Sequence Independent Cloning and Post-translational Modification of Repetitive Protein Polymers through Sortase and Sfp-mediated Enzymatic Ligation

Journal:	*Biomacromolecules*
Manuscript ID	bm-2015-01726r.R1
Manuscript Type:	Article
Date Submitted by the Author:	25-Feb-2016
Complete List of Authors:	Ott, Wolfgang; Ludwig-Maximilians-Universitat Munchen Nicolaus, Thomas; Ludwig-Maximilians-Universitat Munchen Gaub, Hermann; LMU, Nash, Michael; Ludwig-Maximilians-University of Munich, Physik

Sequence Independent Cloning and Post-translational Modification of Repetitive Protein Polymers through Sortase and Sfp-mediated Enzymatic Ligation

Wolfgang Ott[†,‡,§], *Thomas Nicolaus*[†,‡], *Hermann E. Gaub*[†,‡], *and Michael A. Nash*[†,‡,⊥,¶,*]

[†]Lehrstuhl für Angewandte Physik, Ludwig-Maximilians-Universität München, 80799 Munich, Germany.

[‡]Center for Nanoscience (CeNS), Ludwig-Maximilians-Universität München, 80799 Munich, Germany.

[§]Center for Integrated Protein Science Munich (CIPSM), Ludwig-Maximilians-Universität München, 81377 Munich, Germany.

[⊥]Department of Chemistry, University of Basel, 4056 Basel, Switzerland.

[¶]Department of Biosystems Science and Engineering, Eidgenössische Technische Hochschule (ETH-Zürich), 4058 Basel, Switzerland.

Keywords: Golden Gate assembly, elastin-like polypeptides, sequence independent cloning, post-translational protein ligation, bioconjugation

ABSTRACT

Repetitive protein-based polymers are important for many applications in biotechnology and biomaterials development. Here we describe the sequential additive ligation of highly repetitive DNA sequences, their assembly into genes encoding protein-polymers with precisely tunable lengths and compositions, and their end-specific post-translational modification with organic dyes and fluorescent protein domains. Our new Golden Gate-based cloning approach relies on incorporation of only type IIS *BsaI* restriction enzyme recognition sites using PCR, which allowed us to install ybbR-peptide tags, Sortase c-tags, and cysteine residues onto either end of the repetitive gene polymers without leaving any cloning scars. The assembled genes were expressed in *Escherichia coli* and purified using inverse transition cycling (ITC). Characterization by cloud point spectrophotometry, and denaturing polyacrylamide gel electrophoresis with fluorescence detection confirmed successful phosphopantetheinyl transferase (Sfp)-mediated post-translational N-terminal labeling of the protein-polymers with a Coenzyme A-647 dye (CoA-647), and simultaneous Sortase-mediated C-terminal labeling with a GFP domain containing an N-terminal GG-motif in a one pot reaction. In a further demonstration, we installed an N-terminal cysteine residue into a 60 pentapeptide ELP that was subsequently conjugated to a single chain poly(ethylene glycol)-maleimide (PEG-maleimide) synthetic polymer, noticibly shifting the ELP cloud point. The ability to straightforwardly assemble repetitive DNA sequences encoding ELPs of precisely tunable length, and to post-translationally modify them specifically at the N- and C- termini provides a versatile platform for the design and production of multi-functional smart protein-polymeric materials.

INTRODUCTION

Repetitive polymers of controlled length and tunable phase-transition behavior are urgently needed for a variety of applications in the nano/biosciences, including drug delivery[1,2], and medical diagnostics[3]. Such stimuli-responsive polymeric materials are of high interest for fundamental investigations into biomolecules under the influence of mechanical, thermal, and chemical denaturants using biophysical methods such as single-molecule AFM force spectroscopy[4,5] and microscale thermophoresis[6]. Elastin like polypeptides (ELPs) are artificial proteins derived from naturally occurring elastomeric proteins[7,8]. Recombinant ELPs consist of repeats of the amino acid sequence Val-Pro-Gly-Xaa-Gly, where Xaa represents any amino acid except proline. ELPs exhibit a reversible lower critical solution temperature (LCST), and undergo a phase transition that can be triggered by temperature[9]. Other environmental stimuli like pH or ionic strength can also be used to collapse ELPs under isothermal conditions. The guest residue (Xaa) can be used to influence the pH/thermal phase transition properties of the resulting protein-polymers. Insertion of acidic residues such as glutamate or aspartate in the guest residue position results in pH-responsive behavior. The transition temperature is strongly dependent on the concentration and molecular weight, with longer ELP sequences collapsing at lower temperatures. One can also tune the cloud point by changing several environmental parameters at once (e.g., temperature, pH, salt), thereby shifting the transition to lower or higher temperatures as desired[10].

These unique properties of ELPs make them attractive for a variety of applications and scientific investigations[11]. Chromatography free protein purification, for example, can be performed by producing a target protein as an ELP fusion, and precipitating it from cellular extracts, avoiding the need for affinity tags. This method allows for purification of recombinant proteins under mild

conditions. Moreover, it is reported that in combination with maltose binding proteins, ELPs can improve the solubility of fusion domains and thereby improve expression yields[12–14].

In the field of biomaterials science, ELPs represent a viable option as a scaffold material for tissue engineering, or as carriers for drug molecules. Applications for *in vivo* systems demand high predictability and controllability of the biophysical behavior of the molecules. Since ELPs consist only of amino acids, they are competitive in terms of biocompatibility and biodegradation *in vivo* as compared to their synthetic organic polymer counterparts[15,16]. ELPs possess the added advantage of complete monodispersity. More fundamentally, the phase transition characteristics of ELPs have served as an ideal model system for theoretical calculations and modeling studies[17–21]. Additionally, conjugates between ELPs and synthetic polymers (e.g., PEG) are of high interest, and would benefit from site-specific conjugation approaches[22,23].

In order to fully leverage the versatility of repetitive protein-polymers such as ELPs, modular and straightforward approaches to cloning and site-specific post-translational modification are highly desirable. Standard solid-phase gene synthesis methods are so far not able to produce the long (>600 bp) strands of repetitive DNA required for encoding thermally responsive elastin-like polypeptides (ELPs) with lengths > 200 amino acids. Typically rationally designed ELPs are constructed using recursive directional ligation (RDL), which requires plasmid amplification, and restriction digestion, and imposes certain restrictions (i.e., the absence of restriction sites)[24]. Larger ELP genes can also be obtained with the OERCA (overlap extension rolling circle amplification) method, which generates a distribution of unspecified lengths of repetitive DNA sequences[25].

Compared to the RDL method our Golden Gate approach presented here avoids cloning scars due to the use of type IIS restriction enzymes, and is able to cut scarlessly within the coding region[24,26]. The PRe-RDL (RDL by plasmid reconstruction) method relies on several type IIS restriction enzymes and requires certain modifications of the backbone beforehand[27].

Our method is applicable to a broad spectrum of plasmids, since the only limitation is one type IIS restriction enzyme with a recognition site not present in the backbone. Along with this advantage, it is likewise ideal for adding ELPs to an existing gene-containing plasmid to create fusion proteins with different length ELPs. The combinational possibilities also do not rely on a plasmid library, but can be designed using a bottom up block assembly approach. Our approach can also be used in a complimentary way with the existing RDL and OERCA methods, for example, by easily generating fast and reliable plasmid libraries which can then be further extended by combining with RDL or OERCA methods.

We present a sequence independent approach based on the Golden Gate technology employing polymerase chain reaction (PCR) amplification of short ELP repeats and ligation into a plasmid backbone to produce repetitive ELP genes with specific peptide tag end groups for covalent post-translational modification. A single type IIS restriction enzyme is used to create unique ends and guarantee the order of DNA block assembly. Using this method, repetitive DNA sequences up to hundreds of nm in length (i.e., 120 pentapeptide repeats of ELPs) can be rationally designed and created. 5' and 3' tags for post-translational modifications were readily incorporated during the cloning workflow, providing many further possibilities for downstream conjugation and labeling.

Due to the modularity of the system, we were also able to readily install peptide tags (i.e., ybbR[28] and Sortase sequences) to the ELP, enabling enzyme-catalyzed ligation of the peptides to fluorescent proteins and organic dyes (as shown below). Our approach builds on the prior method shown by Huber *et al.* which demonstrated fusion of different kinds of repetitive DNA to create chimeras of ELPs, silk peptides and similar proteins[29]. Our methodology is also compatible with their approach with the advantage of using only one type IIS restriction enzyme.

Alternatively, it is possible to modify the carrier plasmid in the first amplification round and add ELP flanking tags or protein domains easily. Since the reaction starts new every three fragments, one can easily define block patterns that build up an overall sequence. For example, pH responsive blocks can be interspersed with pH-insensitive blocks. In regards to user-friendliness, the presented method is advantageous because it relies on the same ELP gene inserts, which can be reused. Once successful amplification and purification of the sequences is achieved, the PCR amplicons can be stored and used again as needed. This way it is possible to create a whole library of gene sequences and, if desired shuffle these each ligation cycle. Post-translational fusion of ELPs using Sortase ligation circumvents the known issue of low protein yields for N-terminally located ELP domains in fusion proteins[30,31]. Instead of optimizing expression conditions for proteins of low yield, a protein of interest can be produced in its native state and fused afterwards post-translationally with the ELP domain. To the best of our knowledge, this represents the first report using a Sortase-based recognition sequence to fuse ELP proteins to other proteins[13,32].

MATERIALS AND METHODS

All used reagents were of analytical purity grade and were purchased from Sigma-Aldrich (St. Louis, MO, USA) or Carl Roth GmbH (Karlsruhe, Germany).

Monomer Gene Synthesis. A synthetic gene encoding 150 nucleotides (10 pentapeptide repeats) for the $(VPGVG)_5$-$(VPGAG)_2$-$(VPGGG)_3$ peptide (Centic Biotech, Heidelberg, Germany) served as starting material (see Supporting Information, DNA Sequence 1 and Protein Sequence 1).

Cloning. Golden Gate cloning was employed to create the different rationally designed ELP constructs[26]. PCR (Backbone: 98°C 2 min, (98°C 7 s, 72°C 2 min 30 s) x30, 72°C 5 min; Insert: 98°C 2 min, 98°C 7 s, 60°C 7s, 72°C 5 s) x30, 72°C 5 min) was performed with a Phusion high fidelity polymerase master mix. A typical 20 µl PCR mix contained 10 µl Phusion high fidelity polymerase master mix (Thermo Fisher Scientific Inc., Waltham, MA, USA), 0.5 µl per forward and reverse primer (10 µM), 1.5 µl DMSO, 1 ng of template and water. All primers (biomers.net, Ulm, Germany) used in this study are listed in Tab. 1.

