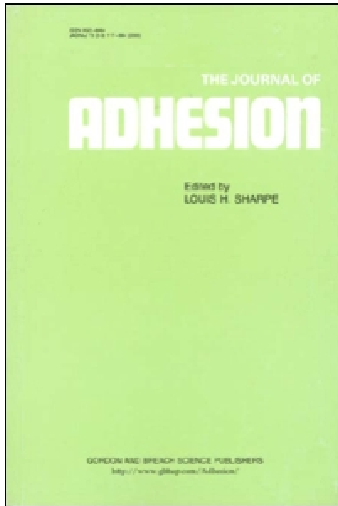
This article was downloaded by:
On: *21 January 2011*
Access details: *Access Details: Free Access*
Publisher *Taylor & Francis*
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



The Journal of Adhesion
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713453635

Glueomics: An Expression Survey of the Adhesive Gland of the Sandcastle Worm

Betsy J. Endrizzi[a]; Russell J. Stewart[a]
[a] Department of Bioengineering, University of Utah, Salt Lake City, Utah, USA

To cite this Article Endrizzi, Betsy J. and Stewart, Russell J.(2009) 'Glueomics: An Expression Survey of the Adhesive Gland of the Sandcastle Worm', The Journal of Adhesion, 85: 8, 546 — 559
To link to this Article: DOI: 10.1080/00218460902996457
URL: http://dx.doi.org/10.1080/00218460902996457

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.informaworld.com/terms-and-conditions-of-access.pdf

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

The Journal of Adhesion, 85:546–559, 2009
Copyright © Taylor & Francis Group, LLC
ISSN: 0021-8464 print/1545-5823 online
DOI: 10.1080/00218460902996457

Taylor & Francis
Taylor & Francis Group

Glueomics: An Expression Survey of the Adhesive Gland of the Sandcastle Worm

Betsy J. Endrizzi and Russell J. Stewart
Department of Bioengineering, University of Utah,
Salt Lake City, Utah, USA

Random clones were sequenced from a cDNA library constructed from the adhesive gland of Phragmatopoma californica*, a marine polycheate that builds protective shells by gluing together sand grains and biogenic mineral fragments. As many as 14 new proteins and two phenoloxidase enzymes were found that may be structural components of or involved in processing the bioadhesive. Glue protein classification was based on the following criteria: (i) the presence of predicted secretion signal peptides, (ii) low complexity sequences, (iii) strongly skewed amino acid compositions enriched with G, Y, K, H, A, or S, (iv) repeating peptide motifs, and (v) homology to known glue proteins, other structural proteins, or enzymes. The new genes provide probes for further characterization of the adhesive gland as well as potential biotechnological resources and insight.*

Downloaded At: 19:54 21 January 2011

Keywords: Bioadhesive; Glycine-tyrosine rich; Matrix proteins; *Phragmatopoma californica*; Polycheates

1. INTRODUCTION

The polychaetes are a diverse and successful taxon within the segmented worm phylum (Annelida) numbering more than 8,000 species and occupying nearly every niche in the oceans [1]. Some polychaetes are fast and agile predators; others are stationary filter feeders that live within mineralized tubular shells. This latter group can be divided

Received 5 December 2008; in final form 20 March 2009.
One of a Collection of papers honoring J. Herbert Waite, the recipient in February 2009 of *The Adhesion Society Award for Excellence in Adhesion Science, Sponsored by 3 M*.
Address correspondence to Russell J. Stewart, Department of Bioengineering, 20 S. 2030 East, Room 506C, University of Utah, Salt Lake City, UT 84112, USA. E-mail: rstewart@eng.utah.edu

into genera that secrete calcium carbonate with an organic matrix to form solid shells and genera that have entirely foregone synthesis of the mineral phase of their shell and instead live in tubular concretions of adventitiously gathered mineral particles [2]. *Phragmatopoma californica*, also called the Sandcastle worm, is a polychaete in this latter category which lives in intertidal zones off the coast of California. The worm deploys a crown of ciliated tentacles for capturing and transporting food and particulates from the water column to its mouth where the captured materials are evaluated with ciliated lips. Food is ingested, unsuitable particles are cast away, while particles judged to be of the right size, shape, composition, and surface chemistry for incorporation into the tube are passed to a pair of dexterous palps, the so-called building organ, located immediately ventral to the mouth. Two minute spots of proteinaceous adhesive are applied before the particles are pressed into place by the building organ onto the end of the tube. The worm wriggles the newly placed particles until the adhesive sets, which takes less than 30 s under cold salty water [3].

Technological interest in the *P. californica* adhesive stems in large part from its ability to bond a diverse range of wet substrates in one step with minimal surface preparation using an apparently self-contained mechanism to trigger the underwater setting reaction [4]. The adhesive is secreted as a colloidal suspension [3] with low initial viscosity and interfacial tension that allows it to spread readily over the surface of wet substrates, yet it is sufficiently cohesive that it does not disperse into the ocean. What is more, the water-borne bioadhesive can apparently displace surface bound water from hydrophilic minerals, a prerequisite for strong interfacial adhesion [5]. Synthetic underwater adhesives meeting all of these specifications have yet to be invented. The *P. californica* adhesive is a valuable paradigm for the design and synthesis of new water-borne, underwater adhesives. Characterization of the protein composition and biological processing of the natural adhesive is a necessary step toward synthesizing effective mimics.

