



Published in final edited form as:

Biomacromolecules. 2007 May ; 8(5): 1417–1424. doi:10.1021/bm060849t.

Improved Non-Chromatographic Purification of a Recombinant Protein by Cationic Elastin-like Polypeptides

Dong Woo Lim, Kimberly Trabbic-Carlson, J. Andrew MacKay, and Ashutosh Chilkoti*

Department of Biomedical Engineering, Box 90281, Duke University, Durham, North Carolina 27708-0281

Abstract

This paper reports an improvement in the purification of thioredoxin (Trx) expressed from *E. coli* by inverse transition cycling (ITC) using cationic elastin-like polypeptides (ELPs). Two ELP libraries having 2 and 5 % lysine residues and molecular weights ranging from 4 to 61.1 kDa showed greater salt sensitivity in their inverse transition behavior than purely aliphatic ELPs. Expression yield of Trx-ELP fusions was an unpredictable function of guest residue composition, but reducing the molecular weight of the ELP tag generally increased Trx yield. A cationic 4.3 kDa ELP is the shortest ELP used to purify any protein by ITC to date. A 15.9 kDa ELP with a guest residue composition of K:V: F of 1:7:1 was found to be the optimal cationic tag to purify Trx, as it provided 50% greater Trx yield and only required one-fifth the added NaCl for purification of Trx compared to previously used aliphatic ELP tags.

Keywords

Elastin-like polypeptides; environmental sensitivity; non-chromatographic protein purification

Introduction

A variety of affinity based chromatographic purification schemes have been developed to simplify protein purification at the laboratory scale in a serial format.¹ By expressing recombinant proteins fused to a peptide or protein having moderate affinity and high specificity for a particular ligand, recombinant proteins can be purified in a single step by binding to resins to which the ligand has been immobilized.² In practice, however, affinity chromatography is costly, requiring specialized equipment and customized resins, and scale-up to industrial scale is empirical, expensive, and often results in significant purification losses.

We have previously developed inverse transition cycling (ITC) of elastin-like polypeptide (ELP) fusion proteins as an alternative to affinity-based purification schemes.^{3,4} ELPs are artificial biopolymers comprised of the pentapeptide repeat motif, Val-Pro-Gly-Xaa-Gly (VPGXG) that is derived from the hydrophobic domain of tropoelastin. At low temperatures, ELPs are soluble in aqueous solution, but as the solution temperature is raised, they become insoluble and aggregate at a critical temperature, termed the inverse transition temperature (T_i).⁵⁻⁷ This process is reversible so that subsequent cooling of the solution below the T_i results in the resolubilization of the ELP. The inverse transition can also be isothermally triggered by

*To whom correspondence should be addressed: Department of Biomedical Engineering, Box 90281, Duke University, Durham, NC 27708; phone, (919) 660-5373; fax, (919) 613-9116; e-mail, chilkoti@duke.edu
dong.lim@duke.edu, kcarlson@duke.edu, jam54@duke.edu, chilkoti@duke.edu

the addition of salt, and the response of ELPs to different salts follows the Hofmeister series.^{5,8-10}

We were the first group to demonstrate that the inverse transition behavior of ELPs was imparted to ELP fusion proteins,³ and exploited this finding to develop a protein purification process that we have termed inverse transition cycling (ITC).^{3,4,11-15} This approach for protein purification has subsequently been validated by other groups.¹⁶⁻²⁴ The ITC method is inexpensive, as it requires no specialized equipment or resins; instead, it uses inexpensive reagents such as sodium chloride to trigger the inverse phase transition and readily available centrifuges to separate the ELP fusion protein from other cellular contaminants. Multiple rounds of ITC increase the purity of the ELP fusion protein, and target proteins can be liberated from their ELP purification tag either by proteolytic cleavage at engineered polypeptide sequences^{3,4,14,15} or by self-cleaving inteins.^{24,25}

Although previous studies have shown that ITC purification of ELP fusion proteins is a promising alternative to chromatography for purification of recombinant proteins, the ELP tags most commonly used for ITC have not been optimized for maximal protein expression or purification efficiency. To date, nearly all published examples of ITC purification of ELP fusion proteins have only utilized two types of ELP tags that were introduced in our original articles on ITC.^{3,4} These tags were not optimal, because they required at least ~2 M NaCl to induce the phase transition. This concentration of salt was necessary to trigger the inverse phase transition of ELP fusion proteins because these ELP tags had aliphatic guest residues that are only modestly sensitive to changes in salt concentration.^{3,4,10} Previous studies have shown that incorporation of ionizable guest residues in an ELP increases its sensitivity to salt.^{5,26} Thus, we hypothesized that the addition of ionizable groups in the ELP tag should allow the purification of ELP fusion proteins from *E. coli* soluble lysate with less added NaCl than is needed to purify ELP fusions containing only aliphatic guest residues.

In this study, we report on the phase transition behavior of charged ELPs and their thioredoxin (Trx) fusions. These cationic ELPs have guest residues composed of Lys (K), Val (V), and Phe (F) in 1:2:1 and 1:7:1 ratios and span a range of molecular weights (MWs) from 7.7 to 61.1 kDa. We chose ionizable Lys residues to enhance the salt sensitivity of these ELPs compared to purely aliphatic ELPs and hydrophobic Phe residues to lower the T_t of the ELP, and thereby counterbalance the effect that charged Lys residues have on the T_t . The systematic investigation of the salt sensitivity, phase behavior, and expression yield of cationic ELPs and their fusions with Trx reported in this study provide significant new information to optimize the molecular design of ELPs as protein purification tags for ITC.

Materials and Methods

ELP Nomenclature

All ELPs are named using the following notation: ELP[X_iY_jZ_k-n] where the bracketed capital letters are single letter amino acid codes of a guest residue, their corresponding subscripts denote the ratios of that guest residue in the monomer, and *n* indicates the number of pentapeptides in the ELP. For example, ELP[KV₇F-36] is an ELP that contains 36 repeats of the VPGXG pentapeptide, in which one pentapeptide contains K at the 4th, guest residue position (X), another pentapeptide has F at the 4th position, and 7 pentapeptides have V at the 4th position.

Monomer ELP gene design, synthesis, and oligomerization

Standard molecular biology protocols were used for all DNA manipulation steps in ELP gene cloning.¹² All restriction endonucleases and T4 DNA ligase were purchased from New

England Biolabs, (Beverly, MA), and calf intestinal alkaline phosphatase was obtained from Gibco BLR-Life Technologies (Grand Island, NY).

Monomer ELP genes of ELP[KV₂F-4] and ELP[KV₇F-9] (Figure 1A) were synthesized by ligating double stranded oligonucleotide cassettes that were created by either annealing 5' phosphorylated synthetic oligonucleotides (Integrated DNA Technologies, Coralville, IA) or were excised from previously synthesized ELP genes in pUC19.²⁶ Monomer genes were oligomerized by recursive directional ligation (RDL)¹² to create genes for ELP[KV₂F-8, 16, 32, 64, 128] and ELP[KV₇F-18, 36, 72, 144] that encode ELPs with molecular weights (MWs) ranging from 7.7 to 55.7 kDa and from 8.4 kDa to 61.1 kDa, respectively. ELP[KV₂F] and ELP[KV₇F] genes were ligated separately into the SfiI site of a modified pET-25b(+) vector (Figure 1B), as previously described.³ The same ELP genes were also ligated into the SfiI site of a modified pET32a(+) vector (Figure 1B), as previously described,³ to express the corresponding series of Trx-ELP fusion proteins.

