent DNA cleavage by nonspecific endonucleases. *Nucleic Acids Res* 35(2):584-594.
10. Malhotra A (2009) Tagging for protein expression. *Methods Enzymol* 463:239-258.

A **Price/Performance (PP) Comparison**

Affinity System			PP[2] ($/mg)	Company
Type[1]	Tag	Column Ligand		
FLAG	Pept	Prot	180/760	SinoBiol/ Sigma
CalBD	Pept	Prot	220	Sigma
ChBD	Prot	SMol	2	NEB
Pr-A	Prot	Prot	13-20	GE
Strep	Pept	Prot	10/12	Qiagen/GE
Halo	Prot	SMol	10 (disposable)	Promega
MBP	Prot	SMol	8-10	GE
GST	Prot	SMol	2	GE
His	Pept	Ni^{2+}	0.8/1.5	GE/Sigma
CL7	Prot	Prot	~0.8-1.1	Lab Prep

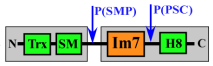
B **Estimated costs of the Im7-column production**

Expression/Purification (lab scale) :
(~25 g cells per 1 L culture) → (~0.8 g pure Im7 per 1 L culture)

Immobilization:
~13-15 mg of Im7 to 1 ml Sulfo-Link beads

Cost (over)estimate (lab scale):
1) (800 mg/L)x(12 L) / (15 mg/1 ml beads) = 640 ml Im7-beads[3]
2) Efforts ($1,000/day x 5 days) = $5,000[4]
3) Sulfo-Link price ($8.7/1 ml) x 640 ml = $5,600[5]

Price: Im7-charged beads ($10,600/640 ml)= ~$16.5/1 ml
Capacity: (300-400 mg protein/20 ml beads) = ~15-20 mg/1 ml[6]
Price/Performance (**PP** = Price/Capacity) = ~$0.8-1.1/1 mg

Further cost reduction:
1) 2 step → 1 step Im7 purification[7] → 

2) Scale-up cell growth/purification[8]
3) Self prep of the Sulfo-Link beads[9]

Fig. S1. The Price/Performance (PP) factors of the affinity systems. *(A)* Comparison of the PPs of the available commercial and Im7 affinity purification approaches. *(B)* Estimate of the PP for the Im7 purification technique (laboratory scale). Prot – protein; Pept – short peptide; SMol – small molecule; PP = (Price of 1 ml beads)/(Amount of protein bound to 1 ml beads); Trx – thioredoxin; SM –SUMO domain; Im7 – Im7 immobilization unit; H8 – 8 histidine tag; P(SMP)/P(PSC) – cleavage sites of the SUMO or PreScission proteases.

[1] - **Protocols** (include recommended Salt Loading Conditions; **SLC**) and dissociation constants (K_D):

FLAG
http://www.sinobiological.com/Anti-DYKDDDDK-Affinity-Resin_p227780.html (binding buffer SLC - 137mM NaCl); K_D (10) (Pages 5162 & 5167).

CalBD
http://wolfson.huji.ac.il/purification/PDF/Expression_Systems/STRATAGENE_Calmodulin_Manual.pdf (SLC - Page 23); K_D (1) (Page 4).

ChBD
https://www.neb.com/~/media/Catalog/All-Products/21A73B351DD24F94BC584FAED2A83A0F/Datacards%20or%20Manuals/manualE6901.pdf (SLC - Page 20, but see also (**Fig. S3**); K_D (11) (Page 464).

Pr-A
https://www.gelifesciences.com/gehcls_images/GELS/Related%20Content/Files/1314787424814/litdoc52208300_20161014083613.pdf (SLC - Page 2, Tris-saline Tween 20; TST); K_D (12) (Page 840).

Strep – protocol (13) (SLC - Page 1530); K_D (13) (Page 1529).

Halo
https://www.promega.com/~/media/files/resources/protocols/technical%20manuals/101/halotag%20mammalian%20protein%20detection%20and%20purification%20systems%20tm348.pdf (SLC – Page 22).

MBP - http://wolfson.huji.ac.il/purification/PDF/Tag_Protein_Purification/Maltose/AMERSHAM_MBPHiTrapII.pdf (SLC - Page 6, see also **Figs. S2B and 2C**); K_D (14) (Page 13667; D-Maltose - K_D = 3.9x10-6).

GST - https://www.gelifesciences.com/gehcls_images/GELS/Related%20Content/Files/1465571389954/litdoc18115758_20161013165930.pdf (SLC – Page 44; see also **Figs. S2C, S3B and S3C**); K_D (15) (Page 330).

His - https://www.auburn.edu/~duinedu/manuals/HisTrapHP.pdf (SLC - Page 8); K_D (16) (Pages 270 & 275)

(2) - The Price/Performance (PP) comparison assumes that an entirely pure protein is produced through a one-step of respective purification. However, if the additional steps are required, as is the case for many complex proteins (**Fig. 2, Figs. S5 and S6),** the PP factor will correspond roughly to the sum of the PPs of the techniques used, and will additionally increase due to a likely loss of a protein at each step and extra efforts/time required.

(3) – A lab shaker allows us to grow 12L (6 x 2L flasks) bacterial culture in one run (1 day, ~22 hours).

(4) – We calculated the efforts with the upper (over) estimate.

Personnel - salary $180K/year or $180K/260 working days = **$692/1 working day.** (In fact, the column preps in our lab are carried out by a single Research Associate at $64K/year).

Equipment & Supplies - $500K/5years = $100K/year or $100K/365 = **$274/day** for equipment, reagents, Ni^{2+}-charged beads, empty commercial columns, protease prep, etc. (In fact, the equipment/supplies we currently use for a column prep are probably ~2-3 times lower, and equipment is normally functioning not for 5 but at least for 8-10 years).

Thus, **one day** of column prep would cost $692 + $274 = $966 or ~**$1,000**.

Prep Time - 5 working days for column prep (1 day for cell cultivation/disruption; 2 days for purification, 1 day for immobilization, and 1 day in excess to account for some potential problems/delays).

(5) – Commercial price: https://www.thermofisher.com/order/catalog/product/20404

(6) – The capacity of the Im7-column is evaluated based on purification of a model, CL7M protein (**Fig. 1D**).

(7) – The current 2-step purification of the Im7 unit may be reduced to 1-step (2 → 1 day) by using the modified vector shown. The expression (Trx-SUMO) tag may be cleaved during the cell lysis with an addition of the cells expressing the GST-tagged SMP (**Fig. S3B**) at ~1:15 ratio to the Im7 cells. This mixed lysates worked well in our practice. The His-tag may be cleaved on-column or after purification by the GST-tagged PSC, which can be then eliminated by the heat step, as the Im7-unit resists ~90°C heating.

(8) – The cell prep may be easily scaled up (10 times or more) using the fermentation facilities. Even assuming that the larger scale would double a time of a column prep (to 10 days), still there will be ~4-5 times advantage in overall cost.

(9) – The price for the inactivated or amino-activated agarose beads is 3-6 times less than that of the commercial iodo-acetyl resin. Given that reagents (carbodiimide & iodo-acetic acid) required to prepare the Sulfo-Link type matrix are also very inexpensive and the protocol of activation seems to be quite simple, an in-house production of the iodo-acetyl beads may further reduce the price of the Im7-column prep by a factor of 2 or more in the industrial environment.

