A designed, phase changing RTX-based peptide for efficient bioseparations

Oren Shur¹, Kevin Dooley¹, Mark Blenner², Matthew Baltimore¹, and Scott Banta¹

¹Department of Chemical Engineering, Columbia University, New York, New York and ²Department of Chemical and Biomolecular Engineering, Clemson University, Clemson, South Carolina

BioTechniques 54:197-206 (April 2013) doi 10.2144/000114010

Keywords: repeat-in-toxin domain; beta roll domain; stimulus-responsive peptides; precipitation

Typically, chromatography is the most costly and time-consuming step in protein purification. As a result, alternative methods have been sought for bioseparations, including the use of stimulus-responsive tags that can reversibly precipitate out of solution in response to the appropriate stimulus. While effective, stimulus-responsive tags tend to require temperature changes or relatively harsh buffer conditions to induce precipitation. Here we describe a synthetic peptide, based on the natural repeat-in-toxin (RTX) domain that undergoes gentler calcium-responsive, reversible precipitation. When coupled to the maltose binding protein (MBP), our calcium-responsive tag efficiently purified the fusion protein. Furthermore, when the MBP was appended to green fluorescent protein (GFP), β-lactamase, or a thermostable alcohol dehydrogenase (AdhD), these constructs could also be purified by calcium-induced precipitation. Finally, protease cleavage of the precipitating tag enables the recovery of pure and active target protein by cycling precipitation before and after cleavage.

Non-chromatographic purification techniques are of significant interest since chromatography is typically the most expensive step in protein purification (1). Alternative approaches often rely on targeted precipitation of the protein of interest. One approach is metal chelate affinity precipitation, where thermoresponsive copolymers can be used to specifically precipitate out poly-histidine tagged recombinant proteins (2,3). Another purely protein-based approach is the use of thermoresponsive elastin-like peptides (ELPs) that consist of tandem repeats of the sequence VPGXG and precipitate with small temperature increases(4,5). ELPs undergo an inverse phase transition and aggregation, which is thought to be driven by the exposure of hydrophobic patches in the peptides upon heating (6). As part of a purification system, ELPs have been coupled to intein domains that have been genetically engineered into minimal self-cleaving units (7). When coupled, the ELP-intein system allows for a simple two-stage purification scheme. In the first step, precipitation of the ELP is triggered and the fusion protein is purified. Then, the intein is induced to cleave off the target protein and the ELP is again precipitated, leaving behind pure target protein in solution (8). While effective for many purification applications, the necessary heating of samples or the alternative use of high salt concentrations (9) can be problematic in many situations. Another protein-based non-chromatographic purification scheme developed by Ding et al. relies on calcium-dependent precipitation of an annexin B1 tag (10). As with ELPs, a self-cleaving intein is also incorporated in the fusion protein to remove the tag following purification.

Our interest in alternatives to chromatography for purifying proteins is a product of discoveries made while exploring repeat scaffolds for protein engineering applications. Repeat scaffolds are of interest to protein engineers as their repetitive, predictable secondary structures make them ideal for studying folding and engineering novel functions (11,12). There are examples of repeat scaffolds being engineered for biomolecular recognition, most notably the ankyrin repeats (13). In order to improve the engineerability of these scaffolds, efforts have been made at consensus design. Consensus design seeks to identify the core repeating peptide unit. Once this sequence is identified, multiple repeats can be concatenated as necessary for the desired application. Consensus design approaches have been successfully used for a number of repeat scaffolds, including ankyrin repeats, tetratricopeptide repeats, and armadillo repeats (14–17). The ability to alter the size of a scaffold is of particular use when engi-

Method summary:
A new calcium-responsive tag based on a consensus sequence found in the natural repeat-in-toxin (RTX) domain is presented. This calcium-responsive tag works under gentler reaction conditions than existing approaches and can be removed through protease cleavage, resulting in a pure, active target protein.
neering binding, as the interface size can be tuned to the particular target.