The sequences of four *P. californica* adhesive protein genes have been reported [6]. The sequences of the first two, referred to as Pc-1 and Pc-2, were largely known from microsequencing of tryptic peptides derived from two proteins isolated from the worm's thorax, which contains the adhesive glands [7]. Their genes were cloned from adhesive gland cDNA using degenerate PCR primers corresponding to the known amino acid sequences [7]. Pc-1 is comprised mostly of just three residues, glycine (45 mol %), lysine (14 mol %), and tyrosine (19 mol %), which occur as several repeats with only conservative variations of the sequence VGGYGYGGKK. Pc-2 consists of several degenerate copies of the sequence HPAVHKALGGYG. Both proteins

Downloaded At: 19:54 21 January 2011

are basic (pIs > 9) and approximately two-thirds of the tyrosine residues are modified to 3,4-dihydro-L-phenylalanine (DOPA) [7]. The sequences of Pc-1 and 2 are not highly homologous to the comparatively well-studied adhesive proteins of the mussel genus *Mytilis*, although they do have several general features in common: signal peptides expected for secreted proteins, highly repetitive sequences based on multiple copies of short peptide motifs, and DOPA residues. DOPA is thought to play multiple roles in mussel adhesion [8], which suggests *P. californica* may use some similar adhesive mechanisms despite phylum-level separation between polychaetes and mussels. The similarities to known mussel adhesive proteins and isolation of the proteins from the adhesive gland-containing thorax, though not definitive, are the best evidence that Pc-1 and -2 are present in the secreted adhesive.

The presence of a serine-rich adhesive protein or proteins was surmised from the observation that the *P. californica* adhesive contains nearly 30 mol % serine [7,9]. Degenerate PCR primers encoding five consecutive serines were used to clone two closely related serine-rich proteins, Pc-3A and -3B, and several minor variants of each from adhesive gland cDNA [6]. The Pc-3A sequence consists almost entirely of runs of 4–13 serines punctuated with single tyrosine residues. Pc-3B contains a similar but shorter tyrosine punctuated polyserine segment fused to a non-repetitive C-terminal domain. Since at least 95% of the serines in the *P. californica* adhesive are phosphorylated [4], the Pc-3 proteins are actually comprised of polyphosphoserine and are therefore extremely acidic. The abundant presence of phosphorus evenly distributed throughout the secreted adhesive is clear evidence that the Pc-3 proteins are integral components of the set adhesive [4]. Notwithstanding the handful of scattered phosphoserines in the interfacial mussel foot protein, Mefp5 [10], the extensive runs of polyphosphoserine in the Pc-3s are unique amongst known bioadhesive proteins and suggest novel bonding mechanisms at work in the *P. californica* adhesive.

The *P. californica* adhesive when set is by nature insoluble, resistant to being taken apart, and, therefore, mostly impervious to conventional biochemical analysis of its components. In this report, we describe an alternate tack—expression analysis of the *P. californica* adhesive glands to provide a panoramic survey of adhesive biosynthesis. This approach efficiently identified structural protein candidates, including minor constituents, as well as enzymes potentially responsible for critical post-translational modifications, which may be valuable resources for producing synthetic or semi-synthetic adhesive mimics.

Downloaded At: 19:54 21 January 2011

Downloaded At: 19:54 21 January 2011

2. MATERIALS AND METHODS

2.1. Animals

Soccer ball-sized fragments of *P. californica* colonies were collected off the coast of Goleta, California, and maintained in a laboratory seawater tank.

2.2. Cryosectioning and Arnow Staining

Worms were removed from their tubes, chilled to 4°C, anesthetized with $MgSO_4$, then fixed in 4% paraformaldehyde in seawater (pH 7.2) at 4°C overnight. The fixed worms were cryoprotected by equilibration in 30% sucrose in seawater (pH 7.2) for 24 h. Sections (Leica Microsystems Cryostat, Bannockburn, IL, USA) of frozen worms were post-fixed in 4% paraformaldehyde in seawater at room temperature for 20 min. For Arnow staining [11], the sections were rinsed with seawater three times, transferred to 0.5N HCl for 30 s, then Nitrite-molybdate reagent for 60 s, and 1N NaOH solution for 60 s, where upon DOPA-containing proteins turn red.

2.3. Construction of Adhesive Gland cDNA Library

The adhesive gland was identified in the thorax using Arnow's reagents which react with DOPA-containing proteins. The red reaction product appeared in the extensive white glandular tissue associated with the outside wall of the coelom in the three parathoracic segments (Fig. 1). The red staining extended up through the building organ palps. The parathoracic region was dissected from approximately 15 worms and ground in liquid N_2 in a cold mortar and pestle. The frozen powder was homogenized in TRIzol Reagent (Invitrogen, Carlsbad, CA, USA) and total RNA was isolated according to the manufacturer's instructions. mRNA was isolated and converted into cDNA using the Cloneminer cDNA library construction kit (Invitrogen, Carlsbad, CA, USA) following the manufacturer's instructions.

2.4. Sequencing

Plasmids were purified from 280 random *E. coli* colonies with a plasmid mini-prep kit (Qiagen, Valencia, CA, USA). DNA sequencing was performed with Applied Biosystems BigDye v.3.1 (Foster City, CA, USA) terminator chemistry and capillary electrophoresis using standard methods and conditions at the University of Utah DNA Sequencing and Genomics core facility.