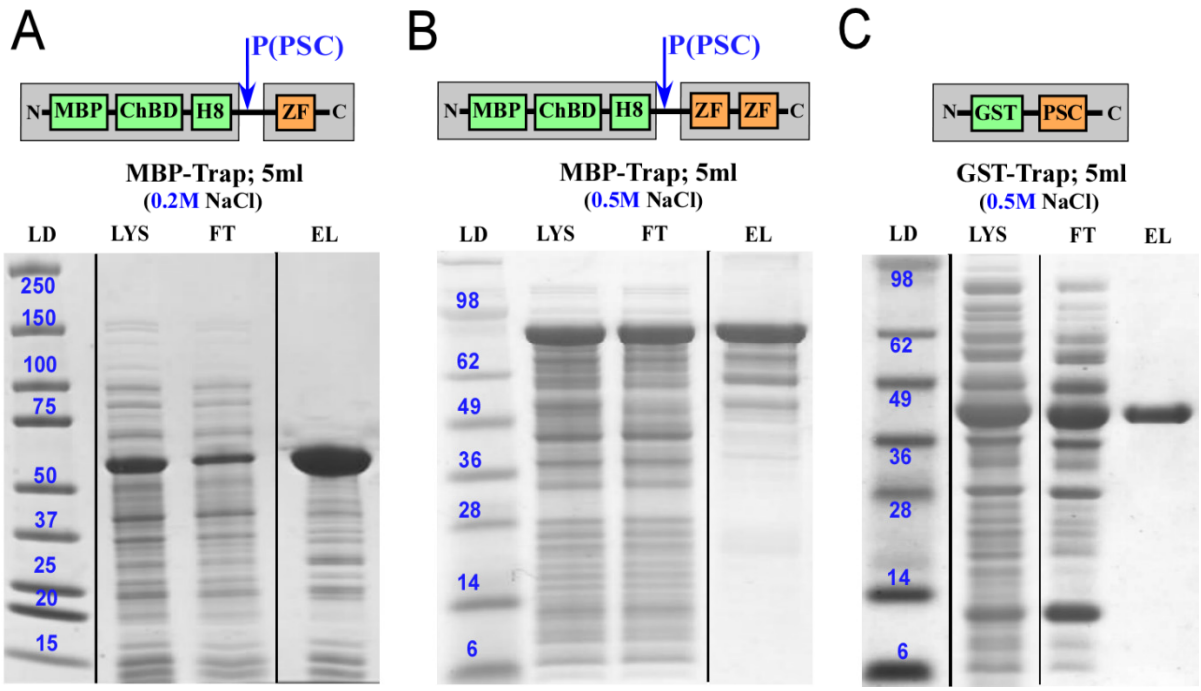


Fig. S2. Effect of salt concentration on the MBP-tag and GST-tag purification systems.

(A, B) Purification of the two MBP-tagged DNA-binding proteins. The proteins contain a single *(A)* or tandem *(B)* zinc finger DNA-binding domains, for which crystal structure in complex with the cognate DNA fragment was previously determined (2). Loading of the protein with a single (ZF) domain on a column in low salt (0.2 M NaCl) demonstrated reasonably good binding but the purified sample was heavily contaminated *(A)*. In contrast, the construct with the tandem ZF domains *(B)* loaded in a higher-salt (0.5 M NaCl) buffer showed less impurities (though still far from an appropriate purity level) but very poor retention to the column (90+ % in flow-through, FT). Importantly, in both cases, contaminants could not be eliminated despite of a high-salt (0.8-1 M NaCl) wash after loading. In addition, both eluted, ZF *(A)* and tandem ZF-containing *(B)* proteins were significantly contaminated by nucleic acids as revealed by a poor OD(260/280) ratio of ~0.95; this ratio should normally be in a 0.55 - 0.65 range for a pure, DNA-free protein. These results suggest that ***(i)*** the primary conditions, at which the lysate is loaded on a column, are crucial for obtaining high purity samples, and ***(ii)*** even 0.5 M NaCl concentration is not sufficient to eliminate DNA-related contaminants in some DNA-binding proteins. *(C)* Purification of the GST-tagged PSC protease. The lysate (LYS) was loaded on the column in presence of 0.5 M NaCl. The eluted sample (EL) was very pure but ~90% of the tagged protein did not bind to the column and remained in the FT fraction though the lysate was loaded on a column quite slowly (0.3 ml/min). In contrast, the same target was successfully purified, when the NaCl concentration during loading was decreased to 0.15 M, allowing an

efficient (~90%) binding of the tagged protein to the column (**Fig. S3C**). We also note that our trials to use Strep-Tag fused to multi-subunit *T. thremophilus* RNA polymerase provided very similar results to those in panels *(B)* and *(C)*. Practically no protein was eluted (bound to the column) when the lysate was loaded on a column in 0.5 M NaCl. GST – glutathione-S transferase; MBP – maltose binding protein; ChBD – chitin-binding domain; H8 – 8 histidine tag; PSC – preScissoin protease; P(PSC) - PSC cleavage site; LD – ladder (kDa); LYS – lysate; FT – flow through; EL – eluate.