In an effort to explore novel scaffolds for protein engineering, we have sought to identify a repeat scaffold that is also stimulus-responsive. To this end, we investigated the calcium-responsive repeat-in-toxin (RTX) domain. The RTX domain is found in proteins secreted through the bacterial type 1 secretion system (18). The domain consists of repeats of the consensus amino acid sequence GGGXGDXUX, where X is variable and U is a hydrophobic amino acid. One of the most well characterized RTX domains is the block V RTX domain from the adenylate cyclase toxin (CyaA) of *B. pertussis*. The domain is intrinsically disordered in the absence of calcium and forms a β roll structure (Figure 1A) in the presence of calcium (19). Of note, the block V RTX domain retains its reversible calcium-responsiveness even when expressed separately from the larger protein (20,21). Previous attempts have been made to use RTX domains in protein engineering, including incorporation into mesh networks, design of synthetic RTX peptides, and generation of hydrogel-forming RTX domains (22–25).

Our original objective was to design consensus RTX domains. Specifically, we identified the frequency of amino acids at each position of the nine amino acid repeat unit from a set of RTX-containing proteins (Figure 1B). This led to identification of the consensus sequence GGAGNDTLY. We then sought to create a library of consensus RTX constructs consisting of 5, 9, 13, or 17 repeats of the consensus unit. Upon purification of a number of these constructs, we observed that many of them precipitated upon cation of a number of these constructs, we decided to explore the possibility of using these consensus β roll tags (BRTs) as a tools for protein engineering, as the interface size can be tuned to the particular target.

In an effort to explore novel scaffolds for protein engineering, we have sought to identify a repeat scaffold that is also stimulus-responsive. To this end, we investigated the calcium-responsive repeat-in-toxin (RTX) domain. The RTX domain is found in proteins secreted through the bacterial type 1 secretion system (18). The domain consists of repeats of the consensus amino acid sequence GGGXGDXUX, where X is variable and U is a hydrophobic amino acid. One of the most well characterized RTX domains is the block V RTX domain from the adenylate cyclase toxin (CyaA) of *B. pertussis*. The domain is intrinsically disordered in the absence of calcium and forms a β roll structure (Figure 1A) in the presence of calcium (19). Of note, the block V RTX domain retains its reversible calcium-responsiveness even when expressed separately from the larger protein (20,21). Previous attempts have been made to use RTX domains in protein engineering, including incorporation into mesh networks, design of synthetic RTX peptides, and generation of hydrogel-forming RTX domains (22–25).

Our original objective was to design consensus RTX domains. Specifically, we identified the frequency of amino acids at each position of the nine amino acid repeat unit from a set of RTX-containing proteins (Figure 1B). This led to identification of the consensus sequence GGAGNDTLY. We then sought to create a library of consensus RTX constructs consisting of 5, 9, 13, or 17 repeats of the consensus unit. Upon purification of a number of these constructs, we observed that many of them precipitated in the presence of calcium. Therefore, we decided to explore the possibility of using these consensus β roll tags (BRTs) as a tools for bioseparation, similar to the ELP system. Here, we report the use of BRTs to purify recombinant proteins. We first purified a maltose binding protein (MBP)-BRT<sub>17</sub> fusion as a proof of principle. Then, this MBP-BRT<sub>17</sub> construct was fused to green fluorescent protein (GFP), which was used as a reporter during initial purification experiments. We have also fused β-lactamase and a thermostable alcohol dehydrogenase (AdhD) to demonstrate the feasibility of purifying enzymatic proteins. Finally, a specific protease site was engineered downstream of the tag to show that target proteins can be fully purified by protease cleavage while retaining their activity.

