In silico analysis of adhesive foot proteins of mussels

Muthukumar Singaravelu1,2, Kirubagaran Ramalingam2 and Sharmila Anishetty1*

1Centre for Biotechnology, Anna University, Guindy, Chennai – 600 025, Tamil Nadu, India.
2National Institute of Ocean Technology, MoES, Pallikaranai, Chennai - 600 100, Tamil Nadu, India.
*[E-mail: s_anishetty@yahoo.co.uk , s_anishetty@annauniv.edu]

Received 27 June 2014; revised 25 August 2014

Byssal threads made up of foot proteins help the mussels to colonize new habitats facilitating attachment in aqueous environments. In the current study, we have analyzed mussel adhesive proteins, foot protein 1 (fp1) and foot protein 3 (fp3) in terms of their amino acid composition and dipeptide and tripeptide preferences. Further, analysis of sequence around the post translationally modified fp1 shows a sequential arrangement of decapeptide or octapeptide patterns. The pattern that was derived from four species of Mytilus was "Y-X(5)-Y-X(3)", while the three sequences from Perna viridis show "W-X(2)-W-X(6)" pattern and the four sequences from Perna canaliculus show octapeptide "Y-X(3)-Y-X(3)" pattern repeats. In silico analysis of amino acid composition using PERL program confirms the significant role of amino acid residues. The dipeptide and tripeptide analyses of cement proteins of barnacles and polychaetes were also performed. Dipeptide and tripeptide analysis provided an overview of compositional bias which is prevalent among marine invertebrate adhesive protein complexes. The analyses of three groups of frequent macrofoulers (mussels, barnacles and polychaetes) bring out the importance of individual amino acids and dipeptides and tripeptides in underwater adhesion.

[Keywords: Adhesion, Foot proteins, Cement proteins, Dipeptide and Tripeptide patterns]

Introduction

Underwater adhesion by many marine organisms notably molluscs and barnacles has attracted attention over the past several decades. Understanding wet adhesion not only paves way for development of antifouling agents but also helps in facilitating bioinspired engineering. Several authors have contributed immensely and provided valuable insights into the nature of the adhesive proteins and their role in adhesion. In the current study, we address the foot proteins of molluscs and cement proteins of polychaetes and barnacles in relation to amino acid preferences and dipeptide and tripeptide preferences. We also evaluate the presence of motifs in terms of aromatic amino acids in fp1 sequences and address the rigidity of fp1 sequence of Perna viridis in terms of rigid tripeptide occurrences. While single residue preferences and repeated motifs have been addressed in literature, to our knowledge, a comprehensive survey of dipeptides and tripeptides as has been done in the current study is not available.

Materials and Methods

The sequences of foot proteins 1 and 3 from Mytilus edulis, Mytilus galloprovincialis, Mytilus coruscus, Mytilus californianus, Perna viridis and Perna canaliculus were obtained in FASTA format from UniProtKB. Sequences of cement proteins 1 and 2 and cement protein 3A and 3B of Sabellaria alveolata and Phragmatopoma californica and cement protein 100 kDa and 20 kDa from Amphibalanus amphiitrite were also retrieved from the UniProtKB.

Amino acid composition, dipeptide and tripeptide composition of all the sequences was computed using in-house developed PERL scripts. For motif generation and amino acid composition of fp1, sequences bearing the following UniprotIDs: (Q2TCK5, Q2TCK6, Q2TCK7, Q2TCK8, A1X158, A1X159, B5B0G9, Q25460, Q27409, Q2TCK9, Q2TCL0 and Q25434) were used. For amino acid composition of fp1 of three species Mytilus californianus, Perna viridis and Perna canaliculus, after computing the individual amino acid composition of different fp1 sequences in each species, average amino acid composition for each species was taken. Sequences which were complete were used for dipeptide and tripeptide analysis. For the dipeptide and tripeptide evaluation, a sliding window of one residue was used on the sequence. The motifs were manually deciphered based on the presence of aromatic amino acids Tyrosine and Tryptophan and validated using Scanprosite tool. For the rigid tripeptide analysis, the consensus repeats of fp1 of Perna viridis were checked for the
Results
Motifs of fp1 sequences of byssal thread
Fp1 is a protein with repeats of decapeptides and octapeptides. We derived patterns by identifying aromatic amino acids Tyrosine (Y) and Tryptophan (W) which appeared repeatedly at fixed positions. The importance of post-translational modification of Y to 3,4-dihydroxyphenylalanine (DOPA) and W to hydroxytryptophan in these proteins is well known. The appearance of these residues at fixed intervals may have some significance in the function of these fp1 sequences. Four sequences from Perna canaliculus showed repeats of the octapeptide pattern "Y-X(3)-Y-X(3)" as seen in Fig. I A. Three sequences of Perna viridis had repeats of the pattern "W-X(2)-W-X(6)" as shown in Fig. I B. Six sequences from four Mytilus species were classified into Group 1 (G1) (Mytilus edulis and Mytilus galloprovincialis) and Group 2 (G2) (Mytilus californianus and Mytilus coruscus). These had motif repeats "Y-X(5)-Y-X(3)" as seen in Fig. I C. In all these motifs, "X" represents any amino acid. It should be noted that in many cases, the residues corresponding to "X" may be conserved across the sequences which have been analyzed here, but we chose to emphasize the positional preference of aromatic amino acids alone in our motifs and therefore designated all other positions as "X".