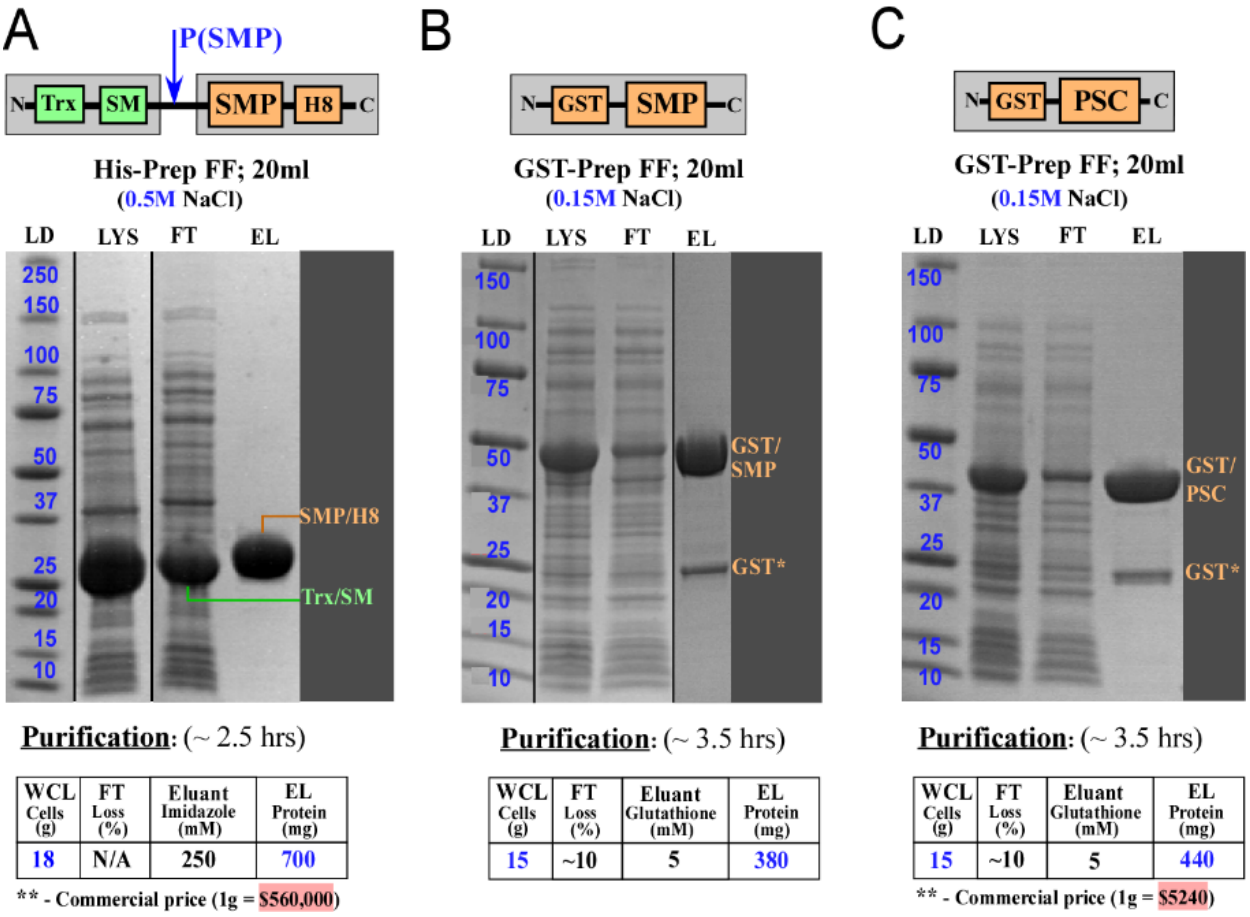
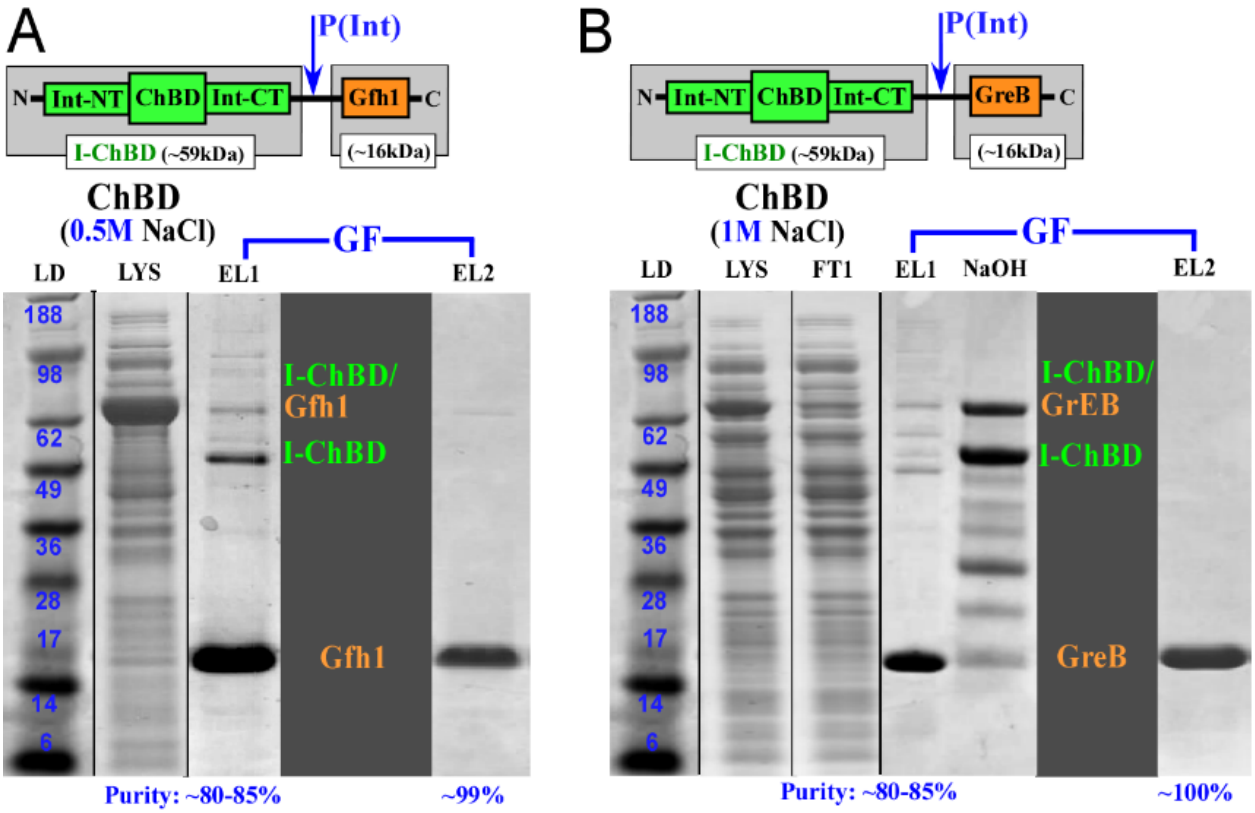


Fig. S3. Expression and purification of SUMO and PreScission proteases. *(A, B)* Expression vectors and purification of the His-tagged *(A)* and GST-tagged *(B)* SUMO protease (SMP). The His-tagged variant has the (Trx-SUMO) fusion N-terminal tag to improve the expression level (compare to panel B). The fusion is self-cleaved by SMP, likely during cell cultivation, generating Trx or SM. *(A)*. *(C)* Expression vector and purification of the GST-tagged PreScisson protease (PSC). For GST-tagged variants of SMP and PSC, the successful purification (with no substantial amount of the target proteins in flow-through, **Fig. S2C**) was achieved with low-salt (0.15 M NaCl) in the lysis/loading buffer. After loading, the column was washed with the alternate cycles (~2-3 column volumes each cycle) of the high (1 M NaCl) and low (0 M NaCl) salt-containing buffers. Trx - thioredoxin; SM – SUMO domain; H8 – 8 histidine tag; SMP/PSC – SUMO or PreScission proteases; P(SMP) – SMP cleavage site; LD – ladder (kDa); LYS – lysate; FT – flow through; EL – eluate; FF – Fast Flow.

GST* - these bands likely correspond to the GST fragment of the constructs, which might be produced either through translational truncation or cleavage by cellular proteases in the flexible linker between the GST-tag and the target proteins.

**** - prices of the respective commercial proteases in market (2).



ChBD beads; 20ml → GF

WCL Cells (g)	FT1 Loss (%)	Eluant DTT (mM)	UN Cleaved (%)	EL1 Amount (mg)	FT2 Loss (%)	EL2 Amount (mg)
5	N/A	50	~45	11	55	5

ChBD beads; 20ml → GF

WCL Cells (g)	FT1 Loss (%)	Eluant DTT (mM)	UN Cleaved (%)	EL1 Amount (mg)	FT2 Loss (%)	EL2 Amount (mg)
7	N/A	50	~45	3.3	60	1.3

Fig. S4. Purification of the *T. thermophilus* and *E. coli* transcription factors, Gfh1 and GreB using the chitin binding approach. *(A, B)* The two-step, chitin-binding (ChBD) → gel filtration (GF) purification of Gfh1 (3, 4) *(A)* and Greb (4, 5) *(B)*. The Gfh1 and GreB genes were cloned in the commercial pTYB12 vector (NEB, Ipswich, MA, USA) at the C-terminus of intein, which contains the chitin-binding domain insertion, so that the target protein can be eluted after purification upon the intein self-cleavage reaction induced by the reducing agent, dithiothreitol (DTT). The Gfh1 and GreB proteins are transcription factors that recognize their cognate nucleic acid substrates when bound to RNAP in the transcription complexes (3, 5), and therefore are likely to have some intrinsic nucleic acid-bidning affinities. For this reason, based on our previous experience (**Figs. S2A and S2B**), we used the highest possible salt concentrations at the stage of the lysate loading on the chitin beads to avoid potential nucleic acid contamination. However, though GreB successfully bound to the beads in 1 M NaCl with no significant flow-through, FT *(B)*, substantial

amount (~60%) of Gfh1 went into FT at the same conditions. To obtain a better yield of Gfh1, we, therefore, used medium, 0.5 M NaCl for its loading onto the column (4) *(A)*.

