The TrpA protein of *Trichodesmium erythraeum* IMS101 is a non-fibril-forming collagen and a component of the outer sheath

Simara Price,1 Siobhan Toal2 and Shivanthi Anandan1

1Department of Biology, Drexel University, 3245 Chestnut Street, Philadelphia, PA 19104, USA
2Department of Chemistry, Drexel University, 3245 Chestnut Street, Philadelphia, PA 19104, USA

Collagen molecules are structural in nature and primarily found in eukaryotic, multicellular organisms. Recently, a collagen-like protein, TrpA, was identified and characterized in the marine cyanobacterium *Trichodesmium erythraeum* IMS 101, and it was shown to be involved in maintaining the structural integrity of the trichomes. The TrpA protein contains one glycine interruption in the otherwise perfectly uninterrupted collagenous domain. In this study, we used phylogenetic analysis to determine that the TrpA protein sequence is most closely associated with non-fibril-forming collagen proteins. Structural modelling and circular dichroism data suggest that the glycine insertion decreases the stability of TrpA compared to uninterrupted collagen sequences. Additionally, scanning electron microscopy revealed that TrpA is expressed entirely on the surface of the trichomes, with no specific pattern of localization. These data indicate that the TrpA protein is part of the outer sheath of this organism. As such, this protein may function to promote adhesion between individual *T. erythraeum* trichomes, and between this organism and heterotrophic bacteria found in the same environment.

INTRODUCTION

Collagens are the most abundant structural protein found in vertebrates. There are currently over 28 different types of collagen (Shoulders & Raines, 2009). Despite this variety, they all have a common feature, a triglycine repeat region in the primary structure leading to the formation of a triple helical tertiary structure. This characteristic sequence consists of (Gly-Xaa-Yaa), where X is most commonly proline and Y can be any amino acid but is usually hydroxyproline. The high proline content of collagen chains results in a polyproline II (PPII) left-handed helical structure (Brodsky & Ramshaw, 1997). The triple helical collagen molecule can be composed of three identical (homotrimer) or similar (heterotrimer) collagen chains (Hulmes, 2002).

Interruptions in the triglycine repeat regions introduce a perturbation into the structure of the polypeptide that prevents the perfect triple helical structure. It is important to note that interruptions are not representative of a mutation that negatively affects the function of the collagen structure that it forms; it is only when the interruption is a mutation in a fibril-forming collagen that pathology is observed (Bella *et al.*, 2006). These interruptions introduce flexibility into the structure of the triple helix that is critical to the supermolecular structures that non-fibrillar collagens form (Hwang & Brodsky, 2012). These include network-forming collagens, beaded-filament-forming collagen, anchoring fibrils, fibril-associated collagens, endostatin-producing collagen and transmembrane collagens (Kadler *et al.*, 2007). For example, the anchoring fibril formed from type VII collagen contains a non-helical domain in the centre of its sequence that allows the fibres to form an arch structure as opposed to a straight fibril (Bächinger *et al.*, 1990; Burgeson, 1993). The arch occurs where the non-helical domain is found in the polypeptide (Bächinger *et al.*, 1990). These arches surround and entrap collagen fibrils, securing them in the basement membrane (Burgeson, 1993).

Previously believed to be limited to vertebrates, collagen and collagen-like proteins have been identified in invertebrates and prokaryotes. The presence of both fibrillar and non-fibrillar collagens in invertebrates including sponge, worm and abalone has provided evidence that a hypothetical ancestor of metazoans contained collagen (Exposito *et al.*, 2002). Collagenous domains in prokaryotes, such as the BcA protein in *Bacillus anthracis* and the ScA protein in *Streptococcus pyogenes*, could also indicate that evolution of the collagen protein predates the divergence of prokaryotes and eukaryotes and aided in the formation of multicellular organisms. However, horizontal gene transfer from eukaryotes into prokaryotes is also a possible pathway explaining the presence of this eukaryotic protein in prokaryotes (Exposito *et al.*, 2002).
The function of collagen-like proteins in prokaryotes is not abundantly clear; however, they have been described as being involved in the adhesion of pathogenic bacterial cells to their host human cells (Sylvestre et al., 2002). *S. pyogenes*, a bacterial pathogen involved in a wide range of human diseases, attaches to cell surfaces and contains the collagen-like proteins ScA and ScIB, which are found on the outer surface of the cell membrane (Rasmussen et al., 2003; Whatmore, 2001). The BcA protein in *B. anthracis*, the causative agent of anthrax, is found on the surface of the exosporium layer of the spore. These spores interact and bind with cells within mammalian hosts and it is possible that the BcA protein is involved in this attachment (Boydston et al., 2005).

