Chapter 6

Common Challenges in Studying the Structure and Function of Bacterial Proteins: Case Studies from *Helicobacter pylori*

Daniel A. Bonsor and Eric J. Sundberg

Abstract

Employing biophysical and structural methods is a powerful way to elucidate mechanisms of molecular recognition in bacterial pathogenesis. Such studies invariably depend on the production of pure, folded and stable proteins. Many proteins that can be expressed recombinantly ultimately fail to meet one or more of these criteria. The *cag* proteins from *Helicobacter pylori* form a secretion system that delivers the oncoprotein, CagA, into human gastric epithelial cells through an interaction between CagL and host cell integrins, where it can cause gastric adenocarcinoma. Expression of full length CagA and CagL is problematic as CagA undergoes rapid degradation during purification and CagL is thermally unstable. Here, we describe a method for the purification of CagA that results in the production of the full length protein through coexpression with its endogenous chaperone, CagF, and its subsequent separation from its chaperone. Furthermore, we detail the production of CagL and the use of differential scanning fluorimetry to identify how CagL is thermally stabilized by reduced pH, which led to a new crystal form of CagL and novel insight to pathogenic mechanisms. The methods described here for the production of stable *cag* proteins can be applied to a wide range of proteins involved in bacterial pathogenesis.

Key words Chaperones, Coexpression, Protein purification, Differential scanning fluorimetry

1 Introduction

Helicobacter pylori infection is widespread. Approximately half of the world's population is infected [1]. Most infected individuals will remain asymptomatic for life; however, around 20% will develop complications during their lifetimes including gastritis, peptic ulcers, lymphomas involving the mucosa-associated lymphoid tissue (MALT lymphoma) and gastric adenocarcinomas [2]. Infection by *H. pylori* is believed to occur during childhood through ingestion. Once in the stomach, *H. pylori* utilizes several virulence factors to achieve colonization of the boundary between the mucosa and surface of the gastric epithelial cells including urease to neutralize the stomach acid, flagella to burrow into the

Pontus Nordenfelt and Mattias Collin (eds.), *Bacterial Pathogenesis: Methods and Protocols*, Methods in Molecular Biology, vol. 1535, DOI 10.1007/978-1-4939-6673-8_6,

mucosa and away from the stomach acid, modified O-antigens to mimic host cell Lewis antigens and adhesins to anchor *H. pylori* onto epithelial cells [3]. Once attached to the host gastric epithelial cell, several virulence factors are secreted. Vacuolating cytotoxin A (VacA) binds to the host cells where it is internalized and causes the formation of large vesicles and in certain cases cell death [4]. More importantly for gastritis, ulcers, lymphomas and adenocarcinomas, cytotoxic associated gene A (CagA), one of the few known oncogenic bacterial proteins, is injected into the host cell cytoplasm though a Type IV secretion system (T4SS) [5].

CagA is a large protein with a variable molecular weight ranging from 120 to 150 kDa due to different strains of *H. pylori* producing CagA with a diverse number and type of tyrosine phosphorylation motifs (TPMs) near the C-terminus [5, 6]. The number of TPMs present in CagA from a particular *H. pylori* strain correlates with gastric cancer risk [7]. CagA is a five domain protein that includes a stable and structured N-terminal 100 kDa region composed of three distinct domains [8–10], an intrinsically disordered domain containing a variable number of TPMs [8] and a C-terminal domain containing a signal peptide that is recognized by the T4SS [11]. CagA translocation through the T4SS is dependent on the recognition of the C-terminal signal peptide of CagA and, prior to that, its interaction with its chaperone protein CagF inside the bacterial cell [11, 12]. CagF is a 32 kDa protein, which contains a putative coiled-coil domain and dimerizes with an association constant of approximately 200 μM [12, 13]. It has been shown using peptide array assays and isothermal titration calorimetry measurements that monomeric CagF contacts all five domains of CagA [13]. The production of milligram quantities of full length CagA for biophysical and structural work has been problematic due to the rapid proteolytic degradation of CagA during expression and purification with the loss of the C-terminus resulting in a protein with a molecular weight of 100 kDa [14], consisting of only the N-terminal structured domains. A *cagA* library, from which ~18,000 expression constructs were screened, identified several constructs corresponding to proteins of approximately 100 kDa in molecular weight that resulted in milligram-level expression of purified CagA [9, 15] and eventually led to the high-resolution X-ray crystal structures of CagA [8, 9]. However, no construct greater than 110 kDa in molecular weight was identified in this screen [15]. As the extreme C-terminus contains both the TPMs, which are important for CagA-host protein interaction, and the secretion peptide that is recognized by the T4SS, production of full length CagA for both biophysical and structural characterization is essential for fully elucidating the function of this oncoprotein.

Upon CagA binding of CagF and recognition by the T4SS, CagF is removed and CagA shuttled into a pore than spans the

periplasm to the pilus [12]. The pilus is comprised of CagC with a diameter of ~70 nm with the surface decorated with at least three other *cag* proteins—CagH, CagI, and CagL [16, 17]. Each of these proteins are believed to attach to the pilus through a highly similar C-terminal hexapeptide sequence ([S/T]-K-[I/V]-I-V-K) [17]. Deletion of CagH, CagI, or CagL results in *H. pylori* that produce either no or stunted pili [17]. Integrins, which are presented on the surface of gastric epithelial cells, were identified as a receptor involved in the secretion of CagA. The $\alpha_5\beta_1$ integrin was initially found to be responsible for CagA translocation into host gastric epithelial cells, although further investigations have indicated that $\alpha_V\beta_3$, $\alpha_V\beta_5$, and $\alpha_V\beta_6$ integrins also permit CagA translocation [17–21]. Most integrins recognize ligands containing an arginine–glycine–aspartic acid (RGD) motif [22]. Sequence analysis of the *cag* pathogenicity island revealed a single RGD motif located within CagL and mutation of this RGD motif prevented CagL binding of integrins and the subsequent translocation of CagA into host cells [18]. Although CagL triggers the translocation of CagA across the host cell plasma membrane, the mechanism by which this is achieved remains unknown. Several X-ray crystal structures of CagL have been solved, revealing that the RGD motif is located in an α-helix, which is unique—all other proteins for which structures have been determined contain their RGD motifs in loop regions [23–25]. Furthermore, each CagL structure revealed both minor and gross conformational changes, such as the sliding of a neighboring α-helix relative to the α-helix in which the RGD motif resides, as well as a helix–turn–helix motif that rearranges to form a single elongated helix [24, 25]. These conformational changes have been shown to occur in response to a change in pH [25]. The relative sliding of neighboring α-helices has been proposed to bury the RGD motif in order to prevent CagL from binding to integrins from shedding cells that have been discarded in the stomach where the pH is acidic [25]. By regulating the CagL–integrin interaction through a pH-induced mechanism, the *H. pylori* T4SS could engage host gastric epithelial cells along the stomach lining where the pH is more accommodating [25].