Table 1. Overview of employed primers

Primer	Sequence 5' – 3'
(1a) FW ELP I ybbR	TATATAGGTCTCCTGGC**TGTGCCGGG**AGAAGGAGTCCCTGG **TGTCGGTGTCCCAGGCG**
(1b) REV ELP I	GGTCTCCTCCTT**CACCCGGAACGCCACCCCCCGGAACACC GCCGC**
(2a) FW ELP II	TATATAGGTCTCAAGGA**GTACCAGGCG**AA**GGCGTGCCGG GTGTC**
(2b) REV ELP II	ATATATGGTC**TC**A**CC**CT**CACCCGGAACG**CC**ACCCCCCGGA ACACCGCCGC**

(3a) FW ELP III	TATATAGGTCTCGAGGGT**GTACCAGGCG**AA**GGGGTGCCGG GTGTC**
(3b) REV ELP III LPETGG	ATATATGGTCTCCGGCAGACCTT**CACCCGGAACGCCACC**CC **CCGGAACACCGCCGC**
(4) REV ELP III	ATATATGGTCTCC**ACC**TT**CACCCGGAACGCCACC**CCC**CGG AACACCGCCGC**
(5) FW backbone LPETGG	ATATATGGTCTCCTGCCGGAAACCGGCGG**CTAACTCGAGTA AGATCCGGCTGC**
(6) REV backbone ybbR	ATATATGGTCTC**AGCCA**GTTTAGAAGCGATGAATTCCAG
(7) FW backbone ybbR	**GACTCTCTGGAATTCATCGCTTCTAAACTGGC**TGGTCTCC AGG**TGTGCCGGGA**
(8) FW ELP II ybbR	TATATAGGTCTCCTGGCG**GTACCAGGCG**AA**GGGGTGCCGG GTGTC**
(9) FW ELP III ybbR	TATATAGGTCTCCTGGCG**GTACCAGGCG**AA**GGC**GTGCCGG **GTGTC**
(10) FW ELP N Cys	**GACTCTCTGGAATTCATCGCTTCTAAACTGGCTG**GTCTCC TGCG**TGCCGGGAGAAGGAG**
(11) REV backbone	**CCCGGCACAGCCAGTTTAGAAGCGATGAATTCCAGAGAG TC**GGTCTCA**CATA**TGTATATC

Primers 1 – 7 are employed for the cloning of the ELPs with 3 fragments growth every cycle. 1-4 are the primers necessary for insert amplification, 5-7 for the amplification of the backbone. Primers 8 and 9 are only important for ELP cloning procedures with the addition of one or two fragments. Primers 10 and 11 were used to change the 5' flanking site of the ELP gene from the gene for the ybbR-tag to a cysteine. DNA Sequence is styled in different ways: **bold** (annealing region), underlined (*Bsa*I recognition site), highlighted in grey (*Bsa*I restriction site)

147 In the first round of PCR (see backbone PCR above, 55°C 7s annealing), the backbone of a
148 modified pET28a vector (Merck KGaA, Darmstadt, Germany) was linearized. The PCR product
149 contained at the 5' end the sequence for a ybbR-tag (DSLEFIASKLA) and at the 3' end a C-
150 terminal Sortase recognition sequence (LPETGG)[33,34]. Sequences of all PCR fragments
151 (backbone, ELP I, II, III, IV) and a description for primer design (see Supporting Information,

Primer 12) based on an original pET28a vector are attached in the supporting information (Fig. S1 – S9, DNA Sequence 1-6 and Fig. S14-S18).

The superfolder GFP (sfGFP) plasmid was created with Gibson Assembly[35]. The gene (Addgene ID: 58708)[36] was amplified with overlaps to match a linearized vector containing sequences encoding N-terminal HIS_6-tag, a TEV protease cleavage site and two glycines (compare the PCR program above; 55°C annealing and an extension time of 1 min. 30 s; see Supporting Information, DNA Sequence 8 and Protein Sequence 4).

All PCR products were digested (37°C, 1 - 12 h) with FD-*Dpn*I (Thermo Fisher Scientific Inc., Waltham, MA, USA) and purified either with QIAquick PCR purification kit or gel extraction kit (Qiagen, Hilden, Germany) (Supporting Information, Fig. S10, Fig. S13). *Dpn*I was added to digest the methylated plasmids serving as starting material (template) in the PCRs, to reduce number of false positive clones in the following transformation.

Typically, a 25 µl Golden Gate reaction (2.5 µl CutSmart buffer (10x), 1.25 µl T7 ligase, 1.25 µl *BsaI*-HF and 2.5 µl ATP (10 mM), New England Biolabs, Ipswich, MA, USA) was set up. The inserts were added in 10-fold molar excess to the backbone (ratio of 0.1 pmol insert to 0.01 pmol backbone). The reaction was performed in a thermo cycler (25x 37°C 2 min, 25°C 5 min; 37°C 10 min; 80°C 10 min). For the Gibson Assembly reaction, 10 µl of the master mix (2x, New England Biolabs, Ipswich, MA, USA) were mixed with 0.01 pmol vector and 0.1 pmol insert. The reaction was incubated for 1 h at 50°C. For the replacement of the ybbR-tag with cysteine, the PCR linearized product was first digested with *BsaI*-HF together with FD-*Dpn*I (1h, 37°C, 5

min, 80°C). The reaction was supplied with 1 µl of dNTPs (10 mM, New England Biolabs, Ipswich, MA, USA), 1 µl of Klenow Fragment (10 U/µL, Thermo Fisher Scientific Inc., Waltham, MA, USA), and incubated (37°C, 15 min, and 75°C, 10 min). After a gel extraction, the corresponding band was excised and a blunt end reaction (6.5 µl PCR product, 1 µl ATP (10 mM), 1 µl CutSmart buffer (10x), 0.5 µl PEG-6000, 1.0 µl T4 Polynucleotide Kinase, 1.0 µl T4 Ligase) was set up (37°C 15 min, 22°C 45 min, 80°C 7 min).

In case of the Golden Gate reaction, 10 µl and in case of the Gibson Assembly or the blunt end ligation 2 µl were used to transform DH5α cells (Life Technologies GmbH, Frankfurt, Germany; 30 min on ice, 42°C 1 min, 1 h 37°C). The transformed culture was plated on appropriate antibiotic LB-Agar plates. A small number (<10) of clones were analyzed by colony PCR, or analytical restriction digestion (FD-*Eco*RI, Thermo Fisher Scientific Inc., Waltham, MA, USA) followed by sequencing (Supporting Information, Tab S1).

Protein Expression. For ELP expression, chemically competent *E. coli* NiCo21(DE3) (New England Biolabs, Ipswich, MA, USA) were transformed with 50 ng plasmid DNA[37]. The cells were incubated in kanamycin containing, autoinducing ZYM-5052 media (supplemented with an amino acid mix 0.1 mg/ml) 24 h at 25°C[38–40]. After harvesting, ice cooled cells were lysed using sonication (Bandelin Sonoplus GM 70, Tip: Bandelin Sonoplus MS 73, Berlin, Germany; 40 % Power, 30 % Cycle 2x 10 min). The supernatant of the lysate (15000 g, 4°C, 1 h) was heated to 60°C for 30 min to denature most of the *E. coli* host proteins. In a second step, the collapsed ELPs within this clouded solution were rehydrated by incubating under continuous mixing for 2 h at 4°C. This allowed the resolubilization of the ELPs while the precipitated host proteins remained insoluble. A centrifugation step (15000 g, 4°C, 30 min) was used to separate the

soluble ELPs and remaining proteins from precipitated cell debris. The clear supernatant turned immediately cloudy after adding 1 M acetate buffer (final concentration 50 mM, pH 3.5), and 2 M NaCl in crystalline form. The mixture was incubated for 30 min at 60°C. The collapsed ELPs were collected by centrifugation (3220 g, 40°C, 75 min). The obtained pellet was re-solubilized in 50 mM TRIS-HCl (pH 7.0) and incubated over night at 4°C. The remaining precipitated debris were removed by a final centrifugation step (3220 g, 4°C, 60 min). The supernatant was mixed again with acetate buffer and sodium chloride to collapse the ELPs. After the heated incubation and centrifugation step, the pellet was resolubilized in buffer (50 mM TRIS-HCl, pH 7.0)[14,41].

The purity of the ELP was confirmed by SDS-PAGE (Any kD™ Mini-PROTEAN® Stain-Free™ Gels, Bio-Rad Laboratories GmbH, Hercules, CA, USA), in order to detect any remaining contaminant host proteins. The ELPs were labeled with CoA-647 (New England Biolabs (Ipswich, MA, USA) and Sfp (37°C, 1 h, 5 mM $MgSO_4$) to visualize them. After labeling, the ELPs were mixed with 6x Loading buffer and heated to 95°C for 10 min[42]. Usually a purity grade of >95% was obtained. Purity analysis was performed by overlaying the UV active Stain-Free™ technology from Bio-Rad (labeling all tryptophan side groups of *E. coli* host proteins) and a fluorophore specific red channel for the CoA-647-ELP constructs (Supporting Information Fig. S11). MALDI-TOF analysis of ELP samples $ELP_{30\text{-}50}$ was performed to increase confidence in the high purity of the samples (Supporting Information, Fig. S19). ELPs were stored at 4°C in 50 mM TRIS-HCl, pH 7.0.

The final ELP concentration was photometrically determined at 205 nm (Ultrospec 3100 pro, Amersham Biosciences (Amersham, England) and TrayCell (Hellma GmbH & Co. KG, Müllheim, Germany))[43].

For the expression of HIS_6-TEV-GG-sfGFP, 50 ng plasmid DNA was used to transform *E. coli* NiCo21(DE3) cells. Kanamycin containing, autoinducing ZYM-5052 growth media was inoculated with an overnight culture[38]. After 24 h incubation at 25°C the cells were harvested, lysed and centrifuged as described above. The supernatant was applied on a HisTrap FF (GE Healthcare Europe GmbH, Freiburg, Germany). After washing five times with wash buffer (25 mM TRIS-HCl pH 7.8, 300 mM NaCl, 20 mM Imidazole, Tween 20 0.25 % (v/v), 10 % (v/v) glycerol) the bound protein was eluted (25 mM TRIS-HCl pH 7.8, 300 mM NaCl, 300 mM Imidazole, Tween 20 0.25 % (v/v), 10 % (v/v) glycerol).

HIS_6-TEV-GG-sfGFP fusion protein (TEV cleavage site: ENLYFQG) was dialyzed immediately after elution with the TEV protease (4°C, 50 mM TRIS-HCl, pH 7.0) overnight. The cleaved product was separated from the uncleaved construct by applying the reaction mix on a HisTrap FF 5 ml column. The successfully cut fragment in the flow through was collected. The fraction was dialyzed against 50 mM TRIS-HCl, pH 7.0 and stored in 50 % (v/v) glycerol at -80°C. The purity of the elution and the cleaved fraction was analyzed via a SDS-PAGE analysis. The specific extinction coefficient of GFP at 485 nm was used to determine the concentration of GG-sfGFP.

Turbidity measurements. For the turbidity measurements, a photometer with a Peltier heating element was used (JASCO V-650, JASCO Germany GmbH, Gross-Umstadt, Germany). The turbidity was determined at 350 nm while the temperature was ramped at a rate of 2°C/min. Measurements were taken every 0.5°C between 20 and 80°C. ELPs were dialyzed against double

distilled water, mixed with a buffer stock solution (1 M TRIS-HCl, pH 7.0), sodium chloride stock solution (3 M or 5 M) and adjusted with water to the desired final concentration.

