PLEASE NOTE: This manuscript was (and is) scheduled to be published as:

Reeck, G., and Balthazor, J. (2016) Approaches to identifying proteins of aphid saliva: past, present and future. *University of Development Alternative (UODA) Studies Part A,* 4, 13-19.

Printing of the issue of the journal that will contain this paper is delayed due to serious health issues of the editor.

APPROACHES TO IDENTIFYING PROTEINS OF APHID SALIVA: PAST, PRESENT AND FUTURE

Gerald R. Reeck[1,*] and James Balthazor[2]

[1]Department of Biochemistry and Molecular Biophysics, Kansas State University, Manhattan KS 66506

[2]Department of Chemistry, Fort Hays State University, Hays KS 67601

*Author for correspondence (grreeck@gmail.com)

ABSTRACT

From both basic and applied viewpoints, studying in detail the roles and functions of proteins of aphid saliva is an important problem and also a challenging one. Here we briefly review the logically prior (and, in that sense, more fundamental) issue of simply identifying aphid saliva-proteins, leading eventually to identifying *all* proteins of aphid saliva. We highlight proteomic and transcriptomic methods that

have proven particularly fruitful in recent years, but also mention other approaches, of the past and future. All of these approaches, when combined, will ultimately provide reliable and comprehensive catalogs of proteins of aphid saliva, catalogs that can then form the basis for work on saliva-proteins as a system rather than simply as individual entities.

REVIEW

The importance of the problem

Aphids (and the plants they feed on) offer an extraordinarily rich biological landscape for scientific investigation, with important goals or pay-offs, both basic and applied. This richness stems from the fact that there are as many as 5,000 aphid species, all of which, one assumes, evolved from an individual primordial aphid species that is believed to have appeared about 220 million years ago, somewhere on the supercontinent Pangea (Grimaldi and Engel 2005). It is worth noting that aphids arose approximately 100 million years before the appearance of angiosperms and therefore that the first plant-feeding aphids must have fed on gymnosperms.

Insights from studies on aphids ultimately feed into our understanding of the very broad issue of the co-evolution of insects and plants –- in other words, into an understanding of plant defense systems and insects' defeat of those systems (or failure to defeat them). In the case of aphid/plant interactions, nature has, in a sense, provided incomparable experimental opportunities for scientific research, because each aphid species feeds on a restricted set of plant species. Moreover, the host-preference range varies widely, with some aphids being much more restricted in the range of host species than is the case for other, more generalist aphid species. The greatest restriction in host-plant range is perhaps the agriculturally relevant (and agriculturally induced) biotypes of species such as the Russian wheat aphid (*Diuraphis noxia*) and the greenbug (*Schizaphis graminum*).

In recent years, an analogy has been drawn between (some) proteins of aphid saliva, and "effector proteins" of fungal and bacterial plant pathogens (Bos et al. 2010; Carolan et al. 2011; Hogenhout and Bos 2011; Elzinga and Jander 2013). This is not to suggest a substantial overlap in the sets of aphid saliva-proteins and the sets of effector proteins of classical plant pathogens. Instead, the notion of saliva proteins as effectors can serve to emphasize the importance of the proteins of aphid saliva and also provides a useful conceptual framework in investigating those proteins and their interactions with plant components.

Clearly, to study the importance of and the detailed functions of individual aphid proteins of aphid saliva, these proteins must first be identified, and in what follows we will briefly review the increasingly effective efforts toward that end. In essentially all approaches, identifications take advantage of genomic or

transcriptomic information from various aphid species, but particularly from the pea aphid, *Acyrthosiphon pisum* (International Aphid Genomics Consortium 2010 and AphidBase (www.aphidbase.com)).

Three levels of analysis that in principle could contribute to identification of proteins of saliva

One can imagine identifying proteins of saliva by analysis at three levels: the protein level; or the level of transcripts encoding such proteins; or, potentially, the level of the genome, for direct identification of genes encoding saliva-proteins.