Gene expression and purification of ELPs and Trx-ELP fusion proteins

The *E. coli* strain, BLR (DE3) cells (Novagen Inc., Milwaukee, WI) were transformed with the modified pET-25b(+) and pET32a(+) expression vectors containing the ELP and Trx-ELP genes. *E. coli* were cultured and purified as previously described.^{3,4} Briefly, 50 ml of transformed cells were cultured in triplicate for 24 h, without induction, at 37 °C with orbital shaking at 200 rpm in CircleGrow™ media (Qbiogene, Carlsbad, CA) supplemented with 100 µg/ml of ampicillin.¹⁴ ELP and Trx-ELP fusions were purified by ITC, as previously described.³ Briefly, cells from 50 ml cultures were harvested by centrifugation at 4 °C, resuspended in PBS buffer (137 mM NaCl, 2.7 mM KCl, 4.2 mM NaH₂PO₄, 1.4 mM KH₂PO₄, pH 7.3) supplemented with a protease inhibitor cocktail (Roche catalog number 1697498) and lysed by ultrasonic disruption at 4 °C. The lysate was centrifuged in order to remove insoluble cell debris. Nucleic acids were precipitated by the addition of polyethyleneimine (0.5 % w/v final concentration) and were removed by centrifugation. The inverse phase transition of the ELP fusion proteins were triggered by the addition of NaCl (0.25-3 M), and the aggregated ELP proteins were separated from the lysate by centrifugation at moderate temperatures (25-37 °C). The aggregated fusion proteins were resolubilized in cold PBS buffer, and the resolubilized protein solutions were centrifuged at 4 °C to remove any particulate contaminants. This aggregation and resolubilization process was repeated 3-4 times until purity of the fusion proteins was approximately 95%, as ascertained by SDS-PAGE.

Characterization of ELPs and Trx-ELP fusion proteins

The concentration of ELPs and Trx-ELP fusion proteins was determined by their molar extinction coefficients at 280 nm (5690 M⁻¹cm⁻¹ for ELPs and 19870 M⁻¹cm⁻¹ for Trx-ELPs) calculated from their primary amino acid sequence with the software program, Protean (DNA Star). Their absorption at 280 nm was measured by UV-visible spectrophotometry (UC-1601, Shimadzu Scientific Instruments). Purity and molecular weight of ELPs and Trx-ELP fusion proteins were characterized by SDS-PAGE (BioRad, Inc., Hercules, CA) with copper staining and matrix-assisted laser desorption/ionization mass spectroscopy (MALDI-MS), performed on a PE Biosystems Voyager-DE instrument equipped with a nitrogen laser (337 nm). The ELP samples for MALDI-MS measurements were prepared in an aqueous 50% (v/v) acetonitrile solution containing 0.1 % (v/v) trifluoroacetic acid, using a sinapinic acid matrix.

The effect of increasing temperature on the turbidity of the ELP solutions was measured at 350 nm on a Cary 300 UV-visible spectrophotometer equipped with a multicell thermoelectric temperature controller (Varian Instruments, Walnut Creek, CA) between 10°C and 90°C at a rate of 1°C/min. The inverse transition temperature (*T_i*) of each ELP fusion protein was defined as the temperature at which the first derivative of the turbidity as a function of temperature was

the maximum. The sensitivity of T_t to ionic strength was defined by the slope of a linear fit of the T_t as a function of NaCl concentration ($\Delta T_t/M$ NaCl) in phosphate buffered saline (PBS).

Statistical Analysis

To quantitatively evaluate the linearity of the ELP and Trx-ELP salt sensitivity trends, the T_t data for each ELP as a function of salt was fit by linear regression. The residual at each data point was determined from the difference between the observation and the best-fit line, and the squared residuals were then \log_{10} transformed to meet equality of variance assumptions. The transformed fit residuals were compared by a three-way ANOVA using SPSS version 14.0.0 (Chicago, IL). Three variables were considered: (1) the ELP library (KV₇F or KV₂F); (2) the presence of Trx (yes or no); and (3) the polymer length (short, medium, or long).

Results and Discussion

Purification of cationic ELPs and their Trx fusions

Figure 2 shows copper stained SDS-PAGE gels of purified ELP[KV₂F] and ELP[KV₇F] of differing molecular weights as well as their corresponding Trx-ELP fusion proteins. The gels show that the purified ELPs and their Trx-ELP fusions migrate to positions that are approximately consistent with molecular weights 20% greater than the molecular weights calculated from their engineered DNA sequences; this observation is typical of previously characterized ELPs.^{12,27} However, MALDI-MS indicated that for all ELPs and their Trx fusions, experimentally measured molecular weights were within 0.1-0.5% of their calculated molecular weights.

ELP[KV₂F-8] and ELP[KV₇F-9], with only 40 and 45 amino acids (4.3 kDa and 4.6 kDa), respectively, are the shortest tags, to date, that have been used to purify an ELP fusion protein by ITC; however their corresponding free ELPs could not be purified by ITC. Although the exact reasons for this remain unclear, two problems -the decrease of ELP expression (data not shown) and the increase in ELP T_t with decreasing molecular weight- compound the purification of very low molecular weight ELPs (Table 1). These combined effects are likely responsible for the observations that the inverse phase transition of ELP[KV₂F-8] and ELP [KV₇F-9] could not be detected in cell lysate and no ELP could be retrieved by ITC. However, Trx fusions to these very short ELPs were successfully purified by ITC, which is consistent with the facts that Trx is an over-expressed protein in *E. coli*, and fusion to Trx lowers the T_t of these ELPs, such that the addition of 3 M NaCl is enough to trigger the phase transition of the fusion protein.

The copper-stained SDS-PAGE gel in Figure 2B shows several smaller MW bands in the purified Trx-ELP[KV₂F-32, 64, and 128] products. Analysis of MALDI-MS data indicates that all of the Trx-ELP[KV₂F] fusions, except Trx-ELP[KV₂F-8], contain minor contaminants, the largest of which is consistently 13.4 ± 0.45 kDa smaller than the Trx-ELP fusion protein from which it is derived. This difference in size approximately corresponds to the MW of Trx that would be liberated from the ELP by enzymatic cleavage with thrombin (13.9 kDa), suggesting that there are two or three sites within the SSGLVPRGS linker region of the Trx-ELP fusion protein that are susceptible to enzymatic attack by *E. coli* proteases. Because this minor contamination product likely consists of the cleaved ELP fragment of the fusion, it is co-purified with the Trx-ELP fusion through multiple rounds of ITC and hence, is present in the final purified product. Furthermore, we hypothesize that ELP[KV₂F] fusions are more susceptible to proteolysis than ELP[KV₇F] fusions because of conformational differences caused by different interactions between Trx and ELP domains in the fusion proteins. Clues to these conformational differences can be elucidated from analysis of the thermal properties of these ELPs and their Trx fusions (*vide infra*).

Thermal Characterization of ELPs and Trx-ELP fusion proteins

Figure 3 and Table 1 show the T_t s of all expressed ELPs and Trx-ELP fusion proteins at a concentration of 25 μ M in PBS over the experimentally accessible temperature range of 20 to 90 $^{\circ}$ C. The T_t s of a number of ELPs and Trx-ELP fusions could not be accurately measured under these solution conditions because they were higher than 90 $^{\circ}$ C (Table 1). In general, the T_t s of members of the ELP[KV₂F] series are 16 to 42 $^{\circ}$ C higher than those of the ELP[KV₇F] series of similar MW, reflecting their > 2-fold greater lysine content. ELP[KV₂F] series ELPs have T_t s that are very similar to ELPs of similar MW having a guest residue composition of 50% valine, 20% alanine, and 30% glycine (ELP[V₅A₂G₃]),^{12,13} the most widely used ELP fusion protein purification tag.^{3,4,14,15} In contrast, the ELP[KV₇F] series exhibit T_t s approximately 20 $^{\circ}$ C lower than their size-matched ELP[V₅A₂G₃] counterparts.