For NaCl titration, 100 µM of the ELP constructs were tested in a range of 0 - 3 M sodium chloride. The 6x ELP construct was also probed in a concentration range of 25 µM - 200 µM with different NaCl concentrations.

For pH titrations, stock solutions of 0.1 M phosphate-citrate buffer at different pH values were mixed with solutions of water solubilized ELPs. Hereby a final concentration of 0.05 M of the phosphate-citrate buffer was obtained.

Data analysis of the transition temperature curves (for NaCl, pH, concentration dependency and PEG-ELP fusions) was performed by fitting the measured data points with a four-parameter logistic function to obtain the corresponding transition temperature.

Sortase and Sfp-mediated protein ligation. For highest ligation efficiencies, enhanced Sortase (eSortase) was used in the reaction[44]. The reaction conditions for both Sfp and eSortase enzymes were chosen according their reported reaction maxima to achieve highest activities[28]. ELPs in excess were added to a solution containing 50 mM TRIS-HCl, pH 7.5, 15 µM ELP, 0.5 µM GG-sfGFP, 0.2 µM eSortase, 1 µM Sfp, 5 mM $CaCl_2$, 5 mM $MgCl_2$, 5 µM CoA-647. The ligation reaction was incubated for 2 h at 37°C.

Cysteine-Maleimide bioconjugation reaction. Cysteine-containing ELPs were reduced with 5 mM *tris(2-carboxyethyl)phosphine* (TCEP, (Thermo Fisher Scientific Inc., Waltham, MA, USA)). After the removal of TCEP with Zeba^{TM} Spin Desalting Columns 7K (Thermo Fisher Scientific Inc., Waltham, MA, USA) cysteine-ELPs were mixed with $Alexa_{647}$-C2-Maleimide (Thermo Fisher Scientific Inc., Waltham, MA, USA) and incubated for 1 h at 37°C (100 mM TRIS-HCl, pH 7.0) (Supporting Information, Fig. S20).

PEG (MW: 20,000 Da, α-Methoxy-ω-maleimide, Rapp Polymere GmbH, Tübingen, Germany) was used in different molar ratios in the bioconjugation reaction with cysteine-ELP_{60} or ELP_{60}. 75 µM of the reduced ELPs were mixed with TRIS-HCl (pH 7.0, 100 mM), PEG, and incubated for 1 h at room temperature. After that they were mixed with 5 M NaCl and to a final concentration of 3 M NaCl and their cloud point was determined as described above.

RESULTS AND DISCUSSION

Our sequence independent Golden Gate-based method provides an easy way to create defined repetitive DNA sequences[26]. We designed and produced gene cassettes encoding repetitive proteins several hundreds of amino acids in length. Fig. 1 outlines the principle of primer design and the following logical and stepwise workflow. The sequence of the starting synthetic gene was designed in such a way that the codon usage within the first and last 15 nucleotides was unique within the otherwise repetitive 150 bp sequence. This was necessary to ensure specific annealing of primers at the 5' and 3' end. Desired modifications were introduced by overhangs of the primers at their 5' end (i.e. *BsaI* recognition site) or at their 3' region (codon shuffling of nucleotides). It was then possible to create 150 bp ELP genes with different flanking regions

from the same template (Primers 1-3; Fig. 1 A) using PCR primers that annealed at the 5' and 3' ends of the synthetic gene.

In the first amplification and linearization reaction of the plasmid, primers annealed at the desired ELP gene insertion site, i.e. at the opening location on the plasmid during the first PCR. In our case this was downstream of the T7 promoter and upstream of the T7 terminator (see Supporting Information Fig. S8 and S9). However, due to the freedom of primer design and plasmid choice, the insertion site can in principle be anywhere in the plasmid. The primers linearized the plasmid and introduced tags at the 5' (ybbR-tag) and 3' (Sortase c-tag) prime ends, as well as *BsaI* recognition sites (Primers 5, 6; Vector A; Fig. 1 A). In our case, a modified pET28a vector, already containing a ybbR-site downstream of the T7 promotor immediately following the start codon AUG, served as template. Hence, only the Sortase c-tag was newly introduced (see Supporting Information for primers for the standard pET28a vector). The continuing general ELP expansion principle relies on having three different PCR amplified ELP fragments (I, II, III) with different codon usages at their 5' and 3' end, within the *Bsa*I-restriction site (Supporting Information Fig. S1-S6). This design made logical and block-wise gene assembly possible. The selected primers introduced a shuffled 3' end which matched the 5' end of the subsequent fragment. In the first ELP assembly round, the 5' end of fragment I matched the ybbR-tag of the linearized backbone. The 3' end of fragment III had compatible sticky ends with the Sortase c-tag of the linearized plasmid (Fig. 1 B, 1st round). After successful annealing of sticky ends, the T7 ligase covalently linked the three ELP fragments seamlessly into the plasmid without any undesirable cloning scars in between.

The forward primer (Fig. 1 A, 2nd: Primer 7) for the following plasmid linearization rounds annealed at a different site within the ELP-containing plasmid, compared to the initial linearization round (Fig. 1 A, 1st: Primer 5). It annealed at the ybbR-tag and the 5' end of the ELP gene. Right in between the two coding regions, a non-annealing loop encoding a *BsaI* recognition site was introduced (Fig. 1 A 2nd, and Fig. 1 B 2nd round) with the primer. The annealing at the ybbR-tag was necessary to ensure high temperature-dependent primer annealing specificity at the very 5' end of the ELP gene, otherwise the primer would anneal at every fragment I throughout the whole assembled ELP gene cassette. High annealing temperatures minimize undesired PCR side products, i.e. only partly ELP-containing, linearized vectors. The reverse primer was the same for all plasmid linearization reactions (Fig. 1 A; Vector B). After the restriction digestion reaction, the linear plasmid now had a Sortase c-tag sticky end at the 3' end and an ELP fragment I sticky end at the 5' end.

Now only the last ELP fragment (Fig. 1 A, Insert IV) had to be amplified with a different reverse primer (Fig. 1 A, Primer 4) to yield a PCR product with a compatible 3' end to the already existing ELP cassette. The growing ELP insert in the plasmid always started with fragment I. This made the reuse of the amplified insert sequences (I, II, IV) for every following expansion cloning round possible (Fig .1 B, >3rd rounds).

This method not only allows a logical assembly of repetitive gene patterns, but also makes the modification of flanking regions or mutation of the first base pairs at 5' end 3' end possible. For example, we introduced two glutamates in each of the fragments at their 5' and 3' ends by changing the codon from the 'X' guest residue at the 5' and 3' end of the VPGXG motif to a

glutamate (VPGEG). The primers did not align completely with the template and introduced the glutamate mutation during PCR amplification. The chemically synthesized sequence also had some minor mistakes at the 3' end, which were corrected with primers within the initial PCR. The final ELP substructure of all ELPs used in this study consisted of 10 pentapeptide repeats ($VPGXG_{10}$, X being $[EV_4A_2G_2E]$). For the rest of the manuscript this motif is referred as ELP_n with n being the number of pentapeptide repeats of this motif (see Supporting Information, DNA Sequence 2, Protein Sequence 2 and DNA Sequence 7, Protein Sequence 3).

We ligated three 150 bp fragments with a linearized vector of choice in one step. It was possible to modify the 5' and 3' ends of the fragments with overhang primers prior to ligation, in our case with an N-terminal ybbR and a C-terminal Sortase tag (Fig. 1 B). Overall seven different ELP constructs were used in this study for biophysical characterization of the peptide sequence, while ten were successfully cloned. The largest ELP gene contained 120 pentapeptide repeats. All ELP constructs were built with the four different ELP PCR products from the same batch. PCR gels from the fragments and an overview of cloning efficiencies can be found in the supporting information (Fig. S9 and Tab. S1). Typical yields after the purification were 56 – 138 mg Protein/l culture, while the ELP_{10} repeat had the lowest yield (2 mg Protein/l culture).

Tab. 2 shows biophysical characteristics of the ELPs characterized in this study. Each ELP was produced with a ybbR-tag at the N-terminus and a Sortase c-tag at the C-terminus. In the bottom right corner of the schematic (Fig. 1), FD-*Eco*RI digested plasmids are shown on an agarose gel. The gel analysis shows the successful construction of plasmids containing 10 to 120 pentapeptide repeats.

Table 2. Biophysical properties of the characterized ELP constructs.

ELP repeats $(5)_x$	ε_{205} [1/M cm][43]	Molecular weight [Da][45]	Glutamate residues in ELP repeat	Isoelectric point[45]	Amino acids in ELP repeats (total)	Total Length [nm][46]
10	196690	5893.7	2	3.91	50 (68)	24.82
20	335690	9908.2	4	3.77	100 (118)	43.07
30	474690	13922.8	6	3.67	150 (168)	61.32
40	613690	17937.3	8	3.59	200 (218)	79.57
50	752690	21951.9	10	3.53	250 (268)	97.82
60	891960	25966.4	12	3.47	300 (318)	116.07
Cys-60	855980	24894.2	12	3.20	300 (308)	112.42

Following successful cloning, expression and purification, we tested the functionality of the attached terminal tags. Fig. 2 A shows the scheme for post-translational protein ligation reactions. The ELPs of varying lengths contain an N-terminal ybbR-tag and a C-terminal Sortase recognition sequence (i.e., LPETGG). Subpanel 2 B to 2 C show an SDS-PAGE image of the same gel with different excitation and emission filters. Using a reaction catalyzed by Sfp, it was possible to fuse a fluorescently labeled CoA-647 to the ELP (N-terminal ybbR-tag). Results of the specific excitation for the CoA-647 dye are shown in 2 B. Brightest are the CoA-647-ELP fusions proteins, but also the CoA-647-ELP-sfGFP fusion proteins are visible above the bright monomer band. Fully denatured proteins appear slightly higher in the gel due to their different

running behavior. The Sortase-tag was simultaneously utilized for fusion of different proteins to the ELP sequences (C-terminal LPETGG). A GG-sfGFP was fused to the ELPs, which was excited with blue LED light and detected within the green emission of sfGFP (Fig. 2 C). Non-ligated and non-denaturated GFP appears at the top of the gel, since it does not run according its molecular weight in its native (i.e., correctly folded) state (see Supporting Information, Fig. S12). No GFP fluorescence is visible in the heated samples due to complete denaturation of the GFP chromophore. Panel 2 D shows an overlay of B and C, visualizing the successful post-translational ligation of GG-sfGFP and CoA-647 to the different ELP peptides within a one-pot reaction. The ligation efficiency of the Sortase never goes to 100 % completion. Due to the Sortase reaction mechanism, a dynamic equilibrium is eventually reached and complete fusion of GG-sfGFP to ELP is therefore not to be expected[47].