This chapter describes the production of full length CagA by coexpression with the effector protein CagF, mimicking that which occurs in *H. pylori* cells. We describe the purification of the CagA–CagF complex and the removal of CagF through partial denaturation to yield highly purified milligram quantities of full length CagA. We also describe the refolding and purification of CagL and its crystallization using a differential scanning fluorimetry assay to identify a buffer system that increased the melting temperature (T_m) and led to a new crystal form and the discovery of the above mentioned conformational changes. These protocols are adaptable to other proteins and could aid in their production, purification and stabilization in order to make them more amendable to biochemical, biophysical and structural studies.

2 Materials

2.1 Bacterial Strains and Media

All strains are grown at 37 °C (unless otherwise stated) either on LB agar plates or in LB broth, supplemented with ampicillin and/or kanamycin at their stated concentrations below.

1. *E. coli* strains: *TOP10* (F- *mcrA* Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ *lac*X74 *rec*A1 *ara*D139 Δ(*araleu*)7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G) and *BL21(DE3)* (F^- *ompT* *hsd*S_B(r_B^- m_B^-) *gal dcm* (DE3)).
2. Media: LB Broth is prepared by dissolving 10 g tryptone, 10 g NaCl, and 5 g yeast extract in a total volume of 1 L of deionized H_2O and sterilizing by autoclaving at 121 °C for 20 min. LB agar is prepared by adding 15 g of agar in 1 L of LB Broth before sterilization.
3. Ampicillin stock: 100 mg/mL ampicillin sodium salt in H_2O, filter-sterilized with 0.22 μm pore-size syringe filter and store at −20 °C. Use at a final concentration of either 100 μg/mL for single plasmids or 25 μg/mL for two plasmids.
4. Kanamycin stock: 30 mg/mL kanamycin sulfate in H_2O, filter-sterilized with 0.22 μm pore-size syringe filter and store at −20 °C. Use at a final concentration of 15 μg/mL.
5. Isopropylthiogalactoside stock (IPTG): IPTG is dissolved in deionized H_2O to a final concentration of 1 M, filter-sterilized with 0.22 μm pore-size syringe filter and stored at −20 °C.

2.2 Plasmids, Plasmid Construction, and Cloning

1. TBE buffer: Dissolve 10.8 g Tris base, 5.5 g boric acid in 900 mL H_2O, add 20 mL 0.5 M EDTA, pH 8.0 and adjust the pH to 8.4. Dilute to a final volume of 1 L.
2. Ethidium bromide stock: 10 mg/mL in H_2O kept at room temperature.
3. pRSFDuet™-1 (Novagen) is modified to remove the C-terminal S-tag in the second multiple cloning site and is replaced with a decahistidine (10×His) tag. This vector is subsequently used for the expression of CagA with an N-terminal hexahistidine (6×His) and a C-terminal 10×His.
4. pET21-d (Novagen) is used for expression of CagL with a C-terminal 6×His.
5. pGEX-5×-2 (GE Healthcare Life Sciences) is used for expression of CagF with an N-terminal GST tag.
6. DNA modifying enzymes: *BamHI*, *NcoI*, *PacI*, *SacI*, and *XhoI* restriction enzymes, Pfu Turbo, T4 Polynucleotide kinase, and Quick T4 DNA ligase with their respective buffers are commercially available.

7. Low-melt agarose gels (1% w/v): 0.5 g of low-melt agarose is added to 25 mL of TBE. The solution is heated in a microwave oven with occasional swirling until the agarose is dissolved. A further 25 mL of TBE is added followed by 2 μl of Ethidium Bromide and the gel is poured into a sealed gel mold with comb and left to set. The gel is placed in an electrophoresis tank and covered with TBE. The samples are loaded and the gel is run at a constant voltage of 100 V for 30 min.

2.3 Protein Expression and Purification

1. GST Binding Buffer (GSTBB): 1 mM KH_2PO_4, 10 mM Na_2HPO_4, 137 mM NaCl, 2.7 mM KCl, pH 7.4.
2. GST Elution Buffer (GSTEB): 20 mM Tris–HCl pH 7.5, 50 mM NaCl, 10 mM reduced glutathione.
3. Glutathione agarose (e.g., GE Healthcare Life Sciences).
4. Nickel Binding Buffer (NiBB): 20 mM Tris–HCl pH 7.5, 500 mM NaCl.
5. Nickel Elution Buffer (NiEB): 20 mM Tris–HCl pH 7.5, 500 mM NaCl, 400 mM imidazole.
6. Ni-NTA agarose (e.g., Thermo Scientific).
7. Partial Denaturation Wash (PDW): 20 mM Tris–HCl pH 7.5, 2.0 M urea.
8. Inclusion Bodies Lysis Buffer: 50 mM KH_2PO_4–K_2HPO_4 pH 7.5, 200 mM NaCl.
9. Inclusion Bodies Wash Buffer: 10 mM Tris–HCl, 1 M NaCl, pH 8.0.
10. Denaturation Buffer (DB): 50 mM KH_2PO_4–K_2HPO_4 pH 7.5, 200 mM NaCl, 6.0 M guanidine hydrochloride.
11. Refolding Buffer (RB): 50 mM Tris–HCl pH 8.3, 20 mM NaCl, 0.1 mM KCl, 1 mM EDTA, 2 mM reduced glutathione, 0.2 mM oxidized glutathione.
12. Gel Filtration Buffer (GFB): 20 mM Tris–HCl pH 7.5, 200 mM NaCl, 1 mM EDTA.

2.4 Differential Scanning Fluorimetry

1. SYPRO Orange 5000x Stock (Sigma).
2. 96-well White TempPlate with semi-skirt (USA Scientific).
3. Buffer Screen: A 24-well screen was devised to screen a pH range from 4.0 to 9.5 in intervals of 0.5 pH units according to Table 1 and aliquoted to positions A1-B12 of a 96-deep well block. The 24 conditions are repeated for positions C1-D12 with the addition of 500 mM NaCl. All 48 conditions are then duplicated for positions E1-H12.
4. Optical Sealing Tape.
5. iQ5 Multicolor Real Time PCR Detection System (Bio-Rad) or similar system.