We have previously reported that the T_t of a set of aliphatic ELPs could be predicted by a simple equation which accounts for the ELP concentration and its length, where the T_t is inversely proportional to the length of the ELP.¹³ Figure 3 shows that the T_t s of these cationic ELPs and Trx-ELP fusions also exhibit a linear relationship with 1/length of ELP pentapeptides. ELP[KV₂F] polypeptides and Trx fusions have steeper slopes than their corresponding ELP[KV₇F] counterparts, which indicates that the fraction of Lys residues in these cationic ELPs dominates their change in T_t with ELP length. Although we were only able to measure the T_t at 25 μ M in PBS for a small subset of the ELP and Trx-ELP proteins ($T_t < 90^{\circ}$ C), limiting our ability to measure the goodness of the linear fits, these data suggest that this linear correlation is not limited to aliphatic ELPs. It appears that cationic ELPs and their Trx fusions also exhibit linear correlations with 1/ELP length as well. Thus, the T_t s of ELPs and Trx-ELPs can be reasonably approximated by these correlations, allowing the T_t s for new ELP lengths to be predicted for these amino acid compositions. This knowledge provides a predictive tool to control the phase transition behavior of ELPs and their fusion proteins for different applications.

Figure 3 and Table 1 also show the effect that fusion of Trx to these ELPs has on the ELP T_t . In general, fusion of Trx to ELP[KV₂F] and ELP[KV₇F] results in a depression in the ELP T_t relative to the free ELP (negative fusion ΔT_t parameter), and the magnitude of this depression depends on the ELP sequence. ELP[KV₂F] sequences show a greater fusion ΔT_t parameter than ELP[KV₇F] sequences. Only ELP[KV₇F-144] exhibits a positive fusion ΔT_t parameter, as the T_t of its Trx fusion is elevated by approximately 2 $^{\circ}$ C. The negative fusion ΔT_t parameter, observed for most of these ELPs, contrasts with the results of a previous study of Trx-ELP fusion proteins, in which fusion of Trx to an ELP having only aliphatic guest residues resulted in an elevation of the ELP T_t by $\sim 6^{\circ}$ C.¹⁵

These previous studies showed that the fusion ΔT_t parameter exhibits a linear negative correlation with the fraction of hydrophobic surface area of the fused protein. These studies were performed using a single ELP tag, ELP[V₅A₂G₃-90] and different fusion partners.¹⁵ In this study, however, we have a single protein, Trx, but the composition and the hydrophobicity of the ELP are altered by variations in both guest residue composition and ELP length.

There are two plausible explanations for the observed changes in the sign of the fusion ΔT_t parameter: enhanced interaction between Trx and ELP domains via hydrophobic contacts or electrostatic attraction between positively charged Lys residues in the ELP and negatively charged acidic residues on the surface of Trx. We note that the latter possibility is unique to the cationic ELPs studied here, as the aliphatic ELP[V₅A₂G₃-90] in the previous study only had Val, Ala and Gly at the guest residue position. The T_t s of ELP[KV₇F] proteins are 16 to 42 $^{\circ}$ C lower than ELP[KV₂F] polypeptides indicating that for the same concentration and similar molecular weight they are more hydrophobic; however, when fused to Trx, ELP[KV₂F] polypeptides exhibit a greater negative fusion ΔT_t parameter, indicating that in this

case it is not the net hydrophobic character of the ELP that is responsible for the changes in the fusion ΔT_t parameter. Instead, the sign and magnitude of their fusion ΔT_t parameter are consistent with electrostatic attraction between the positively charged lysine residues on these ELPs and negatively charged Asp and Glu residues present on the surface of Trx. The isoelectric point of Trx, calculated from its primary sequence is 4.5, indicating that at pH 7, each Trx molecule has a formal charge of nearly -5, so that it presents a negatively charged surface with which positively charged ELPs may favorably interact. Electrostatic interaction between the cationic ELPs and Trx then results in the neutralization of charged moieties in the ELP, which lowers the T_t of the fused ELP, resulting in a negative ΔT_t parameter. As ELP [KV₂F] tags have more than twice the fraction of charged lysine residues, their T_t s are depressed to greater extent by fusion to Trx than ELP[KV₇F], so that they have greater negative ΔT_t parameters.

Salt sensitivity of cationic ELPs and their Trx fusion proteins

Figure 4 shows the effect of NaCl concentration on the T_t of the ELPs (Figures 4A and 4C) and their Trx fusions (Figures 4B and 4D). As is typical of all ELPs, the T_t decreases with NaCl concentration, and this trend holds true for their Trx fusions as well. Comparison of panels A and C in Figure 4 shows that the T_t of ELP[KV₂F] proteins are more greatly affected by NaCl concentration than ELP[KV₇F] proteins, as indicated by their steeper slopes. These trends can be quantitatively evaluated by fitting linear correlations to the data in Figure 4A-D.

Figure 5A plots the slopes of these linear correlations as a function of ELP length for both the ELPs and their Trx fusions, and Figure 5B and 5C show the average of the squared fit residuals for each correlation for ELP[KV₇F] and ELP[KV₂F] proteins with and without the presence of Trx. Only proteins whose T_t could be measured ($T_t < 90^\circ\text{C}$) in PBS for both the ELP and corresponding Trx-ELP protein have been plotted and statistically analyzed to determine the effects of ELP composition and added Trx on the error in the linear fit.

Free ELP[KV₂F] polypeptides exhibit the greatest sensitivity to NaCl, followed by Trx-ELP [KV₂F], ELP[KV₇F], and Trx-ELP[KV₇F]. Although the slopes plotted in Figure 5A are derived from linear fits to the data in Figure 4, it is clear from Figure 4 as well as Figures 5B and 5C that not all the ELPs and their Trx fusions exhibit a linear change in T_t with NaCl concentration. The salt sensitivity of ELP[KV₇F] polypeptides and their Trx fusion proteins are better approximated by a linear fit than the ELP[KV₂F] and their Trx fusions. Figure 5B and 5C shows the average squared fit residuals for the ELPs and their corresponding Trx fusions for which T_t could be measured ($T_t < 90^\circ\text{C}$) in PBS without any added salt. Figure 5C shows that ELP[KV₂F] proteins have average squared fit residuals between 20 and 45 compared to ELP[KV₇F] proteins (Figure 5B), which have average squared fit residuals of less than 1 for the proteins of similar length. When similar length proteins are directly compared, the average squared fit residuals for ELP[KV₂F] proteins are as much as 270 fold higher than ELP[KV₇F] proteins with ELP[KV₂F] proteins. The deviation from linearity is correlated with the net charge on the ELP, as the more highly charged ELP[KV₂F] polypeptides having 5% cationic residues are more poorly fit by a linear function, than the ELP[KV₇F] polypeptides, which have only 2% cationic content.