After confirming the biochemical accessibility and functionality of the terminal ybbR- and Sortase-tags, we characterized the phase behavior of the modified ELPs. Fig. 3 presents an overview of the lower critical solution temperatures (LCSTs) of the characterized ELPs under various conditions. First the temperature dependence of the $ELP_{10\text{-}60}$ constructs were probed against different sodium chloride concentrations, at neutral pH (50 mM TRIS-HCl, pH 7.0) (Fig. 3 A). The 10 pentapeptide repeat ELP did not collapse below 80°C, which is in agreement with the remainder of the data set if one looks at the increasing transition temperature with decreasing size of the construct. The 20 pentapeptide ELP repeat, for example, only collapsed with 3 M of sodium chloride at 60°C. Fig. 3 B clarifies the correlation between salt concentration, molecular mass and transition temperatures. Only the longest ELP construct collapsed across all given sodium chloride concentrations in the temperature range from 20 to 80°C. Salt-induced cloud point shifts are a well known characteristic of ELPs[15,24,48].

The incorporation of two glutamates per ten pentapeptides resulted in pH-dependent transitions. ELPs with glutamates were expected to show pH-responsiveness. Above their pK_a the ELPs have a relatively high transition temperature, since the glutamates are deprotonated and ionized and therefore electrostatically repel each other. Below or close to their corresponding pK_a, the transition temperature significantly decreases due to protonation and neutralization of the negative charge (Fig. 3 C). The decreasing influence of salt at lower pH is similar to that demonstrated by MacKay *et al.*[49]. Fig. 3 D illustrates the dependence of transition temperature on the ELP concentration. At concentrations above 100 µM, the 60 pentapeptide ELP (150, 200 µM) already collapsed at room temperature, hence it was not possible to determine an exact transition point. The ligated product between the 60 pentapeptide ELP repeat and the sfGFP did not show any transition compared to the pure 60 pentapeptide ELP (data not shown). This concentration dependence is also a well-known characteristic of ELPs[10].

This PCR based method can also be employed to change the flanking sequences of the ELP very quickly. Fig. 4 A shows the underlining principle of the cloning procedure used to install cysteine as an end residue with no cloning scar. Due to the repetitive structure of the ELP gene it was necessary to design primers which anneal at the site of replacement. A *BsaI* recognition loop between ELP annealing and deletion annealing site was necessary to remove the deletion site again afterwards. *BsaI* digestion left incompatible 5' and 3' sticky ends, therefore a Klenow Fragment was employed to fill the ends. A standard blunt end ligation circularized the linear plasmid (Fig. 4A and Supporting Information: Fig. S13-S18). This procedure provided an N-terminal cysteine that could be used for bioconjugations to various (macro)molecules (see Supporting Information, DNA Sequence 9, Protein Sequence 5). The cysteine in the ELP is able

to form disulfide bonds with different cysteine containing proteins, but also is able to be clicked to other reactive groups like maleimide (i.e. a maleimide-PEG (Fig. 4 B)). The cloud point determination of Fig. 4 C shows the influence of PEG conjugation on the ELP cloud point, confirming a shift towards higher temperatures (Fig. 4 C, CYS-ELP_{60}) due to conjugation of the hydrophilic synthetic polymer. However, the same PEG added to a solution of the same ELP that lacked the cysteine functionality did not significantly influence the cloud point (Fig. 4 C, ELP_{60}).

CONCLUSION

The presented approach shows an alternative way to create fast and convenient functional ELPs with sequence lengths up to 600 amino acids, or hundreds of nm in stretched contour length. It allows a straightforward fusion of gene sequences encoding the ELP repeats without any prior vector modifications. We used this approach to demonstrate facile incorporation of functional peptide tags as end groups into ELPs. We demonstrate how this approach was useful for developing end-labeled ELPs through enzyme-mediated site-specific ligation to organic dyes and fluorescent proteins, and show how terminal cysteine incorporation expands the versatile toolbox of bioconjugation opportunities. Since we used a PCR and primer-based approach, our method is essentially sequence independent and does not leave cloning scars. In the future we anticipate that such a tool for straightforward end-group modification of ELPs will prove useful for developing custom engineered macromolecular systems.

ASSOCIATED CONTENT

Supporting Information

Additional information including sequence data (DNA and Protein sequences), extended cloning procedures and gel pictures of PCR products and protein purification steps.

This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

e-Mail: *michael.nash@lmu.de

Author Contributions

W.O., M.A.N. and H.E.G. designed the research; W.O. and T. N. performed experiments; W.O. performed data analysis; W.O., and M.A.N. co-wrote the manuscript.

Notes

The authors declare no competing financial interests.

ACKNOWLEDGMENT

We gratefully acknowledge funding from an advanced grant of the European Research Council (Cellufuel Grant 294438), SFB 863 and the Excellence Cluster Center for Integrated Protein Science Munich. M.A.N. acknowledges funding from Society in Science - The Branco Weiss Fellowship program administered by ETH Zürich, Switzerland. We thank the systems biophysics group of Professor Dieter Braun (Ludwig-Maximilians-Universität München) for the access to the JASCO V-650. The authors thank Anna Krautloher for initial concentration determinations of the ELPs at 205 nm. We acknowledge Markus Jobst for his advice on the ongoing manuscript. We thank the following people for providing material to this study: Ellis Durner (eSortase),

Angelika Kardinal (TEV protease) and Diana Pippig (Sfp), and Arne Goldenbaum for assistance in lab work. We are grateful for the MALDI-TOF analysis of ELP samples by the protein analysis group of the Ludwig-Maximilians-Universität München (Professor Axel Imhof, Dr. Andreas Schmidt, Dr. Ignasi Forné and Pierre Schilcher).

REFERENCES

(1) Pack, D. W.; Hoffman, A. S.; Pun, S.; Stayton, P. S. *Nat. Rev. Drug Discov.* **2005**, *4* (7), 581–593.

(2) Onaca, O.; Enea, R.; Hughes, D. W.; Meier, W. *Macromol. Biosci.* **2009**, *9* (2), 129–139.

(3) Nash, M. A.; Waitumbi, J. N.; Hoffman, A. S.; Yager, P.; Stayton, P. S. *ACS Nano* **2012**, *6* (8), 6776–6785.

(4) Nash, M. A.; Gaub, H. E. *ACS Nano* **2012**, *6* (12), 10735–10742.

(5) Urry, D. W.; Hugel, T.; Seitz, M.; Gaub, H. E.; Sheiba, L.; Dea, J.; Xu, J.; Parker, T. *Philos. Trans. R. Soc. B Biol. Sci.* **2002**, *357* (1418), 169–184.

(6) Wolff, M.; Braun, D.; Nash, M. A. *Anal. Chem.* **2014**, *86* (14), 6797–6803.

(7) Urry, D. W.; Haynes, B.; Harris, R. D. *Biochem. Biophys. Res. Commun.* **1986**, *141* (2), 749–755.

(8) Tatham, A. S.; Shewry, P. R. *Trends Biochem. Sci.* **2000**, *25* (11), 567–571.

(9) Urry, D. W.; Haynes, B.; Zhang, H.; Harris, R. D.; Prasad, K. U. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85* (10), 3407–3411.

(10) Meyer, D. E.; Chilkoti, A. *Biomacromolecules* **2004**, *5* (3), 846–851.

(11) Urry, D. W. *J. Phys. Chem. B* **1997**, *101* (51), 11007–11028.

(12) Bataille, L.; Dieryck, W.; Hocquellet, A.; Cabanne, C.; Bathany, K. *PROTEIN Expr. Purif.* **2015**, *110*, 165–171.

(13) Bellucci, J. J.; Amiram, M.; Bhattacharyya, J.; McCafferty, D.; Chilkoti, A. *Angew. Chemie - Int. Ed.* **2013**, *52* (13), 3703–3708.

(14) Meyer, D. E.; Chilkoti, A. *Nat. Biotechnol.* **1999**, *17* (11), 1112–1115.

(15) Gagner, J. E.; Kim, W.; Chaikof, E. L. *Acta Biomater.* **2014**, *10* (4), 1542–1557.

(16) Kojima, C.; Irie, K. *Biopolymers* **2013**, *100* (6), 714–721.

(17) Christensen, T.; Hassouneh, W.; Trabbic-Carlson, K.; Chilkoti, A. *Biomacromolecules*

2013, *14* (5), 1514–1519.

(18) McDaniel, J. R.; Radford, D. C.; Chilkoti, A. *Biomacromolecules* **2013**, *14* (8), 2866–2872.

(19) Rousseau, R.; Schreiner, E.; Kohlmeyer, A.; Marx, D. *Biophys. J.* **2004**, *86* (3), 1393–1407.

(20) Qin, G.; Glassman, M. J.; Lam, C. N.; Chang, D.; Schaible, E.; Hexemer, A.; Olsen, B. D. *Adv. Funct. Mater.* **2014**, *25* (5), 729–738.

(21) Glaves, R.; Baer, M.; Schreiner, E.; Stoll, R.; Marx, D. *ChemPhysChem* **2008**, *9* (18), 2759–2765.

(22) Wang, H.; Cai, L.; Paul, A.; Enejder, A.; Heilshorn, S. C. *Biomacromolecules* **2014**, *15* (9), 3421–3428.

(23) Van Eldijk, M. B.; Smits, F. C. M.; Vermue, N.; Debets, M. F.; Schoffelen, S.; Van Hest, J. C. M. *Biomacromolecules* **2014**, *15* (7), 2751–2759.

(24) Meyer, D. E.; Chilkoti, A. *Biomacromolecules* **2002**, *3* (2), 357–367.

(25) Amiram, M.; Quiroz, F. G.; Callahan, D. J.; Chilkoti, A. *Nat. Mater.* **2011**, *10* (2), 141–148.

(26) Engler, C.; Kandzia, R.; Marillonnet, S. *PLoS One* **2008**, *3* (11), e3647.

(27) McDaniel, J. R.; MacKay, J. A.; Quiroz, F. G.; Chilkoti, A. *Biomacromolecules* **2010**, *11* (4), 944–952.

(28) Yin, J.; Lin, A. J.; Golan, D. E.; Walsh, C. T. *Nat. Protoc.* **2006**, *1* (1), 280–285.

(29) Huber, M. C.; Schreiber, A.; Wild, W.; Benz, K.; Schiller, S. M. *Biomaterials* **2014**, *35* (31), 8767–8779.

(30) Christensen, T.; Amiram, M.; Dagher, S.; Trabbic-Carlson, K.; Shamji, M. F.; Setton, L. A.; Chilkoti, A. *Protein Sci.* **2009**, *18* (7), 1377–1387.

(31) Beerli, R. R.; Hell, T.; Merkel, A. S.; Grawunder, U. *PLoS One* **2015**, *10* (7), e0131177.

(32) Qi, Y.; Amiram, M.; Gao, W.; McCafferty, D. G.; Chilkoti, A. *Macromol. Rapid Commun.* **2013**, *34*, 1256–1260.

(33) Mazmanian, S. K.; Liu, G.; Ton-That, H.; Schneewind, O. *Science* **1999**, *285* (5428), 760–763.