Recently, a gene encoding a collagen-like protein was identified in the genome of the marine cyanobacterium *Trichodesmium erythraeum*. Using immunogold labelling, it was determined that collagen was being labelled on the surface of the *T. erythraeum* cells (Layton et al., 2008). This organism is a marine cyanobacterium found in subtropical oceans that contributes significantly to the fixed carbon and nitrogen of its environment (Capone et al., 1997; Roe et al., 2012). It has been shown recently that this collagen-like protein, now named TrpA, is expressed in *T. erythraeum* cells (Price & Anandan, 2013). TrpA is localized to the extracellular space between adjacent cells along the filament in conjunction with other proteins involved in maintaining attachment of the cells along the filament (Flores et al., 2007; Mariscal et al., 2011; Price & Anandan, 2013). In this study, we used phylogenetic analysis to determine the relatedness of TrpA to other collagen proteins to enable us to determine its structure. In addition, model synthetic peptides were used to determine if the structure of the TrpA protein could form PPII helices. Based on these data, we suggest that TrpA be classified as a non-fibrillar collagen. Scanning electron microscopy (SEM) with silver enhancement was used to resolve the localization of TrpA to the outer surface of the cell membrane (Whatmore et al., 2001). The BcA protein in *B. anthracis*, the causative agent of anthrax, is found on the surface of the exosporium layer of the spore. These spores interact and bind with cells within mammalian hosts and it is possible that the BcA protein is involved in this attachment (Boydston et al., 2005).

Alignment and analysis. All 80 fibrillar and non-fibrillar collagens and the TrpA collagen-like protein were aligned using Seaview (Gouy et al., 2010). After the sequences were aligned, phylogenetic trees were computed by maximum-likelihood (ML) using the program PhyML version 3.0 (Guindon et al., 2005). The ML tree was midpoint rooted using the web-based tool Interactive Tree of Life due to the absence of a good outgroup (Letunic & Bork, 2011).

Structural studies

Peptides. A peptide 13 amino acids in length was constructed based on the amino acid sequence for TrpA including the interruption in the sequence: Gly-Pro-Ala-Gly-Pro-Ala-Gly-Gly-Pro-Ala-Gly-Pro-Ala. Two additional peptides 12 amino acids in length were constructed: a mutant of the TrpA sequence without the interrupting glycine (Gly-Pro-Ala)₄ and a control peptide (Gly-Pro-Pro)₄. The peptides were acetylated at the N-terminal and amidated at the C-terminal and synthesized with >98 % purity (GenScript).

Circular dichroism (CD). The structural conformations of the peptides were determined using circular dichroism (CD). Peptides were solubilized in methanol/acetic acid solvent (85:15, v/v) (Pécher et al., 2009) to a final concentration of 3 µM. Temperature-dependent UV–CD spectra were obtained using a Jasco J-810 spectropolarimeter as described previously (Toal et al., 2011). Changes in peak value were calculated by dividing the difference in the absolute peak value by the maximum absolute peak.