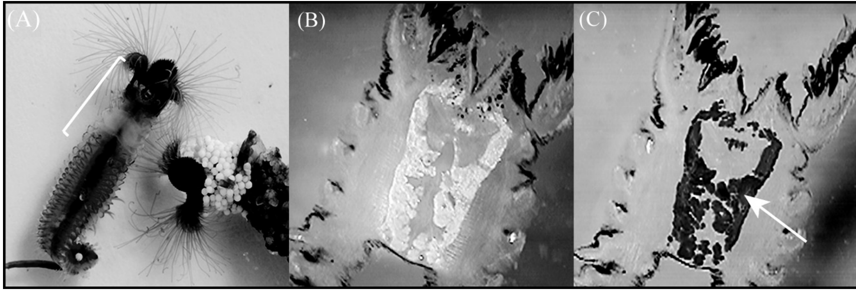


FIGURE 1 (A) *P. californica* out of its tube. The parathoracic segments indicated by the white bracket form a white band below the cephalic tentacles. (B) Coronal cryosection through thoracic region. (C) Coronal cryosection stained with Arnow's reagents. The arrow indicates the dark red-stained portion of the adhesive gland, which appears black in the figure.

Downloaded At: 19:54 21 January 2011

2.5. Sequence Analysis

Open reading frames were identified using the ORF finder tool and the protein sequences were compared with the non-redundant protein sequence database for homologies using pBLAST at the National Center for Biotechnology Information (NCBI, Bethesda, MD) [12]. Protein sequences were further analyzed using the proteomics and sequence analysis tools at ExPasy proteomics server (Lausanne, Switzerland): signal peptide prediction with SignalP 3.0 [13], subcellular location with PSORTII [14], and amino acid composition, MW, and pI with ProtParam [15].

2.6. Database Deposition of Sequences

The sequences were submitted to dbEST at the NCBI and assigned the following accession numbers: Pc-4 (GH160602), Pc-5 (GH160603), Pc-6 (GH160604), Pc-7 (GH160605), Pc-8 (GH160606), Pc-9 (GH160607), Pc-10 (GH160608), Pc-11 (GH160609), Pc-12 (GH160610), Pc-13 (GH160611), Pc-14 (GH160612), Pc-15 (GH160613), Pc-16 (GH160614), Pc-17 (GH160615), Pc-18 (GH160616), tyrosinase (GH160617), and laccase 2 (GH160619).

3. RESULTS AND DISCUSSION

3.1. Expression Survey

Sequences were sorted into broad categories to provide an overview of adhesive gland expression (Fig. 2A). Approximately 50% of the genes

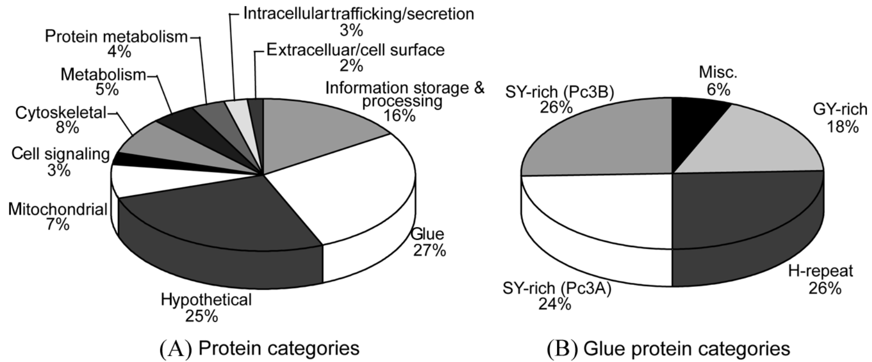


FIGURE 2 Gene frequencies. (A) Functional categorization of total genes. (B) Sub-categorization of glue genes.

Downloaded At: 19:54 21 January 2011

had well-known homologs involved in cellular housekeeping. The largest of this category (16%) were genes involved in storage, retrieval, and translation of genetic information. Another large category (25%) was hypothetical genes containing apparent open reading frames but which had no recognizable homologs in the genetic databases and, therefore, inscrutable functions. All four previously published genes, Pc1-3AB, were found in the first 20 randomly selected adhesive gland cDNAs, which demonstrated the efficiency of the approach for identifying and cloning potential glue protein genes.

In total, about 25% of the randomly selected genes were classified as encoding known or potentially new glue proteins. Classification was based on the following criteria: (i) the presence of a predicted secretion signal peptide since glue proteins are obviously secreted, (ii) low complexity sequences, (iii) strongly skewed amino acid compositions especially enriched with G, Y, K, H, A, or S, (iv) repeating peptide motifs, and (v) homology to known glue proteins or other structural proteins. The putative glue proteins thus identified were further sub-categorized into four groups: (i) GY-rich, with Pc-1 as the prototypical member; (ii) H-repeat, as exemplified by Pc-2; (iii) SY-rich variants of Pc-3A and -3B; and (iv) an eclectic miscellaneous category of previously unidentified proteins meeting the criteria but which did not fall into the first three categories (Fig. 2B).

3.2. SY-Rich Category

The SY-rich was by far the most abundantly expressed category. Pc-3A and -3B were expressed in equal quantities and accounted for 50%

of the total glue gene expression. The high level of expression of the Pc-3s suggested polyphosphoserine may be the most abundant component in the glue. However, expression levels may have limited utility as a predictor of glue composition. Multiplying the amino acid composition of the putative glue proteins by their relative expression levels, then summing to predict mol % of each amino acid in the secreted glue resulted in over-estimation of S by nearly two-fold and under-estimation of G by half compared with the empirically determined amino acid composition (Table 1). Most other amino acids were under-estimated. The amino acid ratios have stayed fairly constant as the database has grown so it seems unlikely G is underestimated because the database is statistically skewed by too few cDNAs. It is more likely that S is overestimated because it was not accessible to amino acid analysis in the insoluble and crosslinked glue. Indeed, it has not been possible to biochemically isolate the Pc-3 proteins from either the cement gland or the glue [6,7].