Table 1
Composition of the 96 differential scanning fluorimetry screen solutions

Well	Buffer
A1/E1 (C1/G1)	Water (500 mM NaCl)
A2/E2 (C2/G2)	200 mM sodium acetate-HCl pH 4.0 (500 mM NaCl)
A3/E3 (C3/G3)	200 mM sodium citrate-HCl pH 4.0 (500 mM NaCl)
A4/E4 (C4/G4)	200 mM sodium acetate-HCl pH 4.5 (500 mM NaCl)
A5/E5 (C5/G5)	200 mM sodium citrate-HCl pH 5.0 (500 mM NaCl)
A6/E6 (C6/G6)	200 mM sodium acetate-HCl pH 5.0 (500 mM NaCl)
A7/E7 (C7/G7)	200 mM sodium citrate-HCl pH 5.5 (500 mM NaCl)
A8/E8 (C8/G8)	200 mM Bis-Tris-HCl pH 5.5 (500 mM NaCl)
A9/E9 (C9/G9)	200 mM K/Na phosphate pH 6.0 (500 mM NaCl)
A10/E10 (C10/G10)	200 mM Bis-Tris-HCl pH 6.0 (500 mM NaCl)
A11/E11 (C11/G11)	200 mM MES-NaOH pH 6.5 (500 mM NaCl)
A12/E12 (C12/G12)	200 mM sodium cacodylate-HCL pH 6.5 (500 mM NaCl)
B1/F1 (D1/H1)	200 mM Bis-Tris-HCL pH 6.5 (500 mM NaCl)
B2/F2 (D2/H2)	200 mM Bis-Tris propane-HCl pH 6.5 (500 mM NaCl)
B3/F3 (D3/H3)	200 mM K/Na phosphate pH 7.0 (500 mM NaCl)
B4/F4 (D4/H4)	200 mM Hepes-NaOH pH 7.0 (500 mM NaCl)
B5/F5 (D5/H5)	200 mM Hepes-NaOH pH 7.5 (500 mM NaCl)
B6/F6 (D6/H6)	200 mM Tris-HCl pH 7.5 (500 mM NaCl)
B7/F7 (D7/H7)	200 mM K/Na phosphate pH 8.0 (500 mM NaCl)
B8/F8 (D8/H8)	200 mM Hepes-NaOH pH 8.0 (500 mM NaCl)
B9/F9 (D9/H9)	200 mM Tris-HCl pH 8.0 (500 mM NaCl)
B10/F10 (D10/H10)	200 mM Tris-HCl pH 8.5 (500 mM NaCl)
B11/F11 (D11/H11)	200 mM Tris-HCl pH 9.0 (500 mM NaCl)
B12/F12 (D12/H12)	200 mM CHES-HCl pH 9.5 (500 mM NaCl)

The concentrations given are the stock solutions which are mixed 1:1 with the protein–SYPRO Orange mixture

3 Methods

The production of proteins that are pure, stable, and crystallizable can be problematic due to contamination and degradation of the protein sample, as well as the low success rate of protein crystallization. Below, we describe the production of the unstable full length

protein CagA through coexpression with CagF and stabilization of CagL by identifying buffer conditions that increase its melting temperature leading to the production of a new crystal form.

3.1 Protein Stabilization by Coexpression

Unstable proteins that display degradation during purification are common. Typically, alternative constructs are designed as the separation of degradation products from full length proteins is problematic and often produces low protein yields. CagA exhibits substantial degradation during expression, resulting in an approximately 100 kDa product [14, 15]. Expression of the target protein with known binding partners has often been used where proteins fail to express or are found in inclusion bodies. Other forms of coexpression involve the expression of chaperones (e.g., GroES-GroEL) or foldases (e.g., peptidyl prolyl *cis/trans* isomerases or disulfide isomerases) [26–29]. We have found that CagA coexpression with CagF, the binding partner and chaperone of CagA, suppresses CagA degradation during expression and purification [13]. However, a major problem with coexpression of known binding partners for the target protein is the separation of the protein of interest and the coexpressed proteins. We have modified the commercial pRSFDuet-1 vector (Fig. 1a) that contains two multiple cloning sites, one with

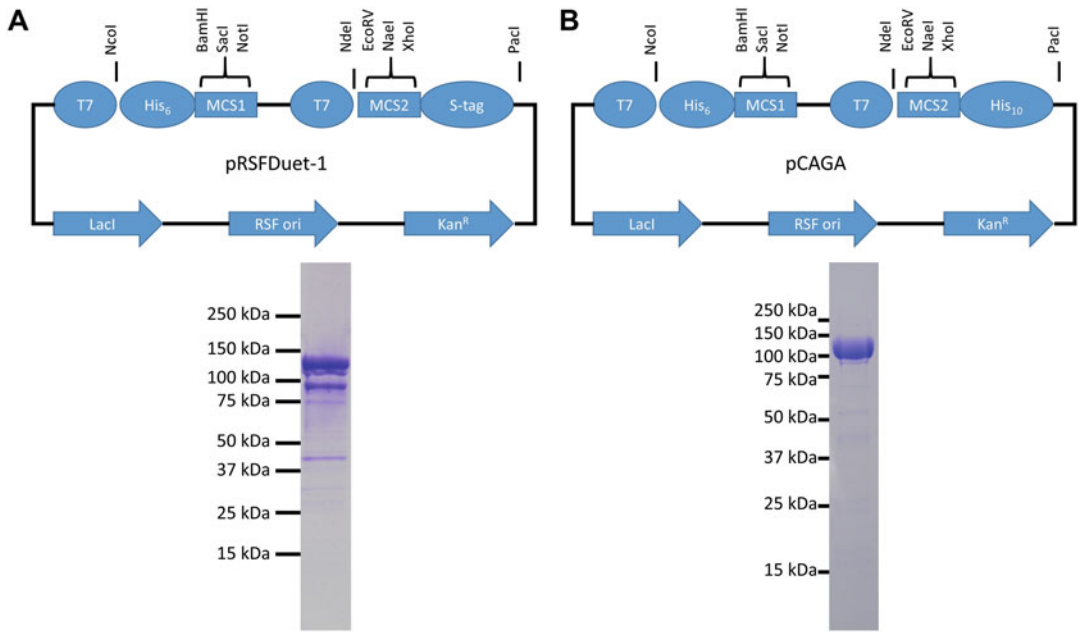

Fig. 1 Construction of a CagA expression vector and its purification. (**a**) Vector map of pRSFDuet. CagA was originally ligated into MCS1. (**b**) Purification of CagA from 6 L growth of BL21(DE3) coexpressing GST-CagF using pRSFDuet shows excessive degradation and low yields. (**c**) Removal of S-tag from MCS2 generated pCAGA. CagA was ligated into a SacI/XhoI cut pCAGA. (**d**) Expression and purification of CagA from 2 L of BL21(DE3) coexpressing GST-CagF using pCAGA show protein of high purity