TrpA modelling

Protein structure prediction. The I-TASSER server was used to create a three-dimensional structure of a portion of the TrpA protein (Zhang, 2008). The 13 amino acid sequence (residues 723–735) used to create the peptide for CD from the TrpA sequence containing the glycine interruption was input into the I-TASSER server. A second, longer (31 amino acids) sequence from the TrpA protein containing the same glycine interruption (residues 711–741) was also input into the I-TASSER server. Five models were created for each sequence that was input into the server; the first model was chosen as the structural model in both cases because the modelling quality for lower-rank models was much weaker than that of the first model given (Zhang, 2008). Predicted protein structure models were viewed as provided by I-TASSER and/or in DeepView (Guex & Peitsch, 1997).

Homology modelling. Based on the structure that I-TASSER ranked as the top template for protein prediction, Protein Database (PDB) code and chain identifier 2kwaA was used as the template for the Automatic Modelling Mode in the SWISS-MODEL server (Schwede et al., 2003). The input target sequence used was the same 31 amino acid sequence that was used for the I-TASSER protein structure prediction. The PDB code 2kwaA encodes the structure of a synthetically constructed collagen heterotrimer that was resolved by solution NMR. The chain identifier for chain A is 2kwaA and is one of three chains. The sequence identity between the template and target sequence was 64.284 %, which is greater than the rule of thumb of 50 % minimum making it an appropriate submission for the ‘Automatic Modelling Mode’ (Arnold et al., 2006). Visual inspection of the model and template structure was done using DeepView (Guex & Peitsch, 1997). All peptide sequences used in this study for peptide studies and modelling are compiled in Table 1.

Culture conditions. *T. erythraeum* IMS101 cells were grown at 26 °C with a 14:10 light–dark cycle at 30–40 µE m⁻² s⁻¹ in amended seawater RMP medium in either 250 or 500 ml polycarbonate flasks as described previously (Price & Anandan, 2013).

Scanning electron microscopy

Immunohistochemistry. *T. erythraeum* IMS101 filaments were collected through filtration onto sterile 40 µm pore, nylon mesh
Table 1. Peptide sequences used for circular dichroism and protein prediction/homology modelling. Protein Database (PDB) code 2kwlA sequence is represented by (GPK)_{10}.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Length (aa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(GPP)_{14}</td>
<td>12</td>
</tr>
<tr>
<td>(GPA)_{14}</td>
<td>12</td>
</tr>
<tr>
<td>(GPA)<em>{12}G(GPA)</em>{12}</td>
<td>13</td>
</tr>
<tr>
<td>(GPA)<em>{16}G(GPA)</em>{14}</td>
<td>29</td>
</tr>
<tr>
<td>(GPK)_{10}</td>
<td>30</td>
</tr>
</tbody>
</table>

cell strainers (BioExpress). The cells were incubated with polyclonal anti-collagen type I–V primary antibody (1:200 dilution) (ab24117; Abcam) at room temperature for 30 min followed by incubation with 5 nm colloidal gold anti-rabbit IgG secondary antibody (G7277; Sigma-Aldrich) for 30 min. Cells were rinsed with RMP medium four times for 30 s each rinse after both antibody incubations. The labelled cells were treated for silver enhancement according to the manufacturer’s protocol (Sigma). Cells were then fixed in 2.5 % glutaraldehyde for 1.5 h, filtered onto 1 μm Nuclepore polycarbonate filters (Fisher Scientific) and post-fixed in 0.1 % osmium tetroxide for 10 min. Cells were rinsed in 0.1 M HEPEs four times for 30 s each rinse after incubation in each fixative.

Sample preparation and imaging. Fixed cells were prepared for SEM imaging as previously described (Marga et al., 2005) except the sample was coated with carbon. Samples were examined in a Supra50VP field-emission scanning electron microscope (Zeiss) at 2 and 7.5 kV using the SE2 or In Lens secondary electron detector (Zeiss). X-ray microanalysis was used to identify the elemental constituents of the sample. Peaks align to specific X-ray lines corresponding with specific elements (Oxford Instruments).