Amino acid Composition of fp1 sequence
Amino acid composition of foot proteins and their role have been analyzed in various studies reported in literature. In the current study, amino acid composition of 12 fp1 sequences from the six species of mussels were analyzed. In G1 and G2 sequences, Proline, Lysine, Tyrosine, Threonine and Serine are the predominant amino acids. This is followed by Alanine, Leucine and Isoleucine. Tryptophan is absent in G1 and G2 sequences. Cysteine and Phenylalanine are absent in Mytilus edulis. All the other amino acids including Cysteine and Phenylalanine occur with a low frequency in all G1 and G2 sequences.

In Perna viridis, Proline is the most predominant residue followed by Alanine, Tryptophan, Threonine, Lysine and Glycine. Cysteine and Glutamine is absent in Perna viridis. This is followed by Valine, Arginine, Histidine, Tyrosine, Isoleucine and Asparagine. All other amino acids occur with a low frequency. In Perna canaliculus, Lysine is followed by Proline, Tyrosine, Valine, Cysteine and Glycine. This is followed by Serine, Asparagine and Leucine. Tryptophan and Glutamate are absent. All other amino acids occur with a low frequency. Details are provided in a graph (Fig. II).

Dipeptide Composition of fp1 sequences
Based on fp1 dipeptide profiling we were able to group the fp1 sequences of six species. Group 1 (G1) comprised of Mytilus edulis and Mytilus galloprovincialis and Group 2 (G2) comprised of Mytilus californianus and Mytilus coruscus. In Mytilus species, G1 shared all their top ten dipeptides (KP, YK, TY, YP, SY, PP, PT, PS, AK and KA), while G2 shared nine (KP, YK, TY, YP, SY, PP, PT, PS, and PK) out of their top ten dipeptides. Sequences of G1 and G2 also shared eight major dipeptides. Perna viridis and Perna canaliculus were distinct from each other and from G1 and G2 in their dipeptide profile. Perna viridis had AW, WT, WK, TA, PW, PP, PK, KP and KA whereas Perna canaliculus had KP, VK, YV, PY, DY, PC, KC, CV, PK and NP in its top ten dipeptides. Taking into account, the cutoff criterion of dipeptides which make up at least 5% of the total dipeptides in any individual fp1 sequence that has been processed, we obtained nineteen distinct dipeptides. These were found to have eight amino acids in various combinations. Proline is found to be present in seven dipeptides out of the nineteen. All the fp1 sequences had KP in them. Barring Mytilus edulis and Mytilus galloprovincialis, all others also had PK in the top ten. Tryptophan containing dipeptides AW, WT, WK and PW were a feature of Perna viridis while Cysteine and Proline containing dipeptides PK, PC, KC, CV, DY and NP were present in Perna canaliculus.