an N-terminal 6×His tag and the other with a C-terminal S-tag, and replaced the S-tag with a 10×His tag (Fig. 1b). CagA is cloned through both cloning sites such that it contains both an N-terminal 6×His tag and a C-terminal 10×His tag. CagF is expressed as a GST fusion from the pGEX-5×-2 vector. Coexpression of both proteins in BL21(DE3) is permitted as the pRSFDuet and pGEX-5×-2 vectors have different origins of replication (RSF and BR322 origins, respectively) and antibiotic resistances (kanamycin and ampicillin, respectively), thereby preventing plasmid instability. As CagA degradation occurs close to the C-terminus [14], washing of the nickel affinity chromatography column with 150–200 mM imidazole removes the N-terminal 100 kDa degradation product, while the full length protein (~130 kDa) and the C-terminal degradation product (~30 kDa) is retained until a further more stringent elution step with 300–400 mM imidazole. The C-terminal degradation product is subsequently separated from full length CagA by size exclusion chromatography.

3.1.1 Preparation of the Modified pRSFDuet Vector (pCAGA)

1. The primers RSFfp (5′-TCGAGCATCACCACCATCAT CACCACCATCACCATTAAAT-3′) and RSFrp (5′-TTAAT G G T G A T G G T G G T G A T G A T G G T G G T G A TGC-3′) are diluted with deionized H_2O to a concentration of 25 μM) and are each phosphorylated using a reaction mix containing: 12 μl primer, 32 μl deionized H_2O, 5 μl 10× T4 Polynucleotide kinase buffer (NEB), and 1 μl T4 Polynucleotide kinase. The reaction is carried out for 30 min at 37 °C, the phosphorylated primers are then mixed and heated to 95 °C for 5 min by water bath or heating block. The power source is removed and the primers are cooled to room temperature gently (*see* **Note 1**).
2. 2 μg of pRSFDuet is double digested with *XhoI/PacI* for 2 h at 37 °C. The restriction enzyme digestion products are analyzed on a 1% w/v low melt agarose gel. The cut vector is visualized with UV-light and carefully excised using a clean razor blade and placed in a microfuge tube. The cut vector is purified using a commercial agarose gel purification kit (e.g., QIAgen).
3. The phosphorylated annealed primers are ligated into the gel purified *XhoI/PacI* pRSFDuet vector. Specifically, 1 μl of purified cut vector is mixed with 1 μl of annealed primers, 2 μl deionized H_2O, 5 μl 2×Quick Ligation Reaction Buffer, and 1 μl Quick T4 DNA Ligase are incubated at RT for 5 min before subsequent transformation into chemically competent *E. coli* Top10, plated on LB agar plates supplemented with kanamycin (30 μg/mL) and incubated overnight at 37 °C (*see* **Note 2**).
4. Single colonies are picked and grown overnight at 37 °C in 3–5 mL LB containing kanamycin (30 μg/mL) and plasmid DNA minipreps are conducted using a commercial kit (e.g., QIAgen).

5. Conformation of insert is verified by DNA sequencing using either the commercial available universal T7 Terminator primer or the ACYCDuetUP1 primer (Novagen, 5′-GGAT CTCGACGCTCTCCCT-3′).

3.1.2 Cloning and Ligation of CagA and CagF for Expression

1. Genomic DNA of the *H. pylori* strain 11637 was purchased from ATCC and used as a template for gene amplification of full length CagA and CagF using the following primers: CagAFP-SacI (5′-GCGCGCCTCGAGAGATTTTTGGAAAC CACCTTTTGTATTAACA-TTTTTG-3′), CagARP-XhoI (5′-GCGCGCGAGCTCGATGACTAACGAAACT ATTGACCAACAACCAC-3′), CagFFP-BamHI (5′-GCG G A T C C C G G A A A A C T T G T A T T T C C A G GGCATGAAACAAAATTTGCGTGAACAAAAATT-3′) and CagFRP-XhoI (5′-GCGCGCCTCGAGTCAATCGTTAC TTTTGTTTTGATTTTTTTGATCG-3′) (*see* **Note 3**). PCR reactions are carried out in a final volume of 50 μL, 10 ng of template, 5 μL 10× Cloned Pfu DNA Polymerase Reaction Buffer, 1 μL 10 mM dNTP mixture (final concentration of 200 μM of each dNTP), 1 μL of each 5 μM primer (100 nM final concentration), and 1 μL of Pfu Turbo. After an initial denaturation of 2 min at 95 °C, the PCR reactions proceed for 30 cycles of 30 s of denaturation at 95 °C, 30 s of annealing at 60 °C, and 3 min 45 s for CagA and 50 s for CagF extensions at 72 °C followed by 6 min at 72 °C.
2. Amplification of *cag* genes are analyzed on a 1 % w/v low melt agarose gel and confirmed by UV light. PCR products are purified (using a commercial kit, e.g., QIAgen) and digested with *SacI/XhoI* (CagA) or *BamHI/XhoI* (CagF) restriction enzymes for 2 h at 37 °C before purification.
3. The vectors pCAGA and pGEX-5x-2 are digested with *SacI/XhoI* and *BamHI/XhoI*, respectively at 37 °C for 2 h. The restriction enzyme digestion products are analyzed on a 1 % w/v low melt agarose gel, visualized with UV-light and extracted using a clean razor blade, placed in a microfuge tube and purified from the gel using a commercial agarose gel purification kit.
4. The digested inserts are ligated into their respective digested vectors. Specifically, 1 μl of purified digested vector is mixed with 1 μl of digested insert, 2 μl deionized H_2O, 5 μl 2×Quick Ligation Reaction Buffer, and 1 μl Quick T4 DNA Ligase and incubated at RT for 5 min before transformation into chemically competent *E. coli* Top10. Cells are plated on LB agar plates supplemented with either kanamycin (30 μg/mL) for the pCAGA-CagA ligation or ampicillin (100 μg/mL) for the pGEX-5×-2-CagF vector, and are incubated overnight at 37 °C.