RESULTS

Bioinformatics

As a strategy to investigate the collagen type that TrpA belongs to, multiple alignments of fibrillar and non-fibrillar collagen protein sequences were performed. Phylogenetic distribution of fibril-forming and non-fibril-forming collagens supported the classification of each collagen type. All of the fibril-forming collagens clustered together as did all of the non-fibrillar collagens with the exception of invertebrate fibrillar collagens that were found clustered with the non-fibril-forming collagens (Fig. 1). TrpA was found within the portion of the phylogenetic tree that contained the non-fibril-forming collagens. Within that section, TrpA was most closely associated with invertebrate fibrillar collagens and the two prokaryotic collagen-like sequences that were used in this analysis (Fig. 1).

Circular dichroism (CD)

We then used UV–CD spectroscopy to investigate the PPII nature of regions of the TrpA protein. Synthetic peptides were synthesized to represent regions of the TrpA protein, and UV–CD spectra were obtained for each peptide (GPP)_{14}, (GPA)_{14} and (GPA)_{12}G(GPA)_{12} between the temperatures 5 and 95 °C (Fig. 2). All three peptides exhibit PPII preference with a peak at 215 nm and a negative maximum at 195 nm. An isodichroic point was apparent at 205 nm of each spectrum, representing the two states of each peptide characteristic of PPII peptides. The peptide (GPP)_{14} (Fig. 2) exhibited the most stable PPII conformation with the highest positive peak at 215 nm. When the proline in the Y (GPP)_{14} position was changed to an alanine, (GPA)_{14}, the PPII content decreased. The presence of a glycine insertion, (GPA)_{12}G(GPA)_{12}, exhibited a further decrease in the PPII content, demonstrating the least stable PPII conformation.

Protein structure prediction and homology modelling

To investigate how the glycine insertion found in the TrpA sequence would affect the secondary structure of the peptide, the 13 amino acid peptide sequence (GPA)_{2}G(GPA)_{2} and the 12 amino acid peptide sequence (GPA)_{14}, the latter of which was designed to be a mutation of the TrpA protein without the glycine insertion, were used to predict the structure (Fig. 3). The secondary structures predicted for both sequences were coils and aligned to the same collagen protein (PDB code 2kwlA). The C-score, an estimate of the quality of the predicted model, was −0.80 for (GPA)_{2}G(GPA)_{2} and −0.62 for (GPA)_{14}, confirming that models are correctly folded (>−1.5) (Roy, 2010). This C-score value indicates that both the false positive and false negative rates are below 0.1. The TM-score for both proteins also indicated that they are useful structural analogues (>0.5) that can be used to determine the protein family the sequence is related to; (GPA)_{2}G(GPA)_{2} had a TM-score of 0.61 ± 0.14 and (GPA)_{14} had a TM-score of 0.63 ± 0.13 (Roy et al., 2010). The TM score assesses the topological similarity of protein pairs, and a TM score of >0.5 indicates that the protein pairs have similar folds. The same analysis was done with a longer, 31 amino acid long sequence (TrpA _31), from TrpA extended on either side of the original (GPA)_{2}G(GPA)_{2}, because the length of the collagen chain (PDB code 2kwlA) identified as the best template for the 13 amino acid sequence is 30 amino acids in length (Fig. 4). The C-score and TM-score were −1.07 and 0.58 ± 0.14, respectively.

Using the TrpA _31 sequence as the target sequence and collagen chain 2kwlA as the template, a comparative protein model was constructed (Fig. 5). The identity between the two sequences was 64.28 %, verifying that it is an acceptable target template for alignment and well above the ‘twilight zone’ (>30 %) in which model quality becomes unreliable (Bordoli et al., 2008; Schwede et al., 2003). The glycine insertion in the TrpA _31 target protein creates an obvious puckering or kink in the secondary structure compared to the template structure (Fig. 5a). The region between the first glycine (residue 16) before the glycine insertion in TrpA _31 and the corresponding glycine in the 2kwlA (residue 16) to next consecutive glycine (TrpA _31, residue 20; 2kwlA, residue 19) also displays an increase in length of 1.66 Å (TrpA _31, 10.99 Å; 2kwlA, 9.33 Å). The region of the protein models from the first glycine before the glycine...
insertion and the corresponding glycine in the template to
the final glycine in both models (TrpA_31, residue 16 to
residue 29; 2kw2A, residue 16 to residue 28) exhibits increase
in length, of only 0.01 Å (TrpA_31, 20.13 Å; 2kw2A,
20.12 Å) (Fig. 5b). The angle was measured between glycine
residues 16 and 19 in 2kw2A and residues 16 and 20 in TrpA.
The angle was increased by 16.62° (TrpA, 141.8°; 2kw2A,
125.18°). The change in angle due to the glycine insertion
distorts the helical structure in the target model (Fig. 5c).
We then used SEM with silver enhancement of the sec-
ondary gold colloidal antibodies for visualization of the
collagen-like TrpA protein on the surface of T. erythraeum.