Overall, both purifications showed very similar results that can be summarized as follows. ***First***, ChBD purification provided reasonably pure samples, but their purities (80-85%) were not appropriate to perform crystallographic analysis. In particular, apparent leakage of the bound, ChBD containing constructs (I-ChBD) was observed upon the cleaved protein elution. ***Second***, the patterns of the Gfh1 and GreB impurities were quite similar and could not be eliminated even upon alternate washing (2-3 columns volumes each cycle) with buffers containing the extra-high (2.0 M NaCl) and low (0 M NaCl) salt concentrations. These results suggest that contaminants are not target- but rather column- and/or tag-dependent (i.e. associated with a cleaved tag leaking from the column). A perfect purity of the proteins was achieved only after a second, GF step. ***Third***, in both cases, ~45% of protein remained uncleaved, even when the induced reactions continued for ~40 hours and at room temperature, as indicated by the NaOH strip (not shown for Gfh1) of the columns after a target elution *(B)*. The strip fraction also showed that a notable amount of a cleaved target protein remains bound to the column. ***Fourth***, though a second, GF step provided high-purity proteins it was also characterized by a very significant (55-60%) loss of the proteins. Together with a limited capacity of the first-step ChBD approach, the overall yield of highly purified protein was modest (Gfh1) or poor (GreB). In particular, while the amount of purified Gfh1 was suitable for crystallographic studies (3), crystallization of GreB (~1 mg from a single purification run), was quite problematic. We, therefore, switched to the different, His-tag construct. With this new construct, we could produce substantially larger (~50-fold) amount of protein with essentially the same purity through a very similar, two-step (His-Trap-GF) protocol (3). The yield allowed for a straightforward crystallographic analysis (5). We concluded that His-Trap is more productive both in terms of efforts and yield than the ChBD approach. This conclusion is also consistent with the fact that while ChBD purification was used as a first step of purification of substantially more complex multi-subunit *E. coli* RNAP in early studies (6) it was eventually replaced by a more efficient His-Trap approach (7). Int-NT/Int-CD – N- or C-terminal portions of intein; ChBD – chitin-binding domain; I-ChBD – intein-ChBD fusion; P(Int) – intein cleavage site; GF – gel filtration; LD – ladder (kDa); LYS – lysate; FT – flow through; EL – eluate; NaOH – column strip with 0.5 M NaOH.

Purification protocols for crystallized membrane proteins

#	Protein	Organism		Purification			PA	PDB	Year	Journal
		SRC	EXP	Steps	Protocol	days	mg/L	ID		
1	BacNAv(voltage gate)	sp	ec	6*	His->[PSC]->His->IEQ->GF	4**	-	5HJ8	2016	Cell
2	σ1 receptor	hm	ins	4	FLAG->[PSC]->GF	3	-	5HK1	2016	Nature
3	Sterol Transporter	hm	yeast	6	His->CalBD->PSC->CalBD->GF	4	-	5DO7	2016	Nature
4	TRPV6 (Calcium Channel)	hm	hm	4	Strep->[THR]->GF	3	1	5IWK	2016	Nature
5	Adenosine A2A receptor	hm	ins	5	His->[TEV]->His->GF	4	2	5G53	2016	Nature
6	Ton complex	ec	ec	6	His->[TEV]->IEQ->GF1->GF2	5	-	5SV0	2016	Nature
7	Chloride channel	nm	ec	3	His->GF	2	-	5G28	2016	Nat Comm
8	TPC1 (voltage channel)	at	yeast	4	His->[THR]->GF	3	-	5E1J	2016	Nature
9	SWEET transporter	at	yeast	4	His->[PSC]->GF	3	-	5CTG	2015	Nature
10	Dopamine transporter	dm	hm	4	His->[THR]->GF	3	-	4XP1	2015	Nature
11	Adiponectin receptors	hm	ins	6	FLAG->IEQ->[TEV]->His->GF	4	-	3WXV	2015	Nature
12	Na^+ pump	ke	ec	5	His->[TEV]->His->GF	4	-	3X3B	2015	Nature
13	P2Y1 receptor	hm	ins	5	His->[PSC]->His->GF	4	-	4XNW	2015	Nature
14	TRAAK K+ channel	hm	yeast	4	His->[PSC]->GF	3	-	4WFE	2014	Nature
15	TMEM16 lipid scramblase	nh	yeast	5	His->[TEV]->His->GF	4	0.14	4WIS	2014	Nature
16	Sterol reductase	ma	ec	3	His->GF	2	-	4QUV	2014	Nature
17	GPR40 receptor	hm	ins	4	His->FLAG->His	3	-	4PHU	2014	Nature
18	Glutamate receptor	hm	ins	3	His->GF	2	-	4OO9	2014	Nature
19	NMDA receptor	xl	hm	4	FLAG->[PSC]->GF	3	-	4TLL	2014	Nature
20	Yidc (used in this work)	bh	ec	5	His->[TEV]->His->GF	4	0.3	3WO6	2014	Nature
21	NRT1.1 nitrate transporter	at	ins	3	His->GF	2	-	4OH3	2014	Nature
22	NapA (proton antiporter)	tt	ec	5	His->[TEV]->His->GF	4	-	4BWZ	2013	Nature
23	MATE multidrug transporter	pf	ec	5	His->[TRN]->His->GF	4	-	3WBN	2013	Nature
24	GIRK2-beta gamma	ms	yeast	6	His->[PSC]->IEQ->GF->IEQ	5	0.9	4KFM	2013	Nature
25	Nitrate/nitrite exchanger	ec	ec	4	His->[THR]->GF	3	-	4JR9	2013	Nature
26	VCX1 calcium exchanger	yeast	yeast	4	His->[PSC]->GF	3	-	4K1C	2013	Nature
27	TrkH ion channel	vp	ec	4	His->[TEV]->GF	3	-	4J9V	2013	Nature
28	PAR1 receptor	hm	ins	5	His->FLAG->[TEV]->GF	4	-	3VW7	2012	Nature
29	TatC core	aa	ec	5	His->[TEV]->His->GF	4	-	4B4A	2012	Nature
30	Multi-Drug Transporter	ce	yeast	5	His->[PSC]->GST->GF	3	-	4F4C	2012	Nature
		Average***		4.5 (2.7)		3.4				
		1st Step % used		H-trap 86; Others 14						