Tripeptide Composition of fp1 sequences
In terms of tripeptides, G1 group shared all the top ten tripeptides whereas G2 group shared nine out of ten of their top ten tripeptides. Making a comparison across G1 and G2 and taking into consideration, the top ten tripeptides, we find that six tripeptides YPP, PTY, SYP, PSY, TYK and PPT were shared between G1 and G2. Three tripeptides KAK, YKA, and AKP of G1 and three tripeptides YKP, KPK, and TYP of G2 were unique among the top ten tripeptides of the corresponding groups. KPS was in the top ten tripeptides of M. galloprovincialis, M. edulis and M. californianus. PKI was one of the distinct tripeptides of M. coruscus appearing in its top ten. The top ten tripeptides of Perna viridis were distinct: WTA, TAW, WKA, AWK, PWT, KP, AT, PW, TP, and KT.
and TPK. Four tripeptides, YVK, VKP, KPY and PYV made the bulk of the fp1 sequences of *P. canaliculus*. These are part of the octapeptide repeats. KPK and PKP are the other minor tripeptides. Eighteen tripeptides fulfill the criterion of making up 5% of the composition of any of the sequences considered in this study. Majority of these had Tyrosine and Proline in them.

We addressed these decapeptides in terms of rigid, intermediate and non-rigid tripeptides. Anishetty *et al.* had analyzed good resolution PDB structures and classified tripeptides into three categories rigid, non-rigid and intermediate based on relative structural rigidity between Cα and Cβ atoms of tripeptides.9

When two consecutive decapeptide consensus sequences were analyzed in terms of the tripeptides, we find that at the interface of the two repeats, there are 2 consecutive rigid tripeptides followed by an intermediate tripeptide followed by 2 consecutive rigid tripeptides. The peptide at the interface of the two consensus repeats "ATPKPWTAWK" and "APPPAWTAWK" was taken as "AWKAPP" shown underlined. Given the notation "R" for rigid, "I" for intermediate, the tripeptide pattern at the interface of the two repeats in terms of rigid and non-rigid tripeptides is "RRIRR". Giving a one residue sliding window and correlating this with rigid tripeptides, we find that AWK and WKA are rigid, while KAP is intermediate. This is followed by two rigid peptides APP and PPP giving a pattern "RRIRR". Given this, the appearance of four rigid tripeptides at the interface of repeats in the *Perna viridis* fp1, is suggestive of a significant role in providing strength to the byssal cuticle. However, it should be noted that the rigidity that we present here relates to the tripeptide by itself and does not take into account the role of any post-translational modification.

### Dipeptide and Tripeptide profiles of Fp3

In terms of dipeptide composition of fp3 sequences, taking the top tripeptides in terms of occurrence, we find that YG, YN and GY are shared by all four *Mytilus* species. While RR and RY is a feature of *M. edulis* and *M. galloprovincialis*, KG, NG, NK, NN and YY are shared by *M. californianus* and *M. coruscus*. GG and NY are shared by three species, the G1 group and *Mytilus edulis* while GW is shared by G2 group and *Mytilus edulis*. There are some dipeptides which are distinct to one or more species. The details are tabulated in **Table 1**.

**Table 1**: The number of sequences which share the top dipeptides of fp3 among the 4 *Mytilus* species.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>YG</td>
<td>19*</td>
<td>2*</td>
<td>8</td>
<td>13</td>
</tr>
<tr>
<td>GY</td>
<td>5</td>
<td>1</td>
<td>11*</td>
<td>14</td>
</tr>
<tr>
<td>YN</td>
<td>4</td>
<td>1</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td>RY</td>
<td>14</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RR</td>
<td>10</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NG</td>
<td></td>
<td>8</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>KG</td>
<td></td>
<td>5</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>YY</td>
<td></td>
<td>2</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>NK</td>
<td></td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>NN</td>
<td></td>
<td>4</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Dipeptides shared between 3 species</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GG</td>
<td></td>
<td>10</td>
<td>2*</td>
<td>1</td>
</tr>
<tr>
<td>GW</td>
<td></td>
<td>6</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>NY</td>
<td></td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Dipeptides shared between 2 species</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VL</td>
<td></td>
<td>8</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>GR</td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>NR</td>
<td></td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RG</td>
<td></td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dipeptides shared by all species</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GF</td>
<td></td>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>FF</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

* denotes present in all the sequences. G1 (*Mytilus edulis* and *Mytilus galloprovincialis*); G2 (*Mytilus californianus* and *Mytilus coruscus*).”