Fig. 1. Maximum-likelihood phylogenetic tree re-created from a multiple alignment of the collagen proteins. The different
collagen types, I–V, VII and invertebrate/prokaryotic, are indicated by different grey and black shaded blocks. The T. erythraeum
collagen is indicated with an asterisk. Collagen classification, fibril-forming and non-fibril-forming, is indicated on the right-hand
side of the tree based on the classification of the majority of collagens in that region. The tree is midpoint rooted. Protein
sequences used for this analysis are given in Table S1.
cells (Fig. 6). The distribution of the TrpA protein on the surface of trichomes, as determined by silver enhancement, indicates that there is no specific pattern of localization. This means that the protein is localized over the entire surface of the trichome. Untreated controls for silver enhancement were negative for labelling.

**DISCUSSION**

Collagens have evolved to provide a variety of structures such as fibrils, microfibrils and meshwork (Kielty & Grant, 2003). Classification of collagen into subfamilies, fibril-forming or non-fibril-forming, takes into account a number of factors including the primary sequence, the length of the collagen chains and the supramolecular quaternary structures that form. We have combined these factors to make a prediction about the structural classification of TrpA, a collagen-like protein, as a strategy to investigate the relationship between the structure of this protein and its function.

A number of collagen proteins contain short collagenous domains that can be either continuous or interrupted and form structures including fibril-associated collagens with interrupted triple helical domains (e.g. fibril-associated collagens with interrupted triple helices), transmembrane collagens and multiplexin collagens (Kielty & Grant, 2003).

**Fig. 2.** Temperature-dependent UV–CD spectra of peptides (a) (GPP)₄, (b) (GPA)₄ and (c) (GPA)₂G(GPA)₂ in methanol/acetic acid solvent (85:15, v/v) are shown. Arrows indicate increasing temperature from 5 to 95 °C; res, residue.

**Fig. 3.** Three-dimensional protein prediction from I-TASSER of (a) (GPA)₄ and (b) (GPA)₂G(GPA)₂ viewed in DeepView. Solid, three-dimensional Van der Waals representations of the backbone are shown. Glycine residues are highlighted in red. The additional glycine found between consecutive Gly-X-Y is highlighted in orange.

**Fig. 4.** Three-dimensional protein prediction from I-TASSER of (a) PDB code 2klwA, (GPK)₁₀ and (b) TrpA_31, (GPA)₆G(GPA)₄ viewed in DeepView. Solid, three-dimensional Van der Waals representations of the backbone are shown. Glycine residues are highlighted in red. The additional glycine found between consecutive Gly-X-Y is highlighted in orange.
The triglycine repeat length of the TrpA protein found in *T. erythraeum* is approximately 10% longer than that from any other sequenced collagen from eukaryotic or prokaryotic organism (Layton et al., 2008). Based on length alone, TrpA can be excluded from creating a structure that is typically formed by short collagen domains.