Fig. S5. HHH-purification of the crystallized significant membrane proteins. A major criterion for selection of proteins for this analysis was their biological significance, which in many cases commanded non-trivial technical challenges in respective studies. We, therefore, inspected and listed only the PDB entries for those published in the top ranked Journals (mostly *Nature*) in a reversed chronological order. No other preferences were included. A few proteins with unique, target-specific purification protocols were excluded from considerations. This analysis revealed several interesting similarities. ***First***, no purification was achieved in 1 chromatographic step and/or in 1 day. Notably, in 87% of studies, over 3 purification steps (and most likely over 2 days) were required to obtain the crystallization quality samples. ***Second***, in the overwhelming majority of studies, His-Trap (H-Trap) was used as a first purification step (86%). Moreover, in 73% of purifications, it was the only affinity approach. ***Third***, the affinity approaches other than His-tag constituted only ~11% (9 out of 81) of all chromatographic steps used. ***Fourth***, a gel-filtration chromatography (GF) was used in all but one cases (96%) and as a last step in 93% of studies. These statistics suggest that GF is nearly universally essential for successful high-purity purifications of the membrane proteins. Despite the universal application of GF, it possesses no specificity to a target apart from its size/shape and, therefore, may be quite sensitive to a variety of the target-specific contaminants. For example, even small impurities may induce dynamic aggregation of a sample through non-specific

contacts with a target protein, detrimental to the GF approach. We observed such effect in our trials with ttRNAP purification, in which GF was also used as a last step (see **Fig. 2** and example #2 in the **Fig. S6**). Even when a small amount of DNA remained in the RNAP sample loaded onto a GF column, it resulted in a smeared multi-disperse peak corresponding to essentially unpurified enzyme and a huge overall loss (~85-95%) of protein. These samples were usually discarded because of their little practical value to either structural or functional studies. Another notable limitation of the GF system is a systematic and significant (40-70% in our experience) loss of the protein samples even in successful GF runs.

Organisms: SRC – source; EXP – expression. sp - *Silicibacter pomeroyi*; ec – *E. coli*; hm – human; ins – insect cells; nm - *Nonlabens marinus;* at - *Arabidopsis thaliana*; dm - *Drosophila melanogaster*; ke - *K. eikastus*; nh - *N. haematococca*; ma – *M. alcaliphilum*; xl - *Xenopus laevis*; bh – *B. halodurans*; tt – *T. thermophilus*; pf - *P. furiosus*; ms – mouse; vp – *V. parahaemolyticus*; aa - *A. aeolicus*; ce - *C. elegans*;

Columns: PA – protein amount; His – His-Trap; IEQ – anion exchange; FLAG – anti-FLAG anti-body; GF- gel filtration; CalBD – calmodulin; Strep – streptavidin; GST – glutathione.

Tag Cleavage: PSC – PreScission protease; THR – thrombin; TEV – TEV protease; TRN – trypsin. We consider protease digestion as a purification step since in many cases it is a time consuming process (16 - 24 hrs).

* - in all these studies, ultracentrifugation (UCF) was the very first and essential purification step, which in our experience, may take almost a full working day (5-7 hours) assuming a two-step conventional protocol (UCF1→Solubilization→UCF2).

** - Purification time was estimated based on our own experience with similar protocols assuming, in particular, that an over-night dialysis is normally used between the different chromatographic steps.

*** - an average number of the total (including UCF, protease cleavage, etc.) purification steps. An average number of chromatographic steps is shown in brackets.

Purification protocols for crystallized nucleic acid-binding proteins

#	Protein	Organism		Purification			PA	PDB	Year	Journal
		SRC	EXP	Steps	Protocol	days	mg/L	ID		
1	I-E Cascade	ec	ec	4	His->[PSC->]His->GF	3*	-	5H9E	2016	Nature
2	RNAP (used in this work)	tt	HOST	6	PEI->AS->IES->HEP->IEQ->GF	8	0.5	5D4D	2016	Nature
3	TET2-5fC complex	hm	ec	3	His->IEQ->GF	3	-	5D9Y	2015	Nature
4	AlkD DNA glycosylase	bc	ec	5	His->[PSC]->HEP->His->GF	4	-	5CL3	2015	Nature
5	Cas1-Cas2 Complex	ec	ec	4	His->[TEV]->MBP->GF	3	-	5DS4	2015	Nature
6	B12 photoreceptor	tt	ec	2	His->IEQ	2	-	5C8E	2015	Nature
7	Tus-Ter	ec	ec	5	AS->IES->PC->AS->GF	3	3-4	4XR0	2015	Nature
8	HIF-2a:ARNT Complex	mm	ec	3	His->IEQ->GF	3	-	4ZPK	2015	Nature
9	RNAP (Pol II)	yeast	HOST	6	AS->HEP->AS->IEQ->WG16->IEQ	7	-	4Y52	2015	Nature
10	DNA polymerase β	hm	ec	4	IEQ->PC->DNC->IES	3	3	4UAW	2015	Nature
11	Cas9 Endonuclease	sp	ec	4	His->[TEV]->IES->GF	3	-	4UN3	2014	Nature
12	AddAB Helicase	bs	ec	5	His->HEP->[TEV]->His->IEQ	4	-	4CEH	2014	Nature
13	Tet-like dioxygenase	ng	ec	5	His->[SMP]->IEQ->IES->GF	4	-	4LT5	2014	Nature
14	Human Aprataxin	hm	ec	4	His->[THR]->GF->IES	3	-	4NDF	2014	Nature
15	cGMP AMP synthase	ss	ec	4	His->[TEV]->IES->GF	3	-	4KB6	2013	Nature
16	HNF4α Nuclear Receptor	hm	ec	3	His->IES->GF	3	-	4IQR	2013	Nature
17	Phage Mu transpososome	ev	ec	4	AS->HEP->HEP->IES	3	-	4FCY	2012	Nature
18	DNA polymerase eta	hm	ec	3	His->IES->GF	2	-	3O1M	2012	Nature
19	Replication initiator DnaA	ae	ec	5	His->HEP->[TEV]->His->GF	4	-	3R8F	2011	Nature
20	ISW1 Chromatin Remodeling	yeast	ins	3	His->HEP->GF	3	-	2Y9Z	2011	Nature
21	DNA Repair Dioxygenase	ec	ec	3	IEQ->IES->IES	2	3	4OH3	2010	Nature
22	PVF Strand Transfer Compl.	sfv	ec	4	His->[PSC]->HEP->GF	3	-	3OS0	2010	Nature
23	DNA glycosylase AlkD	bc	ec	4	His->[SMP]->HEP->GF	3	-	3JX7	2010	Nature
24	GYRASE	sa	ec	5	IEQ->HEP->IEQ->GF->IEQ	5	0.5	2XCT	2010	Nature
25	LexA Repressor	ec	ec	2	His->IES	2	-	3JSO	2010	Nature
26	Argonaute protein	tt	ec	4	His->[SMP]->His->GF	3	-	3HJF	2009	Nature
27	Alkyltransferase	yeast	ec	2	His->GF	2	-	3GYH	2009	Nature
28	Meganuclease	cr	ec	2	His->Strep	2	-	2VBJ	2008	Nature
29	SRA protein UHRF1	mm	ec	5	GST->[SMP]->IES->HEP->GF	4	-	2ZKD	2008	Nature
30	RecA Helicase	ec	ec	6	PEI->AS->His->[TEV]->IEQ->GF	4	-	3CMT	2008	Nature
	Average**			4.0 (3.2)		3.3				
	1st Step % used			H-trap 73; Others 27						