5. Single colonies are plucked to inoculate 3–5 mL of LB containing either kanamycin (30 μg/mL) or ampicillin (100 μg/mL) and grown overnight at 37 °C.
6. DNA plasmid minipreps are performed and insertion of clones is confirmed by DNA sequencing. Specifically, CagA is sequenced using the T7 Terminator primer, the ACYCDuetUP1 primer (Novagen, 5′-GGATCTCGACGCTCTCCCT-3′) and three internal primers to completely cover the sequence (11637Int1 5′-CCGCCTGAATCTAGGGATTTGCTTG ATG-3′, 11637Int2 5′-GTCCTGATAAGGGTGTAG GCGTTACAAATG-3′, and 11637Int3 5′-GCGACC T T G A A A A T T C C G T T A A A G A T G T G A T C A TC-3′). CagF is sequenced using the commercial 5GEX and 3GEX universal primers.

3.1.3 Coexpression and Purification of Full Length CagA

1. 40 μg of pCAGA-CagA and pGEX-5×-2-CagF are combined into a sterile microfuge tube and is transformed into chemical competent *E. coli* BL21(DE3) (*see* **Note 4**). Cells are plated onto LB agar plates supplemented with both kanamycin and ampicillin (15 and 25 μg/mL, respectively) and incubated overnight at 37 °C (*see* **Note 5**). In a separate sterile microfuge tube, 40 μg of pGEX-5×-2-CagF is transformed into *E. coli* BL21 (DE3), plated onto LB agar plates containing ampicillin (100 μg/mL) and incubated overnight.
2. A single colony of each BL21 (DE3) transformant is used to inoculate 100 mL of LB broth containing appropriate concentrations of antibiotics and grown overnight on an orbital shaker at 37 °C.
3. A 1:100 dilution of BL21(DE3) + pCAGA-CagA/pGEX-5×-2-CagF is used to inoculate typically 4 × 1 L of LB broth containing ampicillin and kanamycin. A 1:100 dilution of BL21(DE3) + pGEX-5×-2-CagF is used to inoculate 0.5 L of LB containing ampicillin. Cells are grown at 37 °C until an A_{600nm} of ~ 0.4–0.5 is reached. The flasks are transferred to an orbital shaker set at 18 °C and left to grow for 20 min before induction with a final concentration of 1 mM IPTG. Cells are left to express the protein overnight.
4. Cells were harvested (5500*g*, 4 °C, 12 min) and the BL21(DE3) + pGEX-5×-2-CagF cells were resuspended in ~ 35 mL ice-cold GSTBB. They were disrupted by sonication on ice using a Branson Sonifier 450 with ½ in. stud probe and 60 × 0.7 s pulses at 70 W (0.3 s spacings between pulses). This cell extract is used to resuspend the BL21(DE3) + pCAGA-CagA/pGEX-5×-2-CagF cells and are diluted to ~ 70 mL with ice-cold GSTBB (*see* **Note 6**). The mixture is disrupted further by sonication (2 × 60 × 0.7 s at 70 W with 0.3 s spacings between pulses). Cell debris was removed by centrifugation (20,000 × *g*, 4 °C, 30 min) and the soluble cell extract was loaded onto

~5 mL of Glutathione agarose in a gravity column equilibrated with GSTBB. Unbound protein was washed out with 25 mL of GSTBB before elution with 20 mL of GSTEB.

5. The elution fraction is diluted 2-fold with NiBB and loaded at 2 mL/min onto a 5 mL Ni-NTA agarose column. The column is washed with 50 mL of NiBB before washing with 1.5 L of PDW at a rate of ~1–2 mL/min, which is typically left to run overnight at room temperature (*see* **Note 7**). The beads are washed with 50 mL of NiBB to remove the urea, 30 mL of 1:1 NiBB + NiEB (final imidazole concentration 200 mM) to remove degradation products and CagA is eluted with 15 mL of NiEB. The purified protein is concentrated to ~2–3 mL and can be used as is for biophysical characterization or further purified by size exclusion chromatography (*see* **Note 8**).

3.2 Protein Stabilization for Crystallization

Crystallization of purified protein for X-ray diffraction studies has a low success rate due to the vast number of variables that exist. Typically, after a target protein has either failed to crystallize or produces crystals with poor or no diffraction, several options are available that can be performed with a construct as is, prior to designing an altered construct for expression, purification and crystallization. These include, but are by no means limited to, matrix microseeding, limited proteolysis, lysine methylation, and molecular imprinted polymers which have all shown to have rescued "non-crystallizable" proteins [30–33]. An alternative method that we have had success with is to use differential scanning fluorimetry to identify buffers that stabilize the protein of interest, which may be unstable in the crystallization storage buffer (typically low buffer and salt concentrations) [25]. The differential scanning fluorimetry assay screens pH in the presence and absence of salt to identify a new crystallization storage buffer in which the protein is soluble, folded and exhibits an increased melting temperature (T_m) [34–36]. This is achieved by the addition of fluorescent dye, typically SYPRO Orange, which strongly fluoresces when the protein unfolds and exposes the hydrophobic core as the temperature increases. We present a method for the expression and purification of CagL, and the identification of a crystallization buffer through differential scanning fluorimetry that we used successful to crystallize CagL in a new crystal form.

3.2.1 Cloning and Ligation of CagL for Expression

1. Genomic DNA of the *H. pylori* strain 26695 was purchased from ATCC and used as a template for gene amplification of CagL residues 21-237 using the following primers: CagLFP-NcoI (5′-GCG-CGCCCATGGAAGATATAACAAGCGGTT TAAAGCAACTGG-3′) and CagLRP-XhoI (5′-GCGCGC CTCGAGTTTAA-CAATGATCTTACTTGATTGCCTTTCT TG-3′). The PRC reaction is carried out in a final volume of 50 μL, 10 ng of template, 5 μL 10× Cloned Pfu DNA

Polymerase Reaction Buffer, 1 μL 10 mM dNTP mixture (final concentration of 200 μM of each dNTP), 1 μL of each 5 μM primer (100 nM final concentration), and 1 μL of Pfu Turbo. After an initial denaturation of 2 min at 95 °C, the PCR reactions proceed for 30 cycles of 30 s of denaturation at 95 °C, 30 s of annealing at 60 °C, 40 s for CagL extensions at 72 °C, followed by 6 min at 72 °C.