Based on whole genome sequencing, *T. erythraeum* contains only one gene encoding a collagen-like protein predicting that only a homotrimeric helix would form. Previous work from this laboratory confirms that there is only one protein in this organism that cross-reacts with human collagen type I–V antibody (Price & Anandan, 2013). Using amino acid sequences of the 80 vertebrate, invertebrate and prokaryotic collagens, we gained insight into what structure this collagen-like protein may form based on its grouping with other collagen proteins. We used a ML approach because it provides the highest aggregate probability, which most likely reflects the true phylogenetic tree (Krane & Raymer, 2003). *T. erythraeum* collagen grouped with the clade that included non-fibrillar collagens (Fig. 1). Interestingly, all of the invertebrate fibrillar collagens also grouped amongst the non-fibrillar collagens in this section. The reason the invertebrate fibrillar collagens grouped with the non-fibrillar collagens is because invertebrate fibrillar collagen sequences contain imperfections. The presence of an additional glycine is

---

**Fig. 5.** Homology model of TrpA 31, (GPA) 5 G(GPA) 4 , using PDB code 2klwA (GPK) 10 viewed in DeepView. Solid, three-dimensional Van der Waals representations of the backbone are shown. Glycine residues are highlighted in red. The additional glycine found between consecutive Gly-X-Y is highlighted in yellow. 2klwA (top) and TrpA 31 (bottom) side by side (a) front view and (b) side view. (c) Perturbation caused by the addition of glycine can be seen.

---

**Fig. 6.** Scanning electron microscopy of carbon coated silver enhancement treated trichomes of *T. erythraeum* that were immunogold labelled and probed for collagen. Panels (a) ×500, (b) ×3000, (c) X-ray confirmation of silver labelling. Untreated *T. erythraeum* filament is shown as insert in (a). Scale bar, 5 μm.
characteristic of invertebrates and lower vertebrate chains. For example, the sequences Gly-X-Gly of Riftia, a hydrothermal vent tube worm, or Gly-X-Y-Z seen in sponges, have been shown not to affect molecular stability and fibrillar integrity in these proteins (Sicot et al., 1997; Exposito & Garrone, 1990). In the z-chain of Riftia, there is a single replacement of a glycine with an alanine in the regular Gly-X-Y repeat that should cause a disturbance in the triple helical rigidity. It does not prevent triple helical folding although it does decrease thermal stability (Mann et al., 1992). The presence of imperfections in invertebrate collagens does not impair fibril formation (Mann et al., 1992).

The low environmental temperature at which these organisms that contain the imperfect fibril-forming collagen sequences are found has been implicated as a factor in preventing these imperfections from affecting the structure (Exposito et al., 2002). Considering T. erythraeum is found in similar temperature environments (25–27 °C), we thought it possible that although phylogenetically the TrpA protein is in a clade of non-fibrillar collagens, the imperfection found in the sequence may not impair fibril formation similar to the fibril-forming invertebrate collagens it groups with in this clade (Fig. 1).

To test and determine if the glycine imperfection in the T. erythraeum collagen sequence would prevent triple helical conformation, synthetic peptides were synthesized based on the TrpA sequence. The peptide sequence (GPP)₄ was used to represent a sequence more similar to that of vertebrate collagen, however without the hydroxyproline in the Y position. In vertebrates, the presence of hydroxyproline is a stabilizing force in triple helices. However, upon the identification of collagen-like sequences in prokaryotes it has been shown that hydroxylation of the proline residue is not a requirement for triple helical structure (Mohs et al., 2007). Collagen-like proteins in B. anthracis and S. pyogenes have been shown to self-assemble and form triple helical structures in the absence of hydroxyproline (Mohs et al., 2007). The sequence (GPA)₂G(GPA)₂ is representative of the collagen sequence from residue 723 to residue 735 of the TrpA protein, while the (GPA)₄ sequence is a mutation of the T. erythraeum sequence without the imperfection. The CD spectra revealed that (GPA)₂G(GPA)₂ has a lower PPII content and is less stable than both (GPP)₄ and (GPA)₄, in which there is no glycine insertion.