**Fp1 and Rigid Tripeptides**

Fp1 of green mussel *Perna viridis* has 42 decapeptide sequences with the consensus "ATPKPWTAWK" and "APPPAWTAWK". Adhesion in this species is said to be different from the DOPA based adhesion for other foot proteins. Studies suggest that the hydroxylation of tryptophan to 7-hydroxy tryptophan makes it a mimic of DOPA, the 7 hydroxy tryptophan, when oxidized forms an o-quinonimine and the indolic N-H and phenolic groups can chelate metal ions. The mannosylation of tryptophan residues can make the protein more resistant to degradation by exo- and endo- proteases and makes it less prone to aggregation and makes it water accessible.16, 17
Fig. I - Octapeptide repeats in *Perna canaliculus* (A) decapeptides repeats in *Perna viridis* (B) and of fp1 proteins (C) in *Mytilus* sp. generated using Scanprosite tool.

Fig. II - Fp1 amino acid composition among six mussel species
Di and Tripeptide profiles of Cement proteins of Barnacles

In the case of cement protein 100 kDa of *Amphibalanus amphitrite*, the top dipeptides in terms of percentage composition are SV, QL, LL, LP, LA, LS, SL, GL, VI, VP and IS. Most of them had Leucine in them. There were more number of dipeptides which were represented in this cement protein (i.e.) there seems to be no particular preference for a specific dipeptide. The 20 kDa cement protein sequence from this species was available as fragments in the UniProt database. Taking this into consideration we find that cysteine containing dipeptides CD, CN, DC, PC etc are prevalent. In terms of tripeptides, AAA, LLA, LPK, LQL and PSV are in the top five tripeptides of cement protein 100 kDa from *Amphibalanus amphitrite* while in the 20 kDa cement protein fragments of this species HPC, CDC and SCD were notable.

Di and Tripeptide profiles of Cement proteins of Polychaete

Cement proteins 1 and 2 (cp1 and cp2) from the polychaete *Sabellaria alveolata* were analyzed for their dipeptide and tripeptide preferences. As in the case of fp1 and fp3 of mollusces, we addressed only the top ten occurring dipeptides from cp1 and cp2. We find that in the case of cp1, YG, GA, GY, KG, AY, GK, GG, AK and KA are the top ten while in the case of cp2, YG, YA, GG, AG, HK, KA, GA, GV and VH were the top nine. As is evident, the dipeptides YG, KY, GG, KA and GA are shared between the top ten dipeptides of cp1 and cp2 of *Sabellaria alveolata*. It is also observed that all these five dipeptides are in the top ten of cp2 of *P. californica* while barring KA all the other dipeptides are in the top ten of cp1 of *P. californica*. Cement proteins 3A and 3B of *P. californica* are predominantly made up of dipeptides SS, SY and YS.

In terms of tripeptides GYG, GGY, YGA and YGG are common to cp1 and cp2 of *S. alveolata* and *P. californica* while VHK and HKA are shared by cp2 of these species. KAA is shared by cp1 of *S. alveolata* and cp2 of both the species. In the case of cement protein 3A and 3B of these two species, we find that they are predominantly made up of the tripeptides SSS, SSY, SYS and YSS.

Conclusion

Underwater adhesive proteins share a number of physical and chemical features like the native surface for settlement, post-translational modifications of amino acids, ability to form complexes with metal ions and the ability to displace low affinity ligands from the surfaces to which they adhere. The current study is a survey of the amino acid composition and more specifically dipeptide and tripeptide composition of three macrofoulers. It should be noted that the peptides that were taken into consideration were the top peptides in terms of frequency of occurrence in the corresponding sequences and in some cases, the exclusion of some peptides in the discussion does not necessarily mean that they are absent in the corresponding sequences.

Foot proteins 1 and 3 from six species of *Mytilus* were analyzed and the species were grouped based on the similarity of the shared peptides. In the case of polychaetes, we find that cement proteins cp1 and cp2 of polychaetes are very similar in their dipeptide and tripeptide composition whereas cement protein 3A and 3B are serine-rich and are different from cp1 and cp2 of polychaetes. The peptides of barnacles seem to be very different in their composition. We hope that the current analyses of dipeptides and tripeptides may potentially provide experimental leads for applications involving wet adhesion.

Acknowledgements

The authors thank BTIS, Department of Biotechnology for computational facilities. MS and RK are thankful to Dr. M.A. Atmanand, Director, NIO, for providing all facilities and financial support received from the Earth System Science Organization, Ministry of Earth Sciences, Government of India.

References