Downloaded At: 19:54 21 January 2011

3.3. GY-Rich Category

The GY-rich category is characterized by the presence of repeats of the sequences GYGY, GYGGY, or GYGGGY (Fig. 3). BLAST searches of sequence databases with the GY-rich proteins turned up similarities to a broad array of low-complexity, glycine-rich structural proteins.

TABLE 1 Amino Acid Composition

	Predicted	Measured
Ala (A)	4.7	9.8
Arg (R)	2.9	2.9
Asx (D/N)	1.9	2.8
Cys (C)	1.7	0.4
Glx (E/Q)	0.5	1.4
Gly (G)	13.5	26.2
His (H)	2.4	3.5
Ile (I)	0.9	0.6
Leu (L)	2.6	3.4
Lys (K)	4.3	4.4
Met (M)	0.3	–
Phe (F)	0.9	1.1
Pro (P)	1.4	2.7
Ser (S)	48.7	28.5
Thr (T)	1.6	2.2
Trp (W)	0.6	–
Tyr (Y)	11.3	6.1
Val (V)	3.2	3.4

Pc-1
MW 18, 240 pI = 9.74 7.7%
```
MKVFIVLALVSAA YGCGVGIGC
AGGRCGGACG GKGYGYGG K
LGYGAYGKGG IGGYGYGKGC
VGGYGYGGLG      AGK
LGGYGYGGSK CGGYGYGGQK
LGGYGYGGKK LGGYGYAAKK
VGGYGYGAKK VGGYGYGAKK
VGGYGYGAKK VGGYGYGAKK
VGGYGYGAKK VGGYGYGAKK
VGGYGYGVKK VGGYGYG
```

Pc-7
MW 13, 370 pI 8.72 1.3%
```
MNTFVVLAAIVAVAA
CSGGYDGRQYTYRGR
YNNKCGNDGLYFKDD
KNFXFCSNGNSYVQP
CAPGTRNSG
YNNYKQGSI
YN YRDFCDVNLVDE
GYGVGAKP
GYNK GYNP GYNP
GYG  GYNP GYST
GYGGYKAGPGPYW
```

Pc-9
MW 19, 843 pI 9.21 1. 3%
```
MKVSIAVLIMCCIAAVLADGY
KSKNGGQAGGYGGY
    GSGYGGG
        YGGGYDGGYGGE
KGKSGKGYGDR
KGKSEKGYGNG
KGKGGSGYGGGY    DGGYGGG
KGKSGSGYGGGY    DGGYGGG
KGKSGSGYGGGYDGGYDGGYGGG
KGKSGSGFGGGYDGGYDGGYGGG
KGKSGSGYGGGYDGGYDGGYGGG
KGKSGSGYGGGYDGGYDGGYGGG
KGKSGSGYGGGYDGGYDGRYGGG
KGKSGSG
```

Pc-13
MW 12, 849 pI 7.20 1. 3%
```
MKVFVAALLLCCIAAAAA
EDGYGFGYDGYGSG YGYDGYS
YGGDKGYG YGKGKGYG
YEGGKGYG YEGGKGYG
HEEGKGYG HEGGKGYG
YEGGKGYG YGGGKGYG
HDGGKGYG HDGGKGYG
YGGGKGYG HEGGKGYG
YEGGKGYG KY
```

Pc-11
MW 8,7 64 pI 9.59 2.6%
```
MKLAVFALLVAFAIVYTA
EGLVYGGQ
KGYGYGG
KGYGYG CTGGYGLYGG
KGYGYG
KGYGYG CRGGYG YG
KGYGYGGK Y
RGYGYGNK VGYG YG
QQLGYKNGRK
```

Pc-16
MW 7,2 43 pI 8.64 1.3%
```
MKVFVATLLLCCIAAAAA
 AGYGNG YAGYGSG
YAGYGTG YAGYGSG
YGYDG    YGYGGGKG
YGYGGDKG YGYGGDKG
YGYGGQKG YGYGYGKY
```

FIGURE 3 GY-rich gene category. The frequency of expression of each gene is represented as a percentage of the total number of genes categorized as potential glue proteins.

The GYG motif is the defining feature of several large families of so-called glycine-tyrosine-rich keratin-associated proteins (KAPs) found in mammalian hair, horns, and nails [16–18]. The GY-rich KAPs, along with high-sulfur KAPs, form an amorphous matrix around keratin intermediate filaments in the hair cortex that provides rigidity and toughness to the hair shaft. The GYG motif is also present in several proteins of the pearl oyster shell matrix: prismalin-14, which contains a central GY-rich domain that separates two aspartate-rich calcium-binding terminal domains [19]; a set of lysine-rich matrix proteins (KRMPs) that have a basic N-terminus coupled with a GY-rich C-terminus [20]; and several members of the shematrin family of hydrophobic matrix proteins [21]. Other materials that contain GY-rich proteins include plant cell walls [22], insect cuticle [23,24], eggshell cases [25], tick cement [26], and the pressurized capsules of hydra nematocysts [27]. In common, this diverse range of GY-rich natural materials occur in tough and insoluble extra-cellular structures. In most cases, they seem to form crosslinked, re-enforcing matrices that may provide energy absorbing elasticity to the material. The GY-rich proteins of *P. californica* may have a similar general function.