(34) Yin, J.; Straight, P. D.; McLoughlin, S. M.; Zhou, Z.; Lin, A. J.; Golan, D. E.; Kelleher, N. L.; Kolter, R.; Walsh, C. T. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (44), 15815–15820.

(35) Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, C. J.; Hutchison III, C. A.; Smith, H. O.

Nat. Methods **2009**, *6* (5), 343–347.

(36) Otten, M.; Ott, W.; Jobst, M. A.; Milles, L. F.; Verdorfer, T.; Pippig, D. A.; Nash, M. A.; Gaub, H. E. *Nat. Methods* **2014**, *11* (11), 1127–1130.

(37) Robichon, C.; Luo, J.; Causey, T. B.; Benner, J. S.; Samuelson, J. C. *Appl. Environ. Microbiol.* **2011**, *77* (13), 4634–4646.

(38) Studier, F. W. *Protein Expr. Purif.* **2005**, *41*, 207–234.

(39) Collins, T.; Azevedo-Silva, J.; da Costa, A.; Branca, F.; Machado, R.; Casal, M. *Microb. Cell Fact.* **2013**, *12* (21), 1–16.

(40) Chow, D. C.; Dreher, M. R.; Trabbic-Carlson, K.; Chilkoti, A. *Biotechnol. Prog.* **2006**, *22* (3), 638–646.

(41) MacEwan, S. R.; Hassouneh, W.; Chilkoti, A. *J. Vis. Exp.* **2014**, No. 88, e51583.

(42) Laemmli, U. K. *Nature* **1970**, *227* (5259), 680–685.

(43) Anthis, N. J.; Clore, G. M. *Protein Sci.* **2013**, *22* (6), 851–858.

(44) Dorr, B. M.; Ham, H. O.; An, C.; Chaikof, E. L.; Liu, D. R. *Proc. Natl. Acad. Sci.* **2014**, *111* (37), 13343–13348.

(45) Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M. R.; Appel, R. D.; Bairoch, A. In *The Proteomics Protocols Handbook*; 2005; pp 571–607.

(46) Dietz, H.; Rief, M. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (5), 1244–1247.

(47) Theile, C.; Witte, M.; Blom, A. *Nat. Protoc.* **2013**, *8* (9), 1800–1807.

(48) Catherine, C.; Oh, S. J.; Lee, K.-H.; Min, S.-E.; Won, J.-I.; Yun, H.; Kim, D.-M. *Biotechnol. Bioprocess Eng.* **2015**, *20* (3), 417–422.

(49) MacKay, J. A.; Callahan, D. J.; FitzGerald, K. N.; Chilkoti, A. *Biomacromolecules* **2010**, *11* (11), 2873–2879.

FIGURE CAPTIONS

Figure 1. Cloning schematic

(A) The schematic describes the process of sequence independent PCR amplification of unique inserts (I, II, III) from the same template. The amplification of the first backbone (plasmid A) enables subcloning of the first three inserts, which leads to plasmid B. Plasmid B is linearized at the N-terminal ybbR-tag, as are all the following backbones. The new ELP amplicons can always be inserted upstream of the old ELP repeats. (B) Repetitive rounds of cloning add subsequently more ELP inserts until the desired length is achieved.

Figure 2. Post-translational ligation of the ELP peptide

(A) Schematic of the ELP constructs containing a N-terminal ybbR-tag and a C-terminal Sortase-tag. A post-translational one-pot reaction was used to fuse a CoA-647 fluorescent dye to the N-Termini via an Sfp-catalyzed reaction. In parallel, the eSortase fuses a GG-sfGFP towards the C-terminal LPETGG. (B) An image of a SDS gel obtained following dual labeling of ELPs under different reaction conditions and ELP lengths. The image shows only the red CoA-647 dye (ex: 530/28, em: 695/55 nm). (C) Fluorescent image of the same gel as in B, but this time with blue excitation (ex: 470/530, em: 530/28 nm), hence only the native GFP specific bands are visible. (D) Overlay of B and C plus additional UV illumination which excites tryptophan side group converted fluorophores enabled by the Bio-Rad Stain-Free^{TM} technology.

Figure 3. Cloud point characterization of the 10 – 60 pentapeptide ELP repeats

(A) shows the characteristic decrease in the transition temperature of the ELPs with increasing sodium chloride concentration and ELP length. (B) illustrates the relation between increasing transition temperature and decreasing molecular weight. (C) shows the correlation between pH, NaCl and transition temperature. (D) shows the concentration dependency of the transition temperature for the 60 pentapeptide ELP repeat. Data points for the plots were obtained from triplicates. Error bars represent the standard deviation of the average transition point.

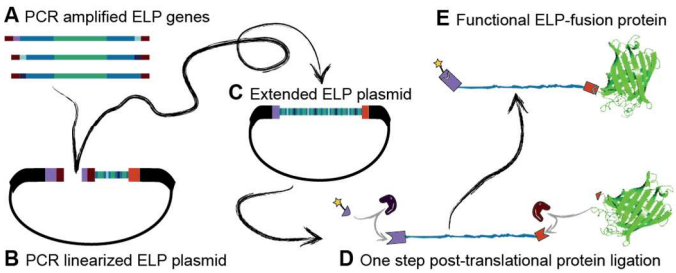
Figure 4. Cloning schematic and bioconjugation of cysteine-ELPs with a maleimide-dye and PEG-maleimide.

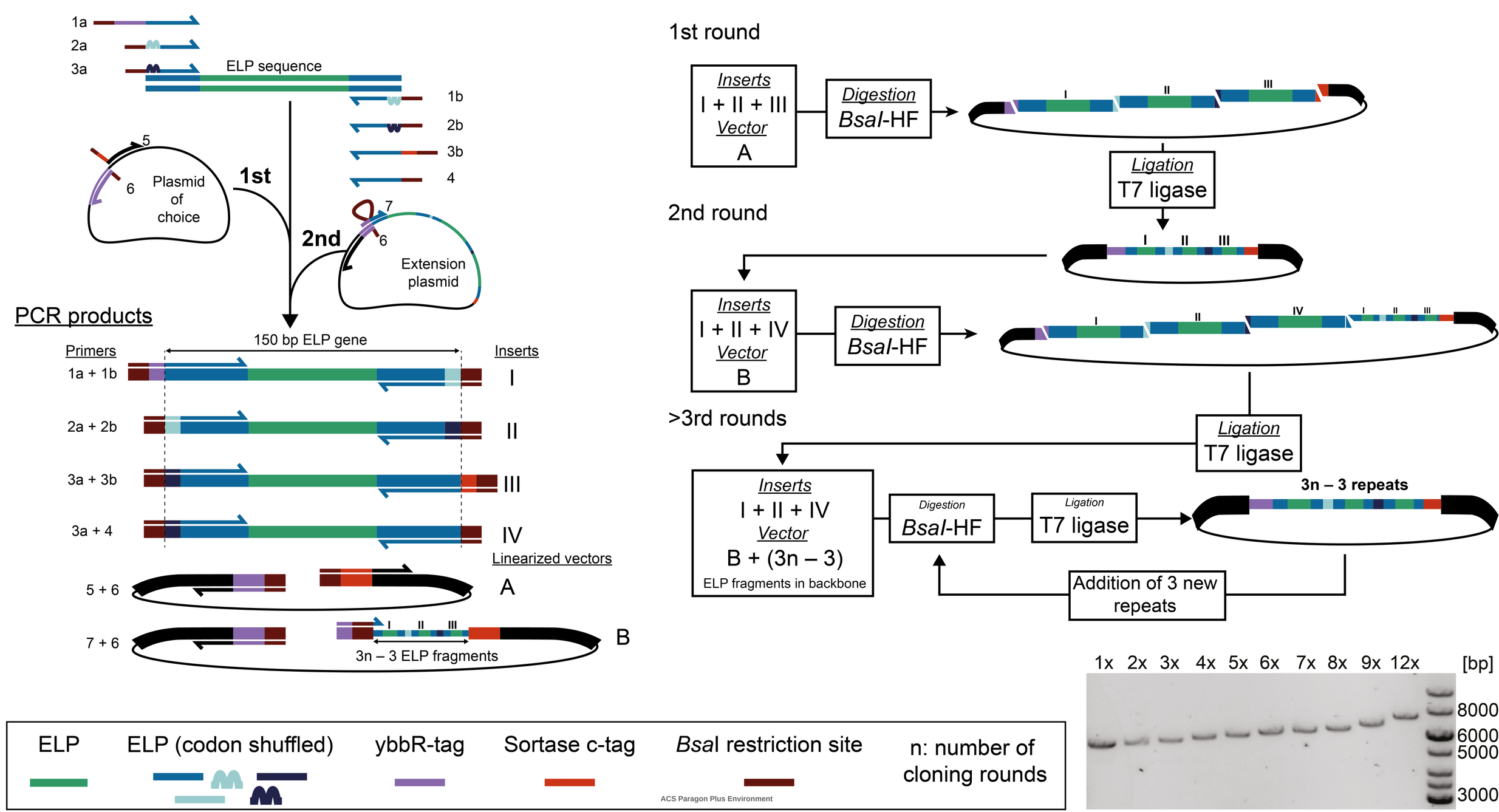
(A) Illustration of the cloning schematic for changing the ELP flanking regions (i.e. replacement of the ybbR-sequence with a cysteine). The ybbR-sequence was deleted *via* a PCR reaction, and a cysteine was introduced (see Primers 10 and 11 in Table 1). The flanking restriction sites were digested with *Bsa*I-HF and the remaining sticky ends were filled in with Klenow Fragment. Finally, the linear product was circularized with T4 ligase. (B) Procedure of the bioconjugation reaction. Cysteine-ELPs were reduced with TCEP and conjugated to a maleimide dye or a PEG-maleimide polymer. A gel image on the right shows the successful conjugation reaction between the dye and the ELP. (C) Bioconjugation of a 20 kDa PEG-maleimide to CYS-ELP_{60} shifted the cloud point up by ~4°C (left panel). The cloud point of ELP_{60} lacking cysteine (middle panel) was not influenced by the addition of maleimide PEG. The right panel shows the cloud point shift (ΔT) due to addition of different concentrations of PEG-maleimide. Data points for the plot were obtained from triplicates. Error bars account for Gaussian error propagation due to calculation of the difference of the average transition point from three samples.