Materials and methods

Cloning
All cloning enzymes were purchased from New England Biolabs (Ipswich, MA). All oligonucleotides were synthesized by Integrated DNA Technologies (IDT) (Coralville, IA) and all sequences are available in Supplementary Table S1. Four differently sized MBP-BRT fusions were prepared consisting of 5, 9, 13, or 17 repeats of the consensus RTX sequence (named BRT<sub>5</sub>, BRT<sub>9</sub>, BRT<sub>13</sub>, and BRT<sub>17</sub>). To generate the DNA fragment for BRT<sub>17</sub>, three oligonucleotides were synthesized: cons<sub>β</sub>1, cons<sub>β</sub>2, and cons<sub>β</sub>3. One ng each of these oligonucleotides was mixed along with the primers cons<sub>1</sub>_AvaI_F and cons<sub>9</sub>_BseRI_HindIII_R. This product was digested with AvaI and HindIII and cloned into the similarly digested pMAL_BRT<sub>5</sub> to yield pMAL_BRT<sub>5</sub>−GFP, pMAL_BRT<sub>9</sub>−βlac and pMAL_BRT<sub>5</sub>−AdhD. The native enterokinase site in the pMAL<sub>c4E</sub> vector, which sits between MBP and BRT<sub>5</sub>, was knocked out in the pMAL<sub>c4E</sub>−βlac and pMAL<sub>c4E</sub>−AdhD plasmids. Two rounds of site-directed mutagenesis were required to change the native recognition site, DDDDK, to DDEEQ, which was shown to be resistant to cleavage. A novel enterokinase recognition site was also engineered downstream of BRT<sub>5</sub> in these constructs to allow for purification of the untagged protein of interest. Full plasmid maps of all cloned constructs are available in Supplementary Figure S1.

Expression and purification
For expression and cloning, Life Technologies (Grand Island, NY) Omnimax T1 E. coli cells were used. One L cultures of TB supplemented with 100 µg/mL ampicillin and 0.2% glucose were inoculated with 10 mL of overnight culture. Cultures were grown at 37°C with shaking at 225 RPM to an approximate OD600 of 0.5 and induced with 0.3 mM IPTG. Cells harboring pMAL_BRT<sub>17</sub> and pMAL_BRT<sub>5</sub>−βlac were allowed to express for an additional two hours and then harvested. Cultures transformed with pMAL_BRT<sub>17</sub>−GFP were transferred to a shaker at 25°C and allowed to express for an additional 16 h and then harvested as no fluorescence was observed when expressed at 37°C. Cultures transformed with pMAL_BRT<sub>5</sub>−AdhD were allowed to express at 37°C for an additional 16 h as previously reported (28). Cells were harvested after expression and resuspended in 1/20 culture volume of 50 mM tris-HCl, pH 7.4 for precipitation purification. For amylose resin purification, cells were resuspended in 1/20 culture volume of MBP column buffer (20 mM tris-HCl, 200 mM NaCl, 1 mM EDTA, pH 7.4). In both cases, cells were subsequently lysed via

**Figure 1. Beta roll structure and sequence logo.** (a) Crystal structure of β roll domain from metallopro tease of S. marcescens (PDB: 1SAT). (b) Amino acid frequencies for single β roll repeat identifying consensus sequence GGAGNDTLY. Height of the letter corresponds to proportion of sequences containing the particular amino acid at that position. Sequence logo was generated using WebLogo (33,34).

*BtoCl and HindIII and then cloned into pMAL_BRT<sub>17</sub> cut with BseRI and HindIII to yield pMAL_BRT<sub>17</sub>_BseRI and pMAL_BRT<sub>17</sub>_HindIII. The β-lactamase gene was amplified from the pMAL_c4E vector using primers βlac_BseRI_F and βlac_HindIII_R. The β-lactamase gene was amplified out of pWUR85 using primers AdhD_BseRI_F and AdhD_HindIII_R (28).

All three of these inserts were digested with BseRI and HindIII and cloned into similarly digested pMAL_BRT<sub>5</sub>−GFP, pMAL_BRT<sub>9</sub>−βlac and pMAL_BRT<sub>5</sub>−AdhD plasmids. Two rounds of site-directed mutagenesis were required to change the native recognition site, DDDDK, to DDEEQ, which was shown to be resistant to cleavage. A novel enterokinase recognition site was also engineered downstream of BRT<sub>5</sub> in these constructs to allow for purification of the untagged protein of interest. Full plasmid maps of all cloned constructs are available in Supplementary Figure S1.
sonication using 15 s pulses for a total of 150 s. Lysate was then clarified by centrifugation at 15,000 g for 30 min at 4°C. For amylose resin purification, clarified lysate was diluted with five volumes of column buffer and purified as previously described (21). All subsequent steps were performed at room temperature.