2. Amplification of the CagL gene is analyzed on a 1% w/v low melt agarose gel and confirmed by UV light. The PCR product is purified (using a commercial kit, e.g., QIAgen) and digested with *NcoI* and *XhoI* restriction enzymes for 2 h at 37 °C before purification.
3. The pET21d vector is digested with *NcoI* and *XhoI* at 37 °C for 2 h. The restriction enzyme digestion products are analyzed on a 1% w/v low melt agarose gel, visualized with UV-light and extracted using a clean razor blade, placed in a microfuge tube and purified from the gel using a commercial agarose gel purification kit.
4. The CagL digested insert is ligated into the NcoI/XhoI digested pET21d vector using the method described above for CagA and CagF. The ligation reaction is transformed into chemically competent Top10 cells which are plated onto LB agar plates supplemented with ampicillin (100 μg/mL). Plates are incubated overnight at 37 °C.
5. Single colonies are plucked to inoculate 3–5 mL of LB containing ampicillin (100 μg/mL) and grown overnight at 37 °C.
6. DNA plasmid minipreps are performed and insertion of the clone is confirmed by DNA sequencing using the universal T7 Forward and T7 Termination primers.

3.2.2 Expression and Purification of CagL

1. 40 μg of pET21d-CagL is placed in a sterile microfuge tube and is transformed into chemical competent *E. coli* BL21 (DE3). Cells are plated out onto LB Agar containing ampicillin and are left to incubate overnight at 37 °C. A single colony is used to inoculate 100 mL of LB broth containing ampicillin and placed on an orbital shaker at 37 °C overnight.
2. A 1:100 dilution of the overnight is used to inoculate 2 × 1 L of LB containing ampicillin and left to shake on an orbital shaker until an A_{600nm} of ~ 0.6 is reached. Cells are induced with a final concentration of 1 mM IPTG and left to grow at 37 °C for 4 h. Cells are harvested by centrifugation (5500 × *g*, 12 min, 4 °C).
3. Cells are resuspended in ~35 mL of Inclusion Bodies Lysis Buffer and are lysed by sonication on ice using a Branson Sonifier 450 with ½ in. stud probe and 2 × 60 × 0.7 s pulses at 70 W (0.3 s spacings between pulses). Inclusion bodies are isolated by centrifugation (20,000 × *g*, 4 °C, 30 min). The supernatant is

discarded and the inclusion bodies are resuspended and broken up by pipette in Inclusion Bodies Wash Buffer. Inclusion bodies are isolated by centrifugation (20,000×*g*, 4 °C, 30 min), supernatant discarded and dissolved in ~5 mL of DB buffer (the inclusion bodies are broken up by pipette and the solution left at room temperature for 1 h). Undissolved cell debris is removed by microcentrifuge (20,000×*g*, 5 min, room temperature).

4. The dissolved inclusion body solution is refolded by manual injected into 400 mL of ice cold RB with stirring (kept at 4 °C) at a rate of ~100 μL/min. The solution is left stirring at 4 °C for 2.5 days before EDTA is quenched with an excess of $MgCl_2$ (final concentration 2 mM) (*see* **Note 9**). Precipitated protein is removed by centrifugation (10,000×*g*, 4 °C, 30 min) and the supernatant is loaded on a 5 mL bead volume of Ni-NTA agarose in a gravity column by a peristaltic pump. The beads are washed with 30 mL NiBB, 20 mL of 1:8 mixture of NiBB and NiEB (imidazole final concentration 50 mM), respectively and CagL is eluted with 10 mL of NiEB. CagL is dialyzed overnight at 4 °C against 2 L of gel filtration buffer. CagL is subjected to size exclusion chromatography on an S200 10/300 GL column (GE Healthcare) equilibrated with gel filtration buffer by 4–6×2mL injections. Folded CagL typically elutes around 16–17 mL. Protein purity is confirmed by SDS PAGE.

3.2.3 Differential Scanning Fluorimetry of CagL

1. CagL is diluted to 1 mg/mL with gel filtration buffer. In limited lighting, 1.375 mL of CagL is added to 2.75 μL of SYPRO Orange in a sterile microfuge tube. 12.5 μL of the CagL-SYPRO Orange mixture is added to each well of a 96-well White TempPlate. Carefully dispense 12.5 μL of the differential scanning fluorimetry solutions to the same plate, seal with Optical Sealing Tape and briefly centrifuge the plate at room temperature (500×*g*, 5 min) to remove any air bubble that may have occurred during the preparation of the plate (*see* **Note 10**).
2. Place the plate in an iQ5 Multicolor Real Time PCR Detection System or other appropriate Real Time PCR machine (e.g., ABi 7900). Run a program which scans the temperature from 25 to 95 °C, at 1 °C/min with a dwell time of 1 min at each temperature (*see* **Note 11**).
3. Export the Relative Fluorescence Units (RFUs) to an EXCEL spreadsheet. Download a copy of the "Transform" for the appropriated Real Time PCR machine and "DSF Analysis" EXCEL spreadsheet (ftp://ftp.sgc.ox.ac.uk/pub/biophysics/). Paste the raw RFUs from your experiment into the "RfU" tab of the Transform spreadsheet. Copy the output data from this and paste into the "Paste in transformed Data" tab of the DSF Analysis file. Visually inspect the graphs produced in the "All Graphs" tab for melting transitions (*see* **Note 12**). In the

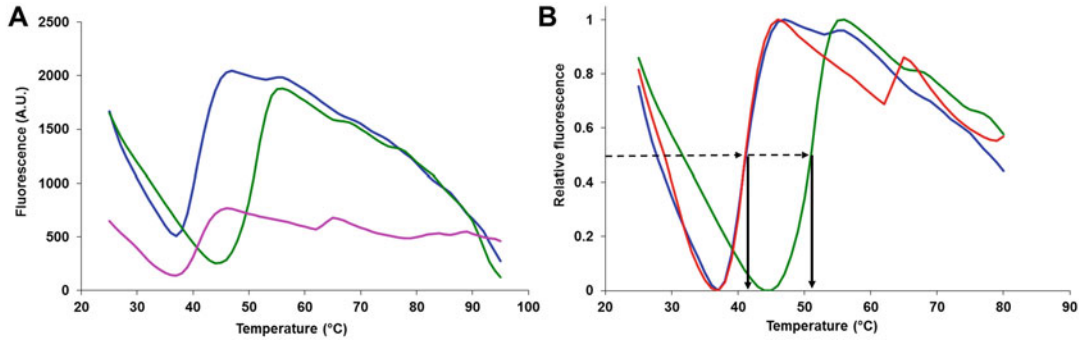

Fig. 2 Differential scanning fluorimetry assay of purified CagL. (**a**) Raw fluorescence of CagL-SYPRO Orange undergoing thermal melting in water (blue), 100 mM Tris, pH 8.0 (magenta) and 100 mM sodium acetate pH 4.0 (green). (**b**) Normalized thermal melt of the same condition. In water and 100 mM Tris, pH 8.0, a T_m of 41 °C is observed. Lowering of the pH (100 mM sodium acetate pH 4.0) causes the T_m to increase by 10 °C

"Custom Graphs" tab several graphs can be superimposed onto a single graph (Fig. 2a), whilst in the "Custom Normalized Graphs" tab, superimposed graphs can be normalized (Fig. 2b). Melting temperatures are calculated where the relative fluorescence is equal to 0.5. CagL should display melting temperatures of ~ 50 °C at low pH (Fig. 2b). As the pH is increased, CagL becomes thermally unstable and shifts to a melting temperature of ~ 40 °C (Fig. 2b).