Before we review approaches of the last decade, it is instructive to look back to an earlier, less productive approach and also to look forward to an approach that will surely become important in the next decade. These two approaches lie at the extreme positions in the information flow in the Central Dogma of Molecular Biology.

On the one end. searches for protein-enzymes of saliva were carried out, and carried out energetically, for at least two decades. This body of work was nicely summarized by Miles (1999). Despite the best efforts of numerous investigators -- and on several aphid species -- the yield of information from enzyme-based studies was modest. The reasons for this are clear: the dilute nature of the samples, an inability to distinguish between different gene products (i.e., different proteins) having the same or overlapping activities, and, of course, the inability to identify any saliva-protein lacking (known or suspected) enzyme activity. We suggest, however, that this approach might yield very useful information with today's improved chromatographic methods and with the development of assays with exquisite sensitivity.

On the other extreme lies the mother-lode of relevant biological information -- nucleotide sequences of aphid genomes. It is reasonable to assume that transcriptional control elements exist in regions upstream of coding regions of many, if not all, genes that encode proteins of saliva. We have in mind, in particular, control elements that govern salivary-gland-specific transcription. At this point, such control elements have not been identified. (The one exception to this the gene encoding the protein Armet (Wang et al. 2015). This is probably an exceptional case, since Armet is bifunctional and its gene's transcription is controlled, at least in part, by mechanisms common to members of the Unfolded Protein Response. A UPR control element is what Dr. Feng Cui and her colleagues have discovered and characterized functionally in the Armet gene (Wang et al., 2015)) As more aphid genomes become available (for instance, the genomes of the green peach aphid (*Myzus persicae),* the Russian wheat aphid, and the greenbug) and as reliable catalogs of genes encoding saliva-proteins become established, searching for such control elements will certainly be undertaken. If such efforts are able to identify control elements involved in regulating transcription of saliva-proteins, the

nucleotide sequences of those control elements could then be used to search for additional protein-saliva genes.

Proteomics-based identifications of proteins of aphid saliva

Numerous aphid species, including the pea aphid (the current genetic model aphid), can be maintained on artificial liquid diets. This fortunate fact, coupled with the fact that aphids salivate continuously while feeding (Cherqui and Tjallingii, 2000), has allowed collection of saliva proteins in liquid diets. Such protein mixtures can be collected, processed and analyzed by proteomic methods, supported by genomic and transcriptomic information. This approach is far from simple, since the saliva proteins are very dilute in the fed-upon diet. As a result, diet fed on by tens of thousands of aphids is typically required for a single proteomic analysis.

Several such studies have now been reported (Harmel et al. 2008; Carolan et al. 2009; Cooper et al. 2010; Carolan et al. 2011; Rao et al. 2013; Vandermoten et al. 2014) on several different aphid species. If we apply the criterion of having an N-terminal secretion signal encoded by a protein's transcript (or gene), we judge that some 30 different proteins of aphid saliva have now been identified by proteomic analysis, with substantial differences in the sets of proteins identified in different aphid species. Of course, the failure to identify a protein in a given sample for a given species does not imply that the protein does not exist in saliva of that species. Thus, there may be more commonalities from species to species than is apparent at first glance from such studies.

Transcriptomics-based identification of proteins of aphid saliva

A somewhat indirect method that has proven very successful is analysis of cDNA libraries prepared with RNA isolated from dissected aphid salivary glands. We will call this, and related approaches, transcriptomics.

It is, of course, a small subset of salivary-gland transcripts that encodes proteins of saliva, and investigators have therefore typically imposed the criterion of the existence of a predicted N-terminal secretion signal. In addition, we have suggested and used an additional criterion (Carolan et al. 2011), namely, enrichment in salivary gland cDNA libraries as compared to whole-body cDNA libraries, using the R-statistic developed for this general purpose by Stekel et al.(2001).