Downloaded At: 19:54 21 January 2011

Pc-1 is the most abundantly expressed of the GY-rich group. Amino acid analysis of Pc-1 precursor protein isolated from cement glands revealed at least two thirds of its tyrosines were post-translationally hydroxylated into DOPA [7]. In the byssal plaque of mussels, DOPA is thought to play two key roles in bonding. First, the unoxidized catechol form has been demonstrated to adhere strongly through chelate bonds to metal oxide surfaces [28]. Second, at the pH of seawater the catechol form can be oxidized by O_2 to form an electrophilic quinone, which can covalently couple with amine and cysteine containing sidechains. The quinone-mediated crosslinking leads to hardening of the adhesive. The reddish-brown color that develops over several hours after secretion and the presence of 5-S-cysteinyl-DOPA residues in the crosslinked glue [6] suggest that similar mechanisms operate in the *P. californica* adhesive.

There are at least two caveats associated with extrapolating the function of the new GY-rich proteins from the function of other DOPA-containing adhesive proteins. First, the set glue contains 4.0 mol % tyrosine and 2.1 mol % DOPA [7]. This ratio is skewed by underestimation of DOPA because DOPA crosslinked with nucleophilic residues are lost to amino acid analysis. Nevertheless, the significant amount of unmodified tyrosine in the secreted glue suggests the tyrosines of some or all of the new GY-rich group members may not be hydroxylated into DOPA as extensively as Pc-1. Second, the new proteins may not be in the adhesive mortar between sand grains. The worm also secretes a reddish-brown organic sheath that lines the inside of its mineral tube. Little is known about the composition, genesis, or area of the worm from which the tube liner is secreted. Some of the new GY-rich proteins may be associated with this tube liner.

Downloaded At: 19:54 21 January 2011

3.4. H-Repeat Category

This category is characterized by the presence of repeating peptides containing one or more H residues, though the categorization is blurred by the presence of the GY motif in some members of the H-repeat group (Fig. 4). BLAST database searches provided little additional insight into potential functions. The proteins have only short regions of homology to other glycine-rich low complexity proteins. Pc-2, the most abundantly expressed of the group, has 7 out of 10 of its tyrosines modified to DOPA. As in the case of the Pc-1-like proteins, the extent of hydroxylation of the Pc-2-like proteins is unknown.

Based on the well-known propensity of H-containing peptides to bind metal ions, the H-repeats may function as metal binding domains. Numerous H-containing sequences have demonstrated

Pc-2
MW 21, 116 pI 9.91 9.0%
```
MKVLIFLATVAAVYG CGGAGG
WRSGSCGG RWG
HPAVHKALGGYGGYGA
HPAVHAAV
    HKALGGYGAGAYGAGAWG
HPAVHKALGGYGAGA     WG
HPAVHKALGGYG GYGA
HPAVHVAVHKALGGYGAGACG
HKTGGYGGYGA
HP V  AV KAAYNHGFNYGA
NNAIKSTKRFG      GYGA
HPVVKKAFSRGLSHGAYAG
SKAATGYGYGSGKAAGGYGY
```

Pc-10
MW 17, 665 pI 11.68 2.6%
```
MKLICLVLLAVCIVAVSA
SSSSGGRRRRVIVIGNRGR
    APARPRSNIHYHMHAPQP
RMMMAP RMMMAP MMMAP
MAMP ATSHVYQSVSYPGA
MYRYGLGSL GGGFISGGL
 GGIVGGGL HGGVVTSGL
HGGVVTSGL HGGVVTSGL
HGGLVSGGW HSGVVS
HGGLIGGGI HTT
YGG      FHKGVV
HGGYTGHYGKRR
```

Pc-5
MW 13, 581 pI 10.29 1.3%
```
MKFLVLLALVASASA
YYPLMGGF
HGGWHAPMVHGGLY
HGGWHAPMVHGGLY
HGGWHAPIVHGG WHAPVF
    HAPAPIHTVSHSVVN
    HVPMMPM  WHHPAP
APAPAPRPGRTIILGGGKY
GPFGKYGGGAGLLALGALG
GNGGFWKRR
```

Pc-4
MW 24, 330 pI 9.49 11.5%
```
MPTLYKKVGKLVILAIIVTVASVASA
GYPTYSPSGGTHSGYNGPHGNVVKK
TYRGPYGAGAAK
AWNGYHGAGYTSVHHGPASTSWHTS
WSNKKGGYGYGLK
   NK-GYGYGLKKVGY
 GVGL       HAAGW
HGVGPYGAGY  HGAGW
NGLGYHGAGYGV HGVGLHGAGYGL
HGVGLHGVGYGL HGVGLHGAGYGL
HGVGLHGAGYGI HGVGLHGAGYGI
HGVGLHGVGYGL HGVGLHGAGYGL
HGVGLHGAGYGI HGVGLHGAGCGI
HKTACY         GVGLHG
HY
```

Pc-18
MW 21, 282 pI 9.33 1.3%
```
LDGGCKPTGGFIKGSVGPCGGYN
HQHVVGPNGA
HGRRVGYGPNGKYSQIYGNGPGGRYS
HTVVYPRVRPYGGYGFKGGYGGY
HGVGYKGGYPGVGYHGGY
HGVSYSGGYPGVGYHGGY
HGVSYSGGYPGVGYHGGY
HGVSYSGGYPGVGYHGGY
HGVRYSGGYPGVGYHGGY
HGVSYSGGYPGVGYHGGY
HGVSYSGGYPGVGYHGGY H
```