Fig. S6. HHH-purification of the crystallized nucleic acid binding proteins. To select crystallographic studies for this analysis we used the same criteria as that described in the **Fig. S5**. The analysis reveals both, general similarities and differences among the approaches in purifying DNA-binding and membrane proteins (MPs; **Fig. S5**). ***First***, similar to MPs, no purifications could be achieved in 1 chromatographic step and/or in 1 day. Also similar, in 84% of studies, over 3 purification steps (and most likely over 2 days) were required to obtain crystallization quality samples. ***Second***, in most studies (albeit to a lesser extent than for the MPs), His-Trap (H-Trap) was used as a first purification step (73%). In 70% of studies, it was the only affinity approach used. ***Third***, the affinity approaches other than His-tag constitute only ~4% (4 out of 96) of all chromatographic steps used. This number is substantially smaller than that of the MPs (~11%). ***Fourth***, a gel-filtration approach (GF) was heavily used (but not universally as for MPs) as a last or intermediate step in the 73% of cases (see discussion of this technique in the **Fig. S5**). ***Fifth***, unlike MPs, where the low-specificity techniques other than GF were used quite rarely (~5%; 4 out of 81 total chromatographic steps), the ion exchange and various nucleic acid mimicking resins were applied in ~87% purifications of the DNA-binding proteins. The performance of most of these techniques is based on the specific and common properties of DNA and/or DNA-binding proteins that bind to the matrixes with

varying affinities. These properties eventually allow for elimination of the DNA and DNA-related contaminants from the target samples. However, usually such purification cannot be achieved through a single run and/or on a single column. In most cases, therefore, a set of these low-specificity columns is used to obtain the high-purity proteins samples. In fact, in our experience, even the order of application of these approaches might differ from one protein to another, as reflected in the selected protocols. Altogether, both this analysis and our own experience suggest that ***(i)*** the DNA-related impurities are the major factor affecting purification of the DNA-binding proteins, and ***(ii)*** the purification protocols for the DNA-binding proteins are quite diverse and should be developed and adjusted for each particular protein. Thus, substantial time and efforts are normally spent before the optimal protocols and conditions for purification are found.

Organisms: bc - *B. cereus*; ec – *E. coli*; tt – *T. thermophilus*; HOST – host organism (no over-expression); hm – human; bc – *B. cereus*; mm – *M. musculus*; sp - *S. pyogenes*; bs – *B. subtilis*; ng – *N. gruberi*; ss – *S. scrofa*; ev - *Escherichia virus Mu*; ae – *A. aeolicus*; ins – insect cells; sfv - *Simian foamy virus;* sa – *S. aureus*; cr – *C. reinhardtii*;

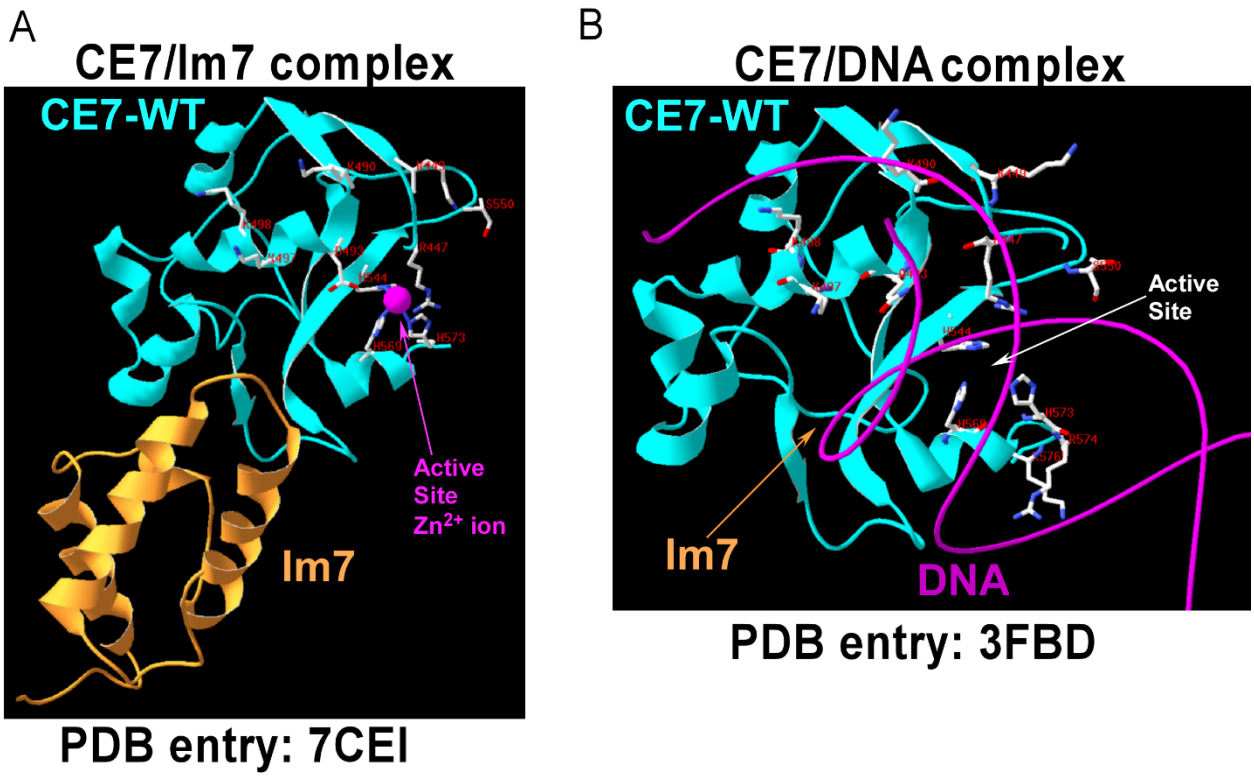
Columns: PA – protein amount; His – His-Trap; HEP – heparin; IEQ/IES – anion or cation exchange; GF- gel filtration; MBP - maltose; PC – phospho-cellulose; WG16 – 8WG16 anti-body Sepharose; DNC – DNA cellulose; Strep – streptavidin; GST – glutathione.

Non-chromatographic steps: PEI/AS – polyethyleneimine or ammonium sulfate precipitations.

Tag Cleavage: PSC – PreScission protease; TEV – TEV protease; THR – thrombin; SMP- SUMO protease. We consider protease digestion as a purification step since in many cases it is a time consuming process (16 -24 hrs).

* - Purification time was estimated based on our own experience with similar protocols assuming, in particular, that an over-night dialysis is normally used between the different chromatographic steps.

** - an average number of the total (including PEI, AS, protease cleavage, etc.) purification steps. An average number of chromatographic steps is shown in brackets.



C

```
CE7-WT   445   SKRNKPGKATGKGKPVNNKWLNNAGKDLGSPVPDRIANKLRDKEFKSFQDFRKKFWEEVS   504
CL7        1   SKSNEPGKATGEGKPVNNKWLNNAGKDLGSPVPDRIANKLRDKEFESFDDFRETFWEEVS    60

CE7-WT   505   KDPELSKQFSRNNNDRMKVGKAPKTRTQDVSGKRTSFELHHEKPISQNGGVYDMDNISVV   564
CL7       61   KDPELSKQFSRNNNDRMKVGKAPKTRTQDVSGKRTSFELNHQKPIEQNGGVYDMDNISVV   120

CE7-WT   565   TPKRHIDIHRGK   576
CL7      121   TPKRNIDIE-G-   132
```

Fig. S7. Structure-based engineering of the CL7-tag, inactive variant of the Colicin E7 DNAse domain (CE7). *(A, B)* The 3D structures of the CE7/Im7 (8) *(A)* and CE7/DNA (9) *(B)* complexes. The common, CE7 domains are shown in the similar orientations. The side chains in the active site histidine residues and DNA-binding residues are depicted. These latter residues were mutated *(C)* to produce the CL7 variant lacking the DNA-binding and catalytic activities. These mutated residues are located remotely to and do not interfere with, the Im7 binding site on CE7. *(C)* Sequence alignment of the wild type CE7 (CE7-WT) and engineered, inactive CL7 proteins. The mutated active site and DNA-binding residues are emphaszied in red and cyan, respectively.