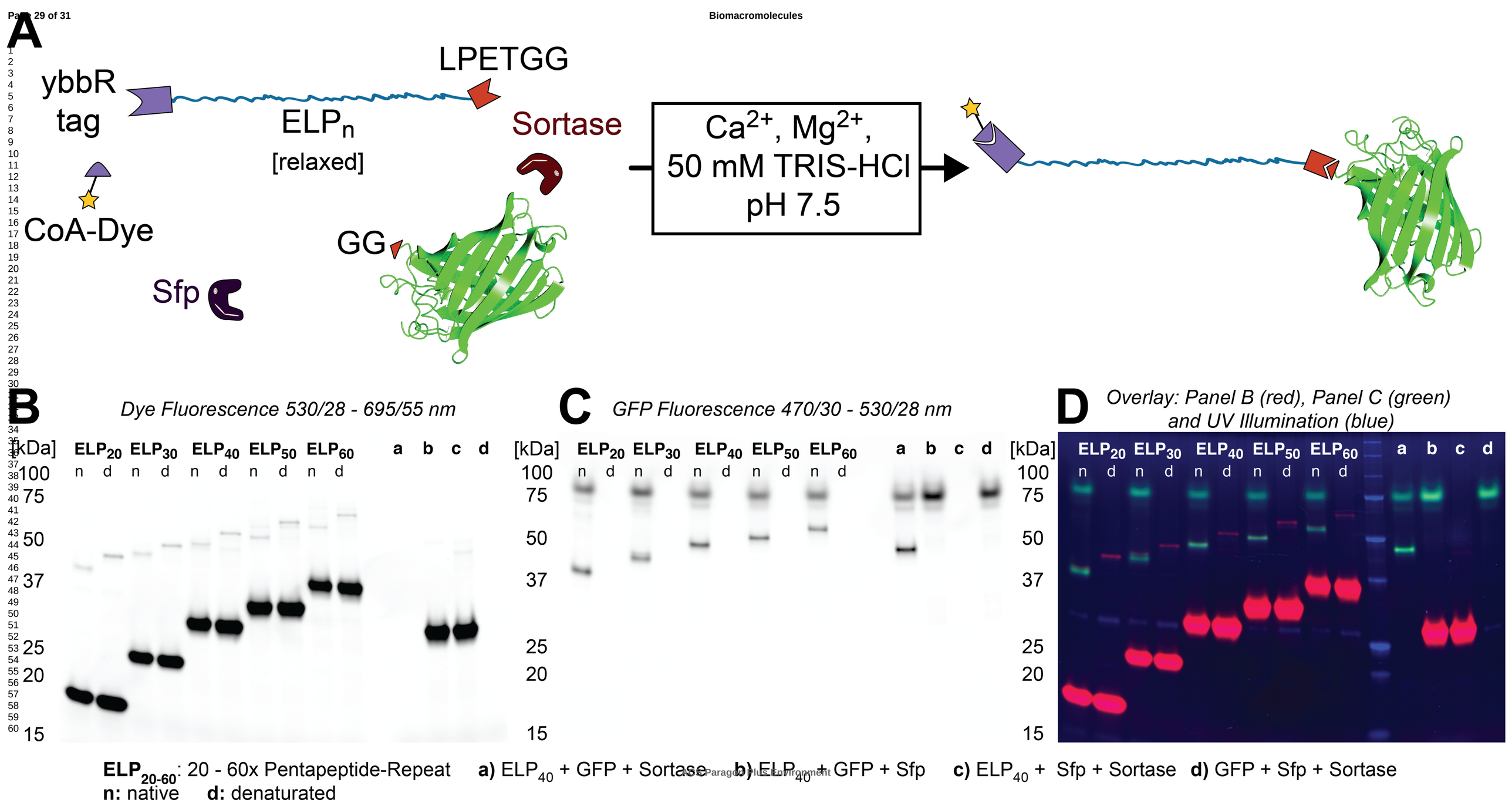
For Table of Contents Use Only

Sequence Independent Cloning and Post-translational Modification of Repetitive Protein Polymers through Sortase and Sfp-mediated Enzymatic Ligation

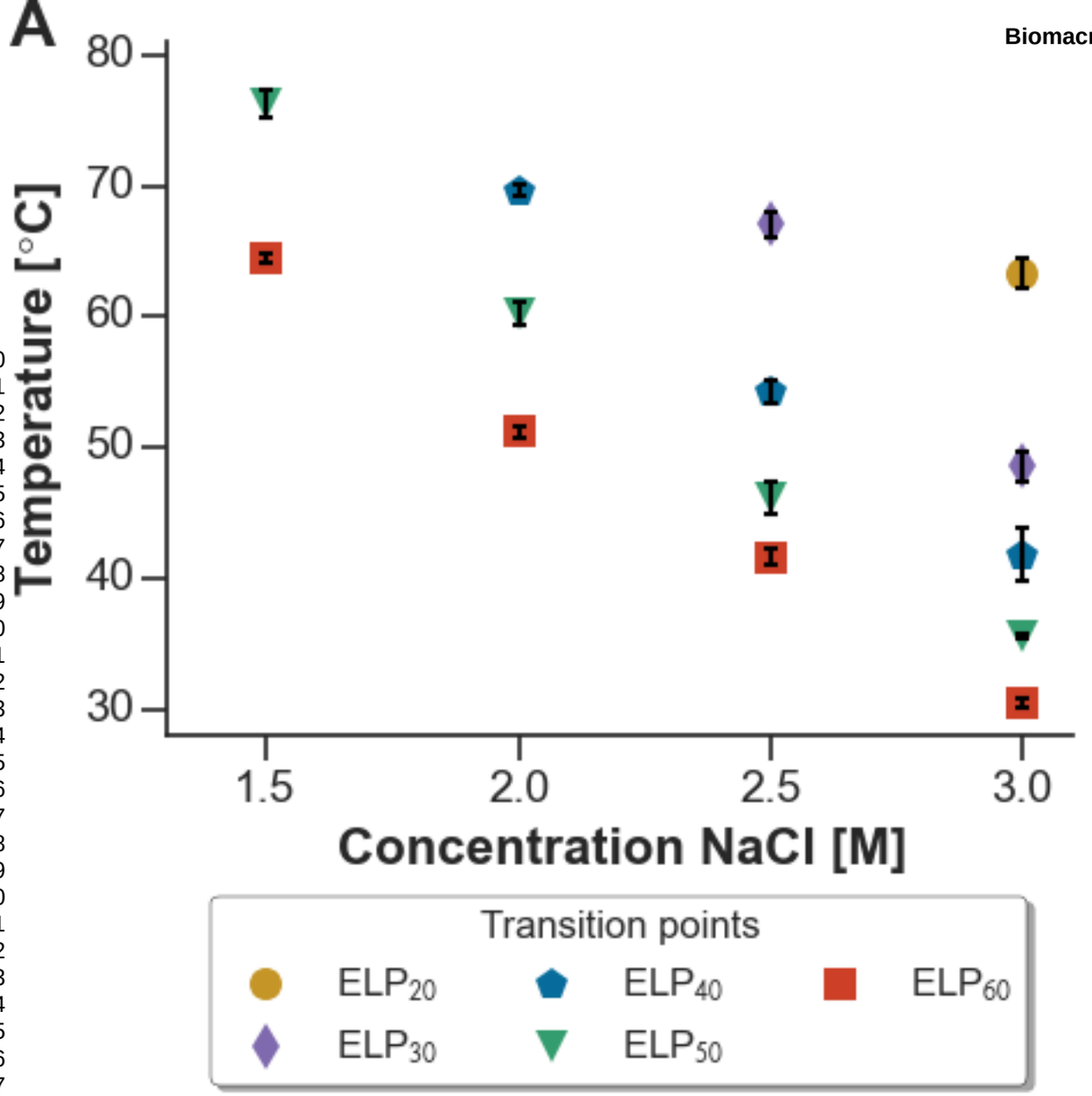
Wolfgang Ott, Thomas Nicolaus, Hermann E. Gaub and Michael A. Nash