We believe that electrostatic repulsion between the ELP and Trx are responsible for the nonlinear changes in T_t with salt concentration. Previous studies of purely aliphatic ELPs (with no ionizable amino acid content) have shown that T_t is linearly correlated with NaCl ($-14^\circ\text{C}/\text{M NaCl}$).¹⁰ ELP[KV₇F-72], which has only 2% positively charged lysine residues, exhibits a $19^\circ\text{C}/\text{M NaCl}$ depression and can be reasonably fit by a linear function having an average squared residual of only 0.84. In contrast, ELP[KV₂F-64], which has more than twice the cationic character (5%), exhibits a $40^\circ\text{C}/\text{M NaCl}$ depression in T_t and is *poorly fit* by a linear model, exhibiting an average squared fit residual of 43. Addition of Trx to both these polymers

not only reduces the slope of the T_t depression to 17 and 26 °C/M NaCl for ELP[KV₇F-72] and ELP[KV₂F-64] proteins, respectively. Furthermore, it also reduces the average squared fit residuals by 1.8 and 7.2 fold, respectively. We believe these changes in slope and linear fit error are a function of Coulombic repulsion, which is a non-linear effect of the solution ionic strength as described by Debye-Huckel screening. The Debye length, which is a measure of the distance over which electrostatic effects are damped out by mobile ions in an aqueous solution, is inversely proportional to the *square root* of the solution ionic strength.²⁸ Thus, in low ionic strength buffer, like charges in the ELP would repel one another over greater distances inhibiting aggregate formation and leading to elevated T_t s. This charge screening hypothesis is consistent with the fact that the greatest non-linearity in T_t as a function of ionic strength is observed for the ELPs with the greatest cationic content and that the addition of the negatively charged ELP reduces the net charge in the fusion leading to increased linear behavior in the T_t with NaCl concentration.

Overall trends in the linearity of fit behavior and the significance of fused Trx to this behavior were determined by statistical analysis of the squared fit residuals. A global 3-way ANOVA was found to significantly fit the entire dataset ($F_{9,30}=12.9$, $p=4.0\times 10^{-8}$, $R^2=0.795$). The statistical analysis pointed to two significant main effects of (1) ELP identity ($F_{1,30}=76.1$, $p=1.0\times 10^{-9}$); and (2) the fusion of Trx ($F_{1,30}=10.2$, $p=0.0033$), confirming the visually observable differences between the thermal behavior ELP[KV₇F] and ELP[KV₂F] polymers. This model also controlled for an expected significant main effect of ELP length ($F_{2,30}=9.42$, $p=0.00067$) and a two-way interaction between ELP and length ($F_{1,30}=11.2$, $p=0.0023$). Since the primary goal of this analysis was to compare deviation from linearity between ELP [KV₇F] and ELP[KV₂F] with and without fused Trx, and since there was no interaction between Trx and ELP, we did not split the dataset to further explore the significance of length on the ELPs and corresponding Trx-ELPs. The main effects from 3-way ANOVA of Trx and ELP on the linearity of T_t with added NaCl indicate two things. First, the charge density of ELP has a significant effect on linearity between T_t and added salt. ELP[KV₂F] was significantly ($p=1.0\times 10^{-9}$) more non-linear than ELP[KV₇F]. Second, the fusion of a negatively charged Trx protein to cationic ELPs overall significantly improved the linearity between T_t and added salt ($p=0.0033$). The significant improvement in linearity upon fusion to Trx suggests that the fusion of the negatively charged Trx to the cationic ELPs effectively reduces the net cationic charge in the fusion protein, promoting T_t salt sensitivity that more closely approximates that exhibited by non-ionic ELPs. This indicates that ELP to fused protein interactions are likely important to the formation of aggregates above T_t , especially at lower ionic strengths. Furthermore, these data suggest that ELP/target protein interactions are improved by matching anionic target proteins with cationic ELPs. By extrapolation, we suggest that cationic proteins should be fused to anionic ELPs in order to improve ELP-protein interactions in the aggregated state.

Because of the heightened sensitivity of charged ELPs and their Trx fusions to NaCl, we would anticipate that Trx-ELP would be purified from soluble *E. coli* lysate with a minimum of added salt by using a relatively short ELP[KV₂F] fusion tag; however, in practice we found that more NaCl was necessary to purify Trx-ELP[KV₂F] proteins from soluble lysate than Trx-ELP [KV₇F] proteins. This unanticipated problem was found to result from the strong dependence of fusion protein yield on the ELP, which altered the ELP concentration in the soluble lysate. Figure 6 shows the quantitative yield of the Trx-ELP fusion protein and the corresponding mass of Trx target protein calculated based on its mass fraction in each fusion for short (30-36 pentapeptides) and long (64-90 pentapeptides) ELPs from three different series: ELP[KV₇F], ELP[KV₂F], and ELP[V₅A₂G₃]. The last series is included because it is the most commonly used ELP tag for purification of recombinant proteins. Figure 6A shows that the yield of Trx-ELP[KV₇F-36] is approximately 2.5 times higher than Trx-ELP[V₅A₂G₃-30] and 3.6 times higher than Trx-ELP[KV₂F-32] proteins, suggesting that the yield of Trx-ELP fusion proteins

is dependent on the stoichiometry of the guest residue position of the fused ELP tag, which may arise from changes in mRNA stability and translation efficiency of the ELP tag.²⁹

In addition, decreasing the length of the ELP tag can be advantageous for fusion protein expression: we previously showed that reducing the length of the ELP purification tag from 90 to 20 pentapeptides, resulted in a ~4 fold increase in the yield of Trx in the ELP fusion.⁴ This is consistent with the data in Figure 6B. The yield of Trx from the shorter fusion proteins is 1.1-1.6 times that of the yield from the corresponding longer ELP fusions. Trx-ELP [KV₇F-36] has more than 1.6-fold greater Trx yield than that from Trx-ELP[KV₇F-72], which is more than 2.5-fold greater than that from Trx-ELP[V₅A₂G₃-90], which utilizes the most commonly used ELP purification tag. Furthermore, the Trx yield from Trx-ELP[KV₇F-32] is approximately 1.5-fold greater than that from Trx-ELP[V-20], the fusion boasting the highest Trx published yield to date.⁴ Although ELP[KV₇F-36] is somewhat longer than ELP[V-20] it has a significant advantage as a purification tag since the Trx fusion was completely retrieved from the *E. coli* lysate without any loss during ITC. In contrast, Trx-ELP[V-20] forms nanoparticles, which cannot be easily recovered by centrifugation, which is a well-documented problem observed in the purification of Trx-ELP[V-20].⁴

To deconvolute the coupled effects that ELP T_t, salt sensitivity, and Trx-ELP expression yield have on the purification efficiency of Trx-ELP fusion proteins, the T_t of Trx-ELP fusion proteins was measured at a fixed concentration in *E. coli* soluble lysate as well as in PBS. Figure 7 shows the difference in T_t between 0.5 M and 1.0 M NaCl ($\Delta T_t(\text{NaCl})$) added to 100 μM Trx-ELP solutions in PBS and soluble *E. coli* lysate for Trx-ELP[KV₇F-36], Trx-ELP [KV₂F-32], and Trx-ELP[V₅A₂G₃-90]. These data indicate that the sensitivity of cationic ELP fusion proteins is maintained in the soluble lysate. For each of the proteins investigated the $\Delta T_t(\text{NaCl})$ in lysate and PBS do not exhibit a significant difference. Trx-ELP[KV₂F-32], which has the highest cationic content in the ELP exhibits the greatest sensitivity to salt in the soluble lysate, similar to its behavior in PBS buffer. However, this sensitivity to salt does not positively impact its purification as an expression product due to its poor expression yield. Conversely, the improved expression yield combined (Figure 6) with similar salt sensitivity (Figure 7) and a lower T_t (Table 1 and Figure 3) allows Trx-ELP[KV₇F-36] to be purified using 1/3 of the salt that is needed to purify Trx-ELP[V₅A₂G₃-90] (the most commonly used purification tag) with a 2.6-fold improvement in the gravimetric yield of Trx. Furthermore, Trx-ELP[KV₇F-36] exhibits a 50% increase in Trx gravimetric yield over that purified from Trx-ELP[V-20], the most successful Trx tag previously published, using only 20% of the added salt without the complication of nanoparticle formation exhibited by the ELP[V-20] tag.⁴