4 Notes

1. The primers can potentially be purchased phosphorylated, annealed directly together and then ligated into the *XhoI/PacI* cut pRSFDuet vector.
2. Ligations can be carried out with T4 ligase, though reaction times are increased.
3. The forward primer for GST-CagF contains a Tobacco Etch Virus (TEV) protease site. We use this to cleave GST from CagF using TEV protease made in-house. pGEX-5×-2 contains a Factor X_A site, which is in reading frame with GST and CagF and can also be used to cleave CagF.
4. Coexpression of GST-CagF and CagA could be achieved by placing them in MCS1 and MCS2 of the modified pRSFDuet-1. However, expression in BL21(DE3) does not produce any protein. Replacing the T7 promoter controlling expression of GST-CagF with a *tac* promoter promotes GST-CagF expression, though no expression of CagA is observed. We only find expression of both CagA and CagF when they are expressed on different vectors. Expression of proteins either from the same

plasmid or on separate plasmids should be considered if no protein expression is observed for the method one chooses to follow.

5. Cells containing multiple plasmids are typically more sensitive to antibiotic selection and concentrations should be halved.
6. Lysing BL21(DE3)+pCAGA-CagA/pGEX-5×-2-CagF cells which have been suspended in a cell extract of lysed BL21(DE3)+pGEX-5×-2-CagF ensures an excess of GST-CagF to bind free CagA and suppress degradation.
7. We have tried several methods to remove CagF from CagA including lower concentrations of urea (1.0 and 1.5 M) and high concentrations of NaCl (3.0 M). However, CagF was not fully removed from CagA after 2 days of washing the Ni-NTA agarose.
8. Typically, we would use the protein that eluted from the Ni-NTA agarose in isothermal titration calorimetry after dialysis and not perform size exclusion chromatography. This is due without CagF present; CagA starts to degrade within 2 days. We note that for unknown reasons, CagA degradation is accelerated in phosphate buffers and should be avoided once CagF is removed.
9. We observed by leaving CagL to refold longer in the REDOX buffer suppresses dimerization though formation of intermolecular disulfides by allowing reshuffling of the disulfides to form intramolecular disulfides.
10. Centrifugation of the plate should be carried out at room temperature to prevent condensation forming on the optical film, which may affect the fluorescence readings in the RT PCR.
11. SYPRO Orange has both a broad excitation and emission. Select an emission filter that encompasses the 600–630 nm range, such as ROX (610 nm) or Cy3 (615 nm).
12. A protein melting curve should show a low initial fluorescence which increases with temperature as the dye binds to unfolding of the protein and begins to fluoresce, reaches a maximum when all the protein is unfolded, and then slowly decrease as the dye dissociates from the unfolded protein. However, several types of curves may present themselves; (1) No transition is observe due to the protein having a higher melting temperature than the experiment; (2) Several transitions may occur suggesting either oligomerization of multi-domain unfolding or; (3) A high fluorescent background with a small transition suggesting that the protein is not folded or hydrophobic patches occur on the surface. In these cases the protein should be either refolded using a different method if you suspect the protein is still unfolded or screen different ratios of protein to

SYPRO Orange. For instance 1 mg/mL of protein to 1× SYPRO Orange if the background is too high or 10× SYPRO Orange if the transition is not clear.

References

1. Peek RM Jr, Blaser MJ (2002) *Helicobacter pylori* and gastrointestinal tract adenocarcinomas. Nat Rev Cancer 2(1):28–37
2. Blaser MJ, Perez-Perez GI, Kleanthous H, Cover TL, Peek RM, Chyou PH, Stemmermann GN, Nomura A (1995) Infection with *Helicobacter pylori* strains possessing cagA is associated with an increased risk of developing adenocarcinoma of the stomach. Cancer Res 55(10):2111–2115
3. Wen S, Moss SF (2009) *Helicobacter pylori* virulence factors in gastric carcinogenesis. Cancer Lett 282(1):1–8
4. de Bernard M, Arico B, Papini E, Rizzuto R, Grandi G, Rappuoli R, Montecucco C (1997) *Helicobacter pylori* toxin VacA induces vacuole formation by acting in the cell cytosol. Mol Microbiol 26(4):665–674
5. Odenbreit S, Puls J, Sedlmaier B, Gerland E, Fischer W, Haas R (2000) Translocation of *Helicobacter pylori* CagA into gastric epithelial cells by type IV secretion. Science 287(5457):1497–1500
6. Fischer W, Puls J, Buhrdorf R, Gebert B, Odenbreit S, Haas R (2001) Systematic mutagenesis of the *Helicobacter pylori* cag pathogenicity island: essential genes for CagA translocation in host cells and induction of interleukin-8. Mol Microbiol 42(5):1337–1348
7. Yamaoka Y, El-Zimaity HM, Gutierrez O, Figura N, Kim JG, Kodama T, Kashima K, Graham DY (1999) Relationship between the cagA 3′ repeat region of *Helicobacter pylori*, gastric histology, and susceptibility to low pH. Gastroenterology 117(2):342–349
8. Hayashi T, Senda M, Morohashi H, Higashi H, Horio M, Kashiba Y, Nagase L, Sasaya D, Shimizu T, Venugopalan N, Kumeta H, Noda NN, Inagaki F, Senda T, Hatakeyama M (2012) Tertiary structure-function analysis reveals the pathogenic signaling potentiation mechanism of *Helicobacter pylori* oncogenic effector CagA. Cell Host Microbe 12(1):20–33
9. Kaplan-Turkoz B, Jimenez-Soto LF, Dian C, Ertl C, Remaut H, Louche A, Tosi T, Haas R, Terradot L (2012) Structural insights into *Helicobacter pylori* oncoprotein CagA interaction with β1 integrin. Proc Natl Acad Sci U S A 109(36):14640–14645
10. Backert S, Tegtmeyer N (2012) *Helicobacter pylori* CagA tertiary structure reveals functional insights. Cell Host Microbe 12(1):3–5
11. Hohlfeld S, Pattis I, Puls J, Plano GV, Haas R, Fischer W (2006) A C-terminal translocation signal is necessary, but not sufficient for type IV secretion of the *Helicobacter pylori* CagA protein. Mol Microbiol 59(5):1624–1637
12. Pattis I, Weiss E, Laugks R, Haas R, Fischer W (2007) The *Helicobacter pylori* CagF protein is a type IV secretion chaperone-like molecule that binds close to the C-terminal secretion signal of the CagA effector protein. Microbiology 153(Pt 9):2896–2909
13. Bonsor DA, Weiss E, Iosub-Amir A, Reingewertz TH, Chen TW, Haas R, Friedler A, Fischer W, Sundberg EJ (2013) Characterization of the translocation-competent complex between the *Helicobacter pylori* oncogenic protein CagA and the accessory protein CagF. J Biol Chem 288(46):32897–32909
14. Couturier MR, Tasca E, Montecucco C, Stein M (2006) Interaction with CagF is required for translocation of CagA into the host via the *Helicobacter pylori* type IV secretion system. Infect Immun 74(1):273–281
15. Angelini A, Tosi T, Mas P, Acajjaoui S, Zanotti G, Terradot L, Hart DJ (2009) Expression of *Helicobacter pylori* CagA domains by library-based construct screening. FEBS J 276(3):816–824
16. Rohde M, Puls J, Buhrdorf R, Fischer W, Haas R (2003) A novel sheathed surface organelle of the *Helicobacter pylori* cag type IV secretion system. Mol Microbiol 49(1):219–234
17. Shaffer CL, Gaddy JA, Loh JT, Johnson EM, Hill S, Hennig EE, McClain MS, McDonald WH, Cover TL (2011) *Helicobacter pylori* exploits a unique repertoire of type IV secretion system components for pilus assembly at the bacteria-host cell interface. PLoS Pathog 7(9):e1002237
18. Kwok T, Zabler D, Urman S, Rohde M, Hartig R, Wessler S, Misselwitz R, Berger J, Sewald N, Konig W, Backert S (2007) Helicobacter exploits integrin for type IV secretion and kinase activation. Nature 449(7164):862–866
19. Conradi J, Huber S, Gaus K, Mertink F, Royo Gracia S, Strijowski U, Backert S, Sewald N