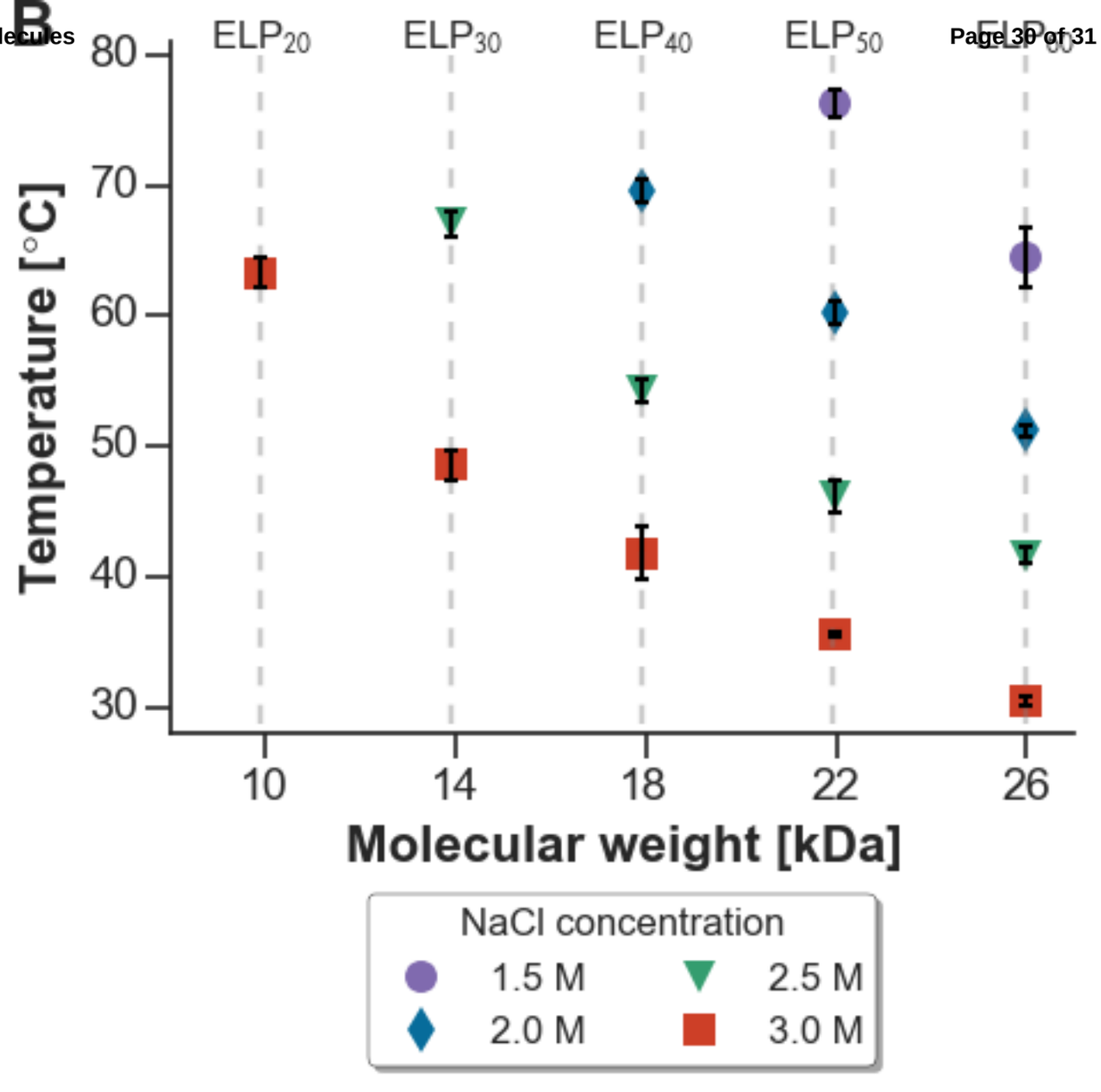


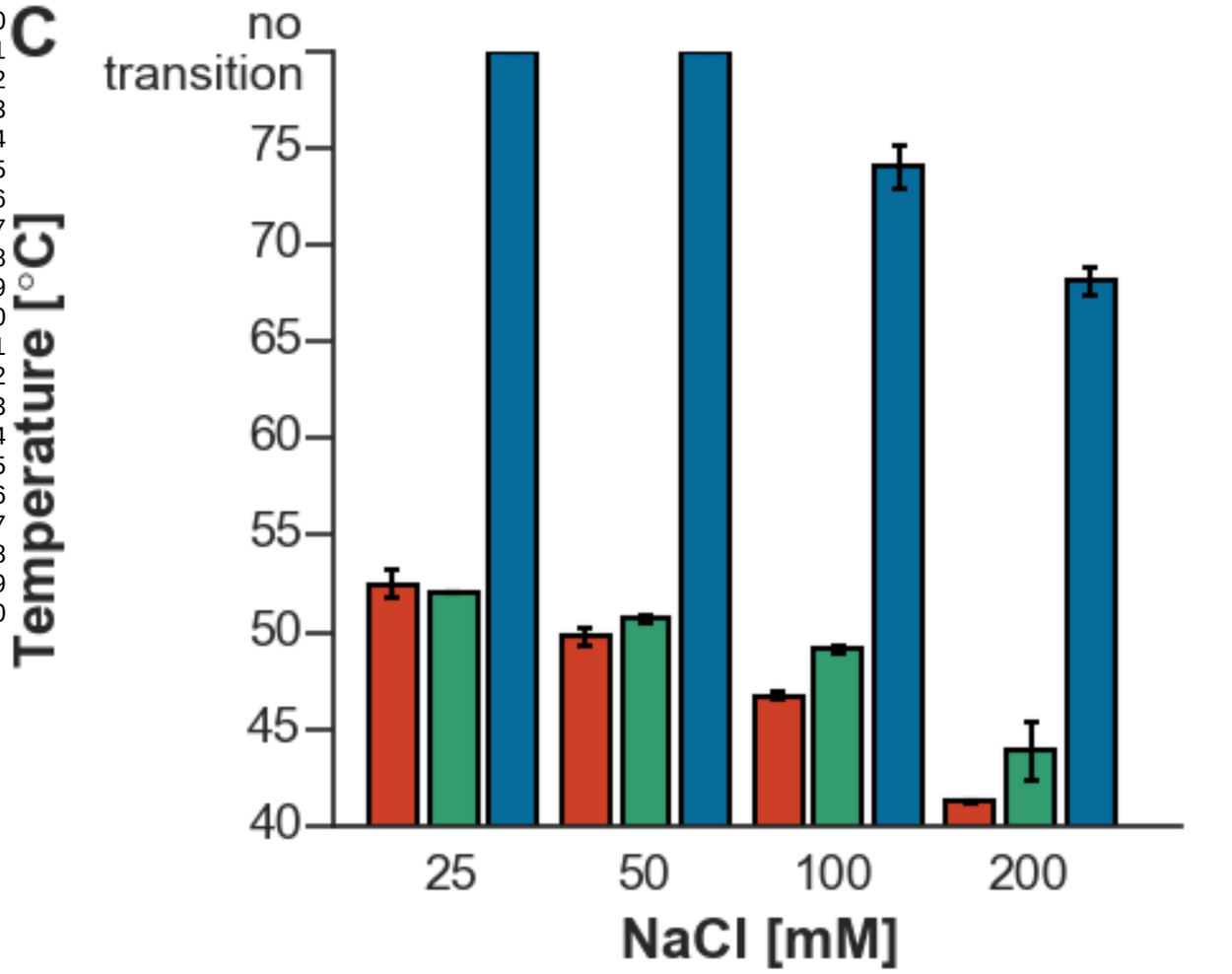

A
Primers and starting material
1a
2a
3a
ELP sequence
1b
2b
3b
4
5
6
Plasmid of choice
1st
2nd
7
6
Extension plasmid
PCR products
150 bp ELP gene
Primers
Inserts
1a + 1b
I
2a + 2b
II
3a + 3b
III
3a + 4
IV
Linearized vectors
5 + 6
A
7 + 6
I II III
B
3n – 3 ELP fragments
B
Cloning procedure
1st round
Inserts
I + II + III
Vector
A
Digestion
BsaI-HF
I II III
Ligation
T7 ligase
2nd round
I II III
Inserts
I + II + IV
Vector
B
Digestion
BsaI-HF
I II IV I II III
Ligation
T7 ligase
>3rd rounds
Inserts
I + II + IV
Vector
B + (3n – 3)
ELP fragments in backbone
Digestion
BsaI-HF
Ligation
T7 ligase
3n – 3 repeats
Addition of 3 new repeats
ELP
ELP (codon shuffled)
ybbR-tag
Sortase c-tag
BsaI restriction site
n: number of cloning rounds
ACS Paragon Plus Environment
1x 2x 3x 4x 5x 6x 7x 8x 9x 12x
[bp]
8000
6000
5000
3000

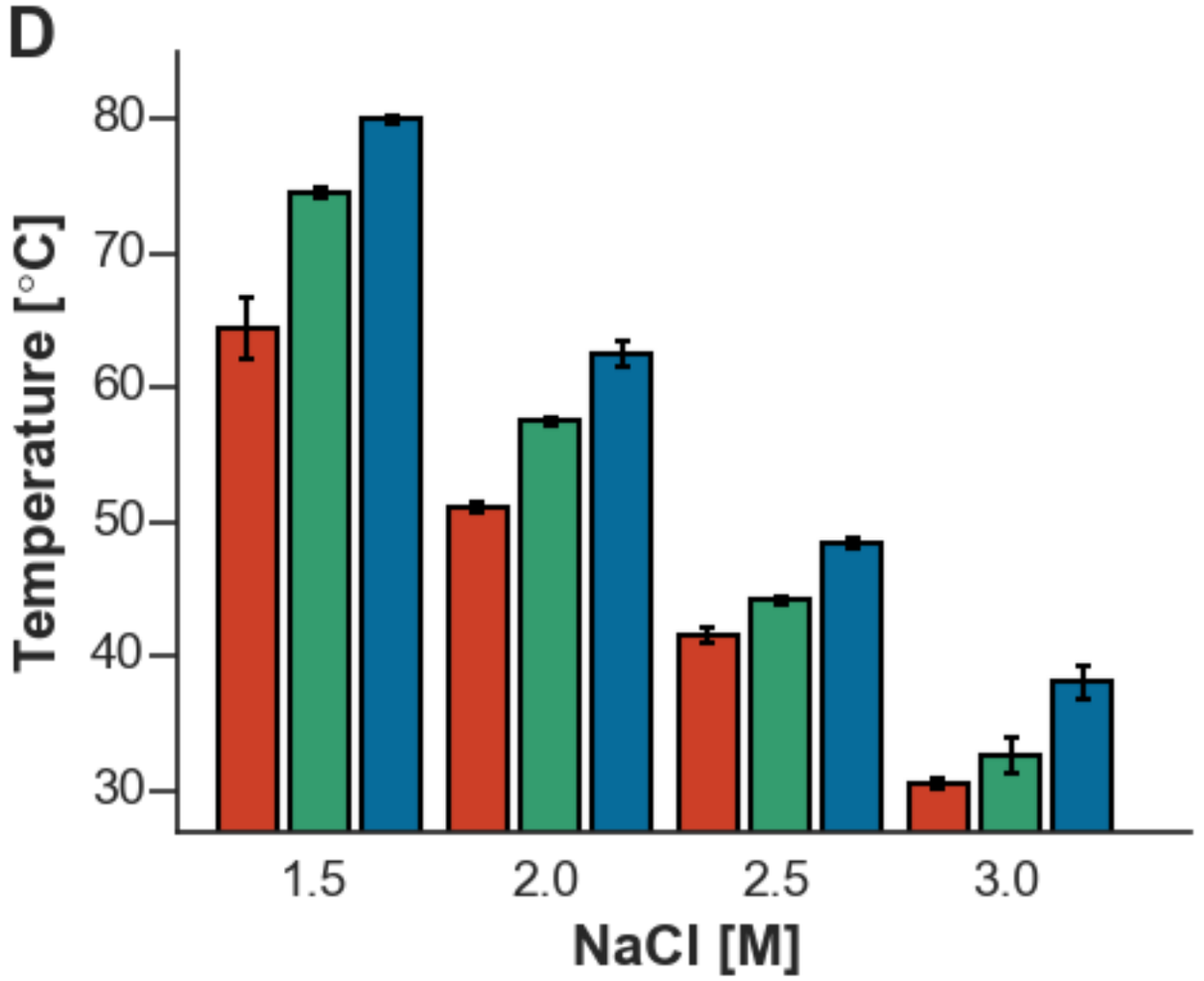


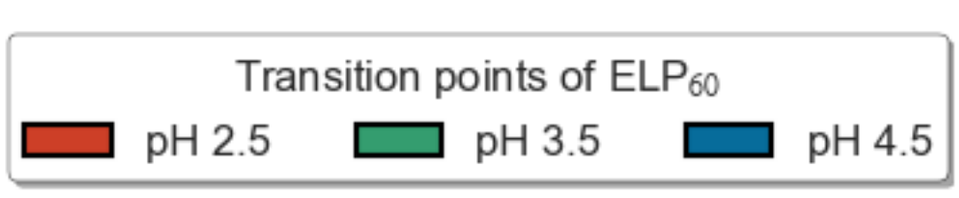
ELP$_{20\text{-}60}$: 20 - 60x Pentapeptide-Repeat **a)** ELP$_{40}$ + GFP + Sortase **b)** ELP$_{40}$ + GFP + Sfp **c)** ELP$_{40}$ + Sfp + Sortase **d)** GFP + Sfp + Sortase
n: native **d:** denaturated

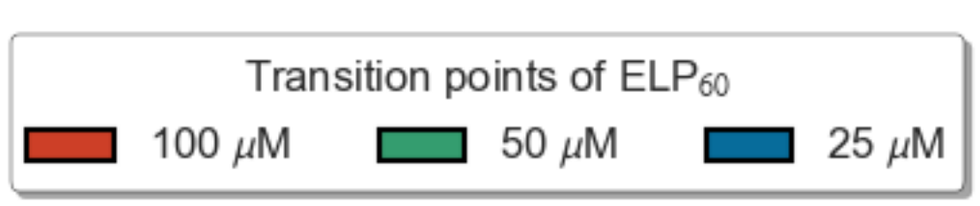

A
Temperature [°C]
80
70
60
50
40
30
1.5
2.0
2.5
3.0
Concentration NaCl [M]
Transition points
ELP20
ELP30
ELP40
ELP50
ELP60