Conclusions

The first important finding of this study is that the ELP tag can be significantly shortened without introducing any complications in the purification of a model protein, Trx. ELP [KV₇F-8] with a MW of ~4.5 kDa is the shortest ELP that has been used to purify an ELP fusion protein to date, and ELP[KV₇F-36], the most optimized tag engineered to date for the purification of Trx, has a MW of only 15.9 kDa. The MW of ELP[KV₇F-8] is similar to that of many peptide affinity tags, while that of the optimized ELP[KV₇F-36], is significantly smaller than or similar to many protein affinity tags.

The study of cationic tags for the purification of Trx also uncovered molecular parameters that will be important for the design of future ELP tags for purification of recombinant proteins. These are: (1) the incorporation of ionizable residues into the ELP sequence enhances the sensitivity of the ELP inverse phase transition to added salt and thereby reduces the amount of salt needed to induce the inverse phase transition in the soluble lysate. (2) The incorporation of ionizable groups must be balanced by the addition of hydrophobic residues in order to

maintain a reasonable T_t of the ELP. The optimal T_t should be above room temperature at all working fusion protein concentrations, but should be designed to be reasonably low so that only a minimum of salt must be added at each round of ITC. (3) The ELP molecular weight should be minimized to improve target protein yield; however, it should be high enough that micellization or nanoparticle formation above the phase transition is not observed. (4) The guest residue composition significantly affects the expression yield of the ELP and thus the ELP fusion protein, so that well-expressed ELP sequences should be used as purification tags. Because the cause of sequence dependent expression is, as yet, unknown, this variable will be difficult to predict and control. (5) Finally, the data suggest that the incorporation of cationic residues into ELPs is effective in promoting ELP-protein interactions in the aggregated state of the anionic Trx-ELP fusion proteins. We believe that oppositely charged Trx and ELPs aid the formation of large aggregates during fusion protein purification. Bearing this in mind, we caution that the design of ELP fusion proteins having like charge characteristics in both the target protein and ELP might inhibit aggregate formation and thereby complicate ELP fusion protein purification.

In conclusion, we believe that for most laboratory scale purification by ITC, optimization of the ELP tag is probably not necessary, as the most commonly used aliphatic ELP tags are sufficiently versatile to provide purified fusion proteins with yields and purities that are typically similar to, if not greater, than those achieved with affinity chromatography.¹⁴ However, the results presented here suggest that large-scale purification of recombinant proteins for pharmaceutical or industrial applications will benefit from optimization of the ELP tag for a specific protein to maximize its expression yield and purification efficiency.

Acknowledgments

We would like to thank Tennyson Liu, for performing preliminary ELP cloning steps necessary to synthesize all the ELP and ELP fusion proteins genes reported in this work. We also thank the National Institutes of Health (GM061232, F32-CA123889) for their financial support of this research.

References

- (1). Nilsson J, Stahl S, Lundeberg J, Uhlen M, Nygren PA. Protein Expression And Purification 1997;11:1–16. [PubMed: 9325133]
- (2). Mondal K, Gupta MN. Biomolecular Engineering 2006;23:59–76. [PubMed: 16527537]
- (3). Meyer DE, Chilkoti A. Nature Biotechnology 1999;17:1112–1115.
- (4). Meyer DE, Trabbic-Carlson K, Chilkoti A. Biotechnology Progress 2001;17:720–728. [PubMed: 11485434]
- (5). Urry DW. Journal Of Physical Chemistry B 1997;101:11007–11028.
- (6). Li B, Alonso DOV, Bennion BJ, Daggett V. Journal Of The American Chemical Society 2001;123:11991–11998. [PubMed: 11724607]
- (7). Li B, Alonso DOV, Daggett V. Journal Of Molecular Biology 2001;305:581–592. [PubMed: 11152614]
- (8). Cacace MG, Landau EM, Ramsden JJ. Quarterly Reviews Of Biophysics 1997;30:241–277. [PubMed: 9394422]
- (9). Zhang YJ, Trabbic-Carlson K, Albertorio F, Chilkoti A, Cremer PS. Biomacromolecules 2006;7:2192–2199. [PubMed: 16827587]
- (10). Luan CH, Parker TM, Prasad KU, Urry DW. Biopolymers 1991;31:465–475. [PubMed: 1868163]
- (11). Chow DC, Dreher MR, Trabbic-Carlson K, Chilkoti A. Biotechnology Progress 2006;22:638–646. [PubMed: 16739944]
- (12). Meyer DE, Chilkoti A. Biomacromolecules 2002;3:357–367. [PubMed: 11888323]
- (13). Meyer DE, Chilkoti A. Biomacromolecules 2004;5:846–851. [PubMed: 15132671]

- (14). Trabbic-Carlson K, Liu L, Kim B, Chilkoti A. *Protein Science* 2004;13:3274–3284. [PubMed: 15557268]
- (15). Trabbic-Carlson K, Meyer DE, Liu L, Piervincenzi R, Nath N, LaBean T, Chilkoti A. *Protein Engineering Design & Selection* 2004;17:57–66.
- (16). Kim JY, Mulchandani A, Chen W. *Biotechnology And Bioengineering* 2005;90:373–379. [PubMed: 15800861]
- (17). Kim JY, O'Malley S, Mulchandani A, Chen W. *Analytical Chemistry* 2005;77:2318–2322. [PubMed: 15828763]
- (18). Kim JY, Mulchandani A, Chen W. *Analytical Biochemistry* 2003;322:251–256. [PubMed: 14596835]
- (19). Kostal J, Mulchandani A, Chen W. *Biotechnology And Bioengineering* 2004;85:293–297. [PubMed: 14748084]
- (20). Kostal J, Mulchandani A, Chen W. *Macromolecules* 2001;34:2257–2261.
- (21). Gao D, McBean N, Schultz JS, Yan YS, Mulchandani A, Chen WF. *Journal Of The American Chemical Society* 2006;128:676–677. [PubMed: 16417330]
- (22). Kostal J, Yang R, Wu CH, Mulchandani A, Chen W. *Applied And Environmental Microbiology* 2004;70:4582–4587. [PubMed: 15294789]
- (23). Shimazu M, Mulchandani A, Chen W. *Biotechnology And Bioengineering* 2003;81:74–79. [PubMed: 12432583]
- (24). Banki MR, Feng LA, Wood DW. *Nature Methods* 2005;2:659–661. [PubMed: 16074986]
- (25). Ge X, Yang DSC, Trabbic-Carlson K, Kim B, Chilkoti A, Filipe CDM. *Journal Of The American Chemical Society* 2005;127:11228–11229. [PubMed: 16089436]
- (26). Trabbic-Carlson K, Setton LA, Chilkoti A. *Biomacromolecules* 2003;4:572–580. [PubMed: 12741772]
- (27). McPherson DT, Morrow C, Minehan DS, Wu JG, Hunter E, Urry DW. *Biotechnology Progress* 1992;8:347–352. [PubMed: 1368456]
- (28). Creighton, TE. *Proteins: Structures and Molecular Properties*. Second ed.. W. H. Freeman and Company; New York: 1993.
- (29). Fexby S, Bulow L. *Trends In Biotechnology* 2004;22:511–516. [PubMed: 15450744]
- (30). Meyer DE, Chilkoti A. *Biomacromolecules*. 2003in review