(2012) Cyclic RGD peptides interfere with binding of the *Helicobacter pylori* protein CagL to integrins αVβ3 and α5β1. Amino Acids 43(1):219–232

20. Wiedemann T, Hofbaur S, Tegtmeyer N, Huber S, Sewald N, Wessler S, Backert S, Rieder G (2012) *Helicobacter pylori* CagL dependent induction of gastrin expression via a novel αVβ5-integrin-integrin linked kinase signalling complex. Gut 61(7):986–996
21. Barden S, Niemann HH (2015) Adhesion of several cell lines to *Helicobacter pylori* CagL Is mediated by integrin αVβ6 via an RGDLXXL motif. J Mol Biol 427(6 Pt B):1304–1315
22. Barczyk M, Carracedo S, Gullberg D (2010) Integrins. Cell Tissue Res 339(1):269–280
23. Barden S, Lange S, Tegtmeyer N, Conradi J, Sewald N, Backert S, Niemann HH (2013) A helical RGD motif promoting cell adhesion: crystal structures of the *Helicobacter pylori* type IV secretion system pilus protein CagL. Structure 21(11):1931–1941
24. Barden S, Schomburg B, Conradi J, Backert S, Sewald N, Niemann HH (2014) Structure of a three-dimensional domain-swapped dimer of the *Helicobacter pylori* type IV secretion system pilus protein CagL. Acta Crystallogr D Biol Crystallogr 70(Pt 5):1391–1400
25. Bonsor DA, Pham KT, Beadenkopf R, Diederichs K, Haas R, Beckett D, Fischer W, Sundberg EJ (2015) Integrin engagement by the helical RGD motif of the *Helicobacter pylori* CagL protein is regulated by pH-induced displacement of a neighboring helix. J Biol Chem 290(20):12929–12940
26. Bross P, Andresen BS, Winter V, Krautle F, Jensen TG, Nandy A, Kolvraa S, Ghisla S, Bolund L, Gregersen N (1993) Co-overexpression of bacterial GroESL chaperonins partly overcomes non-productive folding and tetramer assembly of E. coli-expressed human medium-chain acyl-CoA dehydrogenase (MCAD) carrying the prevalent disease-causing K304E mutation. Biochim Biophys Acta 1182(3):264–274
27. Maskos K, Huber-Wunderlich M, Glockshuber R (2003) DsbA and DsbC-catalyzed oxidative folding of proteins with complex disulfide bridge patterns in vitro and in vivo. J Mol Biol 325(3):495–513
28. Ramm K, Pluckthun A (2000) The periplasmic Escherichia coli peptidylprolyl cis,trans-isomerase FkpA. II. Isomerase-independent chaperone activity in vitro. J Biol Chem 275(22):17106–17113
29. Bothmann H, Pluckthun A (2000) The periplasmic Escherichia coli peptidylprolyl cis,trans-isomerase FkpA. I. Increased functional expression of antibody fragments with and without cis-prolines. J Biol Chem 275(22):17100–17105
30. Ireton GC, Stoddard BL (2004) Microseed matrix screening to improve crystals of yeast cytosine deaminase. Acta Crystallogr D Biol Crystallogr 60(Pt 3):601–605
31. Saridakis E, Khurshid S, Govada L, Phan Q, Hawkins D, Crichlow GV, Lolis E, Reddy SM, Chayen NE (2011) Protein crystallization facilitated by molecularly imprinted polymers. Proc Natl Acad Sci U S A 108(27):11081–11086
32. Rayment I, Rypniewski WR, Schmidt-Base K, Smith R, Tomchick DR, Benning MM, Winkelmann DA, Wesenberg G, Holden HM (1993) Three-dimensional structure of myosin subfragment-1: a molecular motor. Science 261(5117):50–58
33. Wernimont A, Edwards A (2009) In situ proteolysis to generate crystals for structure determination: an update. PLoS One 4(4):e5094
34. Ericsson UB, Hallberg BM, Detitta GT, Dekker N, Nordlund P (2006) Thermofluor-based high-throughput stability optimization of proteins for structural studies. Anal Biochem 357(2):289–298
35. Dupeux F, Rower M, Seroul G, Blot D, Marquez JA (2011) A thermal stability assay can help to estimate the crystallization likelihood of biological samples. Acta Crystallogr D Biol Crystallogr 67(Pt 11):915–919
36. Niesen FH, Berglund H, Vedadi M (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nat Protoc 2(9):2212–2221