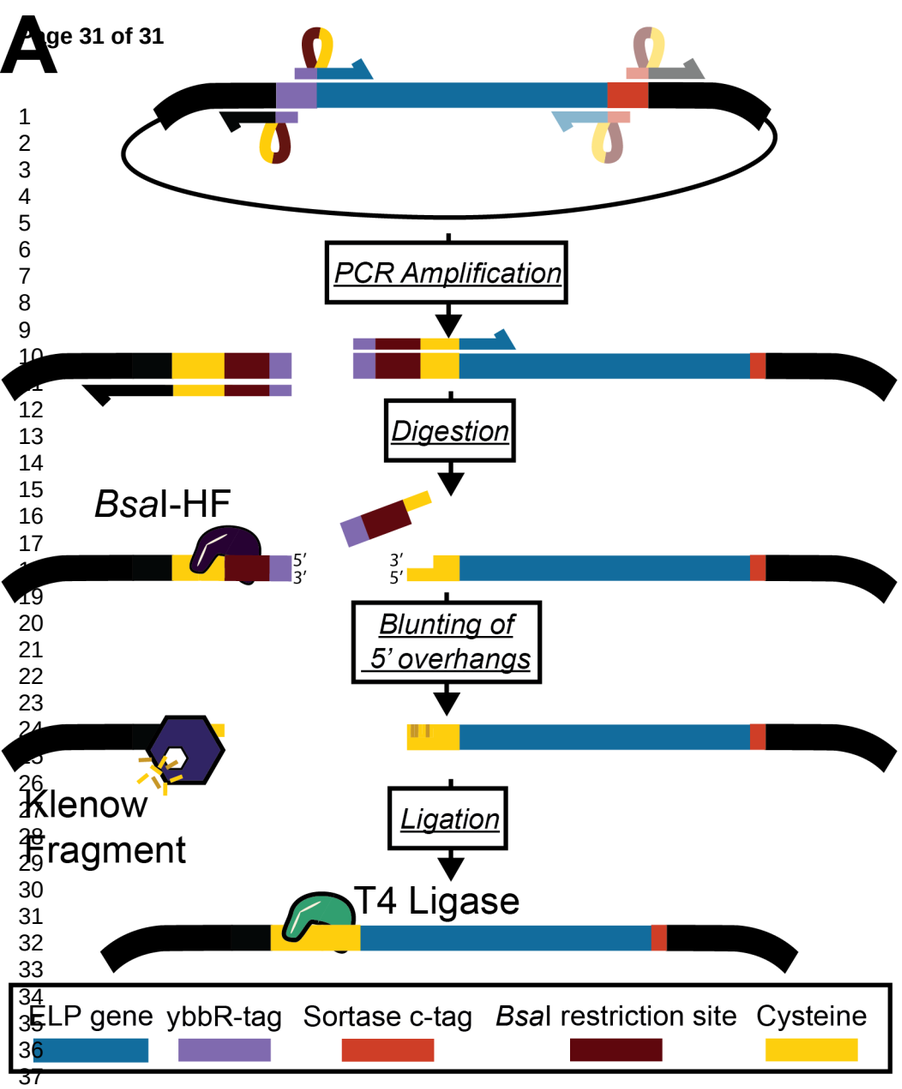

B
ELP20
ELP30
ELP40
ELP50
Temperature [°C]
80
70
60
50
40
30
10
14
18
22
26
Molecular weight [kDa]
NaCl concentration
1.5 M
2.0 M
2.5 M
3.0 M

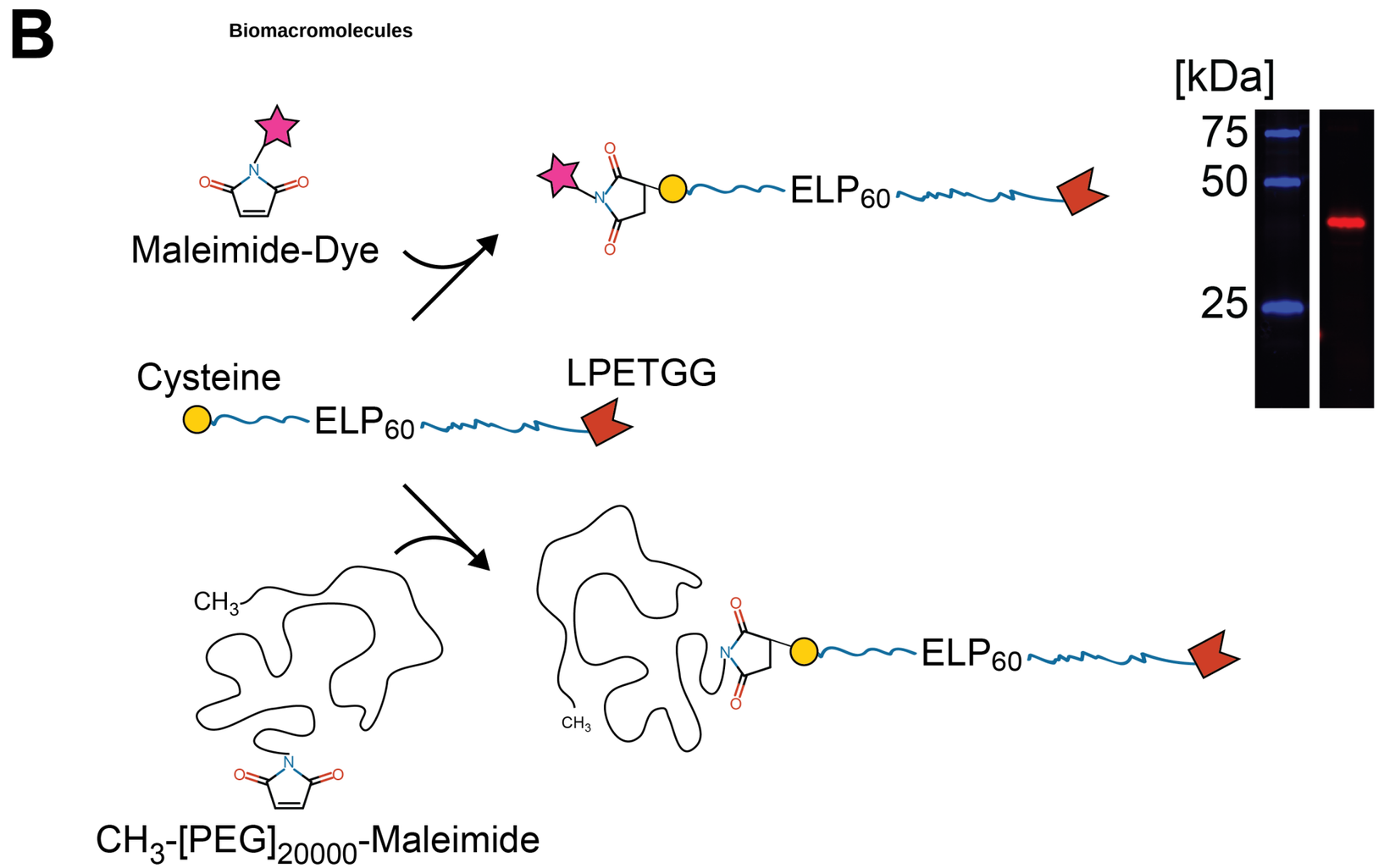

C
no transition
Temperature [°C]
75
70
65
60
55
50
45
40
25
50
100
200
NaCl [mM]

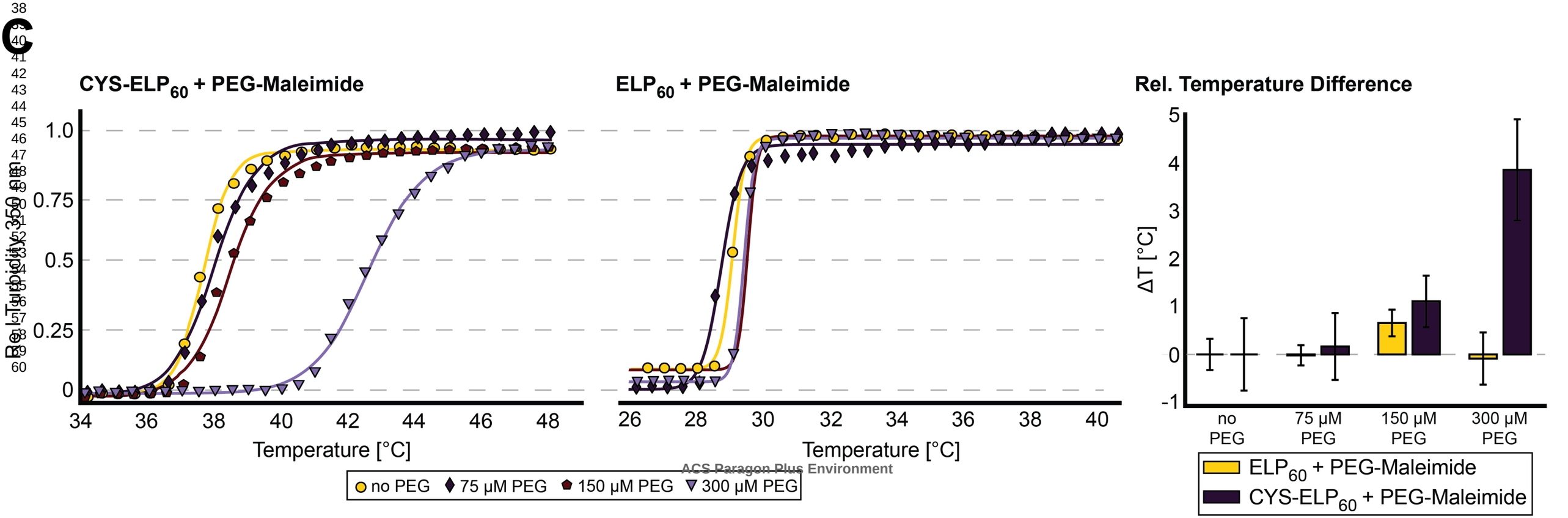

D
Temperature [°C]
80
70
60
50
40
30
1.5
2.0
2.5
3.0
NaCl [M]


Transition points of ELP60
pH 2.5
pH 3.5
pH 4.5


Transition points of ELP60
100 μM
50 μM
25 μM


A
PCR Amplification
Digestion
BsaI-HF
5′
3′
3′
5′
Blunting of 5′ overhangs
Klenow Fragment
Ligation
T4 Ligase
ELP gene
ybbR-tag
Sortase c-tag
BsaI restriction site
Cysteine


B
Maleimide-Dye
ELP60
Cysteine
LPETGG
ELP60
CH3
CH3
ELP60
CH3-[PEG]20000-Maleimide
[kDa]
75
50
25


C
CYS-ELP60 + PEG-Maleimide
ELP60 + PEG-Maleimide
Rel. Temperature Difference
Rel. Turbidity 350 nm
Temperature [°C]
Temperature [°C]
ΔT [°C]
no PEG
75 µM PEG
150 µM PEG
300 µM PEG
ELP60 + PEG-Maleimide
CYS-ELP60 + PEG-Maleimide
ACS Paragon Plus Environment