A. ELP monomer gene sequences

ELP[KV₇F-9]

```

↓EcoRI                               ↓PflMI
A ATT CAT ATG GGC CAC GGC GTG GGC GTA CCG GGT AAA GGT GTT CCT
                               V G V P G K G V P
GGC GTG GGT GTT CCG GGC GTA GGT GTC CCA GGT GTG GGC GTA CCG
G V G V P G V G V P G V G V P
GGC GTT GGT GTT CCT GGT GTC GGC GTG CCG GGC GTG GGT GTT CCG
G V G V P G V G V P G V G V P
                               ↓BglI ↓HindIII
GGT TTC GGC GTG CCG GGC TGG CA
G F G V P G

```

ELP[KV₂F-4]

```

↓EcoRI                               ↓PflMI
A ATT CAT ATG GGC CAC GGC GTG GGC GTA CCG GGT AAA GGT GTT CCT
                               V G V P G K G V P
                               ↓BglI ↓HindIII
GGC GTG GGT GTT CCG GGT TTC GGC GTG CCG GGC TGG CA
G V G V P G F G V P G

```

B. Modified pET expression vector sequences

Modified pET-25b for ELP expression

```

↓Nde I                               ↓Sfi I                               ↓EcoR I
T ATG AGC AAA GGG CCG GGC TGG CCG TGA TAA
(M) S K G P G W P ter ter

```

Modified pET-32b for thioredoxin-ELP fusion protein expression

```

↓MscI
Thioredoxin gene-C CAT ATG CAC CAT CAT CAT CAT TCT TCT GGT CTG GTG
                      H M H H H H H H S S G L V
                      ↓thrombin cleavage site ↓Sfi I ↓Ava I
CCA CGC GGT TCT GGT AAA GGG CCG GGC TGG CCG TGA TAA C
P R G S G K G P G W P ter ter

```

Figure 1.

Cloning scheme for cationic ELPs and their ELP fusion proteins showing coding DNA strand, the corresponding amino acid sequence, the ELP guest residue amino acids in bold, and the location of key endonuclease restrictions sites. PflM I and Bgl I restriction enzymes are used for recursive directional ligation to oligomerize ELP monomer genes in (A) as well as for gene excision for subcloning into pET expression vectors at the SfiI site shown in (B).

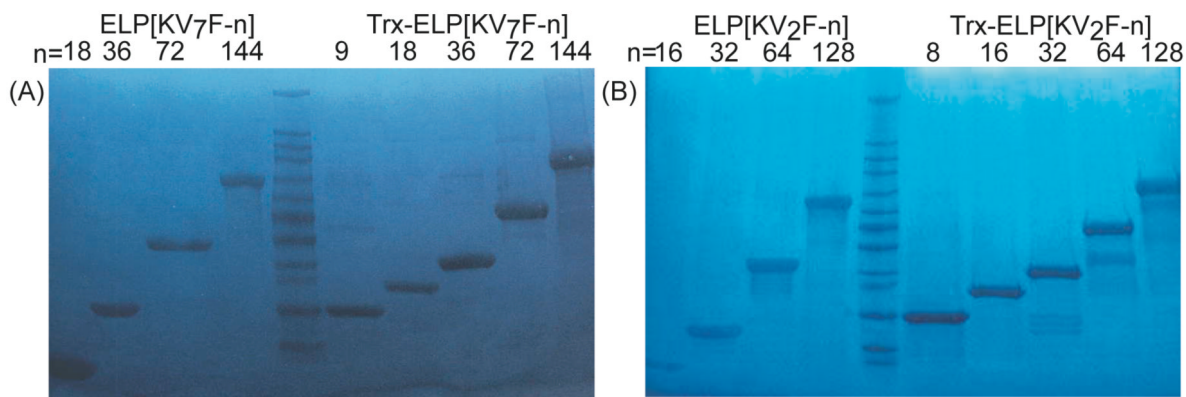


Figure 2.

Copper stained SDS-PAGE gels of purified (A) ELP[KV₇F] (lanes 1-4) and Trx-ELP[KV₇F] (lanes 6-10) and (B) ELP[KV₂F] (lanes 1-4) and Trx-ELP[KV₂F] (lanes 6-10). Lane M (lane 5 on both gels) contains molecular size markers (205, 116, 97, 84, 66, 55, 45, 36, 29, 24, 20, 14, 7 kDa from top to bottom). Each lane is marked with the number of pentapeptide repeats (n) in the ELPs and ELP fusion proteins. Lower MW fragments observed in Trx-ELP[KV₂F] protein samples are the result of proteolytic cleavage within the linker peptide located between Trx and ELP domains.

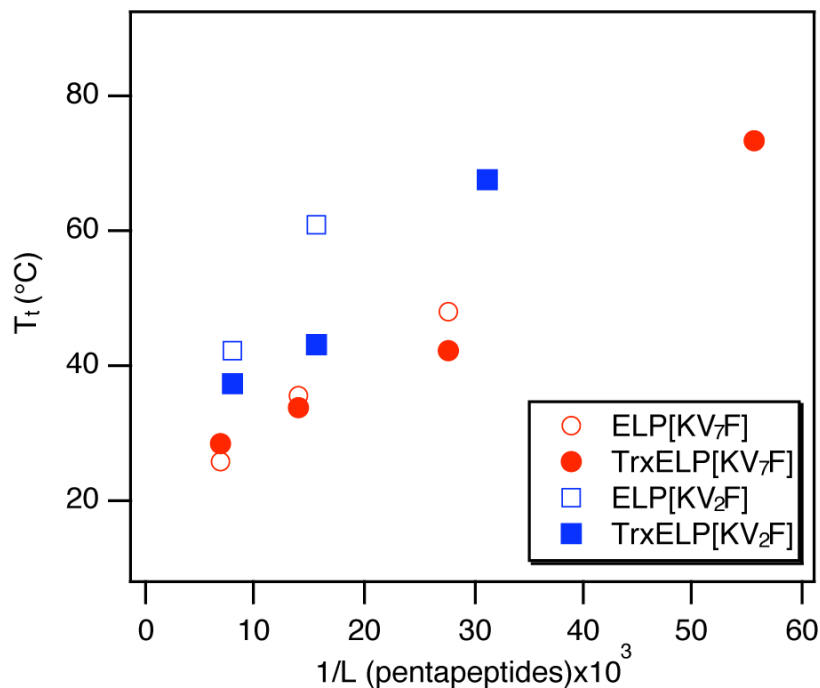


Figure 3.

Plot of transition temperature as a function of $1/\text{length of ELP pentapeptide} \times 10^3$ for ELPs and thioredoxin-ELPs. The transition temperature was measured from $25 \mu\text{M}$ solutions of ELP and Trx-ELP in PBS, and the data supports a linear relationship between T_t and $1/\text{ELP length}$ similar that observed for other ELPs.³⁰ Linear regression shows $r^2=0.986$ for ELP[KV₇F], $r^2=0.983$ for Trx-ELP[KV₇F], and $r^2=0.977$ for Trx-ELP[KV₂F]. Goodness of fit could not be established for ELP[KV₂F] polymers because only two of the four expressed ELP[KV₂F] proteins have measurable T_t (below 90°C) at $25 \mu\text{M}$ in PBS.