For precipitation purification, clarified lysate was added to a concentrated calcium stock according to the data presented in Figure 2. For example, for precipitation of MBP-BRT$_{17}$ lysate in 100 mM CaCl$_2$, 950 μL of clarified lysate was added to 50 μL of 2 M CaCl$_2$ solution. The sample was promptly mixed by gentle pipetting, allowed to sit at room temperature for 2 min and then centrifuged at 16,000 g in a microcentrifuge for 2 min. The supernatant was carefully removed and the pellet was resuspended in the same volume as the lysate from which they were extracted, so signals were compared directly.

For estimation of MBP-BRT$_{17}$-βlactamase recovery, protein was added to a nitrocefin solution and the absorbance at 486 nm was tracked corresponding to the hydrolysis of nitrocefin. 500 μL of nitrocefin solution was prepared by placing three nitrocefin disks (Fluka) in 450 μL 50 mM tris-HCl, pH 7.4 and 50 μL DMSO. In each sample well, 50 μL of this solution was mixed with 90 μL of the same tris buffer and 10 μL of protein sample. For each sample tested, serial dilutions from 1× to 1000× were prepared from lysate and purified protein. Initial rates were determined by measuring the change in absorbance at 486 nm over the first 20% of the change in signal between the starting absorbance and the end absorbance. The same nitrocefin stock solution was used for all samples to account for variations in concentration.

MBP-BRT$_{17}$-AdhD recovery was also evaluated by enzymatic activity using a protocol previously described (28). Since this AdhD was isolated from the hyperthermophile Pyrococcus furiosus, all samples were heat treated at 80°C for 1 h prior to evaluating activity. All assays were performed at saturated conditions of both cofactor and substrate, 0.5 mM NAD$^+$ and 100 mM 2,3-butanediol, respectively. Reaction mixtures containing 2,3-butanediol and protein sample in 50 mM glycine pH 8.8 were incubated at 45°C in a 96 well UV microplate in a spectrophotometer. Reactions were initiated by the addition of NAD+. Initial rates were calculated by following the production of NADH at 340 nm. Specific activity of cleaved AdhD was calculated using an NADH extinction coefficient ($ε$ = 6.22 mM$^{-1}$ cm$^{-1}$).

All spectroscopic measurements were done on a SpectraMax M2 (Molecular Devices; Sunnyvale, CA).

**Results and discussion**

In order to identify the consensus RTX sequence, a database of RTX containing proteins was constructed by searching the UniProt (www.uniprot.org) database for hemolysin-type calcium binding domains. Individual repeats were identified and the frequency of amino acids at each of the nine repeat positions was determined (Figure 1B). From this result, the repeat sequence GGAGNDTLY was identified as the consensus sequence. For a few of these positions, other amino acids were found with nearly equal frequency. However, as this sequence was found to be effective for purification, further investigation on sequence variation was not performed. A variety of synthetic RTX domains of different lengths (BRT$_{3}$, BRT$_{5}$, BRT$_{10}$, BRT$_{15}$) were prepared as fusions to the C terminus of MBP, with subscripts denoting the number of repeats. These lengths were chosen as they reflect the variability of naturally occurring RTX domains. Unexpectedly, we observed that upon the addition of calcium to the purified BRT$_{17}$ construct, there was significant precipitation out of solution, which was reversed upon the addition of the chelating agent EGTA.

In order to more thoroughly characterize the observed precipitation behavior, cells were induced to express the four MBP-BRT constructs. Clarified cell lysates were
Table 1 Recovery data for three constructs tested.