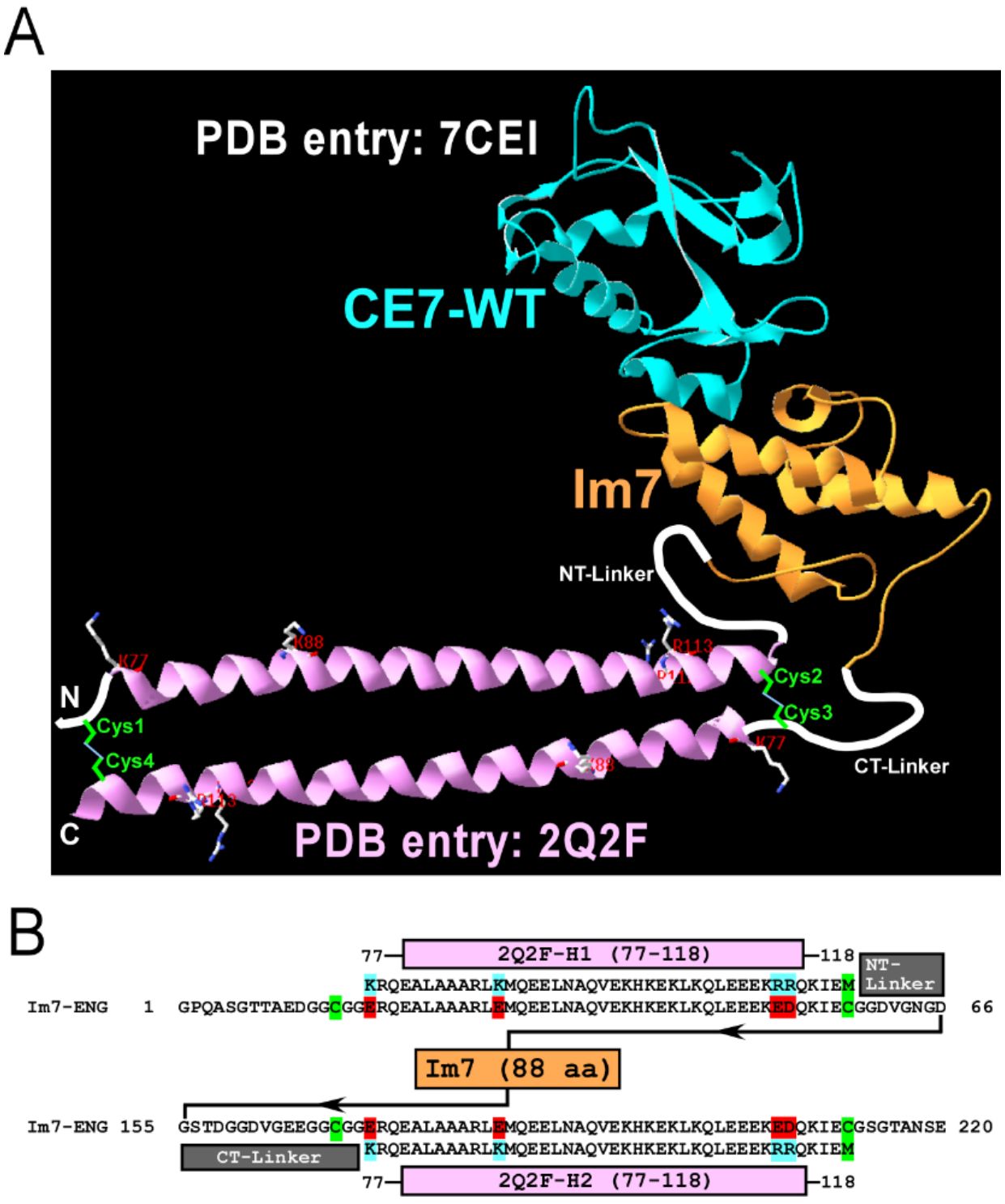


Fig. S8. Design of the Im7-immobilization unit. *(A)* The two α-helices (H1 and H2) of the homodimeric coiled-coil (a fragment of human selenoprotein S, PDB ID 2Q2F) are fused through short N- and C-terminal

linkers to the Im7 protein. The CE7 protein complexed with Im7 is shown for reference to illustrate that the coiled-coil domain fused to Im7 unlikely interferes with the CL7/Im7 binding. The four Cys residues are introduced in the non-functional (coiled-coil/linker) domain for efficient coupling to the iodo-acetyl (Sulfo-Link) beads. Cys2 and Cys4 are located in the structured (α-helical) coiled-coil domain, whereas Cys1 and Cys3 are in the linkers, immediately N-terminal to the respective α-helices of the coiled-coil domain and in close proximity to Cys2 and Cys4 located in the opposite coiled-coil helices. In this design, we assumed that Cys1-Cys4 and Cys2-Cys3 would likely form the intra-molecular disulfide bonds that, in turn, would prevent protein aggregation, which might occur in solution through the inter-molecular cysteine bridges. Indeed, we could concentrate the purified Im7-unit up to ~40-50 mg/ml in the absence of the reducing agents and observed no signs of aggregation. These likely intra-molecular cysteine cross-links are reduced right before immobilization reaction by addition of sulfhydryl-free reducing agent, TCEP (which does not react with the iodo-acetyl groups) to open the Cys side chains for coupling to the Sulfo-Link beads. The four cysteine residues also allow for expedient immobilization (~15-20 min) thereby minimizing potential non-specific coupling through other protein side chains (His, Tyr), the probability of which increases with prolonged incubation. In addition, this relatively bulky non-functional coiled-coil domain may provide some spacing between the active Im7-units on the beads, thereby minimizing their potential steric interference, which might affect the Im7 binding performance (accessibility) to the CL7-tags on the target proteins in the lysates. *(B)* Sequence of the engineered part of the Im7 immobilization unit (Im7-ENG) and alignment of its coiled-coil domain with the wild type (2Q2F) H1 and H2 helices. The basic residues (cyan) with the exposed side chains *(A)* were mutated to the acidic ones (red) to reduce positive charge of the wild type protein. The helices depicted were of the wild type sequences, not the engineered Im7.