Pc-15
MW 21, 005 pI 9.97 2.6%
```
MKVIILLATVAAV
YGCGGWNGGFGGGKAC
  GGGWGAKALGGYGSYNGNGYGA
HPVAVKSAFNKGVSYGA
RSAVKATRGFAYGKGSSYGYGA
HPAVKSAFGNGFKTGAHAAVNGYG
YGAVKSALSGGYGYGSYGA
HPAVKSAYRKGLSYGAK
 SAVKATRGFAYGRSGYGA
HPVVKSAFSNGFKYGA
HAAVKATNGYGYGAV
HPAVKAAVK GGYGYG
NKGGYGAGYAAY
```

FIGURE 4 H-repeat gene category. The frequency of expression of each gene is represented as a percentage of the total number of genes categorized as potential glue proteins. The Pc18 cDNA was incomplete at the 5′ end.

Downloaded At: 19:54 21 January 2011

affinity for metal ions. An obvious example is the H_6-peptide tag widely used for immobilized metal affinity chromatography with a variety of complexed transition metal ions [29]. In a similar vein, peptides selected for transition metal affinity from random peptide libraries invariably contained at least one and usually several H residues [30,31]. As another example, Ni^{2+} and Cu^{2+} are transported in blood as complexes with the DAH tripeptide at the N-terminus of human serum albumin [32]. And the peptide sequence XXH forms redox active complexes with Ni^{2+} capable of catalyzing protein crosslinking through tyrosine residues [33–35]. Experiments are in progress to quantify the Ni, Cu, and other transition metal concentrations in the secreted adhesive. The presence of stoichiometric transition metal concentrations would suggest several potential functions of the repeating H-peptides. They may form redox active complexes that catalyze oxidative crosslinking through DOPA residues. Or, they may form structural domains organized around a complexed metal ion analogous to the H-containing zinc finger domain [36] and numerous other metal-organized domains [37]. Or, they could form

an intermolecular bridging complex between proteins that contributes to cohesion similar to the mechanisms proposed for H-rich mussel byssal plaque proteins [38].

3.5. Miscellany

The third glue protein category is an intriguing potpourri of previously unknown proteins with only limited and scattered homologies with known proteins (Fig. 5). For example, the repeating positively charged RGGR motif of Pc-8 occurs in RNA-binding proteins where it could be expected to associate electrostatically with the phosphodiester backbone of nucleic acids. Pc-8 is not likely to be a nucleic acid binding protein because it has a secretion signal peptide. The C residues flanking the RGGR motif as well as the GY-rich blocks at the N- and C-termini are unique to Pc-8. At greater than 13 mol % it has the highest C content of the potential *P. californica* adhesive proteins. Its C residues may participate in cysteinyl-DOPA covalent crosslinking [6]. *In vitro*, efficient covalent crosslinking between dopaquinone and nucleophilic sidechains including cysteine required the reactants to be in close proximity as part of a complex [39]. In a similar manner, the function of the positively charged CRGGRC motif may be to juxtapose the cysteine residues of Pc-8 with DOPA residues in negatively charged Pc-3 for optimal intermolecular crosslinking. Considering that most of the non-Pc-3 proteins are positively charged, electrostatic

Downloaded At: 19:54 21 January 2011

Pc-6
MW 34, 769 pI 8.27 1.3%

```
MLFYNANFVQKSWVLILLGLAAVVA
CSEYDKGLGGYGRPSY GGRRGY
GGRRGLQYHGKY
QGRCEYDGLYFRDEKSFVYCSNRNSYIQ
PCAPGTRNSPYTKYNRGSKYNYRDFCEVN
LVDSGYVPKPGYLPA PKKAY PTKVYD
LKVDYA PKVDYA   PKVDYA
PKVDYA PKVDYVA  PKASYVP
PKASYVDPTPTYGYEAPFKGGY DK
PSYGKDVDTSYESKTTYTVEKTADK
 GYGKGYGDKEISAKKSYTLTEKRDY
DTGYDNSRSDEDSKEYGYDNDRSESY
ERTESYTDERTDGYGTQKVEYTQQSEY
DRVTRRGIWLHKGTEVEHVLY
```

Pc-12
MW 34, 845 pI 8.96 1.3%

```
     KYYDDEKRDADKYRKPSYNPYNTYKD
YPPKKIYNDDEKRDADQYRISYNPYNTYKD
YPPKKKYYDDEKRDAYKYRNPSYNPYNTYKD
YPPKKIYYDDEKRDADQYRNPSYNPYNTYKD
YPPKKKYYDDEKRDADQYRNPSYNPYNTYKD
YLPKKKYYDDEKRDADQYRKPSYNPYNSYKD
YPPKKKYYDDEKRDADQYRNPSYNPYNTYKD
YLPKKKYYDDEKRDADQYRNPSYNPYNTYKD
YPPKKKYYDDEKRDADQYRNPSYNPYNTYKD
YP
```

Pc-17
MW 9,7 51 pI 7.98 1.3%

```
MKLLLLFALAAVAVA
LPYGYSGKPG YGYDAVDAVY
NRLADKQQAV NRKAEYVGAG
TGTAKYNGVP GANYGYENDR
KYGYDNKGYG GYGDKGYGGY
GDKGLYDGYY
```