The CD results correspond with the predicted model of TrpA from 1-TASSER and the homology model of TrpA from SWISS-MODEL, both of which indicate that the additional glycine would cause a disruption to the secondary structure of the TrpA collagen chain, decreasing its stability. However, (GPA)₂G(GPA)₂ still exhibits PPII content and, thus, does not show any indication that this structure would be incapable of self-assembling to form collagen triple helices. The CD spectra indicate that at lower temperatures the PPII content and stability of the (GPA)₂G(GPA)₂ would increase, aligning with previous research showing that collagen proteins that contain interruptions found in invertebrates growing at lower temperatures form stable, fibril-forming collagens (Exposito et al., 2002).

Previous work from this laboratory has shown that the TrpA protein is localized to the junctions between cells along the multicellular trichome of T. erythraeum (Price & Anandan, 2013). To further resolve the localization and function of the TrpA protein on the surface of the filaments, SEM was used to determine if this protein is found on the cell surface, and to observe its structure on the cell surface. Silver enhancement with SEM revealed that the TrpA protein is found on the cell surface, and is not localized to a specific region of the trichome surface. Though SEM is only able to resolve the image of the surface of the cells, in contrast to our previous scanning confocal microscopy results, which were able to image at various planes through the sample, these SEM results add valuable information about additional localization of the TrpA protein in T. erythraeum (Price & Anandan, 2013). Attempts to pre-treat trichomes with collagenase prior to SEM were unsuccessful, as the treated trichomes were not sufficiently robust after the enzyme treatment to withstand sample preparation for SEM. The pattern of TrpA localization seen with SEM indicates that this collagen-like protein in T. erythraeum is a component of the outer sheath of the trichome (Fig. 6). These data indicate that the presence of the collagen-like TrpA protein may function in the adhesion of trichomes to one another in colony formation. It is also possible that heterotrophic bacteria utilize the presence of this protein on the surface of trichomes for adherence and in the formation of a multi-species biofilm. T. erythraeum is found in oligotrophic regions of subtropical oceans, and the formation of multi-species biofilm communities may aid in its acquisition of mineral nutrients (Capone et al., 1997). In support of this idea, Rubin et al. (2011) have demonstrated that natural puff-shaped colonies of T. erythraeum are very effective at dissolving dust and oxides as a mechanism for the acquisition of dissolved iron compounds. Work by Roe et al. (2012) has also suggested that possible mutualistic relationships between heterotrophic bacteria and T. erythraeum may play a role in iron acquisition by T. erythraeum cells. We suggest, then, that a possible role for the TrpA protein on the outer sheath of trichomes is in promoting colony formation in the natural habitat. Iron is an integral component of the Mo–Fe nitrogenase protein responsible for biological nitrogen fixation and the diazotrophic life style of T. erythraeum. Thus, the formation of T. erythraeum colonies is essential for the physiology of this organism and its success in its preferred environment (Kustka et al., 2003; Lenes et al., 2001). The TrpA protein being a non-fibril-forming collagen probably functions in the adherence of trichomes and heterotrophic bacteria, to create mixed communities within the puff-shaped colonies demonstrated by Rubin et al. (2011).

In conclusion, this study demonstrates the probable structure of the TrpA protein as a non-fibril-forming collagen that is stable at lower temperatures characteristic of the natural environment of T. erythraeum. We also...
demonstrate that this protein is a component of the outer sheath of this organism, which suggests that this protein may participate directly in the adherence of trichomes to form colonies, and indirectly in the formation of colony-associated mixed bacterial communities. The colonies of *T. erythraeum* have been shown to support heterotrophic bacteria and represent active micro-habitats for biogeochemical cycling, particularly iron. The establishment of these colonies therefore impacts the well being of entire microbial communities, and impacts biogeochemical cycles in this environment.

**ACKNOWLEDGEMENTS**

*T. erythraeum* IMS101 was a gift from Dr John Waterbury at Woods Hole Oceanographic Institute. We thank Dr Jacob Russell, Department of Biology, Drexel University, for valuable discussions on the phylogenetic analyses, and Dr Edward Basgall and the Centralized Research Facilities at Drexel University for training on scanning electron microscopy. This research was funded by National Science Foundation grant CMMI-0900101.

**REFERENCES**


Edited by: A. Wilde