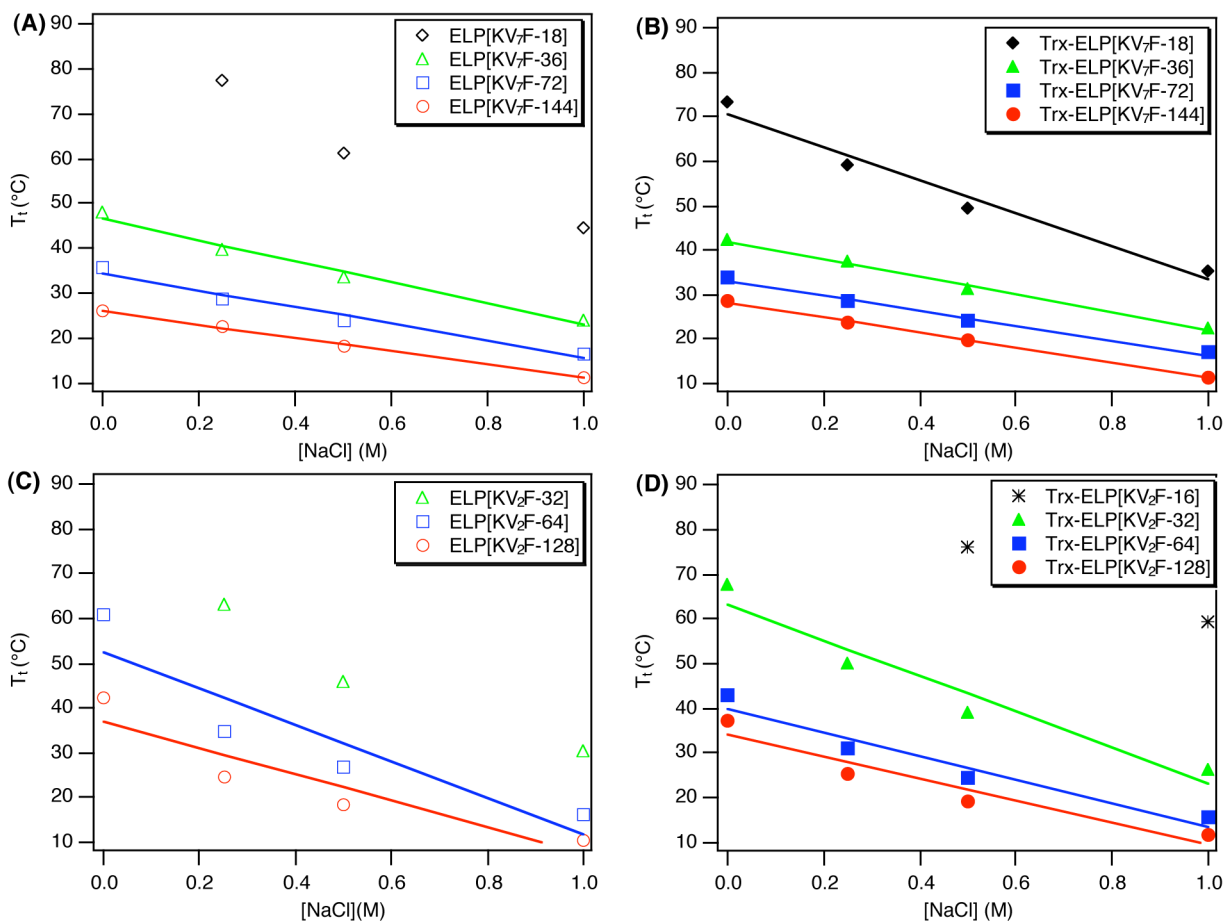
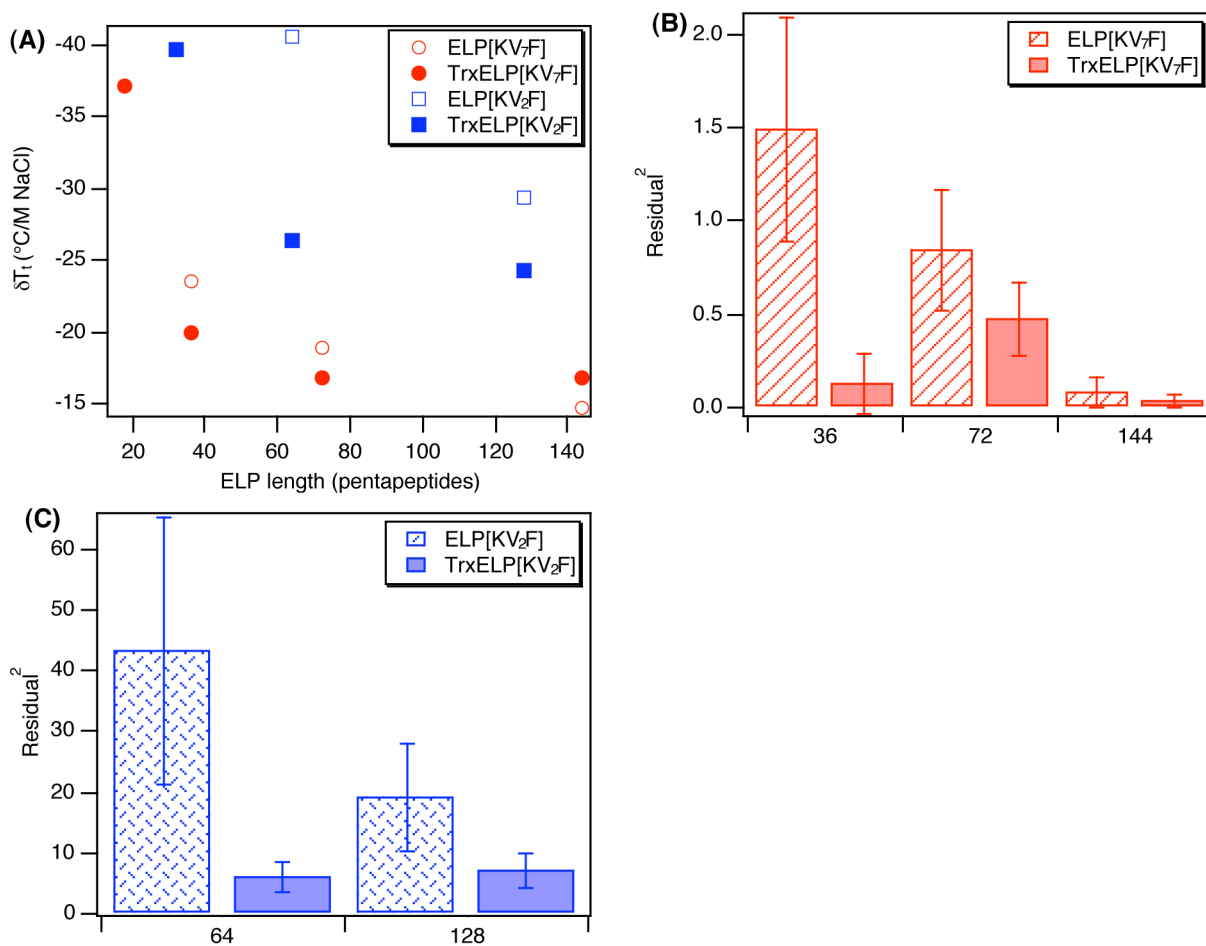


Figure 4.