Pc-8
MW 13, 697 pI 10.52 1.3%

```
KLALLLLVAVCAAVAVNA
CGPLGCS GGYGGVLK
CGVGGCALGGYGGGYSAGIGGYGIK
RLGCRGGRCGLRRRVGCRGGRCGLRG
RLGCRGGRCGLR KLGCRGGRCGLRG
RLGCRGGRCGLRKRLGCRGGRCGRGG
YGGGYGGVCSKGVCGGYPAYGK
```

Pc-14
MW 12, 246 pI 10.4 1.3%

```
GRIVICLLVLVAGA
YGIGCGYGGYGGYGGGF
HGGYIGY  HGGYPGY
SGGFRGY GYPGRVHTNVV
HHNIPVFMPPPMPRRAPAPAPRGR
TIILGGGKYGLFGKKSKNKGFGGL
GVLSLLGGLGGKGGGGIRFLGRK
```

FIGURE 5 Miscellaneous gene category. The frequency of expression of each gene is represented as a percentage of the total number of genes categorized as potential glue proteins. The Pc12 cDNA was incomplete at the 5′ end.

association between basic and acidic proteins may be the fundamental organizing principle of the *P. californica* adhesive [4], positioning redox active DOPA residues near nucleophilic sidechains to maximize intermolecular crosslinking and bond strength.

In addition to potential structural adhesive proteins, the adhesive gland expression survey should turn up genes for enzymes involved in adhesive processing. Vovelle reported phenoloxidase activity in the parathoracic region of another tube building polychaete, *Sabellaria alveolata* [40]. Two candidates for this category are a tyrosinase (accession #GH160617) and a laccase (accession #GH160619), both of which are copper oxidase enzymes that operate on phenolic substrates. Tyrosinases, widespread in plants and animals, catalyze the hydroxylation of tyrosine to DOPA (EC 1.14.18.1) and the further oxidation of DOPA to dopaquinone (EC 1.10.3.1) to initiate the pathway leading to melanin pigments [41]. The *P. californica* tyrosinase, which lacks a leader sequence and is predicted by PSORTII [14] to be cytoplasmic rather than secreted, may generate the DOPA that seems critical to the bonding and curing mechanisms of its bioadhesive. Laccases are found in plants, fungi, insects, and microorganisms. The *P. californica* laccase was somewhat more homologous to insect laccases involved in cuticle schlerotization and pigmentation [42]. Laccases (EC 1.10.3.2), possessing a broader substrate specificity than tyrosinases, oxidize N-acyl-catecholamines to form reactive quinones and quinone methides that crosslink nucleophilic sidechains between adjacent cuticle proteins and chitin microfibrils [43]. The *P. californica* laccase had a leader sequence and was predicted to be secreted [14]. It may be involved in schlerotization and/or pigmentation rather than hardening of the sandcastle glue. An enzyme category conspicuously absent was kinases, which could be reasonably expected based on the abundant expression and dense phosphorylation of the Pc-3 proteins.

Downloaded At: 19:54 21 January 2011

4. CONCLUDING REMARKS

The preliminary expression survey has produced a library of genetic probes for looking more deeply into the inner workings of the *P. californica* adhesive gland. Experiments are underway to localize expression of the new genes by *in situ* hybridization. Confirmation of expression in the adhesive gland will provide additional evidence the proteins may be in the secreted adhesive, but may also reveal compartmentalized expression and evidence of sequential assembly and processing of the secreted adhesive. Ongoing expansion of the survey will provide a still broader vista of the adhesive gland and may provide valuable biotechnological resources. Improved understanding of the

natural adhesive will guide efforts to develop synthetic mimetics [44]. Adhesives comprised of colloidal prepolymers carried in water with potentially benign setting reactions modeled after the Sandcastle adhesive could have profitable applications in medicine.

ACKNOWLEDGMENTS

This work was supported by a grant from the NIH (R01 EB006463). Some of the sequences were generated during an undergraduate laboratory associated with a molecular bioengineering course at the University of Utah.