<table>
<thead>
<tr>
<th>Calcium, mM</th>
<th>MBP-BRT(_{\beta{\text{lac}}}^{-})</th>
<th>MBP-BRT(_{\beta{\text{lac}}}^{-})-GFP</th>
<th>MBP-BRT(_{\beta{\text{lac}}}^{-})-AdhD</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>2.0 ± 0.1</td>
<td>2.8 ± 0.1</td>
<td>1.6 ± 0.1</td>
</tr>
<tr>
<td>50</td>
<td>2.3 ± 0.1</td>
<td>3.7 ± 0.1</td>
<td>4.0 ± 0.1</td>
</tr>
<tr>
<td>75</td>
<td>2.2 ± 0.2</td>
<td>2.8 ± 0.3</td>
<td>3.4 ± 0.1</td>
</tr>
</tbody>
</table>

“Fold versus Resin” denotes protein quantity recovered relative to amylose resin for equivalent loading amount. For MBP-BRT\(_{\beta{\text{lac}}}^{-}\)-GFP, MBP-BRT\(_{\beta{\text{lac}}}^{-}\), and MBP-BRT\(_{\beta{\text{lac}}}^{-}\)-AdhD fluorescence and activity are the respective properties relative to clarified lysate. Errors represent standard deviations. All data were collected in triplicate.
recoveries were calculated as the percentage of fluorescence signal of purified sample compared with lysate (this was normalized against control lysate). Along with total protein recoveries estimated by UV absorbance, recoveries of both MBP-BRT$_{17}$-βlac and MBP-BRT$_{17}$-AdhD were estimated by comparing lysate activity to the activity of the MBP-BRT$_{17}$-βlac recoveries were calculated using activity measured by tracking the absorbance at 486 nm for the hydrolysis of nitrocefin. MBP-BRT$_{17}$-AdhD recoveries were calculated by tracking NADH formation at 340 nm in saturating conditions of both substrate and cofactor. Results of these trials are shown in Table 1. For MBP-BRT$_{17}$, calcium precipitation recoveries about double the amount of protein as compared with amylose resin purification. For MBP-BRT$_{17}$-GFP, we observed up to 86% recovery of fluorescence. MBP-BRT$_{17}$-βlac recovery from the lysate was not as high, but was still 5-fold better than the amylose resin, yielding a significant quantity of protein. Similar results were also observed for MBP-BRT$_{17}$-AdhD, although the yields were not quite as high compared with the resin (2-fold improvement). It should be noted that while the overall values of the activities recovered in Table 1 appear low, they were all larger than the values obtained using the amylose resin purification. It is also possible that measuring activity in crude extracts may introduce error beyond what was accounted for in the measurement of endogenous hydrolysis (β-lactamase) and reduction (AdhD).

Table 2 lists the absolute yield of each fusion protein based on UV absorption at 280 nm. All fusion proteins were shown to be purified in high yields.

To increase the utility of this tag, it would be beneficial to couple our system with a cleavage tag to separate the protein of interest from the BRT. The pMAL-c4E vector we used for these experiments contains a cleavable enterokinase site between the MBP and BRT. This recognition sequence was removed via site-directed mutagenesis. A new enterokinase site was engineered between the BRT and the protein of interest for MBP-BRT$_{17}$-βlac and MBP-BRT$_{17}$-AdhD. Therefore, as a proof of principle, we took purification precipitated MBP-BRT$_{17}$-βlac and MBP-BRT$_{17}$-AdhD and subjected them to overnight cleavage by enterokinase digestion. Calcium can then be added directly to the cleavage reaction to precipitate MBP-BRT$_{17}$ thereby separating the tag from the protein of interest following centrifugation. This is shown in Figure 5 for MBP-BRT$_{17}$-AdhD, showing pure, soluble protein by SDS-PAGE. Recoveries of 93 ± 7% were obtained by tracking UV absorbance at 280 nm, meaning 93% of the AdhD in the precipitation purified sample was recovered after cleavage and reprecipitation of the tag. Specific activity of the purified enzyme was also calculated to be 20.2 ± 1.3 min$^{-1}$, which is similar to what has been previously reported, indicating this system has little to no effect on the activities of purified proteins (28). However, in the case of MBP-BRT$_{17}$-βlac, the cleaved β-lactamase remained in the insoluble fraction following enterokinase cleavage and calcium precipitation. Upon further investigation it was found that β-lactamase will precipitate in high calcium concentrations. As a control experiment, we purchased recombinant β-lactamase and observed similar behavior. In 75mM CaCl$_2$, an insoluble pellet was formed upon centrifugation. Activity assays confirmed a significant amount of active protein in the insoluble fraction (data not shown). This illustrates a caveat of the BRT system. Proteins that naturally precipitate in CaCl$_2$ solutions cannot be efficiently separated from the BRT. For future improvement of this system, the protease used could be fused to the precipitating BRT or a self-cleaveing intein could be incorporated. Fusing the protease to the BRT would enable its removal from the target protein in the final precipitation. A self-cleaving intein would fulfill a similar function. It should also be noted that the BRT can precipitate without being fused to the MBP, suggesting that the MBP is not essential for this system; however, the MBP may be useful for improving protein expression levels.