Surveying results in all papers that use a transcriptomics approach of one sort or another (Ramsey et al. 2007; Mutti et al. 2006, 2008; Bos et al. 2010; Cui et al. 2012; Atamian et al. 2013) and cross-checking in AphidBase for pea aphid orthologs of transcripts initially proposed to encode saliva proteins in other species, we have found (Balthazor et al. 2016) that well over 100 different transcripts encoding

saliva-proteins have been proposed. Some of these, of course, overlap with proteins identified by proteomics.

Over 100 proteins of aphid saliva might strike one as high, but we note that fully 42 transcripts were identified in a single study, namely our transcriptomics work in the pea aphid (Carolan et al 2011), work that imposed the dual criteria of encoding putative signal sequences and being statistically significantly enriched in salivary gland cDNA libraries.

Conclusions

There is, of course, a great deal that could be said about the individual proposed proteins of aphid saliva, but our interest here is a broader one and a largely methodological one – the current, past and future means by which investigators have identified (and will continue identify) proteins of aphid saliva and their encoding transcripts and genes. Our main conclusion is that combining several approaches will provide the most comprehensive and powerful approach to identifying saliva proteins and thus establishing a saliva proteome (that is, a catalog of all proteins in an aphid species' saliva).

We close with a couple of general observations. First, the proteomics and transcriptomics approaches of the last decade have revealed a level of complexity of the saliva-proteome that was previously not known or, we suspect, imagined. This would suggest that additional saliva-proteins remain to be identified. In other words, there is no reliable estimate at this point of the total number of proteins of aphid saliva.

Secondly, a strikingly high number of the proposed members of the aphid saliva-proteome, whether identified from proteomics or transcriptomics, are "unknown" in the sense that homologs are not recognizable outside of aphids. Such unknown (or un-annotatable) proteins are particularly intriguing and perhaps particularly important, as suggested by the accumulating evidence that the prototype of such proteins in aphid saliva, that is, Protein c002 (Mutti et al. 2006), has been shown to be required for feeding on a host plant (Mutti et al. 2008) and, when overexpressed in a host plant, to lead to increased fecundity of aphids feeding on such plants (Bos et al. 2010).

ACKNOWLEDGEMENT

This work was supported by the Kansas Agricultural Experiment Station (contribution number 16-191-J).

REFERENCES CITED

Atamian, H.S., Chaudhary, R., Cin, V.D., Bao, E., Girke, T., and Kaloshian, I. (2013) In planta expression or delivery of potato aphid *Macrosiphum euphorbiae* effectors Me10 and Me23 enhances aphid fecundity. *Molecular Plant Microbe Interactions* 26: 67–74.

Balthazor , J., Chandrasekar, C., Aksamit, A., Cui,F., Caragea, D., Reese, J.C., and Reeck, G.R. (2016) Toward a comprehensive pea-aphid saliva-proteome (submitted).

Bos, J.I., Prince, D., Pitino, M., Maffei, M.E., Win, J., Hogenhout, S.A. (2010) A functional genomics approach identifies candidate effectors from the aphid species *Myzus persicae* (green peach aphid*). PLoS Genetics* doi: 10.1371/journal.pgen.1001216.

Carolan, J.C., Fitzroy, C.I.J., Ashton, P.D., Douglas, A.E., and Wilkinson, T.L. (2009) The secreted salivary proteome of the pea aphid *Acyrthosiphon pisum* characterized by mass spectrometry. *Proteomics* 9: 2457–2467.

Carolan, J.C., Caragea, D., Reardon, K.T., Mutti, NS., Dittmer, N., Pappan, K., Cui, F., Castaneto, M., Poulain, J., Dossat, C., Tagu, D., Reese, J.C., Reeck, G.R., Wilkinson, T.L., and Edwards, O.R., (2011) Predicted effector molecules in the salivary secretome of the pea aphid (*Acyrthosiphon pisum*): A dual transcriptomic/proteomic approach*. Journal of Proteome Research* 10: 1505-1518.