REFERENCES

[1] Rouse, G. W. and Pleijel, F., *Polychaetes*, (Oxford University Press, Oxford, 2001).
[2] Defretin, R., in *Comprehensive Biochemistry: Extracellular and Supporting Structures*, M. Florkin and E. H. Stotz (Eds.) (Elsevier, Amsterdam, 1971), p. 713.
[3] Stevens, M. J., Steren, R. E., Hlady, V., and Stewart, R. J., *Langmuir* **23**, 5045–5049 (2007).
[4] Stewart, R. J., Weaver, J. C., Morse, D. E., and Waite, J. H., *J. Exp. Biol.* **207**, 4727–4734 (2004).
[5] Baier, R. E., Shafrin, E. G., and Zisman, W. A., *Science* **162**, 1360–1368 (1968).
[6] Zhao, H., Sun, C., Stewart, R. J., and Waite, J. H., *J. Biol. Chem.* **280**, 42938–42944 (2005).
[7] Waite, J. H., Jensen, R. A., and Morse, D. E., *Biochemistry* **31**, 5733–5738 (1992).
[8] Waite, J. H., Andersen, N. H., Jewhurst, S., and Sun, C., *J. Adhes.* **81**, 297–317 (2005).
[9] Jensen, R. A. and Morse, D. E., *J. Comp. Physiol. B Biochem. Syst. Environ. Physiol.* **158**, 317–324 (1998).
[10] Waite, J. H. and Qin, X., *Biochemistry* **40**, 2887–2893 (2001).
[11] Arnow, L. E., *J. Biol. Chem.* **118**, 531–537 (1937).
[12] Karlin, S. and Altschul, S. F., *Proc. Natl. Acad. Sci. USA* **87**, 2264–2268 (1990).
[13] Bendtsen, J. D., Nielsen, H., von Heijne, G., and Brunak, S., *J. Mol. Biol.* **340**, 783–795 (2004).
[14] Horton, P. and Nakai, K., *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **5**, 147–152 (1997).
[15] Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., and Bairoch, A., in *The Proteomics Protocols Handbook*, J. M. Walker (Ed.) (Humana Press, Totowa, NJ, 2005), Ch. 52, pp. 571–607.
[16] Rogers, M. A., Langbein, L., Praetzel-Wunder, S., Winter, H., and Schweizer, J., *Int. Rev. Cytol.* **251**, 209–263 (2006).
[17] Wu, D. D., Irwin, D. M., and Zhang, Y. P., *BMC Evol. Biol.* **8**, 241 (2008).
[18] Alibardi, L., Toni, M., and Valle, L. D., *Exp. Dermatol.* **16**, 961–976 (2007).
[19] Suzuki, M., Murayama, E., Inoue, H., Ozaki, N., *et al.*, *Biochem. J.* **382**, 205–213 (2004).
[20] Zhang, C., Xie, L., Huang, J., Liu, X., and Zhang, R., *Biochem. Biophys. Res. Commun.* **344**, 735–740 (2006).
[21] Yano, M., Nagai, K., Morimoto, K., and Miyamoto, H., *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **144**, 254–262 (2006).

Downloaded At: 19:54 21 January 2011

[22] Alexandrov, N. N., Brover, V. V., Freidin, S., Troukhan, M. E., Tatarinova, T. V., Zhang, H., Swaller, T. J., Lu, Y. P., Bouck, J., Flavell, R. B., and Feldmann, K. A., *Plant Mol. Biol.* **69**, 179–194 (2009).
[23] Futahashi, R. and Fujiwara, H., *Dev. Genes. Evol.* **218**, 491–504 (2008).
[24] Suzuki, Y., Matsuoka, T., Iimura, Y., and Fujiwara, H., *Insect Biochem. Mol. Biol.* **32**, 599–607 (2002).
[25] Ruangsittichai, J., Viyanant, V., Vichasri-Grams, S., Sobhon, P., Tesana, S., Upatham, E. S., Hofmann, A., Korge, G., and Grams, R., *Int. J. Parasitol.* **36**, 1329–1333 (2006).
[26] Guilfoile, P. G. and Packila, M., *Exp. Appl. Acarol.* **32**, 103–110 (2004).
[27] Koch, A. W., Holstein, T. W., Mala, C., Kurz, E., *et al.*, *J. Cell Sci.* **111** (Pt 11), 1545–1554 (1998).
[28] Lee, H., Scherer, N. F., and Messersmith, P. B., *Proc. Natl. Acad. Sci. USA* **103**, 12999–13003 (2006).
[29] Porath, J. and Olin, B., *Biochemistry* **22**, 1621–1630 (1983).
[30] Patwardhan, A. V., Goud, G. N., Koepsel, R. R., and Ataai, M. M., *J. Chromatogr. A* **787**, 91–100 (1997).
[31] Matsubara, T., Hiura, Y. and Kawashiro, K., *Int. J. Mod. Phys. B* **17**, 1324–1328 (2003).
[32] Harford, C. and Sarkar, B., *Acc. Chem. Res.* **30**, 123–130 (1997).
[33] Stayner, R. S., Min, D. J., Kiser, P. F., and Stewart, R. J., *Bioconjugate Chem.* **16**, 1617–1623 (2005).
[34] Endrizzi, B. J., Huang, G., Kiser, P. F., and Stewart, R. J., *Langmuir* **22**, 11305–11310 (2006).
[35] Brown, K. C., Yang, S. H., and Kodadek, T., *Biochemistry* **34**, 4733–4739 (1995).
[36] Berg, J. M., *Proc. Natl. Acad. Sci. USA* **85**, 99–102 (1988).
[37] Lippard, S. J. and Berg, J. M., *Principles of Bioinorganic Chemistry* (University Science Books, Mid Valley, CA, 1994).
[38] Zhao, H. and Waite, J. H., *Biochemistry* **45**, 14223–14231 (2006).
[39] Liu, B., Burdine, L., and Kodadek, T., *J. Am. Chem. Soc.* **128**, 15228–15235 (2006).
[40] Vovelle, J., *Arch. de Zool. Exp. Gen.* **106**, 1–187 (1965).
[41] Sanchez-Ferrer, A., Rodriguez-Lopez, J. N., Garcia-Canovas, F., and Garcia-Carmona, F., *Biochim. Biophys. Acta* **1247**, 1–11 (1995).
[42] Arakane, Y., Muthukrishnan, S., Beeman, R. W., Kanost, M. R., and Kramer, K. J., *Proc. Natl. Acad. Sci. USA* **102**, 11337–11342 (2005).
[43] Hopkins, T. L. and Kramer, K. J., *Annu. Rev. Entomol.* **37**, 273–302 (1992).
[44] Shao, H., Bachus, K. N., and Stewart, R. J., *Macromol. Biosci.* **9**, 464–471 (2008).

Downloaded At: 19:54 21 January 2011