It is not completely clear why these consensus RTX constructs are able to function as bioseparation tags. We do observe a correlation between length and precipitation (Figure 2), so size certainly plays a role. However, there has not been extensive work in studying the role of the number of repeats on RTX behavior. We recently studied the impact of altering the number of native RTX repeats in the block V CyaA RTX domain of B. pertussis but no significant size effect was observed and, furthermore, C-terminal capping was required for calcium-responsiveness (31). As for past efforts to design synthetic RTX domains, the synthetic domains created by Scotter et al. consisted of 4 RTX repeats and those prepared by Lilie et al. were weakly calcium-responsive, while those of Scotter et al. were only lanthanum-responsive and formed partially insoluble filaments in the presence of lanthanum. In general, it is fairly well established that β sheets are prone to aggregation and nature uses various strategies to ensure solubility of proteins containing these motifs (32), so perhaps BRTs are a balance between this tendency and the calcium-responsiveness of the β roll. Further investigation will be required to better elucidate the mechanism of BRT functionality, but their use as a tool for protein purification is clear.

The technique described here offers a new stimulus-responsive phase-changing peptide that could be useful in a range of applications similar to those for which ELPs have been used, such as recombinant protein purification or the creation of “smart” biomaterials. This new tag possesses certain advantages over ELPs and annexin B1 since precipitation is simpler to achieve and the BRT peptide is significantly smaller. Additionally, BRT$_{17}$ precipitates in as little as 25 mM CaCl$_2$ at room temperature, compared to the larger ionic strength and higher temperature increases required for ELP precipitation. Precipitation also occurs instantaneously, whereas annexin B1-based systems require a 2 h incubation period at 4°C. Overall, BRTs offer a new tool for rapid purification of recombinant proteins. The protocol described here can be performed to obtain purified fusion protein from lysate in only a few minutes. Further optimization of the BRT system should enable the use of specific proteases to purify target proteins and further improve the precipitation and resolubilization process, greatly enhancing the ability to rapidly purify recombinant proteins.

Acknowledgements

The authors gratefully acknowledge funding from the National Science Foundation.

Competing interests

The authors declare no competing interests.

Table 2 Absolute fusion protein yields.

<table>
<thead>
<tr>
<th>Calcium, mM</th>
<th>MBP-BRT$_{17}$</th>
<th>MBP-BRT$_{17}$-GFP</th>
<th>MBP-BRT$_{17}$-βlac</th>
<th>MBP-BRT$_{17}$-AdhD</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>268 ± 11</td>
<td>333 ± 12</td>
<td>124 ± 3</td>
<td>198 ± 3</td>
</tr>
<tr>
<td>50</td>
<td>305 ± 14</td>
<td>434 ± 17</td>
<td>160 ± 7</td>
<td>273 ± 9</td>
</tr>
<tr>
<td>75</td>
<td>295 ± 26</td>
<td>336 ± 40</td>
<td>176 ± 5</td>
<td>214 ± 6</td>
</tr>
</tbody>
</table>

Amount of protein recovered for each fusion construct after precipitation and washing. Values were determined using UV absorbance at 280 nm and calculated extinction coefficients available in the Supplementary Table S2. All data were collected in triplicate and errors represent standard deviations.
References


Received 27 June 2012; accepted 13 March 2013

Address correspondence to Scott Banta, Department of Chemical Engineering, Columbia University, New York, New York. E-mail: sbanta@columbia.edu

To purchase reprints of this article, contact: biotechniques@fosterprinting.com