Plot of transition temperature as a function of NaCl concentration and molecular weight for (A) ELP[KV₇F], (B) Trx-ELP[KV₇F], (C) ELP[KV₂F], and (D) Trx-ELP[KV₂F]. The transition temperature was measured from 25 μ M ELP and Trx-ELP solutions in PBS with each added NaCl concentration. ELP[KV₇F] (r^2 from 0.980 to 0.998) and Trx-ELP[KV₇F] (r^2 from 0.974 to 0.999) exhibit largely linear trends in T_t as a function of NaCl concentration, while ELP[KV₂F] (r^2 from 0.840 to 0.860) and Trx-ELP[KV₂F] (r^2 from 0.918-0.941) proteins show more non-linear character consistent with their higher ELP charge densities. Best fit linear correlations to salt sensitivity data were plotted only for ELPs and Trx-ELPs where the T_t could be accurately measured at all four tested NaCl concentrations in PBS (0, 0.25, 0.5, and 1 M NaCl). These correlations are summarized in Figure 5.

**Figure 5.**

Sensitivity of the ELP and Trx-ELP phase transition to added NaCl as a function of ELP length. (A) $\delta T_1/M$ of NaCl taken from the slope of a linear fit to the data in Figure 4 plotted as a function of number of ELP pentapeptides for ELPs and Trx-ELPs. Linear fit error is expressed as average of the squared fit residuals for (B) ELP[KV₇F] and (C) ELP[KV₂F] proteins and corresponding Trx fusions. ELP[KV₂F] proteins were found to be significantly more non-linear than ELP[KV₇F] proteins ($p = 1.0 \times 10^{-9}$) in their sensitivity to added salt, and error in the linear fit over all the ELP fusion proteins is significantly reduced ($p = 0.0033$) upon fusion to Trx. Error bars reflect the standard deviation for $n = 4$ salt concentrations per protein.

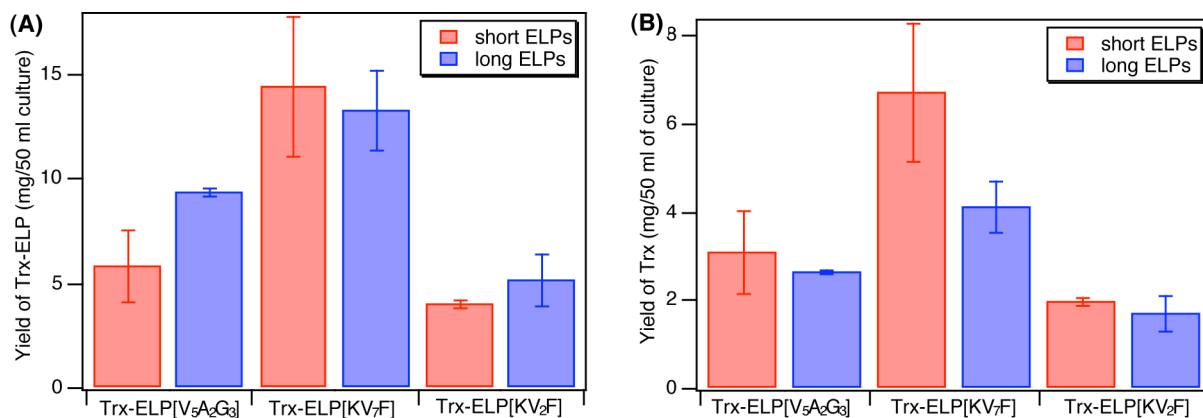


Figure 6.

Yield of (A) Trx-ELP fusion proteins and (B) target Trx protein from 50 mL of *E. coli* culture as a function of the ELP length (30-36 pentapeptides for the short tags and 64-90 pentapeptides for the long tags) and guest residue composition of the ELP purification tag. Protein yield is reported as the mean \pm SD (n=3). The gravimetric yield of Trx is higher for the lower molecular weight tag (30-36 pentapeptides) in the case of all three ELP guest residue compositions and the yield of Trx-ELP is greatest for the ELP[KV₇F] tags. The poorest fusion protein and target protein yield was exhibited by the ELP[KV₂F] proteins suggesting that expression is highly dependent on the guest residue composition and not easily predictable.

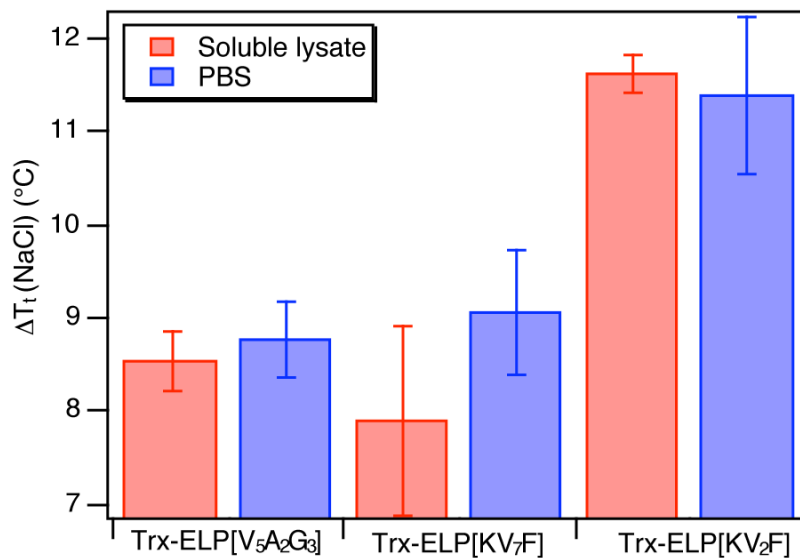


Figure 7.

Transition temperature difference ($\Delta T_t(\text{NaCl})$) between 0.5 and 1M NaCl added to *E. coli* soluble lysate and to PBS solutions of 100 μM Trx-ELP[KV₇F-36], Trx-ELP[KV₂F-32], and Trx-ELP[V₅A₂G₃-90]. There is no significant difference between lysate and PBS in the $\Delta T_t(\text{NaCl})$ in for any of the proteins investigated. The transition temperature difference ($\Delta T_t(\text{NaCl})$) is reported as the difference in the mean T_t s at 0.5M and 1M NaCl ($n=3$ for each salt concentration) \pm the standard deviation calculated by the propagation of errors in the measurements of T_t with 0.5 and 1M NaCl.

Table 1

Molecular weight, lysine (K) and phenylalanine (F) content, transition temperature (T_t) and fusion ΔT_t parameters for purified ELPs and Trx-ELPs at 25 μ M in PBS (N/A*: not available)

ELPs	# pentapeptides	#K or F residues	Purified ELPs		Purified Trx-ELPs		fusion ΔT_t parameter
			Observed MW (kDa)	T_t ($^{\circ}$ C)	Observed MW (kDa)	T_t ($^{\circ}$ C)	
[KV ₇ F]	9	1	N/A*	N/A*	18.5	>90.0	N/A*
	18	2	8.4	>90.0	22.2	73.1	>-16.9
	36	4	15.8	48.0	29.7	42.1	-5.9
	72	8	31.0	35.7	44.9	34.0	-1.7
144	16	61.1	26.1	74.9	28.4	2.3	
[KV ₂ F]	8	2	N/A*	N/A*	18.1	>90.0	N/A*
	16	4	7.7	>90.0	21.5	>90.0	N/A*
	32	8	14.5	>90.0	28.4	67.4	>-22.6
	64	16	28.4	60.7	42.1	43.1	-17.6
	128	32	55.7	42.4	69.6	37.3	-5.1