A

```
(β′ω)-LNK 1 GNGDQGLEVLFQGPQGENSGTSGTDNGSSDGLEVLFQGPAGNSASSG 47
                    PSC4                          PSC5
```

B

E1

```
MBP-WT    1 MGKIEEGKLVIWINGDKGYNGLAEVG              26
MBP-ENG   1 MGKIEEGKQETWNNGDKGYNGRAEDGLEVLFQGPGTSG 38
                                        PSC1
```

E2

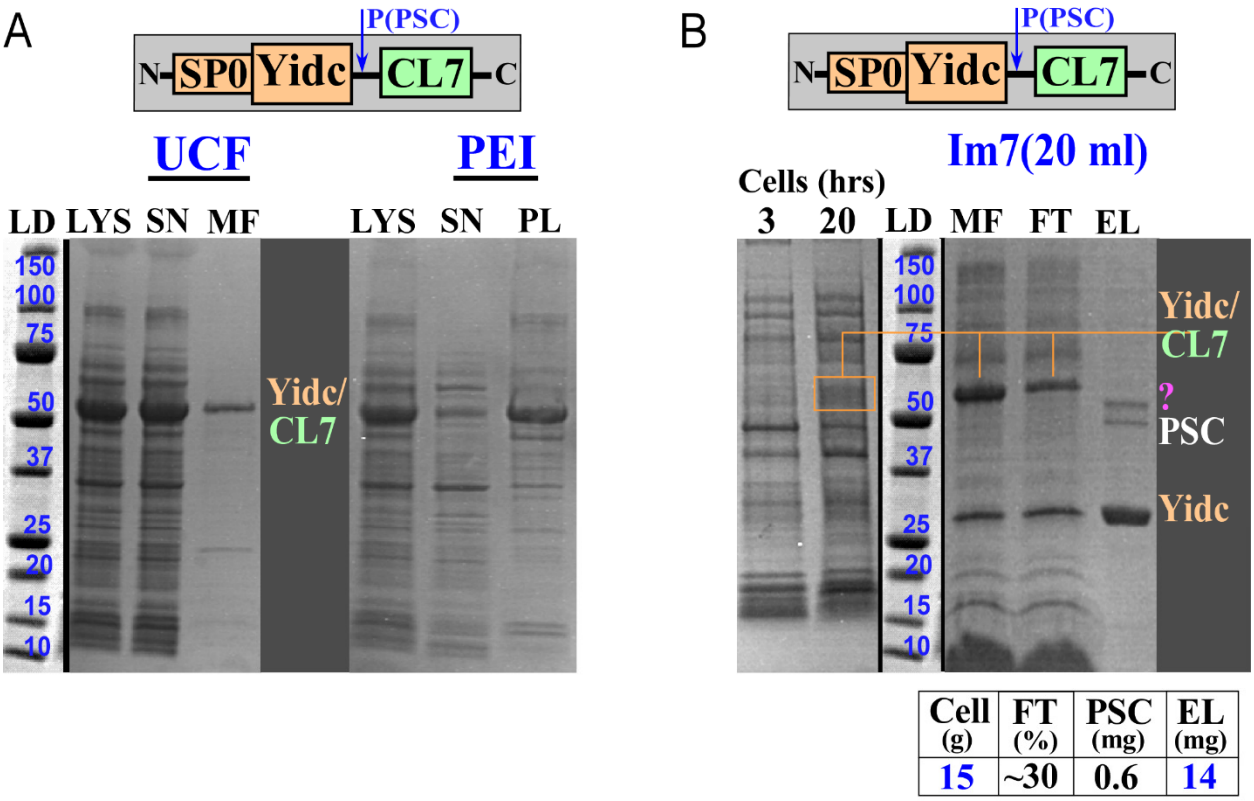
```
Trx-WT    1 MSDKIIHLTDDSFDTDVLKADGAILVDFWAE                31
Trx-ENG   1 MSDKINHQTDDSYDTDVDKADGANQVDYSAELEVLFQGPGQASG  44
                                          PSC2
```

E3

```
NusA-WT   1 MGNKEILAVVEAVSNEKALPREKIFEAL                  28
NusA-ENG  1 MGNKEILGEDEAESNEKAQAGEKNYEALGTLEVLFQGPGDSASG  44
                                      PSC3
```

Fig. S9. Amino acid sequences of the engineered (β′ω)-linker and expression tags (EX) used in the designed multi-subunit vector. *(A)* Flexible (47 residues–long) linker [(β′ω)-LNK] designed to fuse the C-termini of the β′-subunit with the N-termini of the ω-subunit. The two PSC binding sites are incorporated in vicinity to the C- and N-termini of the β′ and ω-subunits, respectively, to excise (rather than cleave) this artificial long linker. The hydrophobic residues are not included in a linker sequence (with exception of the PSC binding sites) to avoid potential interference of the linker residues with folding of the β′ and ω-subunits. *(B)* For expression tags (E1/E2/E3), we used the short (~30 amino acids) N-terminal peptides from three proteins (MBP, thioredoxin and NusA), modified to avoid hydrophobic residues. Each of these full-length proteins is known to improve expression when fused to the N-terminus of a target protein (10). We hypothesized that the N-terminal sequences of these proteins are likely to optimize translation initiation similarly. In addition, all tags (“expression” and purification) can be removed by the same protease in a single chromatographic step. The short N-terminal peptide impurities can be then eliminated through a simple dialysis step. Mutations (ENG), which were introduced in the wild type (WT) sequences of the respective proteins are emphasized (yellow). The major rationale to make these mutations was to remove hydrophobic residues and, at the same time, to avoid significant alterations of the original (WT) nucleic acid sequences that might be crucial for efficient translation initiation. The PreScission protease cleavage sites (PSC, cyan) are numbered according to their locations in the multi-subunit MV2 vector

(beginning from the 5'-end; **Fig. 3**). MBP – maltose binding protein (cytoplasmic variant, with no signal peptide); Trx – thioredoxin; NusA – transcription elongation factor, NusA.



Cell (g)	FT (%)	PSC (mg)	EL (mg)
15	~30	0.6	14

C

Protein	Expression		Purification			PA	Price (Company)
	ORG	System	Steps	Protocol	PT (days)	6g cells (mg)	($/mg)
hmCNX	hm	HOST	?	FLAG->?	?	?	~20,000 (OriGene)
hmCNX	ec	O-EXP	2	PEI->Im7	1	~20.0	1 Run: ~$400,000*

Fig. S10. Purification of the membrane, YidC and CNX proteins. *(A)* Expression (with IPTG induction) of YidC. The gene sequence was adjusted for *E. coli* codons and tagged at C-terminus with CL7. YidC has a WT signal peptide (SP0) sequence. A Cys residue remains at the N-terminus after cleavage by a signal peptidase *in vivo*. Upon enhanced expression, YidC and the other membrane proteins remained in the soluble fraction (SN) after ultra-centrifugation (UCF; left panel) with only a small amount of protein in the insoluble membrane fraction (MF). The protein, however, precipitated with DNA upon polyethyleneimine (PEI) precipitation in ~0.3-0.35 M NaCl (right panel) with no protein in the soluble, supernatant (SN) fraction. The YidC protein could be then released from the PEI pellet with the buffer containing 0.6 M NaCl and 1.5% dodecyl-maltopyranoside, DDM (PL; right panel). *(B)* Purification of YidC from non-induced cells using ultra-centrifugation as a first step. *(C)* Comparison of purification of the CL7/Im7 and available commercial (Origene) protocols of purification of the full-length CNX protein. MF – membrane fraction (after ultra-centrifuge pellet dissolved in 1.5% DDM); Cells (3/20 hrs) - uninduced (no IPTG) cells after

3/20 hrs culturing; LD – ladder (kDa); FT –flow-through; EL – eluate; PSC- PreScission protease; P(PSC) – PSC cleavage site; hm – human; M-step – multiple steps; Fc-tag - immunoglobulin Fc domain (Protein A Sepharose); FLAG – antibody ligand.

* – this value estimate is based on a single Im7 purification run, which yields ~20 mg CNX (**Fig. 6C**), and on the commercial price of the full-length hmCNX.

OriGene - http://www.origene.com/protein/TP300229/CANX.aspx