Cherqui, A. and Tjallingii, W.F. (2000) Salivary proteins of aphids, a pilot study on identification, separation and immunolocalisation. *Journal of Insect Physiology* 46: 1177-1186.

Cooper, W.R., Dillwith, J.W., and Puterka, G.J. (2010) Salivary proteins of Russian wheat aphid (Hemiptera: Aphididae). *Environmental Entomology.* 39: 223–231.

Cui, F., Smith, C.M., Reese, J., Edwards, O., and Reeck, G. (2012) Polymorphisms in salivary-gland transcripts of Russian wheat aphid biotypes 1 and 2. *Insect Science* 19: 429-440.

Grimaldi, D., and Engel, M.S. (2005) Evolution of the Insects. Cambridge University Press, New York, NY.

Harmel, N., Letocart, E., Cherqui, A., Giordanengo, P., Mazzucchelli, G., Guillonneau, F., De Pauw, E., Haubruge, E., and Francis, F. (2008) Identification of aphid salivary proteins: a proteomic investigation of *Myzus persicae. Insect Molecular Biology* 17: 165–174.

Hogenhout, S.A., and Bos, J.I. (2011) Effector proteins that modulate plant-insect interactions. *Curr Opin Plant Biol.* 14: 422-428.

International Aphid Genomics Consortium. (2010) Genome sequence of the pea aphid *Acyrthosiphon pisum. PLoS Biology*. doi: 10.1371/journal.pbio.1000313

Miles, P.W. (1999) Aphid saliva. *Biol. Rev. Cambridge Phil Soc.* 74: 41 – 85.

Mutti, N.S., Park, Y., Reese, J.C., and Reeck, G.R. (2006) RNAi knockdown of a salivary transcript leading to lethality in the pea aphid, *Acyrthosiphon pisum*. *J. Insect Sci.* doi: 10.1673/031.006.3801.

Mutti, N.S., Luis, J., Pappan, L.K, Pappan, K, Begum, K., Chen, M.S., Park, Y., Dittmer, N., Marshall, J., Reese, J.C., and Reeck, G.R. (2008) A protein from the salivary glands of the pea aphid, *Acyrthosiphon pisum*, is required in feeding on a host plant. *Proc. Natl. Acad. Sci. USA* 105: 9965–9969.

Ramsey, J.S., Wilson, A.C., de Vos, M., Sun, Q., Tamborindeguy, C., Winfield, A., Malloch, G., Smith, D.M., Fenton, B., Gray, S.M., and Jander, G. (2007) Genomic resources for *Myzus persicae*: EST sequencing, SNP identification, and microarray design. *BMC Genomics* 8: 423. doi:10.1186/1471-2164-8-423.

Rao, S. A., Carolan, J. C., and Wilkinson, T. L. (2013). Proteomic profiling of cereal aphid saliva reveals both ubiquitous and adaptive secreted proteins. *PLoS ONE*. 8:e57413. doi: 10.1371/journal.pone.0057413.

Stekel, D.J., Git, Y., and Falciani, F. (2000) The comparison of gene expression from multiple cDNA libraries. *Genome Research.* 10: 2055–2061.

Vandermoten, S., Harmel, N., Mazzucchelli, G., De Pauw, E., Haubruge, E., and Francis, F. (2014) Comparative analysis of salivary proteins from three aphid species. *Insect Mol. Biol.* 23: 66 – 77.

Wang, W., Dai, H., Zhang, Y., Chandrasekar, R., Luo, L., Hiromasa, Y., Sheng, C.Z, Peng G.X., Chen, S.L., Tomich, J.M, Reese, J.C, Edwards, O., Kang, L., Reeck, G.R., and Cui, F. (2015) Armet is an effector protein mediating aphid-plant interactions. *FASEB J.* 29: 2032-45.